Music Industry Business Models in the Digital Millennium: An Empirical Analysis of Streaming Music and Multi-Product Profits

---

A Thesis
Presented to
The Established Interdisciplinary Committee for Mathematics and Economics
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Andrew C. Shapiro

May 2009

Approved for the Committee
(Mathematics and Economics)

Lon Peters

Joe Roberts

Abstract

The music industry is undergoing a period of extreme flux. For most of the past century, the high costs of producing, promoting and distributing music served as massive barriers to entry into the music industry. In the last decade, however, new technologies and the proliferation of the internet have all but completely dissolved these barriers to entry. Among other developments, artists may now record and produce their music with relatively inexpensive technologies, market and distribute it via the internet, and incur low fixed costs and almost zero marginal costs in the process. As a result, artists are now generally less dependent on record labels, and more free to maximize their own multi-product profits. Profit-maximizing artists must now choose how to best utilize the imperfectly substitutable forms in which they may distribute their music, balancing the tradeoffs among CD revenues, paid digital download revenues, and the complementary revenues generated by free online music. With industry CD revenues continuing to fall, and alternative sources of music-related revenues growing and proliferating, one would expect a divergent shift in music industry business models away from those designed to maximize CD revenues. The relatively new technology of streaming music offers a valuable vantage point from which to survey these new business models empirically.

Controlling for past and current album sales and radio play, as well as determinants of concert demand such as ticket pricing and previous years' audiences, my thesis seeks to explore the following hypotheses regarding multi-product profit-maximizing firm behavior:

1. Artists that choose to supply more free streaming music are those that choose to offer a larger yearly supply of tickets to their performances.

2. Artists that exhibit higher demand for their free streaming music are those that choose to offer a larger yearly supply of tickets to their performances.

3. Artists that exhibit higher demand for their paid digital-downloads are those that choose to supply less free streaming music.

4. Free streaming music and paid digital-downloads are substitutes, albeit imperfect. We expect the cross-price elasticity of an artist's free streaming music and paid digital-downloads to be positive, ceteris paribus. We also expect the levels of demand for the two goods, across artists, to be inversely correlated, though free streaming music may generate an opposing effect, stimulating digital-download sales by allowing consumers to sample.

Table of Contents

List of Tables

Chapter 1

Background

1.1 Traditional Music Industry Structure

The music industry is composed of a complex variety of players, goods and contracts. Traditionally, four major types of firms have profited from the sales of music and music-related goods. At the center of the industry are musicians. For the purposes of this thesis, any individual or group of individuals that release, record or perform music under a unified name shall be referred to as artists.[1] These artists may write their own music and lyrics, or may purchase them from outside composers. Once copyrighted, compositions become the intellectual property of the musician(s) and/or composer(s) who are their authors, and recordings become the collective intellectual property of the artists that recorded them. The next two types of firms are concert promoters, who organize concert tours primarily by securing concert venues and promoting events, and record labels, which provide the means to produce and market albums. Furthermore, for those artists that write their own music, performing rights organizations (PRO's) license and monitor all outside use of their compositions, such as by radio or television stations, and collect royalties. Finally, there are a wide variety of additional players scattered throughout the industry, such as venue owners and ticket distributors.

The music industry, however, may be more broadly construed as being comprised not only of the markets for live and recorded music. Markets for recorded or live music often generate a variety of ancillary markets for complementary, non-music goods, the most traditional of which is artist-affiliated merchandise, such as the concert t-shirt. The supply side of the industry, therefore, may be viewed as the composite of not only musical artists and composers, but also the variety of firms that together supply

[1] The author would like to note that the designation of "artist" gives undue credit to many of the individuals and groups that have released, recorded and performed music in the last century.

consumers with music and its ancillary goods.

1.1.1 Record Labels & the "Recording Industry" Subsector

Prior to the impact of digitalization, the costs of producing, distributing, and promoting recorded music were sufficiently high that most artists could not independently record and promote their own music. These production and distribution costs reflected primarily the costs of physically producing and distributing each CD. A lesser promotional cost of note, however, is represented by the common practice of record labels paying radio stations for air play. This practice is legal according to US law[2] if, and only if, the payment is acknowledged at the time of the broadcast. Considerable weight, however, is generally given to the impact of the illegal variety of such payments, the exchange of which has been termed "payola."

Hence the development of the "record label." A record label is essentially a firm that amasses the means to record, distribute and promote albums. A label then offers to "sign" certain artists to its roster. If an artist agrees, the label will enable this musical artist to record and promote an album, subject to a contractual agreement. These contracts have traditionally provided much larger CD revenue shares for labels than for musical artists. Such contracts, however, were at least partially necessitated by the high costs of recording and promoting musical artists, combined with the low likelihood that a new musical artist would generate significant revenue from album sales.

Historically, only very few artists were able to adopt business models that generated profits without the support of, and contractual obligation to, record labels. Such artists are epitomized by the Grateful Dead, who were able to generate demand for their music and complementary goods primarily through live performance and without initial label support. The remaining vast majority of artists, however, were able to enter the music industry only under contract to labels. As such, most artists were obligated to maximize their labels' profits from album sales, rather than their own multi-product profits. For this reason, the recording industry has long been the dominant force in the music industry, so much so that the two were virtually synonymous prior to the impact of digitalization.

[2] 47 U.S.C §317 (Announcement of payment for broadcast)

Recording Industry Concentration

The high fixed costs of producing, distributing and promoting recorded music have not only ensured artists' dependence on record labels in the past, but they also constituted significant barriers to new record label entry into the recording industry. The industry, as a result, has long been highly concentrated. Alexander (1994) notes, in fact, that the industry has been highly concentrated for most of the past century. The only periods of exception were during the 1910's and 1950's, when the advent of new technologies significantly lowered production costs and triggered waves of firm entry.

As early as the 1960's, however, a wave of horizontal mergers began in the recording industry.[3] At the same time, the industry began to shift from independent to integrated distribution as the major firms began purchasing independent distributors. The resulting pressure began to drive the remaining independent distributors to bankruptcy over the following decades, a trend that accelerated in the 1980's.[4] By the end of the decade, six major firms held dominant market shares as measured at the distributor level.[5] [6] Since then, the industry has not become any less concentrated. Curien and Moreau (2005) noted in early 2005 that only four firms, Sony/BMG, Universal Music, EMI, and Warner Music, accounted for 80% of music revenues worldwide since a merger between Sony Music and Bertelsmann Music Group (BMG) in November 2003.[7]

Record Labels & Pricing

As a result of its high concentration, the structure of the recording industry over the past two decades has been unquestionably oligopolistic. Theoretical literature, however, generally assumes that record labels price as monopolists. Rob and Waldfogel (2006) note that "because CD's were easily transferable even in the absence of downloading, substantial price discrimination was impracticable. As a result, firms were compelled to price as single price monopolists."[8] They later conclude that, "prior to the advent of unpaid downloading . . . we view the seller of each album as a single-price monopolist."[9] Curien & Moreau (2005) take this conclusion one step further,

[3] Alexander (1994)
[4] Black & Greer (1987)
[5] Alexander (1994)
[6] Business Week (1988)
[7] Curien & Moreau (2005)
[8] Rob & Waldfogel (2006)
[9] Rob & Waldfogel (2006)

choosing to "consider the music industry as a monopoly, a sketchy representation of the majors' oligopoly."[10]

Record labels are assumed to price as monopolists for two reasons. First, albums are an extremely differentiated product and, as such, albums by different artists are very imperfect substitutes. A consumer who intends to purchase a specific album by one artist will not easily be swayed to, instead, purchase a less expensive substitute album by another artist. As technological progress allows more artists to promote their music and consumers are presented with more venues through which to sample music, however, it is worth noting that the likelihood of such substitution may very well be increasing. Regardless, because musical artists have traditionally signed multi-album contracts with record labels, each label essentially holds a monopoly over its artists' albums and are, therefore, generally assumed to price as monopolists.

The second reason that record labels are assumed to price as monopolists is slightly more concrete. There is evidence that the major record labels have, at times, colluded and acted as a cartel. The Universal, Warner, BMG, Sony and EMI record labels, along with three major music retailers, were charged with price collusion over the period 1995-2000.[11] In 2002, these firms agreed to a $143 million settlement, indicating that the major labels had likely colluded and behaved as a cartel. Presumably, it is the combination of these factors that drive Curien & Moreau to view the "majors' oligopoly" roughly as a monopoly.

1.2 The Music Industry Enters the Digital Era

1.2.1 CD Sales Displacement, Napster and the RIAA

In 1999, a college student brought online the most notorious peer-to-peer file-sharing network to date. This network, called Napster, enabled users to share any file on their computers, as well as to download any files shared by other users, and was the first such network to gain widespread use. Exchange via the network was in no way regulated, and many Napster users chose to share and/or download copyrighted materials. The widespread appeal of this exchange would soon be evidenced by massive growth in the network's usage.

While Napster rose to prominence, however, the US recording industry was not experiencing the same success. In 2000, US record sales, which had been growing an

[10] Curien & Moreau (2005)
[11] Curien & Moreau (2005)

average of 10% per year for seven consecutive years,[12] began a sharp decline. Between 2000 and 2003, the number of CD's shipped in the US fell by 20% to 750 million units (RIAA, 2004). The Recording Industry Association of America (RIAA), composed of the six major US record labels, immediately attributed this sales displacement to the exchange of its copyrighted materials via peer-to-peer networks. The RIAA launched a massive campaign against this perceived threat to its revenue, which it termed "piracy."

The RIAA finally began to make significant progress in its battle against Napster in early 2001, when the appeal of a preliminary injunction against the network's online exchange of copyrighted material reached a federal appeals court. During these proceedings, Napster estimated that it had 65 million users overall and 10 million users daily. The RIAA asserted that it held the copyrights to as much as 70% of the materials being shared via Napster at the time. The appeals court upheld the issuance of a preliminary injunction against Napster (albeit in a narrower form) in February of 2001, but this was a small victory for the RIAA. The previous month, Napster's users had downloaded an estimated 3 billion songs and, even in the face of the newly upheld injunction, the network remained online without any additional regulation. Subsequently, however, as a result of the injunction, Napster was forced to go offline entirely.[13]

As Napster's popularity grew, a number of similar networks began to appear. Some of these networks, such as Kazaa, rose to great prominence in the wake of the Napster shutdown. Kazaa, as well as other notable peer-to-peer networks such as Morpheus and Grokster, operated using technology that arose from a venture called FastTrack. Networks that operate using FastTrack technology, which had been downloaded by an estimated 400 million users as of 2005, cannot be centrally shut down in the same manner as Napster.[14] In fact, many networks have been able to sustain the exchange of copyrighted material long since Napster's shutdown, as a result of their structural differences from Napster. In addition to its litigation against peer-to-peer network operators, the RIAA has chosen to pursue an additional course of action and, in 2002, began suing individual users of peer-to-peer programs directly for copyright infringement.

[12] Rob & Waldfogel (2006)
[13] NYT, Richtel (2001)
[14] Billboard, Butler (2005)

1.2.2 New Sources of Digital Music

iTunes

While the major US labels have busied themselves battling piracy via peer-to-peer networks, new digital sources of recorded music are rising to prominence. Recent years have seen drastically increasing sales of digital music downloads by Apple Inc. through its "iTunes music store." As early as March of 2004, 50 million songs had been purchased through iTunes. By the end of 2006, downloads comprised 5% of US music sales and iTunes' average yearly sales reached $1 billion. Less than 2 years later, in February of 2008, downloads comprised 10% of US music sales and iTunes was ranked second only to Wal-Mart as the most popular music retailer in the US.

This growth is not particularly surprising. Online music sales not only offer consumers a more convenient method of shopping and browsing, but they also allow firms to forgo many of the costs associated with physically producing, distributing and selling CDs. Furthermore, the iTunes store offers most popular music at the price of $0.99 per song or $9.99 per album, which are prices far less than those at which CDs have ever been available.

Streaming Music

Since the initial rise of "piracy," another form of free online music has gained prominence. Streaming music is roughly to pirated recordings as radio play is to CD's. Like piracy, streaming music has made copyrighted music widely and easily available at no cost. Music being streamed, however, must be listened to from an open internet browser and hence, the listener is never in possession of a copy of the recording. Many sources of streaming music online, such as "internet radio," do not allow the user to control entirely the music to which they listen. Two sources of free streaming music, however, that do allow the user to select exactly the tracks to which they listen are MySpace.com and YouTube.com. MySpace and YouTube are two of the most trafficked websites on the internet. The former was developed and initially gained popularity as a "social networking tool," and the latter as host for videos that users chose to upload ("user-generated content").

YouTube has long exhibited the widespread usage and universal availability of copyrighted materials that was characteristic of file-sharing in its heyday. Furthermore, as was true of Napster, recordings on YouTube, being user-uploaded, are of variable quality and may, for example, be mislabeled. Furthermore, it seems as though YouTube is finally succumbing to copyright law. If one now browses the site for copyrighted music, one will eventually encounter the message that a song has been removed at the request of the copyright holder. On a potentially related note, YouTube has recently introduced official artist channels, which offer officially uploaded or endorsed videos through a main page where artists may post promotional materials, such as updates and tour dates.

Though MySpace was initially developed as a social networking tool, much like Facebook.com, most artists now have official MySpace pages on which they offer free streaming music. Furthermore, the site recently introduced a new, standardized music player. This player offers a drop-down menu of playlists, which some artists use to offer multiple albums. It initially opens by default, however, to each artist's "featured playlist." The player, therefore, allows artists to choose the recordings to which consumers will receive the greatest exposure and may serve as a useful promotional tool for some artists. Furthermore, for each individual recording that musical artists choose to stream via MySpace, they may choose whether to additionally offer it for free or paid download.

1.2.3 New Industry Business Models

Declining Costs of Recording, Distribution & Promotion

With technological progress and the proliferation of the internet there now exist means by which to digitally produce and distribute music and promote it via the internet, with little fixed costs and essentially no marginal costs. Alexander (1994) predicted well over a decade ago that a "digital delivery highway" for music products would allow firms to distribute these goods, both as promotional samples and for the purpose of sales. Furthermore, he predicted that these methods of promotion, distribution and sales would be different from their traditional equivalent, such as promotional radio airplay and the sales and distribution of music in physical formats, in that these new methods would be less costly and non-exclusionary.

Alexander also hypothesized that the advent of digital media technologies would facilitate firm entry into the music industry and erode the then-current market structure. This was quite astute. As a result of technological progress and the proliferation of the internet, musical artists have little to no remaining need for record labels to serve in their traditional capacity as an intermediary between musical artists and the public. Unsurprisingly, many musical artists are adopting business models in which they have no need for a record label and are choosing to leave their labels altogether. Labels, meanwhile, are struggling in many cases to adapt their contracts with artists to be more inclusive of alternative revenues.

Alternative Revenue Streams

In the current literature, revenues from music-related products other than recorded music are considered to be "alternative revenue streams." In fact, revenues from live performances, and even digital-format recordings, are generally referred to as "alternative." This terminology is misleading, as it is based on the antiquated notion that CD's and other physical recordings are the clear primary sources of revenue in the music industry. In fact, while CD revenues are declining, music related revenues from advertising, ringtone sales, digital-downloads and live performances are all increasing. Furthermore, alternative revenue streams are diversifying as ancillary markets continue to appear. The band Radiohead, for example, recently introduced its own social networking website, akin to the Facebook and MySpace sites. Such diversification is indicative of larger developments in the music industry.

Conclusion

With the influence of record labels declining, and alternative revenue streams growing and proliferating, the variety of multi-product music industry business models with the potential for profits is also growing. This growth is evidenced by the appearance of new and unique business models. These business models, in many cases, may stand to increase their profits through the free distribution of digital music online. Economists, in fact, have dedicated significant effort to examining the developments in the music industry over the last decade.

Chapter 2

Review of Relevant Economic Literature

There is a wealth of economic literature pertaining to musical artists' and record labels' respective multi-product profits in the presence of illegal file-sharing. Free streaming music, like illegal file-sharing, is a source of free, low-quality, imperfect substitutes for purchased music. This literature, therefore, can be generalized to fit the framework of free streaming music.

2.1 Illegal File-Sharing vs. RIAA CD Sales Displacement

Publication of the earliest literature to examine the effects of file-sharing began in 2004. This literature was primarily focused on determining the extent to which consumers substitute illegal downloads for CD purchases. This batch of literature arose in response to the RIAA's claim that illegal file-sharing was entirely responsible for its CD sales displacement that began in 2000. Two of the earliest and most notable empirical studies to investigate this claim were Rob & Waldfogel (2004) and Oberholzer & Strumpf (2004). Rob & Waldfolgel gathered data from a sample of 412 students at four Northeast US colleges and universities. The students were administered surveys regarding their music purchases, illegal downloads and broadband internet access. The study used both OLS and an instrumental variable approach, with broadband access as a source of exogenous variations in downloading behavior, and the two approaches yielded consistent estimates. In the sample, each illegally downloaded album was found to have displaced an estimated .2 album purchases. The subset of "hit albums released 1993-2003," exhibited no significant sales displacement as a result of illegal downloading.

2.1.1 Rob & Waldfogel, Oberholzer & Strumpf

Rob & Waldfogel also obtained additional data from a subsample of 92 students regarding their ex post and ex ante valuations of their music. In this subsample, the amount of deadweight loss eliminated by illegal downloading exceeded the resulting quantity of displaced album sales revenues. This social surplus, they argue, resulted from socially beneficial transactions that occurred via illegal file-sharing, but which otherwise would have been forgone due to record labels' pricing as single-price monopolists. Given the monopolistic behavior of suppliers, therefore, illegal downloading is arguably socially beneficial.

Unlike Rob & Waldfogel, Oberholzer & Strumpf approached the issue using national data. They drew a sample of albums from 2002 Billboard charts and obtained detailed nationwide sales data for these albums from Nielsen SoundScan, the company that gathers the raw data for the Billboard charts. They then gathered data regarding the extent of the illegal downloading of these albums. This data came from two peer-to-peer network servers, which they monitored for approximately four months. Over the sample period they observed a total of 1.75 million downloads, which the authors estimate comprised .01% of total worldwide file-sharing at the time. Furthermore, the authors matched the data from their servers to that of a large sample drawn from FastTrack/Kazaa networks, the most highly used network at the time, and found the two sets to be highly correlated. Hence, they could use their results to estimate file-sharing behavior at the national level.

Endogeneity and Album Sales Displacement

In order to estimate the negative correlation between album sales and levels of illegal downloading, i.e., album sales displacement, one must control for their endogenous relationship. Suppose that one were to simply regress sales of individual albums as a dependent variable on their amounts of illegal downloading and radio airplay, as two explanatory variables. Both illegal downloading and album sales should vary positively with radio airplay. Hence, in this regression there is no way to differentiate a result with one estimated coefficient on radio airplay and a negative coefficient on illegal downloads from another result with an appropriately smaller coefficient on radio airplay and a coefficient on illegal downloads equal to zero. Because album sales and illegal downloading co-vary with many of the same explanatory variables, their relationship is said to be endogenous, as are these explanatory variables. In order to deal with the problem of endogeneity in a regression model, one must specify the model with the proper number of exogenous instruments, i.e., variables that are not endogenous.

The most notable, or at least unique, source of exogenous variation used in Oberholzer & Strumpf's model came from German school holidays. Apparently, Germany had the greatest rate of high speed internet access at the time, as well as the largest internet population in Europe. As a result, one in six downloads from file-sharing networks at the time came from German up-loaders. Such up-loaders were generally schoolchildren, who did so not in school, but at home. Hence, German school holidays, which conveniently have little overlap with American holidays, served as an ideal source of exogenous variation in file-sharing behavior. Though the authors were concerned particularly with downloading in the US, the downloading of materials via peer-to-peer networks is, by nature, dependent on the amount of concurrent uploading of these materials. In the end, this study found the effect of illegal downloading on CD sales to be statistically indistinguishable from zero.

Critiques

Rob & Waldfogel and Oberholzer & Strumpf discuss each other's approaches in their 2006 and 2007 papers.[1] Oberholzer & Strumpf had compared file-sharing and sales data on a weekly basis, for example. Rob & Waldfogel note that CD's are durable goods and, as such, if substitution or displacement occurs, it may very well not occur in the course of a single week. They also note that the correlation that Oberholzer & Strumpf sought to estimate may not necessarily bear on the issue of sales displacement from piracy at all. They construct a model in which consumers are either downloaders or buyers, and choose exclusively one method of obtaining music. In this framework, a negative correlation across albums between sales and illegal downloads would merely indicate a significant difference in musical taste between downloaders and buyers. This seems to be a valid point. Downloaders and buyers may very well be relatively distinct demographics, by age, for example. These demographics, in turn, may have relatively distinct musical tastes. Suppose classical music exhibits high levels of sales and low levels of downloads. Now, suppose that classical music is simply more popular with older consumers than with younger consumers, and that older consumers tend to download less and purchase more than younger consumers.

[1] Rob & Waldfogel (2006), Oberholzer & Strumpf (2007)

Finally, suppose that young consumers were to trend towards consuming more classical music, over time, while, over the same time period, older consumers were to trend away from consuming classical music, due to general demographic shifts in musical taste. In this scenario, the decrease in classical albums' sales and the concurrent increase in their level of downloads would represent a negative correlation attributable to an omitted variable. A finding of an inverse correlation between album sales and file-sharing over this period, therefore, would be inflated by, if not solely indicative of, the demographic shift. Hence, Oberholzer & Strumpf's findings may have been biased by such unobserved effects.

Oberholzer & Strumpf have less criticism of Rob & Waldfogel's approach. They do comment, however, on the latter's finding of less sales displacement for "hit albums." They note that Rob & Waldfogel had examined only music that students had acquired in 2003, and had defined "hit albums" as having sold more than 2 million copies since 1999. Hence these "hit albums," compared to the average album acquired in 2003, were more likely to have experienced more sales earlier in the 1999-2003 period. Since illegal file-sharing was less prevalent earlier in this period, Rob & Walfogel's finding that these "hit albums" experienced less sales displacement may have resulted most directly from a bias in their approach.

2.2 File-Sharing and Music Industry Profits

By 2005, the focus of the economic literature regarding file-sharing had shifted notably. Much of this recent literature examines how file-sharing can be harnessed to stimulate profits. The first step in this direction in 2004 was the introduction of the concept of sampling. The essence of this concept is that when consumers are able to sample music for free, they are able to find music that better suits their tastes. As a result, consumers are more likely to purchase albums in the presence of sampling. Hence, file-sharing could potentially stimulate album sales. There is a rough consensus, however, that this sampling effect did not prevail over the opposing substitution effect in the years 2000-2003, but theoretically this need not be the case.

2.2.1 Peitz & Waelbroeck (2005) *Why the Music Industry May Gain From Free Downloading - the Role of Sampling*

Peitz and Waelbroeck (2005) construct a model to demonstrate that, by allowing consumers to more easily find music that suits their tastes, file-sharing yields the potential for increased profits from album sales. They argue that "sampling appears to be important in the market for recorded music" because, not only is music an experience good, but also "horizontal product differentiation[2] and taste heterogeneity" are crucial components of the market for recorded music. By definition, consumers cannot ascertain the quality or characteristics of an experience good without consuming it. Hence, in the absence of file-sharing, consumers can only identify music that matches their tastes if they have previously heard it, on the radio or through a friend, for example, and must otherwise choose their music purchases at random. This information imperfection makes it difficult for consumers to find the products that match their tastes, and this problem is exacerbated by the fact that both music and musical tastes exist in a wide and distinct variety. Hence, the authors argue, "music labels may actually gain from peer-to-peer networks (and other ways to listen to recorded music for free) and use them to solve a two-sided asymmetric information problem between buyers and sellers."[3] The essence of this argument is that the degree of product differentiation in the market for music does not only make it difficult for consumers to locate the goods that match their tastes, it also makes it more difficult for suppliers to locate the consumers to which their products are best suited. Hence, this problem is "two-sided." Music labels, therefore, in addition to consumers, stand to benefit from solutions to this problem, such as sampling.

The paper begins by assuming that the recording industry consists of a single multi-product monopolist. This treatment is appropriate, the authors argue, because CD prices are generally uniform across albums and labels, as well as because the industry is dominated by a tight oligopoly with a history of price collusion. This monopolist offers N products and incurs no marginal or fixed costs of production, such that its profits are equal to its revenues. In order to "formalize product differentiation," the authors introduce the structure of a "Salop circle," a unit length circle

[2]With regard to the market for recorded music, horizontal product differentiation generally refers to differentiation among albums or musical artists, whereas vertical product differentiation generally refers to differentiation among different formats, such as CD's vs. digital downloads, of the same album.

[3]Peitz & Waelbroeck (2005), pg. 3

along which all products are distributed evenly. Each product i located at point ℓ_i on the circle, is equidistant from its neighboring products such that the distance between each product is $1/N$. Because all products are equidistant, they are all presumed to sell for the same profit maximizing price p. Consumers with a total mass of 1 are uniformly distributed on the circle and will purchase no more than one product. Their hypothetical ideal products are located at ω, such that the product that best suits the taste of consumer j is the ℓ_i located closest to ω_j. If a consumer located at ℓ chooses the product located at ℓ_i, then this consumer's gross surplus is $r - \tau|\omega_j - \ell_i|$, where r is the maximum possible surplus, which occurs if, and only if, the condition $\ell_i = \omega_j$ holds true. The parameter τ is termed the "transport cost" and indicates the degree of substitutability between products, such that the larger τ, the more differentiated the products. As in all other similar literature, purchased products are considered to yield consumers higher utility than free downloads. In this model, a consumer who chooses to purchase his or her product of choice receives a benefit that is greater than the value of its copy by the factor $\alpha/2$.

Initially, consumers do not know where on the circle products are located, and are presented with a two-stage decision. In stage one, they must choose whether or not to download, denoted $d = 1$ if they do and $d = 0$ if they do not. Through downloading, consumers can locate their ideal product, though this process has an opportunity cost s as a result of the inconvenience of using file-sharing networks. In stage two, they then choose whether to buy a product, in which case $b = 1$, or not to, in which case $b = 0$. Each consumer's action is denoted (b, d), and his or her expected utility at stage one is denoted $u(b, d)$. This function is normalized to zero such that $u(0, 0) = 0$. The authors proceed to construct a similar function for consumers' expected utility at stage two. They then explore their model under varying parameter values.

The authors conclude that firm profits are increased by free downloading if consumers' tastes are sufficiently heterogeneous and there is sufficient product diversity (N). A significant consequence of this model is that consumers are willing to pay more when free sampling is available, because they are able to locate products that better suit their tastes.

2.2.2 Gayer and Shy (2006) *Publishers, Artists and Copyright Enforcement*

Gayer and Shy (2006) construct a model in which the recording publisher, i.e., record label, and the musical artist are separate firms. They begin by constructing consumers' demand for recorded media. Consumers may purchase a given recording for price p_r or may obtain an illegal copy for free. These consumers are indexed by $x \in [0, +\infty]$, according to a declining preference for obtaining the given recording, and their utility functions are given by:

$$max\{\alpha(1-x) + \gamma N - p_r,\ \beta(1-x) + \gamma N,\ 0\}, \tag{2.1}$$

where the first term is consumers' utility from consuming a purchased recording, the second from consuming an illegal copy, and the third from choosing not to consume. α and β are parameters for consumers' utilities from obtaining the paid and illegal copies of the recording. N is the total number of consumers in possession of the recording, regardless of whether their copies are legal or illegal. The parameter γ reflects the extent to which the potential utility from obtaining the recording increases with N. Hence, this parameter represents the additional utility that results from a network effect as the recording becoming more popular. The authors assume that legally and illegally-obtained recordings are vertically differentiated such that $\alpha > \beta > \gamma$. The implication of this assumption is that, given that the set of all consumers is continuously distributed along a number line, this continuum is partitioned into three segments at two points, which represent two marginal consumers. One of these consumers is indifferent between purchasing or downloading the recording and the other is indifferent between downloading the recording or not obtaining it at all. Hence, the set of all consumers is partitioned such that, if one imagines the continuum of all consumers vertically, its upper segment contains all consumers who purchase the recording, its middle segment contains all consumers who illegally download the recording, and its bottom segment contains all consumers who do not obtain the recording.

The authors then proceed to assume that live performance has a linear demand function, though they note that they do so only for the sake of simplicity. This demand is given by $q_p = max\{\delta N - p_p, 0\}$, where, p_p is the ticket price and the parameter δ ($\delta > 0$) measures the extent to which the magnitude of N affects live performance demand. Next, the authors assume that musical artists price concert tickets as monopolists, bear no costs and maximize their live performance profit function, $\pi_p = p_p q_p = p_p(\delta N - p_p)$. Ticket price and profits from live performance are therefore given by:

$$p_p = \frac{\delta N}{2} \Rightarrow \pi_p = p_p^2 = \frac{\delta^2 N^2}{4} \tag{2.2}$$

Next the authors note that N is the single determinant of the location of the marginal consumer who is indifferent between consumption of the illegal download and non-consumption. They conclude, therefore, that they may substitute $x = N$ into the condition for this marginal consumer, which yields $\beta(1-N)+\gamma N = 0 \Rightarrow N = \frac{\beta}{\beta-\gamma}$. Next they define $\hat{x}$ as the total number of buyers, and substitute this variable into the condition for the marginal consumer indifferent between downloading and buying. The result is, $\alpha(1-\hat{x}) + \gamma N - p_r = \beta(1-\hat{x}) + \gamma N$, into which they substitute the previous result for N, and solve for $\hat{x}$:

$$\hat{x} = \frac{\alpha - \beta - p_r}{\alpha - \beta} \tag{2.3}$$

The authors then introduce sales of a recorded medium that is priced by the record label at p_r, the profits from which are π_r. The parameter s $(0 < s < 1)$ represents the artist's share of the profit from sales of the recorded medium, such that the artists profits are $s\pi_r$ and the record label's profits are the remaining $(1-s)\pi_r$. The authors also assume that the cost (c) of producing a copy of the recording for sale is less than the increase in utility a consumer experiences from purchasing rather than downloading. Mathematically, $c < \alpha - \beta$. They proceed to determine the price p_r that maximizes record label profits, and finally determine the musical artist's profit function:

$$\pi_a = \pi_p + s\pi_r = \frac{\beta^2\delta^2}{4(\beta-\gamma)^2} + s\frac{(\alpha-\beta-c)^2}{4(\alpha-\beta)} \tag{2.4}$$

The authors proceed to examine the behavior of the model under varying conditions, i.e., different sets of values for the parameters α, β, γ and δ. They conclude that copyright enforcement is only profitable to musical artists when $s > \phi\delta^2$, where ϕ is an intricate function of α, β and γ. These are musical artists whose shares of revenues from recorded music, s, are sufficiently large, and the demand for whose live performances are sufficiently unaffected by variations in N. In most literature, however, there is a consensus that s is, in fact, almost universally small, as well as that, since music is an experience good, N is perhaps the single largest determinant of the demand for live performance. Hence, in most cases, copyright enforcement constrains artist profits and, therefore, offering free streaming music should be a profitable undertaking. The authors also note that their model indicates a conflict of interest between musical artists and record labels. Under all specifications of the model, the presence of piracy diminishes profits from album sales and increases profits from live performance. Hence, for the large majority of realistic parameter values, musical artists' profits are increased by piracy, whereas record labels' profits are diminished.

2.2.3 Grassi (2007) *The Music Market in the Age of Download*

Grassi (2007) closely parallels the work of Gayer & Shy (2006), though the variations between the two papers are of great significance. Grassi begins by examining empirical data from the Italian markets for music and home video. He discerns market trends not only concurrent with increasing levels of file-sharing and internet access, but also market trends concurrent with shifts in the dominant media format, such as those that occurred with the emergence of the CD as the dominant format. He notes that such shifts allow firms to increase their profits, since some consumers will repurchase the same product in a new media format, which he refers to as a "replacement effect."

Grassi places consumers in essentially the same theoretical framework as that of Gayer & Shy, however, the mathematical construction of his model is notably different. Like Gayer & Shy, he assumes that the demand for music comes from a uniformly distributed continuum of consumers. His indexation of these consumers by θ, however, is far more intuitive. Placing $\theta \in [0, 1]$ and introducing one recording available for purchase price p, which is set by a monopolist, consumers' utility is given $U = max\{\theta - p, 0\}$. Grassi then introduces a pirated imperfect substitute good, the utility from which is discounted by the factor β, and which costs consumers some price w, such that $U = max\{\theta - p, \beta\theta - w, 0\}$. The parameter w essentially represents the opportunity cost of illegal downloading as presented in Peitz & Waelbroek (2005), however, Grassi notes that a component of this opportunity cost is fear of litigation. This model omits the network effect on album sales presented by Gayer & Shy (2005), which was γN in their model. Also, where Gayer & Shy constructed a strict inequality, Grassi explores the "corner solutions," such as values of β that annihilate the market for legally purchased music.

Grassi also explores a number of extensions of Gayer & Shy's model. He places the sampling effect in a simple, two-period, inter-temporal framework, for example. The crucial aspect of this framework is the introduction of the parameter λ as the factor by which the demand for illegal downloads in period one increases the demand for CD's in the following period, where clearly $0 < \lambda < 1$. The model can be manipulated to show that, for $w = 0$, record label profits are not diminished by file-sharing if and only if $\lambda \geq 2\beta$. File-sharing, therefore, inevitably diminishes record label profits if the quality of pirate copies is greater than half that of CD's. Grassi also extends the model to include paid download sales of a digital-format good. This good is sold for price q, by the same firm that sells the CD, and yields utility discounted by the factor

α. The paid download clearly must be of higher quality than its illegal equivalent in order to sell, so $\alpha > \beta$, and consumer utility is now given by:

$$U = max\{\theta - p, \alpha\theta - q, \beta\theta - w, 0\} \tag{2.5}$$

Finally, Grassi integrates live performance into the model. The notable aspect of this framework is the parameter ϕ, which is virtually identical to the δ parameter in Gayer & Shy (2006). ϕ is the factor by which the demand for recorded music affects the demand for live music, and seems to equal the exact likelihood that a consumer who has heard an artist's music would choose to attend a free live performance by the artist. In the absence of file-sharing, the demand for the CD is simply $D_{cd}(p) = 1 - p$, and the demand for a live performance, with ticket price p_c, is given by $D_c(p_c) = \phi(1 - p) - p_c$. The model can be manipulated to show that, in the presence of file sharing, the marginal consumer who is indifferent between the pirate copy and nothing at all is located at the point $\frac{w}{\beta}$, and the demand for live performance is given by:

$$D_c(p_c) = \phi(1 - \frac{w}{\beta}) - p_c \tag{2.6}$$

In light of this previous result, Grassi responds to Gayer & Shy's proposed artist-label conflict. He notes that this conflict existed before the advent of digital media and argues that it is, in fact, eliminated in the presence of file-sharing. His basis for this argument is that, as previously demonstrated, the demand for live performance is dependent upon CD pricing only in the absence of file-sharing. This argument, however, is only somewhat valid. It is true that record labels' monopolistic pricing of CD's has long constrained the demand for live performance and, therefore, prevented most artists from maximizing their own profits. Furthermore, it is relatively intuitive that the availability of low-quality pirate copies of recordings should diminish the impact of CD pricing on the aggregate of consumers exposed to an artist's music, which should, in turn, diminish the extent to which musical artist's and record label's incentives conflict. Grassi, however, neglects crucial implications of his model. Most notably, when pirate copies are available, live performance profits in his model are maximized by minimizing $\frac{w}{\beta}$, the ratio of the costs of obtaining pirate recordings to the quality of these recordings. Grassi demonstrated earlier in the paper that, when pirate copies are available, $D_{cd}(p) = 1 - \frac{p-w}{1-\beta}$. The respective demands for CD's and live performance, therefore, vary inversely with values of p and β and, hence, the artist-label conflict is reinstated if musical artists can control the quality and availability of free digital copies of their recordings.

In the case of streaming music, pirated copies of most recordings are available for streaming via YouTube. Via myspace, however, one may stream official copies of recordings, though only those chosen at the artist's discretion, which unlike pirated copies are guaranteed to be full quality, properly labeled, and easily searchable. artists, therefore, may decrease consumers' opportunity costs from searching and mislabels, as well as potentially increase the average quality of their free streaming music available online, and, therefore, diminish $\frac{w}{\beta}$ by offering a larger proportion of their catalog on MySpace. In accordance with Grassi's model, this should increase live performance revenues and overall profits for most artists, but diminish CD revenues and record label profits.

Returning to his introduction, though Grassi discusses the topics of the "replacement effect" and shifts in dominant media formats only briefly and in passing, further discussion of these topics is warranted. In the final paragraph of his paper, Grassi comments that "industries that sold the machines used by the pirates have increased their business, and for example the market of the MP3 players [sic] has been invented from nothing. Probably the 'big enemy' of the recording industry is not the final consumers, that occasionally can act as a pirate [sic], but the industries that are cannibalizing music market profits."[4]

Inferences

It appears that the major US record labels failed to anticipate the shift in dominant format from CDs to mp3s. It is logical that the RIAA observed such a strong correlation between CD sales displacement and piracy via Napster because, as the first major source of pirated digital media, Napster served as an excellent proxy for the unobserved shift in dominant formats. Clearly, if labels had not begun to offer music in newer formats as vinyl or cassette tapes grew obsolete, they would have experienced a similar sales displacement. Furthermore, physical media formats can deteriorate over time, losing quality or breaking entirely. Not only do such phenomena increase potential profits from the replacement effect, but they will no longer occur with new digital formats. The advent of digital media was an opportunity for record labels to increase their profits, however, they failed to capitalize on the replacement effect. Furthermore, they allowed the Apple corporation to dominate the emerging market and gradually absorb their target consumers: music purchasers.

[4] Grassi (2007), pg. 23

2.3 Trends in the Live Performance Subsector

2.3.1 Mortimer & Sorensen (2005) *Supply Responses to Digital Distribution: Recorded Music and Live Performance*

Mortimer and Sorensen (2005) carried out an empirical analysis of the effects of file-sharing on the supply of live music. Their data set contained CD sales and touring data for the nearly 2,000 musical artists for which both Pollstar and Nielsen SoundScan had data over the period 1993-2002. They verified the integrity of this data set using data for RIAA album certifications as well as data drawn from artists' discographies. They also compared their sample to the set of all musical artists listed on Pollstar and found their sample to be unbiased, despite excluding all artists for which SoundScan data was not available.

Upon first examination of the data, the authors concluded that both concert revenues and the number of musical artists performing concerts have increased since the introduction of file sharing. From this conclusion, they proceed to examine whether this trend was, in fact, more pronounced in market segments in which file-sharing was more likely to have had a significant impact. Their final conclusion is that file-sharing has increased the profitability of live performances.

2.3.2 Live Performance

Krueger (2004) outlines a notably different set of trends in the live performance subsector from those of Mortimer & Sorensen. Live performance revenues have continued to increase and comprise the largest share of alternative revenues. While these revenues are increasing, however, ticket sales for live performances are actually falling. Over the period 1981-2003, concert ticket prices increased faster than inflation, and significantly so from 1996 on. Krueger develops his "Bowie theory" to explain these trends.

Krueger proposes that it was once more profitable for musical artists to underprice live performances in order to generate demand for album sales. Consumer sampling, however, made possible by new technologies, has significantly eroded the complementary nature of live performance to album sales. As a result, Krueger concludes, artists have driven the trends in the subsector by choosing to price live performances more like monopolists. It is this monopolistic pricing that has resulted in artists better maximizing their profits, but has also diminished the quantity of tickets sold to live performances.

2.4 Summary

Overall, the literature suggests that musical artists stand to benefit from the drastic changes in the music industry. Even record labels stand to yield higher profits with piracy than without it, if they adopt the proper business model. New technologies and increasing internet access continue to allow an increasing number of music consumers worldwide the opportunity to sample before they buy. Furthermore, firms can increasingly control the nature of this sampling. At the industry level, revenues not only from CD's, but from recorded music in general, continue to fall. Meanwhile, revenues from an increasing array of complementary products continue to increase. With some economists questioning the current state of copyright laws, and the record label now virtually obsolete, one would expect the emergence of new business models in which musical artists utilize recorded music primarily to generate complementary revenues.

Chapter 3

Theory

3.1 The Artist Profit Function

Let's begin by considering the artist profit function. Suppose an artist has four potential sources of revenue: the sales of CD-format recordings, the sales of downloads of mp3-format recordings, the sales of tickets to live performances by the artist, and the sales of cell phone ringtones based on recordings by the artist (master tones). Let p_{cd}, p_{mp3}, p_{tic} and p_{rt} denote the prices of CDs, mp3s, concert tickets and master tones, respectively, where p_{tic} may be manipulated at the artist's discretion and all other prices are constants fixed by industry standards. Suppose that music can be recorded and distributed digitally at no cost. In this case, mp3s and ringtones may be sold at no cost to the artist. In other words, the quantities of mp3s and master tones sold have no effect on the costs incurred by the artist. We shall consider the production and distribution of CDs to have no fixed costs, only the marginal costs of the physical distribution of each CD. Let this marginal cost be denoted c_{cd}, a fixed constant. Artists incur a variety of costs from supplying live performances, most notably those costs associated with each individual performance. For each individual performance supplied, an artist incurs variable costs of travel, accommodations, and rental of the performance venue. Let the sum total of these costs associated with live performance be denoted C_{TotLP}.

Now, let D_{cd}, D_{mp3}, D_{tic} and D_{rt} denote the total demand for CDs, mp3 downloads, live performance tickets and ringtones, respectively. Then the artist's total profit function is given by:

$$\pi = p_{cd}D_{cd} - c_{cd} + p_{mp3}D_{mp3} + p_{tic}D_{tic} - C_{TotLP} + p_{rt}D_{rt} \tag{3.1}$$

3.2 Demand Functions

At this point, we return to the work of Grassi (2007) in order to derive the demand functions on which the artist profit function is dependent. Assume that an artist has one album of recorded music, which is made available to consumers in hardcopy CD, downloadable mp3, and streaming formats. The demand for these goods comes from a continuum of consumers with a uniform distribution, which is illustrated below. This distribution is indexed by $\theta \in [0,1]$, according to a decreasing preference for consumption.

3.2.1 Consumers

Suppose that every consumer in the world were given the CD for free, and each gave a completely accurate rating of his or her enjoyment of the CD on a scale from zero to one. Suppose the consumers were then placed in a line facing left (towards 0), such that every consumer rated the CD more highly than the consumer behind him or her. This would be a (discrete) representation of a continuum of consumers indexed by $\theta \in [0,1]$ according to decreasing preference. Each consumer represents a point on the continuum. The value of θ corresponding to this point is equal to this consumer's rating of the CD. Now, θ indexes a "uniform distribution." The implications of this uniformity, returning to the image of the world's consumers in a line (facing right), are that the rating of the CD by the first consumer in line is 1, and of the last is 0. Furthermore, as one proceeds down the line from each consumer to the next, the corresponding ratings decrease by a constant amount. In other words, no matter which consumer one chooses to examine, the difference between this consumer's rating and that of the consumer in front of him or her, will be the same.

Consumer Utility

Introducing all three goods and their associated prices and costs, we adapt equation 2.5 to the notation of this theory section. Hence, consumers' utility is given by the maximum value of:

$$U_{MAX}(\theta - p_{cd}, \beta_1\theta - p_{mp3}, \beta_2\theta - k, 0) \tag{3.2}$$

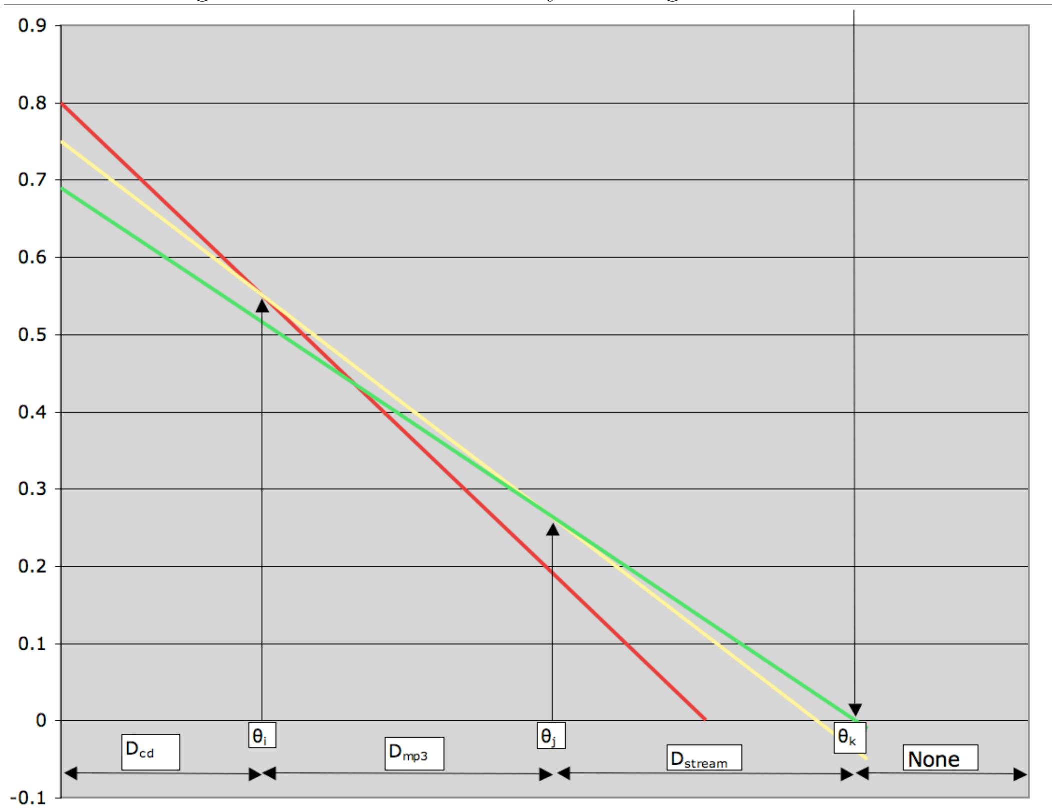
where the first term represents the potential utility from consumption of the CD, the second from consumption of the mp3, the third from streaming consumption of the album, and the final from non-consumption.

Clearly p_{cd} and p_{mp3} denote the price of the CD and the mp3, respectively. Hence, consumer i's potential utility resulting from consumption of the CD is $\theta_i - p_{cd}$. Now, β_1 and β_2 are discount factors, bounded by zero and one, which indicate the fraction of the potential raw utility from consumption of the CD that consumers may enjoy by consuming the album in mp3 or streaming format, respectively. Put simply and ignoring prices for a moment, if $\beta_1 = \frac{3}{4}$, then consumers enjoy consuming the album in mp3 format seventy-five percent as much as they would consuming the album in CD format. Hence, consumer i's potential utility from consuming the mp3 is given by $\beta_1\theta_i - p_{mp3}$. The third term, which represents the potential utility from consumption of the album in streaming format, is essentially the same, except for the introduction of the variable k in lieu of a price variable.

Streaming music is technically free, or is offered at no price, however there are costs to the consumer associated with the consumption of streaming music. Most simply, streaming music offered on YouTube is user uploaded and, hence, may be mislabeled. As a result, a consumer may waste time searching for a desired track. Hence, k represents consumers' opportunity/search costs from finding their desired music in streaming format. Unlike YouTube, streaming music on MySpace is properly labeled, though artists must choose to upload and offer each track individually. An artist, therefore can reduce k by offering more music in streaming format via MySpace.

Returning to the consumer utility function, each consumer chooses only one form of consumption, that which corresponds to the term of his or her utility function that has the greatest value. The subsets of consumers that choose each form of consumption, i.e., the demands for each good, partition the continuum of consumers. The border between any two such subsets that are adjacent is the marginal consumer who is indifferent between the two corresponding forms of consumption. Figure 3.1 illustrates, for an arbitrary set of parameters, the potential utility from each form of consumption across the continuum of consumers, as well as where the marginal consumers lie.

Figure 3.1: Consumer Utility & Marginal Consumers



Marginal Consumers

The marginal consumer indifferent between the CD and the mp3 is located at the point:

$$\theta_i - p_{cd} = \beta_1 \theta_i - p_{mp3} \Rightarrow \theta_i = \frac{p_{cd} - p_{mp3}}{1 - \beta_1} \tag{3.2a}$$

The marginal consumer indifferent between the mp3 and streaming is located at the point:

$$\beta_1 \theta_j - p_{mp3} = \beta_2 \theta_j - k \Rightarrow \theta_j = \frac{p_{mp3} - k}{\beta_1 - \beta_2} \tag{3.2b}$$

And the marginal consumer indifferent between streaming and non-consumption is located at the point:

$$\beta_2 \theta_k - k = 0 \Rightarrow \theta_k = \frac{k}{\beta_2} \tag{3.2c}$$

Hence, the demand for CD's is given by:

$$D_{CD}(p_{cd}, p_{mp3}) = 1 - \theta_i = \frac{1 - \beta_1 - p_{cd} + p_{mp3}}{1 - \beta_1} \tag{3.3}$$

Similarly, the demand for mp3's is given by:

$$D_{mp3}(p_{mp3}, p_{cd}, k) = \theta_i - \theta_j = \frac{(\beta_1 - \beta_2)p_{cd} + (\beta_2 - 1)p_{mp3} + (1 - \beta_1)k}{(1 - \beta_1)(\beta_1 - \beta_2)} \tag{3.4}$$

And the demand for streaming is given by:

$$D_{stream}(k, p_{mp3}) = \theta_j - \theta_k = \frac{\beta_2 p_{mp3} - \beta_1 k}{(\beta_2)(\beta_1 - \beta_2)} \tag{3.5}$$

3.2.2 Live Performance & Ring Tones

Suppose that some fraction ϕ of consumers who listen to the music in any format choose to attend a live performance at cost p_{tic}. We derive our equivalent of equation 2.6, and the demand for concert tickets is given by:

$$D_{tic}(p_{tic}) = \phi_1(D_{CD} + D_{mp3} + D_{stream}) - p_{tic} = \phi_1(1 - \frac{k}{\beta_2}) - p_{tic} \tag{3.6}$$

We next turn to the demand for ring tones. Like concert tickets, ring tones are not a substitute for, but do have a complementary relationship with, recorded music. As such, we treat the demand for ring tones in the same fashion as the demand for concert tickets, and assume it to be given by:

$$D_{rt}(p_{rt}) = \phi_2(D_{CD} + D_{mp3} + D_{stream}) - p_r = \phi_2(1 - \frac{k}{\beta_2}) - p_{rt} \tag{3.7}$$

Ring tone sales, in this model, are equivalent to any number of sources of revenue that are complementary to, and not substitutable for, recorded music. Even a source such as advertising revenue, which is not generated directly by consumers, may be directly related to consumer exposure, and appropriately treated in this manner.

3.3 Behavior of the Model

We may now substitute the demand functions derived above into equation 3.1, which yields:

$$\begin{aligned} \pi \;=\; & p_{cd}\left(\left(\frac{1 - \beta_1 - p_{cd} + p_{mp3}}{1 - \beta_1}\right) - c_{cd}\right) + p_{mp3}\left(\frac{(\beta_1 - \beta_2)p_{cd} + (\beta_2 - 1)p_{mp3} + (1 - \beta_1)k}{(1 - \beta_1)(\beta_1 - \beta_2)}\right) \\ + \; & \left(p_{tic}\left(\phi_1\left(1 - \frac{k}{\beta_2}\right) - p_{tic}\right) - C_{TotLP}\right) + p_{rt}\left(\phi_2\left(1 - \frac{k}{\beta_2}\right) - p_{rt}\right) \end{aligned} \tag{3.8}$$

In order to examine the manner in which artist profits vary with the opportunity costs of obtaining their music in a free streaming format, we take the partial derivative of π with respect to k, which yields:

$$\frac{\partial \pi}{\partial k} = \frac{p_{mp3}}{(\beta_1 - \beta_2)} - \frac{\phi_1 p_{tic} + \phi_2 p_{mt}}{\beta_2} \tag{3.9}$$

Here k represents the opportunity cost of obtaining free streaming music, which declines with the availability of free streaming music. Note that $\frac{\partial^2 \pi}{\partial k^2} = 0$. Hence, artist profits are either unaffected by variations in k, or are strictly increasing or decreasing with k. Profit maximizing artists, therefore, depending on the nature of their revenues, should act either to minimize or to maximize consumers' opportunity costs of streaming. Returning to 3.9,

$$\frac{p_{mp3}}{(\beta_1 - \beta_2)} <,> \; or = \frac{\phi_1 p_{tic} + \phi_2 p_{mt}}{\beta_2}$$

governs whether artists should minimize, maximize, or are unaffected by the opportunity costs of streaming, respectively. The natures of the numerators above are completely intuitive, if one considers that streaming music is a complement to concert tickets and ring tones, but a substitute for mp3 downloads. The likelihood that ease of streaming will stimulate revenue is increasing with the prices of concert tickets and ring tones, as well as with the proportions of consumers who choose to purchase concert tickets or ring tones as a result of exposure to an artist's music, but decreasing with the price of mp3s. The denominators, however, are less intuitive. Apparently, the likelihood that ease of streaming will stimulate profits decreases with the quality of the streaming music, but increases with the quality of mp3 downloads $(\beta_1 - \beta_2)$. While many of the variables in this model are arguably beyond the control of artists, is is likely that artists may act to set ticket prices in order to maximize their profits. The partial derivative of the artist profit function with respect to p_{tic} is:

$$\frac{\partial \pi}{\partial p_{tic}} = \phi \left(1 - \frac{k}{\beta_2}\right) - 2p_{tic} \tag{3.10}$$

and hence the profit maximizing value of p_{tic} is:

$$\frac{\partial \pi}{\partial p_{tic}} = 0 \Rightarrow \frac{\phi}{2}\left(\frac{k}{\beta_2} - 1\right) \tag{3.10a}$$

Substituting the above result into equation 3.8 yields:

$$\begin{aligned}\pi &= p_{cd}\left(\left(\frac{1-\beta_1-p_{cd}+p_{mp3}}{1-\beta_1}\right)-c_{cd}\right)+p_{mp3}\left(\frac{(\beta_1-\beta_2)p_{cd}+(\beta_2-1)p_{mp3}+(1-\beta_1)k}{(1-\beta_1)(\beta_1-\beta_2)}\right)\\ &+ \left(\frac{3}{4}\phi_1^2\left(1-\frac{k}{\beta_2}\right)^2-C_{TotLP}\right)+p_{rt}\left(\phi_2\left(1-\frac{k}{\beta_2}\right)-p_{rt}\right)\end{aligned} \tag{3.11}$$

The partial derivative of π with respect to k is now given by:

$$\frac{\partial\pi}{\partial k}=\frac{p_{mp3}}{(\beta_1-\beta_2)}+\frac{3\phi_1^2}{2\beta_2}\left(\frac{k}{\beta_2}-1\right)-\frac{\phi_2 p_{mt}}{\beta_2} \tag{3.12}$$

Note that

$$\frac{\partial\pi^2}{\partial^2 k}=\frac{3\phi_1^2}{2\beta_2^2}>0$$

Hence, with the exception of artists that experience no increase in ticket sales from consumer exposure to their music ($\phi_1 = 0$), setting 3.13 equal to zero and solving for k,

$$k_{min}=\beta_2\left(\left(\frac{2\beta_2}{3\phi_1^2}\left(\frac{\phi_2 p_{mt}}{\beta_2}-\frac{p_{mp3}}{(\beta_1-\beta_2)}\right)\right)+1\right) \tag{3.13}$$

yields a profit minimizing value. In fact, the growth of artist profits is infinite and explosive with k. This is because neither k, nor mp3 download sales, have any maximum bound, which is neither realistic or appropriate. Again, profit maximizing artists should act to minimize or maximize, or are unaffected by, the opportunity costs of streaming. This choice is determined by the relationship between the artist's profits under $k = 0$ compared to those under the maximum possible value of k.

3.3.1 Conclusions & Critiques

Overall, the behavior of the model supports the hypotheses of this thesis. The incentive for profit-maximizing artists to provide greater ease of streaming increases with their revenue share from touring. The only artists with an incentive to maximize consumers' opportunity costs from streaming are those with large revenue shares from mp3 sales and/or those who experience little increase in ticket sales from consumer exposure to their music. The inclusion of network effects in this model would have served only to increase artists' incentive to provide streaming ease, as would have the inclusion of sampling effects. Furthermore, the inclusion of a sampling effect that served to stimulate mp3 download sales would have served to diminish or completely annihilate the subsection of profit-maximizing artists with the incentive to diminish ease of streaming. It is the opinion of the author that virtually all consistently touring bands with a relatively young, or internet savvy, target audience stand to benefit from minimizing the opportunity costs of streaming, if not necessarily maximizing its quality.

One weakness of this model is that streaming music does not affect album sales unless it annihilates the market for mp3s entirely. As discussed earlier in this thesis, the empirical literature regarding this matter is inconclusive. This uncertainty is definitely expected to influence artists' behavior and diminish their incentive to offer streaming music. The most simple means by which to integrate this uncertainty into the model would be the inclusion of an unknown fraction of consumers who choose only between CDs, streaming and non-consumption, i.e., those for whom $\beta_2 > \beta_1$. Generally, a more realistic construction of the model would treat the various β's and ϕ's not as constants, but as functions that that may be nonlinear, or may vary across consumers.

MySpace & YouTube

The MySpace and YouTube websites are valuable and significant venues through which to examine the theory presented in this thesis. MySpace offers musical acts both the freedom to control free offerings of their materials, as well as a venue for digital download sales. YouTube, on the other hand, offers neither, but does offer what MySpace does not; user-generated content that may violate copyright law. MySpace, furthermore, offers its digital download sales through its MySpace player. The fact that its sales apparatus, therefore, is nested within its free streaming player makes MySpace an valuable venue through which to test the hypothesis that streaming music can potentially serve as a complement to digital download sales via the sampling effect. Finally, the universality and standardization of these sites, in combination with the wide variety of statistics for number of "hits" and linked sites which they offer, make them ideal data sources from which to carry out empirical work, which until now appear to be untapped.

Chapter 4

Data & Methods

4.1 Sampling

The artists in the sample were chosen from two distinct Billboard charts, the "Hot Hip-Hop/R&B Airplay" chart, and the "Hot Modern Rock Tracks" chart. The former ranks 75 artists each week, while the latter ranks only 40. Using the "true random number generator" available on the Random.org website, 800 pairs of random numbers were drawn. The first number in the pair was chosen between 1 and 52, indicating the chart corresponding to a specific week during the previous year. For half of the number pairs the second number was chosen between 1 and 40, corresponding to a rank on the rock airplay chart. For the other half of the number pairs, the second number was chosen between 1 and 75, corresponding to a rank on the hip-hop airplay chart. These 800 number pairs were sorted by chart, then week, then rank, and the artist corresponding to each randomly chosen week - rank pair was manually located and copied to a list. This list was then filtered for duplicates, resulting in a sample of just under 300 artists, with approximately half of these artists having been drawn from each chart.

4.2 Data Collection and Cleaning

The data for the regression were gathered from five online sources: PollstarPro, Billboard.biz, the RIAA Gold and Platinum searchable database,[1] MySpace.com and YouTube.com. Computer programs, commonly referred to as bots, were written to cull data from the source code of four of these five sources. The data from PollstarPro

[1] http://www.riaa.com/goldandplatinumdata.php?table=SEARCH

were gathered manually. The raw data required extensive cleaning and aggregating before summary data were generated using Microsoft Excel pivot tables.

4.2.1 RIAA Data

The RIAA provides a free online searchable database of all historical gold and platinum certifications. A bot was designed to enter each individual artist's name in the *search* field and return a spreadsheet containing the results of this search. Each line in this spreadsheet indicates the certification, such as Gold, Platinum, 2 x Platinum, etc., of an album or single corresponding to the artist. The adjacent columns indicate the corresponding format (album or single), song or album title, and date of the certification. Due to an unfixable problem with this bot, the first data point for each artist had to be manually collected and entered into each spreadsheet. These spreadsheets then had to be manually culled for inappropriate data points, such as a certification for the artist "Nelly Furtado" returned by a search for the artist "Nelly." Such data points did not correspond to the artist of interest and were therefore removed. All certifications corresponding to a format other than single or album, such as music video, were also removed. Manual online searches were executed in order to verify that every artist, for which the bots had returned no data, had not, in fact, received any certifications. Furthermore, each individual certification was missing its *type* (Standard, Digital or Master Tone) and this data had to be collected manually.

Once cleaned, the spreadsheets were then aggregated into one single spreadsheet. This spreadsheet was sorted by RIAA certification, and a column was generated containing the number of sales corresponding to each level of certification; gold = half a million, Platinum = one million, 2 x Platinum = two million, etc. The data were then sorted by title, format, type and date. Recurrences were flagged with an indicator and previous sales were subtracted from the levels of sales indicated by recurrences. For example, if an album was certified gold in 2006, then platinum in 2007, this latter certification would be first flagged. From the one million sales corresponding to the recurrent certification, the half a million sales corresponding to the previous gold certification would be subtracted, yielding half a million sales.

4.2.2 MySpace Data

Another bot was designed to search the music category of MySpace.com for each artist's name and then navigate to the artist's page. The bot then culled the page's source code for the code corresponding to the MySpace music player, and output a "dump" of code in a single file for each artist. This "dump" contained the list of tracks available on the player and their corresponding numbers of total hits and downloads. The first step in treating this data was to properly import it into Microsoft Excel. Titles containing certain characters, such as parentheses or dashes, would incorrectly trigger Excel to split the title into two columns. Hence, the data had to be manually culled for such titles and the error corrected.

The next problem was posed by duplicate track listings. The data was sorted by artist, then by total hits, and then by total downloads. In a number of cases, two or more tracks with similar titles had identical hit counts. These duplicate tracks usually took the form of 'Song A - Playlist A' and 'Song A - Playlist B' or, particularly for hip-hop artists, 'Song A (censored)' and 'Song A (uncensored).' This particular aspect of the MySpace data was the most peculiar aspect of the whole data set. Since duplicate songs listed identical hit counts, it would seem that such songs had a single corresponding hit count, which increased when either of the tracks was clicked. The sum total hit count for each artist, however, listed on the top of the MySpace player, reflected that the hit counts for such duplicate tracks were, in fact, counted twice. Presumably, this method of calculating artists' sum total hit counts reflects an error in the MySpace player's code. This error, however, has an implication that may be worth noting. Hip-hop artists in the sample tended to have more duplicate tracks than rock artists, due to their preponderance of censored and uncensored versions of the same title. Hence, the sum hit count listed on the MySpace player generally inflates the sum hit count of hip-hop artists. Duplicate tracks on MySpace exhibited another peculiarity. Their download counts were generally identical, however in many cases they were not. The MySpace data was cleaned by deleting all but one data point from each set of duplicate tracks and retaining the highest number of downloads corresponding to that set.

4.2.3 YouTube Data

The next bot was designed to execute a search on YouTube for each artist's name. The bot output a spreadsheet for each artist, containing the top 25 results returned by this search, with corresponding video titles and hit counts. This data also required extensive manual cleaning. Consider the search for the artist "10 years after." Of the top 25 videos returned by this search, one was titled "10 years after Matthew Shepard's murder." Clearly this video does not pertain to the artist "10 years after," and was therefore deleted from the data. Every datapoint was manually verified in this fashion. After examination of the cleaned data, the data was reduced to the top 15 videos for each artist, according to hit count.

4.2.4 PollStarPro Data

The free online searchable database of concert dates, Pollstar, provides a paid online service, PollstarPro. This service allows users to search an extensive database of concert records dating as early as 1999. For each artist, a manual search for the artist's name was executed. Such searches direct the user to a summary page of the artist's tour history, at the center of which is a list of all of this artist's historical tour dates. Next to each date is listed the corresponding city and venue, as well as an otherwise empty column, in which an "s" may indicate that on said date, the artist played as a supporting act. In another otherwise empty column, a "bxo," which stands for box office report, may indicate that, for this tour date, the artist reported corresponding box office statistics. These statistics include the number of tickets sold and the gross receipts from ticket sales. In addition to tour dates, each artist's PollstarPro page contains a small table of summary statistics for tour dates over the previous 36 months. These summary statistics include the artist's number of headline dates, number of box office reports, and the average number of tickets sold and average gross receipts for all shows with box office reports.

For each artist, the tour date history and summary statistic data were manually copied and pasted into an Excel spreadsheet. Next, manual searches for venues by starting letter were executed, on PollstarPro, and the capacity of each venue was copied into a new spreadsheet. This process was repeated until a spreadsheet had been generated that contained the name of every venue contained in the PollstarPro database and its corresponding city and capacity. This spreadsheet was aligned with the master spreadsheet of all artists' tour dates, sorted by venue and city, and the capacity corresponding to each venue was copied into a new column of the master spreadsheet. This process provided venue capacities for approximately 75% of the tour dates in the data set. Manual Google searches for official venue websites eventually brought this statistic above 80%. Next, each of over 40,000 tour dates were manually inspected and cleaned. For each tour date, a dummy variable equal to 1 or 0, respectively, was generated to indicate whether said tour date was inside the US and Canada, or at a foreign location. Manual Google searches were executed to rectify confusions regarding venues with multiple rooms of varying capacities. The data was also sorted by venue, then date, and manually inspected for venues with clustered dates. In most cases, Google searches for such "venues" verified that they were, in fact, music festivals. The venue capacities listed for these festivals were generally estimated on the basis of weekend-long attendance. These values would have inflated the sum headline venue capacities of artists that performed at more festivals, and were therefore removed as inappropriate. Instead, an "s" was added in the corresponding "support" column, in order to treat each such observation as akin to a date as a supporting act. This treatment was appropriate because festival performances, like supporting performances, generate exposure, but the capacities of these events are not accurate representations of an artist's "draw."

4.2.5 Billboard Data

The final bot was written to execute searches of the chart archives available from Billboard.biz. By artist name, the bot searched the *Hot Hip-Hop/R&B Airplay* and the *Hot Modern Rock Tracks* singles charts, as appropriate. For each individual artist in the sample, with the exception of quite a few artists for which it failed, the bot output an individual spreadsheet with an observation for each appearance of a song by the artist on any week's chart since 1975. Each observation listed, in three columns, the song title, song rank and chart week. These spreadsheets were manually cut and pasted into an aggregate spreadsheet with two additional columns indicating chart and artist name.

The aggregate spreadsheet was next sorted by artist, then song title. The song title column was exported as a new file and run through a *python* routine designed to indicate with a 1, in a new column, each row different from that above it, and all others with a zero. This new column served to indicate one instance of each song, such that it could be summed to determine an artist's total number of songs, and was inserted into the aggregate spreadsheet.

Next, a new column was generated equal to each rank subtracted from 76 for hip-hop artists and from 41 for rock artists. These values of the "inverted rank" were now such that 1 now represented the lowest rank, and 75 or 40 the highest, depending on the chart. Values of the inverted rank normalized between 0 and 1 were then generated in a new column by dividing the inverted rank by 75 for hip-hop artists and by 40 for rock artists. Two final columns were then generated, one containing the values of the inverted and normalized ranks squared, and the other containing the square roots of these values.

4.2.6 Re-evaluation of the Sample

The artists in the sample were then re-examined. Many artists were missing data from a specific source due to a bot failure. When possible, this data was manually collected. Otherwise, these artists were removed from the sample. Similarly, all artists not listed in the PollstarPro database were removed from the sample. Two artists had to be removed from the sample due to typographical errors resulting in bot failure. The PollstarPro data was examined, and any artist that had submitted no box office reports, or had submitted only one report out of more than fifteen headlines, was removed from the sample. Similarly, the YouTube data was examined, and all artists, for which the YouTube search had returned less than 15 relevant videos, were removed from the sample. Finally, two artists, “Janet” and “Bow Wow,” formerly “Janet Jackson” and“Lil’ Bow Wow,” were removed from the sample due to their confounding changes in artist names. The remaining sample contained 90 artists: 26 hip-hop artists and 64 rock artists.

4.3 Data Treatment

4.3.1 Period Structure

Once the data had been cleaned, it was then grouped by “period.” Each spreadsheet of raw data was sorted by date, and an empty column was created to contain the designated period number corresponding to each date. Period 0 was designated to be the 36 months preceding data collection, 03/01/06 - 03/01/09. Period 1 was designated to be the 12 months preceding period 0, 03/01/05 - 03/01/06, period 2 to be the 12 months preceding period 1, 03/01/04 - 03/01/05, and so forth in this fashion. Two major factors motivated this period structure. First, period 0 was designed to be the 36 months from which the Pollstar BXO summary statistics were generated. Second, the period numbering was designed to be increasingly positive as one moves back through time, in the interest of the ease of applying decay rate functions, which will be discussed in a later section.

4.3.2 Non-Lagged Variables

Once the data had been grouped by period, Microsoft Excel pivot table reports were used to generate statistics grouped by artist and period. For lack of a better means, in order to format this output so it could be imported into STATA, all output data was manually cut and pasted, artist by artist and variable by variable, into a new spreadsheet. The period 0 and non period-based data was then used to generate variables corresponding to each artist. All such non-lagged variables are listed and described in the tables 4.1 - 4.2 below. Summary statistics for these variables are presented in Appendix A, tables A.1 - A.4.

Table 4.1: Non-Lagged Variables

Source Data: RIAA	
Name	**Description**
TotAlb0	The sum total sales of RIAA certified albums during period 0
NumAlb0	The total number of RIAA certified albums during period 0
TotDL0	The sum total sales of RIAA certified digital format downloads during period 0
NumDL0	The total number of tracks that earned digital download format RIAA certifications during period 0
TotSing0	The sum total sales of RIAA certified standard singles during period 0
NumSing0	The total number of standard singles that received RIAA certifications during period 0
TotMT0	The sum total sales of RIAA certified “master tone” (Ring Tone) sales during period 0, the only period during which the data set contained RIAA master tone certifications
Source data: Billboard	
Rock1	Indicator variable equal to 1 for rock artists and 0 for hip-hop artists, determined by the Billboard chart from which the artist was drawn
SumInvRank-BBHH0	The period 0 sum total of the inverted Billboard Hip-Hop chart ranks over all songs and all weeks
SumInvRank-BBRk0	The period 0 sum total of the inverted Billboard Rock chart ranks over all songs and weeks
SumInvRank-Norm0	The period 0 sum total of the inverted and normalized Billboard chart ranks over all songs and weeks
SumInvRank-Norm0sq	The period 0 sum total of the inverted, normalized and squared Billboard chart ranks
SumInvRank-Norm0sqrt-	The period 0 sum total of the square root of the inverted and normalized chart ranks
NumSongs-BB0	The total number of different songs to appear on the genre appropriate Billboard chart during period 0
AvgSong-WksBB0	The average number of weeks each song appeared on a Billboard chart during period 0, taken across all songs

Table 4.2: Non-Lagged Variables (cont.)

Source data: YouTube	
Name	**Description**
YTTotHits	The sum total of the streaming hit counts of the top 15 videos returned by a search for the artist's name
YT SDvHits	The standard deviation of the streaming hit counts of the top 15 videos returned by a search for the artist's name
Source data: MySpace	
MS TotHits Hits	The sum total of the streaming hit counts of all songs offered on the artist's MySpace player
MS SDv-Hits	The standard deviation of the streaming hit counts of all songs offered on the artist's MySpace player
MS Num-Trax	The total number of unique tracks offered for streaming on the artist's MySpace player
MS TotDL	The sum total number of downloads purchased through the artist's MySpace player
MS NumDL	The total number of unique tracks that had at least one instance of being purchased for downloads through the artist's MySpace player
Source data: PollstarPro	
NumHdl-US0	The total number of headline performances in the US or Canada during period 0
NumNoVC HdlUS0	The total number of headline performances in the US or Canada during period 0 for which the data set had no venue capacity recorded
NumSup-US0	The total number of non-headline performances in the US or Canada during period 0
NumFor-eign0	The total number of performances outside of the US and Canada during period 0
SumVC0	The period 0 venue capacity sum total for all headline performances in the US or Canada
AvgVC HdlUS0	SumVC0/NumHdlUS0
SumVC-adj100	$SumVC0 + (NumNoVCHdlUS0) * (AvgVCHdlUS0)$
SumVC adj75	SumVC0 + 0.75*(NumNoVC HdlUS0)*(AvgVC HdlUS0)
SumVC adj50	SumVC0 + 0.50*(NumNoVC HdlUS0)*(AvgVC HdlUS0)
NumBXO	The number of Pollstar box office reports over the 36 months preceding data collection
AvgTix	The average number of tickets to performances sold as calculated from Pollstar box office reports over the 36 months preceding data collection
PTicAvg	The average price of tickets to performances as calculated from the average ticket sales and average gross receipts corresponding to Pollstar box office reports over the 36 months preceding data collection

4.3.3 Lagged Variables and Decay Rates

The treatment of time-specific data is perhaps the most unique aspect of the regression model in this thesis. The model is technically cross-sectional, however much of the data is, in fact, longitudinal. The data on Billboard rankings, RIAA certifications and tour dates are week or date specific over the past four decades. The observations of this data, however, are far too sporadic to warrant time-series analysis in a sample of this size. Hence, the question arises, how should one construct variables in order to best represent longitudinal data in a cross-sectional model?

Consider a simplification of the regression model in which the RIAA certification data is the only explanatory variable. The dependent variable remains the sum of live performance ticket sales over the last three years. Now, suppose artist A received only a platinum certification in 2006, whereas artist B received platinum certifications in 2006 and 1996, and artist C received only a platinum certification in 1996. If one were to simply take the sum of the certifications with no respect for time, the magnitude of the explanatory variable would be equal for artists A and C, and twice as great for artist B. A linear model, therefore, would estimate artists A and C to have equal touring popularity in 2007, and artist B to have double that. This estimate seems highly unlikely. Given an average consumer who purchases, on a given date, a single CD by an artist, the probability that this consumer will purchase a ticket to see a performance by this artist is likely to be higher the following year than ten years later. Hence, a reliable model should estimate the 2007 touring popularity of artist A to be greater than that of artist C.

The converse of the above disregard for time-lapse would be the construction of a variable that represents only recent CD sales. In this case, the model would estimate the 2007 touring popularity of artist A and artist B to be equal, and that of artist C to be equal to that of an artist with no historical CD sales whatsoever. These estimates are no more realistic than the previous, though they represent the opposite end of the spectrum. Clearly an accurate model must lie in the middle ground, and the construction of variables from time-series data, in this model, must somehow adjust for time-lapse.

Decay Rates

Functional forms that adjust for time-lapse are common not only in economics, but throughout scientific modeling. In economics, the ‘discount rate’ adjusts the potential utility from future consumption relative to that of current consumption. In chemistry and physics, the 'half-life' functional form of decay rates was spawned by the exponential increase in the rate at which radioactive particles decay. It has since found a wide variety of applications.

The basis for such functional forms is the treatment of time as a series of periods, represented generally by the variable t. The value of t represents the number corresponding to the present time period, where $t = 0$ represents the time period during which particle decay began. Periods may be any constant length and, hence, $t = 1$ may indicate that a second, minute, hour, day, week, decade, or 8.6 seconds has elapsed since period 0. It should be apparent that this format is forward looking, so to speak, in that period 1 occurs later than period 0. Hence, discount and decay rates adjust for future time-lapse. The regression model being constructed, however, must adjust in the same fashion for past time lapsed from the present period $t = 0$. The period before $t = 0$ is conventionally notated as $t = -1$. This model, however, notates the period before the present as $t = 1$. As previously mentioned, this period structure was designed for the ease of applying decay rates. More specifically, this structure automatically inverts forward-looking decay rates about the y-axis or, in other words, inverts them with respect to time such that they become backwards-looking.

Functions Applied

The following list presents the decay functions applied to the values of the lagged data in this analysis, where x represents each lagged value, t, the corresponding period number, and y, the variable generated.

Lin1 $y = \sum_t \left[x \left(\frac{31 - 3t}{28} \right) \right]$

Lin2 $y = \sum_t \left[x \left(\frac{15 - t}{14} \right) \right]$

Lin3 $y = \sum_t \left[x \left(\frac{29 - t}{28} \right) \right]$

L1 $y = \sum_t \left[x \left(\frac{95 - 2t}{93} \right) \right]$

L2 $y = \sum_t \left[x \left(\frac{94 - t}{93} \right) \right]$

L3 $y = \sum_t \left[x \left(\frac{21 - t}{20} \right) \right]$

INV1 $y = \sum_t \left[x \left(\frac{1}{t} \right) \right]$

INV2 $y = \sum_t \left[x \left(\frac{4}{7t} + \frac{3}{7} \right) \right]$

INV3 $y = \sum_t \left[x \left(\frac{2}{7t} + \frac{5}{7} \right) \right]$

Cos1 $y = \sum_t \left[x \left(\frac{1}{3} \left(2 + \frac{\pi(t - 1)}{31} \right) \right) \right]$

These functions were used to generate the variables listed in Table 4.3. They were selected primarily based on the maximum value of t observed in the data to which they were applied. For example, the maximum value of t observed in the Pollstar data was 8. For $t = 8$, $Lin1 = \frac{3}{4}$, $Lin2 = \frac{1}{2}$, and $Lin1 = \frac{1}{4}$. These three functions were, therefore, selected for application to the Pollstar data in order to preserve a minimum of 75%, 50%, and 25% of lagged venue capacity observations, respectively.

Table 4.3: Lagged Variables

Function Applied	**Source Variable (x)**	**Variable Name (y)**
Lin1	SumAdj50	SumAdj50Lin1
	SumAdj75	SumAdj75Lin1
	SumAdj100	SumAdj100Lin1
Lin2	SumAdj50	SumAdj50Lin2
	SumAdj75	SumAdj75Lin2
	SumAdj100	SumAdj100Lin2
Lin3	SumAdj50	SumAdj50Lin3
	SumAdj75	SumAdj75Lin3
	SumAdj100	SumAdj100Lin3
L1	SumAlbSales	SumAlbSalesL1
	SumStSales	SumStSalesL1
L2	SumAlbSales	SumAlbSalesL2
	SumStSales	SumStSalesL2
L3	SumInverseRank	SumIRL3
	SumInverseRankNormSq	SumIRNSqL3
	SumInverseRankNormSqrt	SumIRNSqrtL3
INV1	SumAdj50	SumAdj50INV1
	SumAdj75	SumAdj75INV1
	SumAdj100	SumAdj100INV1
INV2	SumAdj50	SumAdj50INV2
	SumAdj75	SumAdj75INV2
	SumAdj100	SumAdj100INV2
Lin3	SumAdj50	SumAdj50INV3
	SumAdj75	SumAdj75INV3
	SumAdj100	SumAdj100INV3
Cos1	SumAlbSales	SumAlbSalesCos1

Chapter 5

Analysis

5.1 Pairwise Correlations

In order to construct the regression model in an informed fashion, the relationships between the individual variables to be used in the regression were first examined. Almost universally, the various groups of decayed or adjusted versions of the same variable exhibited relationships with the other variables in the model of equal significance and direction. Presented in Appendix B, Tables B.1 - B.4 are the basic pairwise correlations between the variables in the model, with only one variable chosen to represent each such group. Each correlation printed is significant at the 10% level, with a star indicating significance at the 5% level.

The data set exhibits many of the expected correlations of significance. Examine the top left of table B.1, column 1, rows 2-5, for example. The Billboard rock chart from which the data are drawn has only 40 positions, whereas the hip-hop chart has 75. It is of no surprise, therefore, that rock artists exhibited, with 5% significance, lesser numbers of songs and total song weeks than hip-hop artists. From this correlation, it follows that rock artists, in turn, also exhibited significantly lesser summed chart ranks, even with these ranks normalized.

Moving down this column, of greater interest is the fact that, during period 0, the rock artists in the sample, compared to their hip-hop counterparts, exhibited significantly greater numbers of period 0 support, foreign and US headline performances. In turn, these rock artists also exhibited greater period 0 sum venue capacities, though their average numbers of tickets sold did not exhibit significant deviation from the sample as a whole. Furthermore, the fact that rock artists had significantly more tour dates during period 0, but significantly lower ticket prices, exhibits basic microeconomic theory. Considering the losses that artists incur if they do not sufficiently fill the venues they rent, one may view sum venue capacity not merely as a supply quantity, but as a proxy for quantity exchanged.[1] Hence, the facts that rock artists in the sample exchanged a greater quantity of tickets but at a lower price go hand in hand. Furthermore, hip-hop artists in the sample exhibited greater levels of MySpace downloads and RIAA certified ring tone sales. Overall, these genre specific relationships of significance are not unrelated, but indicate a significant cross-genre variation in revenue structure.

5.2 Choices Among Families of Variables

The number of variables meant to be substituted for one another, such as the various decayed versions of the same variable, is so large that the behavior of these different variables in the variety of regressions that follow could not be rigorously explored and presented in any practical manner. The choices of which such variables to use in these regressions were not made at random, however. These choices were made based on the author's observations, after running a variety of regressions.

The period 0 sum venue capacity variable, for example, was initially replaced by *TotVCadj100* in an attempt to adjust for non-reported venue capacities. This adjusted variable represents artists' predicted sum period 0 venue capacity under the assumption that the capacities of non-reported venues are equal to those of artists' average performance. Regressions run using this adjusted variable, however, behaved extremely differently from those run using the unadjusted variable. Based on these differences, as well as intuition, the author concluded that the adjusted variable was biased to inflate the sum venue capacities of artists with greater numbers of period 0 performances, for which no capacity was reported. Particularly, because many of the venue capacity observations that were initially missing from the data set were filled using Google searches, observations that remained missing at the time of analysis were likely smaller venues. Hence, these venues were likely smaller than, not equal to, the average venues at which artists performed. For this reason, the variables *TotVCadj75* and *TotVCadj50* were generated in order to examine whether it might be more appropriate to assume that non-reported venues had capacities equal to 75% and 50%, respectively, of the average capacity of venues at which artists performed. The author concluded *TotVCadj50* to be most appropriate, though he acknowledges

[1]Note: This treatment is not contrary to the lack of a significant pairwise relationship between the genre indicator and AvgTix variables. The latter variable is constructed from Pollstar BXO data, which is strongly dependent on artists' reporting behavior and is, as a result, unreliable.

that this variable may still exhibit a bias, and that smaller levels of adjustment may be even more appropriate.

After similar examination of the lagged, decayed variables, the author selected certain variables for use in the regressions presented in this chapter. These variables did not always exhibit significant differences from their counterparts, however they were chosen for the following reasons. *SumAlbSalesL1* was chosen, rather than *SumAlbSalesL2*, for its greater level of decay. No difference in the behavior of *SumAlbSalesL1*, could be discerned from that of *SumAlbSalesCos1*, which had the same long term rate of decay, so the former was selected in the interest of simplicity. *SumAdj50Lin2* was chosen for its mid-level rate of decay, compared to its counterparts, as well as in the interest of simplicity, because no differences could be discerned among this family of lagged variables.

L3 was the only decay rate applied to the Billboard data. The intention of generating variables from the squared and square root values of the normalized rank was to test whether the effects of chart rank might not be linear, but might increase or be diminished as one moves up the charts. *SumIRNSqrt*, which inflates the significance of high chart rank relative to number of chart weeks, was deemed most appropriate in early regressions. This choice, however, was later deemed indefensible, and variables based on the unaltered values of inverse chart rank were used in later regressions.

5.3 Regressions

5.3.1 System of Equations: 3SLS Estimation

The demand for music and music related goods is dependent on the variety of ways in which consumers may be exposed to an artist's music. There is a complex endogenous relationship, particularly over time, between the sizes of the populations that consume an artist's music through live performance, streaming, radio airplay and purchase of recorded music. It was the original intention of this thesis to estimate a system of equations treating these population sizes as dependent variables. Stata, however, lacks the capacity to estimate non-linear systems of equations. Hence, the 3SLS estimation in this thesis could not properly treat dependent variables with non-fixed effects.

Many of the sources of exogenous variation in this system should not be treated as having fixed linear effects. For example, consider a regression treating period 0 sum venue capacity as the dependent variable, and the number of period 0 headline performances as an exogenous explanatory variable. Suppose this regression were to estimate a significant coefficient on the number of headline performances, equal to, say, five thousand. This estimation would imply that the period 0 sum venue capacity corresponding to the average artist in the sample increased by five thousand for every additional performance. Such an estimate would be too low for extremely popular artists and too high for unpopular artists. Hence, the estimation of the other dependent variables in the regression will be biased in order to compensate. Some other coefficient will serve to deflate the predicted values of the dependent variables for unpopular artists, and to inflate these predicted values for very popular artists. An appropriate treatment of this regression would be to estimate the relationship between a logarithmic value of sum venue capacity and the number of headline shows, such that the coefficient on the latter would represent a percent increase in sum venue capacity. Such treatment of the dependent variable, however, would be inappropriate for those independent variables, such as total album sales, with which it is expected to have a linear relationship.

Consider the bias of the number of headline shows, as illustrated above. Replacing this variable with the size of the average show switches this bias to inflate predictions for artists with a large number of headline shows, and to deflate predictions for artists with a small number of headline shows. Coefficient estimates that remain relatively consistent and significant as the biases of the model are shifted are more likely to be accurate. For this reason, four otherwise identical systems of equations were estimated, using as sources of exogenous variation the period 0 values of either the number of headline shows or the average venue capacity, as well as either the number of albums or the average album sales. These regressions were run over the entire sample, then over the rock subsample, the results of which are presented in Appendix C, Tables C.1 - C.8 and C.9 - C.16, respectively.

Analysis

These regression results are not particularly illuminating, however a few results are worth note. In particular, MySpace and YouTube total streaming hits are, at best, somewhat consistent in their significance as explanatory variables. With the exception of one sub-equation of one of the eight regressions, however, their coefficient estimates are consistent with regard to sign (+ or -). These estimates are potentially noteworthy because, while the coefficient estimates on MySpace are consistently positive, those on YouTube are consistently negative. Also, while extremely insignificant when regressed over the entire sample, when regressed over the rock subsample, lagged album sales has consistent negative coefficient estimates, significant at the 7% level, as an explanatory variable of MySpace total streaming hits. These estimates may indicate, among other hypotheses, that consumers who stream rock music from MySpace are of a particularly young age demographic, or that rock music has a longer "shelf-life" than hip-hop music, and is therefore more likely to experience streaming displacement as a result of previous album purchases. Tables 5.1 and 5.2 list the dependent variables for which the coefficient estimates remained consistent through all four regressions, with regard to direction and significance at the 5% level.

Table 5.1: Consistent 3SLS Estimates, Entire Sample

Dependent Variable	**Independent Variable**	**Sign**
TotAlb0	SumAdj5Lin2	+
SumVCadj50	SumAlbSalesL1	+
MSTotHits	MSSDvHits	+
	SumInvRankNormSqrt0	+
YTTotHits	YTSDvHits	+
	SumInvRankNormSqrt0	+
SumInvRankNormSqrt0	None	

Table 5.2: Consistent 3SLS Estimates, Rock Subsample

Dependent Variable	**Independent Variable**	**Sign**
TotAlb0	None	
SumVCadj50	SumAlbSalesL1	+
MSTotHits	MSSDvHits	+
	SumInvRankNormSqrt0	+
YTTotHits	YTSDvHits	+
	SumInvRankNormSqrt0	+
SumInvRankNormSqrt0	TotDL0	+

5.3.2 OLS Regressions on Sum Venue Capacity

Next, a variety of simple OLS regressions, treating adjusted period 0 sum venue capacity as the dependent variable, were run over the entire sample, then over the rock subsample, the results of which are presented in tables 5.3 and 5.4, respectively. The most notable of these results, which follow, are the consistently positive and significant coefficient estimates on MSTotHits. These estimates, in the presence of a variety of control variables, appear to exhibit the complementarity of streaming music and live performance hypothesized in this thesis.

Table 5.3: OLS Regression Results: Period 0 Sum Venue Capacity

	(1) SumVC0adj50	(2) SumVC0adj50	(3) SumVC0adj50	(4) SumVC0adj50
MS TotHits	0.00276**	0.00245***	0.00162***	0.00163***
	(2.96)	(3.78)	(3.92)	(3.94)
TotDL0	-0.0613**	-0.0631**	-0.0531**	-0.0503**
	(-2.96)	(-3.18)	(-2.84)	(-2.72)
Alb0Rk	-0.0361	-0.0301		
	(-0.94)	(-0.82)		
Alb0HH	-0.0488	-0.0493		
	(-0.68)	(-0.71)		
MS TotDL	-0.839			
	(-0.40)			
SumInvRankBBHH0	-18.21	-15.47		
	(-1.10)	(-1.05)		
SumInvRankBBRk0	163.8***	165.0***	168.3***	152.6***
	(5.24)	(5.84)	(6.05)	(6.48)
SumAdj.5Lin2	0.302*	0.248**	0.243**	0.222**
	(2.51)	(3.19)	(3.28)	(3.11)
AlbL1Rk	0.0637***	0.0665***	0.0609***	0.0623***
	(3.50)	(5.18)	(4.95)	(5.09)
AlbL1HH	0.00114			
	(0.04)			
SumStSingSalesL1	-0.0672			
	(-0.39)			
SumDLSalesPd1-2	-0.0791			
	(-0.77)			
SumIR L3	-22.11	-25.92*	-22.89*	-17.90*
	(-1.31)	(-2.35)	(-2.30)	(-2.05)
Rock=1	-113278.7	-105152.1	-51872.2	
	(-1.79)	(-1.83)	(-1.05)	
Constant	117937.5*	113342.6*	66662.3	35024.5
	(2.30)	(2.50)	(1.74)	(1.48)
Observations	89	89	89	89

t statistics in parentheses

Table 5.4: OLS Regression Results: Period 0 Sum Venue Capacity,Rock Subsample

	(1)	(2)	(3)
	SumVC0adj50	SumVC0adj50	SumVC0adj50
TotAlb0	-0.0626	-0.0592	
	(-1.26)	(-1.26)	
TotDL0	-0.0726*	-0.0719*	-0.0808**
	(-2.33)	(-2.37)	(-3.02)
MS TotHits	0.00339*	0.00336*	0.00372**
	(2.42)	(2.46)	(3.27)
MS TotDL	2.500	2.419	
	(0.55)	(0.55)	
SumInvRankNorm0	6397.0***	6416.6***	5681.0***
	(4.03)	(4.19)	(4.43)
SumAdj.5Lin2	0.318*	0.303**	0.235*
	(2.11)	(2.80)	(2.47)
SumAlbSalesL1	0.0613*	0.0613**	0.0590***
	(2.60)	(3.19)	(3.55)
SumStSingSalesL1	-0.0595		
	(-0.26)		
SumDLSalesPd1-2	-0.0253		
	(-0.19)		
SumIR L3	-27.93	-28.81	
	(-0.90)	(-0.97)	
Constant	-8494.3	-9198.1	-14361.7
	(-0.22)	(-0.25)	(-0.39)
Observations	61	61	61

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3.3 OLS Regressions on MySpace Data

OLS regressions treating $MSTotDL$ and $MSNumTrax$ as dependent variables were run using a variety of specifications. As previously discussed, MySpace is a prominent source of free streaming music, over which artists have complete control. With regard to informing artist behavior regarding multi-product profit maximization, therefore, these regressions were potentially the most significant aspect of the empirical work in this thesis. The regressions on $MSNumTrax$ failed to yield significant results. The regressions on $MSTotDL$, however, the results of which are presented in tables 5.5 - 5.7, were more successful.

After these OLS regressions were initially run, they were tested for heteroskedasticity using the classic "White test." The results of these tests indicated that all of these regressions were, in fact, heteroskedastic. One of the assumptions underlying the accuracy of OLS estimates is constant variance of the error term across observations, or homoskedasticity. Heteroskedasticity, or non-constant variance of the error term, represents a violation of this assumption. Heteroskedasticity does not affect the magnitude of coefficient estimates. It does, however, tend to deflate the variance and, therefore, standard error associated with variables, which in turn inflates their associated t-statistics and can cause insignificant coefficient estimates to appear significant. For this reason, the regression results presented use "robust standard errors," which correct for heteroskedasticity.

Regressions 1-3 presented in tables 5.5 and 5.6 illustrate how the dependent variables behave as others are removed. Regression 1 includes all independent variables deemed appropriate. Notice that *SumAdj.5Lin2* loses its significance in regression 2, once other insignificant variables are removed, and that when this variable is then removed in regression 3, *SumInvRankBBRk0* loses its significance as well. In regression 4, presented in table 5.7, only significant coefficient estimates remain.

Table 5.5: OLS Regression Results (1-3): MySpace Downloads

	(1) MS TotDL	(2) MS TotDL	(3) MS TotDL
TotAlb0	0.00300		
	(1.44)		
SumAlbSalesL1	0.00195*	0.00171**	0.000952**
	(2.48)	(2.68)	(3.10)
TotDL0	0.00359***	0.00398***	0.00428***
	(3.49)	(4.98)	(5.19)
SumDLSalesPd1-2	0.00399		
	(0.97)		
SumStSingSalesL1	-0.00350		
	(-0.72)		
SumVC0adj50	-0.00128		
	(-0.26)		
SumAdj5Lin2	-0.0111*	-0.00546	
	(-2.11)	(-1.18)	
SumInvRankBBHH0	-3.233*	-2.859**	-2.757*
	(-2.31)	(-2.64)	(-2.56)
SumInvRankBBRk0	-3.152*	-2.730*	-2.232
	(-2.25)	(-2.43)	(-1.86)
SumIRRkL3	-1.865*	-2.128**	-2.489**
	(-2.43)	(-3.05)	(-3.24)
SumIRHHL3	-1.028	-1.072	-1.066
	(-1.05)	(-1.60)	(-1.73)

Table 5.6: OLS Regression Results (1-3): MySpace Downloads (cont.)

YT TotHits	0.0000230		
	(0.48)		
MS NumTrax	8.253		
	(0.30)		
MS TotHits	0.000464***	0.000468***	0.000461***
	(4.57)	(5.81)	(5.80)
MS SDvHits	-0.00360***	-0.00350***	-0.00352***
	(-3.72)	(-5.34)	(-5.54)
Rock1	-10586.8*	-10333.6*	-10746.7*
	(-2.05)	(-2.20)	(-2.29)
Constant	12170.8*	12198.3**	12172.7**
	(2.16)	(2.78)	(2.79)
Observations	89	89	89

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.7: OLS Regression Results (4-5): MySpace Downloads

	(4)	(5)
	MS TotDL	MS TotDL
TotDL0	0.00420***	0.00398***
	(5.02)	(5.52)
SumInvRankBBHH0	-2.882**	-2.686*
	(-2.64)	(-2.17)
SumInvRankBBRk0	-3.309**	-3.039*
	(-2.80)	(-2.59)
MS TotHits	0.000447***	0.000378***
	(5.95)	(3.56)
MS SDvHits	-0.00324***	-0.00248**
	(-5.42)	(-3.01)
Rock1	-9346.6*	-9748.5*
	(-2.35)	(-2.43)
MS NumDL		150.3
		(1.97)
Constant	10850.5**	9140.5*
	(3.03)	(2.17)
Observations	89	89

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Regression 5 adds as an independent variable, $MSNumDL$, to the model in regression 4. Not all songs available for streaming via MySpace are available for download via MySpace. The number of tracks available for download via MySpace represents a significant omitted variable, on which $MSTotDL$ is expected to be strongly dependent. The number of different tracks downloaded from MySpace, $MSNumDL$, should serve as a strong proxy for the number of tracks available for download via MySpace. This proxy, however, is biased in that it is not representative of tracks that are available for download, but have never been downloaded. As a result, coefficient estimates on $MSNumDL$ will inevitably be biased upwards, which, in turn, creates a bias that decreases the coefficients estimated on other independent variables in the regression. Nonetheless, all of the coefficient estimates in regression 4 remained consistent with regard to significance and sign in the presence of the bias in regression

5. It should be noted that, while $MSNumDL$ was not found to be significant at the 5% level, its associated p-value was 0.052.

Interpretations of the Results

In each of the previous regressions, $MSTotHits$, the total number of streaming hits on MySpace, was found to be a significant positive predictor of the number of downloads from MySpace. Its coefficient estimated in regression 4 indicates that, on average, one download sale is associated with approximately every 2,240 (0.000447^{-1}) streaming hits. Furthermore, the number of tracks available for streaming via MySpace was not found to be as significant a predictor as the total number of streaming hits. $TotDL0$ was expected to be the strongest control variable in this regression and, as expected, its coefficient estimates were significant and positive. Regression 4 indicates that one download sale via MySpace is associated with approximately every 238 (0.00420^{-1}) overall download sales. The coefficient estimate on $Rock1$ in regression 4 indicates that hip-hop artists sell an average of approximately 9,347 more downloads via MySpace than equivalent rock artists. The coefficient estimates on the Billboard variables in regression 4, however, indicate that holding the peak rank on the hip-hop chart for one week is associated with an approximate decrease of 216 (2.882*75) download sales via MySpace, whereas holding the peak rank on the rock chart for one week is associated with an approximate decrease of only 132 (3.309*40) download sales via MySpace. Finally, the significant negative coefficient estimates on the standard deviation of MySpace streaming hits is relatively intuitive. The lower this standard deviation, the greater the concentration of streaming hits by individual track. Since only individual tracks are sold for download via MySpace, one would expect artists with a greater concentration of hits to be more likely to sell the tracks on which their streaming hits are concentrated, and to therefore exhibit greater overall sales than artists with more diffuse streaming hits.[2]

Implications

These regressions yielded the most significant findings of the empirical work presented in this thesis. While the lack of significance of the number of tracks offered for streaming on MySpace represents a lack of artists' control over their multi-product profits derived from sampling, the coefficient estimates on $MSTotHits$ tell a different story. These coefficient estimates may represent the potential of the sampling effect,

[2] The converse would be expected, were albums sold via download to be examined.

which has just begun to be tapped. Perhaps MySpace has succeeded in sufficiently nesting its download sales apparatus in the site's streaming offerings, such that the sampling effect can overwhelm the strong substitution effect, between free streaming music and paid digital downloads, and generate revenues. Furthermore, the negative coefficient estimates on the Billboard variables indicate that less popular, or more obscure, artists are more likely to sell downloads via MySpace. Interestingly, hip-hop artists, with whom greater levels of MySpace downloads are associated, exhibit this aforementioned trend to a greater degree. Perhaps the results of these regressions indicate that artists who lack mainstream promotion, particularly such hip-hop artists, are utilizing MySpace both as a promotional tool and as a venue for sales. Such an indication would exhibit the emergence of a new and unique business model, as hypothesized in this thesis, made possible by technological progress and the proliferation of the internet.

Chapter 6

Conclusion

Overall, there exist a wealth of economic principles evident in the music industry today that warrant examination. The extreme changes in the industry over the last decade, affected by technological progress and the proliferation of the internet, represent a veritable playground for econometricians. The implications of this thesis bear not only on the music industry, but on all industries dependent on entertainment media, if not on marketing and sales via the internet in general.

6.1 Recommendations for Further Research

The labor intensive nature of the data treatment necessary for the analysis presented in this thesis prevented the wealth of information contained in the data set from being fully utilized. This problem could clearly be overcome by greater resources. The labor intensity of the data treatment, however, was caused primarily by a single problem, the lack of any efficient means by which to aggregate Excel pivot table output into a form that could be exported to Stata. Hence, one recommendation of this thesis is that in similar empirical analysis in the future, the process of generating analyzable data from the raw data should be automated to a greater extent.

Ideally, the Billboard data should be analyzed to generate additional variables, such as debut, peak and exit ranks, and the number of weeks from debut to peak rank. As explored in the empirical work of Bradlow and Fader (2001), such variables reflect the "path" of a song through the charts over time, and are representative of many characteristics of songs and artists that were not captured by the way manner in which chart ranks were summed in this analysis. In fact, the Billboard.biz chart archive search engine allows searches for debut and peak weeks. Many variables, therefore, could be generated using far fewer observations of the Billboard charts than were used in this thesis. Conversely, however, the Billboard variables used in this thesis could not have been generated without the observations for every week's chart on which artists' songs appeared. Furthermore, as conceived in the treatment of the normalized Billboard chart rank in this thesis, the nature of the effects of chart rank warrants further research.

The box office report aspect of the PollstarPro data also warrants further analysis. Ideally, for each artist, the average number of tickets sold, as reported by PollstarPro, should be divided by the average venue capacity corresponding to box office reports, yielding a proxy variable for percentage concert attendance. The validity of this proxy variable could be assessed by comparing the average box office reported venue capacity to each artist's actual period 0 average venue capacity, as well as by examining the number and percent of headline shows with corresponding box office reports. Furthermore, the analysis in this thesis would have been well served to better control for artists' touring behavior. This would be a complicated task, but, to begin with, the geographic location and variation of performances represent valuable controls that could generated from the Pollstar data.

Another source of untapped potential contained in this data set are the song and album titles corresponding to the RIAA, Billboard, MySpace and YouTube data. Tracing the behavior of songs and albums across the data sources would likely yield interesting results, though such analysis would likely require extensive programming and labor.

It was the intention of this thesis to gather data regarding the pricing and availability of downloads on MySpace. In the end, this was not possible due to programming problems and time constraints. In light of the significant results that arose from the MySpace data, however, future collection and analysis of the data is definitely warranted.

Finally, this data set warrants the application of regression types other than those used in this thesis. Likely, the most accurate analysis would treat the data as panel data. Such analysis could incorporate additional data, such as indicator variables reflecting the emergence and shut down of Napster, variables reflecting the litigation of peer-to-peer users, or variables reflecting CD prices as indicated by record label price fixing settlements and the rise of iTunes and the $9.99 CD. Panel data techniques could also better estimate the significance of decay rates, which would have valuable bearing on economic concepts regarding consumer utility.

6.2 Results and Implications

The results of the empirical analysis generally support the hypotheses explored in this thesis. The sample exhibited significant divergence between rock and hip-hop artists. Particularly, rock artists offered a far greater supply of live performance, while hip-hop artists exhibited greater ring tone sales and download sales via MySpace, as well as a greater inverse correlation between radio airplay and MySpace downloads. Intuitively, this divergence seems indicative of a larger cross-genre variation in dominant revenue streams and business models. The sales of ring tones, as well as downloads via MySpace, are both recent technological advents. As such, hip-hop artists in the sample exhibit the departure from traditional revenue streams and business models that was a central hypothesis of this thesis. Furthermore, the data from MySpace bears strongly on two distinct hypotheses presented in this thesis. First, it supports the complementarity of free streaming music and live performance. Second, while the data overall did not bear particularly on the substitutability of free streaming music and digital downloads, the MySpace data, in particular, yielded a far more significant result.

The significant positive correlation between MySpace streaming hits and paid downloads indicates that free streaming music, if properly harnessed, can serve to stimulate paid downloads. This is a very pleasant result. The average overall period 0 download sales of artists in the sample, according to the RIAA data, was just under 450,000. The average period 0 sales of downloads via MySpace by artists in the sample was just under 7,500, comprising roughly 1.7% of overall download sales. Hence, MySpace appears to be a relatively obscure source of paid digital downloads. It seems safe to assume that consumers do not visit MySpace for the express purpose of being diverted to Amazon.com in order to purchase music.[1] On the other hand, MySpace does not at all appear to be an obscure source of streaming music. The average number of total MySpace streaming hits in the sample was well over 28 million. It seems likely, therefore, that download sales via MySpace were generated by the sampling effect, i.e., consumers, as a result of streaming free music, chose to purchase music that they otherwise may not have. This hypothesis seems to be supported by the inverse correlation between MySpace downloads and Billboard chart rank. Consumers may be choosing to purchase, via MySpace, primarily music by artists that receive less radio airplay, because they use MySpace to sample music

[1] Based on spot checks, this appears to be the primary, if not the only, means of download via MySpace.

to which they would not have otherwise been exposed. MySpace, therefore, increases consumer utility, allowing some consumers to match their tastes to music that they otherwise may not have located, and also generates revenue for the obscure artists who lack exposure and need this revenue the most. Furthermore, not only is the MySpace player free, but it allows the user control over the music to which they listen. Hence, MySpace allows many consumers simply to enjoy free music. Offering free streaming music via MySpace, therefore, increases social surplus and represents the potential for Pareto improvement. The only losers are traditional record labels. More generally, it seems clear that, rather than using resources to battle free offerings of their music, artists should take control of the situation. Instead, artists should offer their music for free, in quality and quantity sufficient to divert consumers from other free sources of their music, and they should do so in a manner that maximizes their profits from consumer sampling.

Appendix A

Summary Statistics

Table A.1: Summary Statistics: PollStarPro Data

Variable	Obs	Mean	Std. Dev.	Min	Max
SumVC0	89	204709.3	325743.8	450	1850623
SumVC0adj50	89	219481.6	344350.8	450	1970702
SumVC0adj75	89	226867.7	353987.5	450	2030741
SumVC0adj100	89	234253.8	363822.8	450	2090780
AvgVCHdlUS0	89	3400.592	4119.787	450	20980.22
NumHdlUS0	89	66.7191	53.35622	1	207
NumNoVCHdl~0	89	10.8764	11.40058	0	55
NumSupUS0	89	50.5618	45.79672	0	196
NumForeign0	89	26.91011	32.56903	0	148
NumHdlSTAT	89	84.95506	89.37649	2	418
NumHdlRpts~T	89	39.8427	37.02545	2	154
AvgTix	89	2144.787	2745.295	185	16727
AvgGross	89	82973.15	141955.2	2111	778746
PTicAvg	89	28.51881	13.36479	9.685715	65.61205
AvgVCBXO	89	3337.745	4164.848	0	20980.22
NumBXO	89	36.44944	35.94409	0	157
NumBXOnoVC	89	2.146067	3.050896	0	16

Table A.2: Summary Statistics: RIAA Data

Variable	Obs	Mean	Std. Dev.	Min	Max
TotAlb0	89	387640.4	752846.1	0	3500000
NumAlb0	89	.4382022	.6901096	0	3
TotDL0	89	449438.2	1162982	0	6000000
NumDL0	89	.4382022	1.033035	0	5
TotSing0	89	5617.978	52999.89	0	500000
NumSing0	89	.011236	.1059998	0	1
TotMT0	89	314606.7	905733.1	0	5000000

Table A.3: Summary Statistics: Billboard Data

Variable	Obs	Mean	Std. Dev.	Min	Max
Rock1	89	.6853933	.46699	0	1
SumInvRan~H0	89	1042.618	2726.66	0	17223
SumInvRan~k0	89	552.6404	791.617	0	3828
SumInvRan~m0	89	27.71758	36.399	0	229.64
SumInvRan~q0	89	18.82121	27.89008	0	169.9831
SumInvRan~t0	89	35.93135	43.3402	0	277.9877
NumSongsBB0	89	2.88764	2.621754	0	16
TotSongWks~0	89	50.30337	54.33951	0	350

Table A.4: Summary Statistics: YouTube & MySpace Data

Variable	Obs	Mean	Std. Dev.	Min	Max
YTTotHits	89	2.26e+07	3.51e+07	312172	2.03e+08
YTSDvHits	89	1676575	2484145	21195.29	1.05e+07
MSNumTrax	89	42.10112	42.33579	4	312
MSTotHits	89	2.84e+07	5.00e+07	837359	3.70e+08
MSSDvHits	89	2085889	2951242	45825.14	1.38e+07
MSNumDL	89	13.96629	17.98197	0	98
MSTotDL	89	7443.27	16040.84	0	95962

Appendix B

Pairwise Variable Correlations

Table B.1: Pairwise Variable Correlations (part 1)

	Rock1	SumIn~m0	NumSon~0	TotSon~0	SumIRL3	TotAlb0	NumAlb0
Rock1	1.0000						
SumInvRan~m0	-0.3083*	1.0000					
NumSongsBB0	-0.4005*	0.8682*	1.0000				
TotSongWks~0	-0.3594*	0.9837*	0.9189*	1.0000			
SumIRL3	-0.3249*	0.3400*	0.5329*	0.3752*	1.0000		
TotAlb0		0.2740*	0.2497*	0.2682*		1.0000	
NumAlb0		0.6633*	0.6619*	0.6613*		0.5333*	1.0000
SumAlbSal~L1			0.1825		0.5557*	0.4953*	
TotDL0		0.3227*	0.1938	0.2669*		0.2627*	0.3323*
NumDL0		0.4314*	0.2575*	0.3551*	0.2347*	0.3197*	0.3971*
SumDLSale~12			0.3138*	0.1941	0.6330*		
TotSing0			0.3726*		0.6987*		
NumSing0			0.3726*		0.6987*		
SumStSingS~1					0.4021*		
TotMT0	-0.4216*	0.5784*	0.7185*	0.6104*	0.5613*		0.3951*
MSNumTrax					0.1790	0.1767	
MSTotHits		0.7239*	0.6400*	0.7076*	0.2293*	0.2771*	0.6402*
MSSDvHits		0.5122*	0.4060*	0.4908*		0.2467*	0.4754*
MSNumDL		0.3166*	0.3222*	0.3117*	0.2627*	0.2230*	0.3125*
MSTotDL	-0.2825*	0.4244*	0.3987*	0.4086*	0.2218*	0.3049*	0.5238*
YTTotHits		0.5963*	0.6648*	0.6065*	0.5186*	0.2675*	0.5154*
YTSDvHits		0.3894*	0.4371*	0.3974*	0.3120*		0.3932*
SumVC0adj50	0.3003*	0.1842			0.1778	0.4842*	0.3156*
NumHdlUS0	0.5072*						
NumSupUS0	0.2979*				-0.2926*		
NumForeign0	0.4225*						
AvgVCHdlUS0		0.2406*	0.2456*	0.2166*	0.3428*	0.6739*	0.4720*
SumAdj5Lin2					0.4659*	0.5112*	
AvgTix		0.1878	0.2076		0.3737*	0.6148*	0.4125*
PTicAvg	-0.3919*	0.3169*	0.4121*	0.3343*	0.5813*	0.3379*	0.3602*

Table B.2: Pairwise Variable Correlations (part 2)

	SumAl~L1	TotDL0	NumDL0	SumDL~12	TotSing0	NumSing0	SumStS~1
SumAlbSal~L1	1.0000						
TotDL0	0.2052	1.0000					
NumDL0	0.3064*	0.8416*	1.0000				
SumDLSale~12	0.5196*		0.2455*	1.0000			
TotSing0	0.3481*			0.6012*	1.0000		
NumSing0	0.3481*			0.6012*	1.0000*	1.0000	
SumStSingS~1	0.4444*				0.3700*	0.3700*	1.0000
TotMT0		0.3632*	0.2943*	0.3860*	0.5546*	0.5546*	
MSNumTrax	0.2768*						
MSTotHits		0.5193*	0.4767*	0.1848			
MSSDvHits		0.5894*	0.4204*				
MSNumDL	0.2886*	0.2034	0.2375*				
MSTotDL	0.1804	0.5664*	0.5442*				
YTTotHits	0.2981*	0.4332*	0.4245*	0.5204*	0.5500*	0.5500*	
YTSDvHits		0.5420*	0.4225*	0.3668*	0.3456*	0.3456*	
SumVC0adj50	0.7094*		0.3436*	0.3552*			
NumHdlUS0							
NumSupUS0	-0.2798*			-0.1851			-0.2245*
NumForeign0							
AvgVCHdlUS0	0.7267*	0.2922*	0.4979*	0.4387*			
SumAdj5Lin2	0.8777*		0.2330*	0.6920*	0.3609*	0.3609*	0.3125*
AvgTix	0.6712*	0.3114*	0.5446*	0.4249*			
PTicAvg	0.5313*		0.3056*	0.3130*	0.2759*	0.2759*	0.3145*

Table B.3: Pairwise Variable Correlations (part 3)

	TotMT0	MSNumT~x	MSTotH~s	MSSDvH~s	MSNumDL	MSTotDL	YTTotH~s
TotMT0	1.0000						
MSNumTrax		1.0000					
MSTotHits	0.6311*	0.1948	1.0000				
MSSDvHits	0.4462*		0.8324*	1.0000			
MSNumDL	0.3418*	0.7809*	0.4806*	0.1833	1.0000		
MSTotDL	0.5692*	0.3028*	0.7502*	0.5137*	0.5767*	1.0000	
YTTotHits	0.6454*		0.6799*	0.6168*	0.2903*	0.4502*	1.0000
YTSDvHits	0.5179*		0.5690*	0.6086*		0.3873*	0.8458*
SumVC0adj50					0.2578*		
NumHdlUS0	-0.2083						
NumSupUS0		-0.1809	0.1775	0.3198*			
NumForeign0							
AvgVCHdlUS0		0.3892*	0.2269*		0.4272*	0.3633*	0.3092*
SumAdj5Lin2		0.2034			0.2135*		0.3240*
AvgTix		0.3240*			0.3906*	0.3049*	0.3442*
PTicAvg	0.3167*	0.3830*	0.1815		0.3451*	0.3229*	0.3783*

Table B.4: Pairwise Variable Correlations (part 4)

	YTSDvH~s	SumVC~50	NumHdl~0	NumSup~0	NumFor~0	AvgVCH~0	S~j5Lin2
YTSDvHits	1.0000						
SumVC0adj50		1.0000					
NumHdlUS0		0.5096*	1.0000				
NumSupUS0			0.4385*	1.0000			
NumForeign0					1.0000		
AvgVCHdlUS0		0.7137*		-0.2248*		1.0000	
SumAdj5Lin2	0.1938	0.7107*		-0.2274*		0.7250*	1.0000
AvgTix	0.2105*	0.6134*		-0.2493*	0.2534*	0.9295*	0.6344*
PTicAvg	0.1848	0.2652*	-0.2489*	-0.4339*		0.5836*	0.4070*

	AvgTix	PTicAvg
AvgTix	1.0000	
PTicAvg	0.6011*	1.0000

Appendix C

3SLS Regression Results

Table C.1: 3SLS Results: Entire Sample, Regression 1 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	**88**	**10**	**464162.2**	**0.6187**	**155.06**	**0.0000**
SumVC0adj50	**88**	**10**	**158151.3**	**0.7883**	**339.80**	**0.0000**
MSTotHits	**88**	**6**	**2.14e+07**	**0.8163**	**382.73**	**0.0000**
YTTotHits	**88**	**6**	**1.41e+07**	**0.8376**	**458.44**	**0.0000**
SumInvRan~t0	**88**	**11**	**11.02834**	**0.9347**	**1257.71**	**0.0000**

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
NumAlb0	627999.7	107756.2	5.83	0.000	416801.4	839198
SumInvRan~t0	1862.618	2153.407	0.86	0.387	-2357.982	6083.219
SumIRNSqrtL3	-5280.754	1567.307	-3.37	0.001	-8352.62	-2208.888
SumVC0adj50	-.6695747	.3347736	-2.00	0.045	-1.325719	-.0134304
SumAdj5Lin2	1.180385	.3085369	3.83	0.000	.5756633	1.785106
SumAlbSal~L1	.0913948	.0466965	1.96	0.050	-.0001287	.1829183
TotDL0	.1038295	.0501758	2.07	0.039	.0054868	.2021723
SumDLSale~12	-.4252654	.2405055	-1.77	0.077	-.8966475	.0461166
MSTotHits	.0019948	.0026558	0.75	0.453	-.0032106	.0072001
YTTotHits	-.0083436	.0033419	-2.50	0.013	-.0148936	-.0017936
_cons	134376.1	78630.86	1.71	0.087	-19737.52	288489.8
SumVC0adj50						
NumHdlUS0	3447.946	462.1053	7.46	0.000	2542.236	4353.656
NumSupUS0	-905.7478	444.8901	-2.04	0.042	-1777.716	-33.77917
NumForeign0	296.1604	605.0496	0.49	0.625	-889.7151	1482.036
TotAlb0	.1331423	.0556048	2.39	0.017	.0241589	.2421258
TotDL0	.0169363	.019497	0.87	0.385	-.0212771	.0551498
MSTotHits	.0014569	.0011151	1.31	0.191	-.0007286	.0036424
YTTotHits	-.0017001	.0014534	-1.17	0.242	-.0045487	.0011485
SumInvRan~t0	3.120841	730.1204	0.00	0.997	-1427.889	1434.131
SumAdj5Lin2	.1243375	.1022867	1.22	0.224	-.0761408	.3248158
SumAlbSal~L1	.0438477	.013097	3.35	0.001	.0181781	.0695174
_cons	-107943.9	36947.83	-2.92	0.003	-180360.3	-35527.46
MSTotHits						
MSSDvHits	10.88004	.9643506	11.28	0.000	8.989943	12.77013
SumInvRan~t0	431072.8	66777.4	6.46	0.000	300191.5	561954.1
TotAlb0	-.986971	5.59442	-0.18	0.860	-11.95183	9.977891
SumAlbSal~L1	-1.072391	1.725322	-0.62	0.534	-4.453959	2.309178
SumVC0adj50	10.62783	12.62598	0.84	0.400	-14.11865	35.3743
SumAdj5Lin2	9.103253	10.74707	0.85	0.397	-11.96061	30.16712
_cons	-1.22e+07	3359283	-3.64	0.000	-1.88e+07	-5635991

Table C.2: 3SLS Results: Entire Sample, Regression 1 (part 2)

YTTotHits						
YTSDvHits	9.231691	.6930501	13.32	0.000	7.873338	10.59004
SumInvRan~t0	308464.4	44215.4	6.98	0.000	221803.8	395125
TotAlb0	-3.033245	3.529666	-0.86	0.390	-9.951263	3.884772
SumAlbSal~L1	.091409	1.123865	0.08	0.935	-2.111326	2.294144
SumVC0adj50	-9.775771	8.41706	-1.16	0.245	-26.2729	6.721363
SumAdj5Lin2	20.24186	7.304467	2.77	0.006	5.925365	34.55835
_cons	-4845482	2245456	-2.16	0.031	-9246494	-444470.4
SumInvRan~t0						
NumSongsBB0	14.22402	.9515316	14.95	0.000	12.35905	16.08898
AvgWksBB0	1.772383	.1740884	10.18	0.000	1.431176	2.11359
SumIRNSqrtL3	-.0346268	.0534357	-0.65	0.517	-.1393588	.0701052
SumVC0adj50	.0000191	7.99e-06	2.39	0.017	3.40e-06	.0000347
SumAdj5Lin2	-4.51e-06	.0000101	-0.44	0.657	-.0000244	.0000154
TotAlb0	3.92e-07	4.97e-06	0.08	0.937	-9.35e-06	.0000101
SumAlbSal~L1	-1.04e-06	1.15e-06	-0.90	0.369	-3.30e-06	1.22e-06
TotDL0	3.13e-06	1.52e-06	2.06	0.039	1.59e-07	6.10e-06
SumDLSale~12	-4.51e-06	7.25e-06	-0.62	0.533	-.0000187	9.69e-06
MSTotHits	6.74e-08	7.82e-08	0.86	0.388	-8.57e-08	2.21e-07
YTTotHits	-8.49e-08	9.37e-08	-0.91	0.365	-2.69e-07	9.87e-08
_cons	-36.64765	3.300468	-11.10	0.000	-43.11644	-30.17885

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: NumAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 NumHdlUS0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.3: 3SLS Results: Entire Sample, Regression 2 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	88	10	341193.9	0.7940	342.59	0.0000
SumVC0adj50	88	10	149841.3	0.8099	377.39	0.0000
MSTotHits	88	6	2.13e+07	0.8176	380.87	0.0000
YTTotHits	88	6	1.39e+07	0.8432	470.34	0.0000
SumInvRan~t0	88	11	11.26591	0.9319	1211.03	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
AvgAlb0	1.007604	.0951949	10.58	0.000	.8210259	1.194183
SumInvRan~t0	2867.904	1559.666	1.84	0.066	-188.9844	5924.793
SumIRNSqrtL3	-2641.491	1281.096	-2.06	0.039	-5152.394	-130.5884
SumVC0adj50	-.5719632	.2452922	-2.33	0.020	-1.052727	-.0911993
SumAdj5Lin2	.3310833	.2534107	1.31	0.191	-.1655925	.827759
SumAlbSal~L1	.0325709	.0340959	0.96	0.339	-.0342558	.0993976
TotDL0	-.0172681	.0403502	-0.43	0.669	-.0963531	.0618169
SumDLSale~12	.0112988	.2035793	0.06	0.956	-.3877094	.4103069
MSTotHits	.0033655	.002039	1.65	0.099	-.0006309	.0073619
YTTotHits	-.0035835	.0024912	-1.44	0.150	-.0084662	.0012991
_cons	93986.48	57529.82	1.63	0.102	-18769.9	206742.9
SumVC0adj50						
NumHdlUS0	3414.315	411.7071	8.29	0.000	2607.384	4221.246
NumSupUS0	-881.4918	426.2104	-2.07	0.039	-1716.849	-46.13472
NumForeign0	387.673	556.3937	0.70	0.486	-702.8386	1478.185
TotAlb0	.0752676	.0363701	2.07	0.038	.0039836	.1465516
TotDL0	.0305844	.018415	1.66	0.097	-.0055084	.0666772
MSTotHits	.0013474	.000942	1.43	0.153	-.0004988	.0031936
YTTotHits	-.0023	.0011544	-1.99	0.046	-.0045626	-.0000373
SumInvRan~t0	523.9377	675.6482	0.78	0.438	-800.3083	1848.184
SumAdj5Lin2	.1888617	.0857352	2.20	0.028	.0208238	.3568996
SumAlbSal~L1	.042192	.0123175	3.43	0.001	.0180502	.0663338
_cons	-105946.2	34725.48	-3.05	0.002	-174006.9	-37885.56

Table C.4: 3SLS Results: Entire Sample, Regression 2 (part 2)

MSTotHits						
MSSDvHits	10.97625	.9548295	11.50	0.000	9.10482	12.84768
SumInvRan~t0	432257.2	66231.09	6.53	0.000	302446.7	562067.8
TotAlb0	-1.49782	4.57925	-0.33	0.744	-10.47299	7.477346
SumAlbSal~L1	-.7371119	1.734079	-0.43	0.671	-4.135845	2.661621
SumVC0adj50	3.478909	13.06013	0.27	0.790	-22.11848	29.0763
SumAdj5Lin2	11.35647	10.71381	1.06	0.289	-9.642209	32.35515
_cons	-1.15e+07	3372842	-3.42	0.001	-1.82e+07	-4934993
YTTotHits						
YTSDvHits	9.313566	.6848786	13.60	0.000	7.971228	10.6559
SumInvRan~t0	296408.9	42635.53	6.95	0.000	212844.8	379973
TotAlb0	-.4192929	2.886128	-0.15	0.884	-6.076	5.237414
SumAlbSal~L1	.1592301	1.114666	0.14	0.886	-2.025475	2.343935
SumVC0adj50	-12.77371	8.615309	-1.48	0.138	-29.6594	4.111988
SumAdj5Lin2	19.25678	7.197184	2.68	0.007	5.150563	33.363
_cons	-4802930	2237501	-2.15	0.032	-9188352	-417507.5
SumInvRan~t0						
NumSongsBB0	14.46313	.9089121	15.91	0.000	12.68169	16.24456
AvgWksBB0	1.844802	.1746205	10.56	0.000	1.502553	2.187052
SumIRNSqrtL3	-.0599916	.0488142	-1.23	0.219	-.1556656	.0356825
SumVC0adj50	.0000159	8.44e-06	1.89	0.059	-6.13e-07	.0000325
SumAdj5Lin2	1.53e-06	8.93e-06	0.17	0.864	-.000016	.000019
TotAlb0	-2.81e-06	3.19e-06	-0.88	0.379	-9.06e-06	3.44e-06
SumAlbSal~L1	-7.92e-07	1.16e-06	-0.68	0.495	-3.07e-06	1.48e-06
TotDL0	3.44e-06	1.52e-06	2.25	0.024	4.49e-07	6.42e-06
SumDLSale~12	-7.46e-06	7.04e-06	-1.06	0.289	-.0000213	6.33e-06
MSTotHits	9.18e-08	7.27e-08	1.26	0.207	-5.06e-08	2.34e-07
YTTotHits	-1.11e-07	8.79e-08	-1.26	0.208	-2.83e-07	6.17e-08
_cons	-37.29971	3.342459	-11.16	0.000	-43.8508	-30.74861

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: AvgAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 NumHdlUS0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.5: 3SLS Results: Entire Sample, Regression 3 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	88	10	558496.2	0.4480	134.02	0.0000
SumVC0adj50	88	10	214294.9	0.6113	168.82	0.0000
MSTotHits	88	6	2.71e+07	0.7064	324.56	0.0000
YTTotHits	88	6	1.72e+07	0.7586	410.62	0.0000
SumInvRan~t0	88	11	11.02447	0.9348	1258.95	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
NumAlb0	378569.9	105800.8	3.58	0.000	171204	585935.7
SumInvRan~t0	1719.706	2147.705	0.80	0.423	-2489.719	5929.131
SumIRNSqrtL3	-2970.479	1455.418	-2.04	0.041	-5823.045	-117.9129
SumVC0adj50	1.037146	.5143675	2.02	0.044	.0290044	2.045288
SumAdj5Lin2	.5850585	.331999	1.76	0.078	-.0656475	1.235765
SumAlbSal~L1	-.0003014	.0514383	-0.01	0.995	-.1011187	.1005158
TotDL0	.0853478	.04514	1.89	0.059	-.0031249	.1738206
SumDLSale~12	-.2966117	.2024471	-1.47	0.143	-.6934007	.1001773
MSTotHits	.0010564	.0025198	0.42	0.675	-.0038822	.005995
YTTotHits	-.0044075	.0032148	-1.37	0.170	-.0107085	.0018934
_cons	-31136.65	91117.97	-0.34	0.733	-209724.6	147451.3
SumVC0adj50						
AvgVC0	20.43188	10.01685	2.04	0.041	.7992222	40.06454
NumSupUS0	1052.092	471.2506	2.23	0.026	128.4582	1975.727
NumForeign0	673.313	653.9951	1.03	0.303	-608.4939	1955.12
TotAlb0	.133772	.0850835	1.57	0.116	-.0329886	.3005327
TotDL0	-.0225672	.0239028	-0.94	0.345	-.0694159	.0242814
MSTotHits	.0013589	.001449	0.94	0.348	-.0014811	.0041989
YTTotHits	-.0017602	.0017985	-0.98	0.328	-.0052852	.0017649
SumInvRan~t0	-503.1378	927.9499	-0.54	0.588	-2321.886	1315.61
SumAdj5Lin2	.1250697	.1249835	1.00	0.317	-.1198934	.3700328
SumAlbSal~L1	.0386067	.0172217	2.24	0.025	.0048528	.0723606
_cons	-16169.15	45937.55	-0.35	0.725	-106205.1	73866.79

Table C.6: 3SLS Results: Entire Sample, Regression 3 (part 2)

MSTotHits						
MSSDvHits	10.39119	1.027178	10.12	0.000	8.37796	12.40443
SumInvRan~t0	396431.8	73903.97	5.36	0.000	251582.7	541281
TotAlb0	19.02111	8.007239	2.38	0.018	3.327207	34.71501
SumAlbSal~L1	.7329359	2.006044	0.37	0.715	-3.198838	4.66471
SumVC0adj50	-48.5899	25.20404	-1.93	0.054	-97.9889	.8091011
SumAdj5Lin2	13.13029	11.8066	1.11	0.266	-10.01022	36.2708
_cons	-7808592	3975339	-1.96	0.050	-1.56e+07	-17071.69
YTTotHits						
YTSDvHits	9.391005	.7383443	12.72	0.000	7.943877	10.83813
SumInvRan~t0	323800.5	47124.76	6.87	0.000	231437.7	416163.4
TotAlb0	-10.84319	5.113357	-2.12	0.034	-20.86519	-.8211983
SumAlbSal~L1	-.8463771	1.278682	-0.66	0.508	-3.352548	1.659794
SumVC0adj50	18.5788	16.45852	1.13	0.259	-13.6793	50.8369
SumAdj5Lin2	17.02083	7.878936	2.16	0.031	1.578395	32.46326
_cons	-7046021	2617186	-2.69	0.007	-1.22e+07	-1916430
SumInvRan~t0						
NumSongsBB0	14.06772	1.179845	11.92	0.000	11.75526	16.38017
AvgWksBB0	1.768345	.2065975	8.56	0.000	1.363422	2.173269
SumIRNSqrtL3	-.0317922	.0597473	-0.53	0.595	-.1488948	.0853104
SumVC0adj50	.00001	.0000187	0.54	0.592	-.0000267	.0000467
SumAdj5Lin2	-3.58e-06	9.87e-06	-0.36	0.717	-.0000229	.0000158
TotAlb0	2.83e-06	8.65e-06	0.33	0.743	-.0000141	.0000198
SumAlbSal~L1	-7.01e-07	1.23e-06	-0.57	0.569	-3.11e-06	1.71e-06
TotDL0	2.82e-06	1.61e-06	1.75	0.080	-3.40e-07	5.97e-06
SumDLSale~12	-4.67e-06	7.43e-06	-0.63	0.529	-.0000192	9.88e-06
MSTotHits	7.55e-08	7.89e-08	0.96	0.338	-7.90e-08	2.30e-07
YTTotHits	-8.70e-08	9.64e-08	-0.90	0.367	-2.76e-07	1.02e-07
_cons	-35.79734	3.290813	-10.88	0.000	-42.24722	-29.34747

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: NumAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 AvgVC0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.7: 3SLS Results: Entire Sample, Regression 4 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	88	10	399301	0.7178	273.13	0.0000
SumVC0adj50	88	10	194995.5	0.6781	194.75	0.0000
MSTotHits	88	6	2.25e+07	0.7974	348.03	0.0000
YTTotHits	88	6	1.40e+07	0.8404	463.21	0.0000
SumInvRan~t0	88	11	11.42898	0.9299	1170.29	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
AvgAlb0	.8603099	.1018427	8.45	0.000	.6607019	1.059918
SumInvRan~t0	1378.614	1689.206	0.82	0.414	-1932.17	4689.397
SumIRNSqrtL3	-2062.99	1214.115	-1.70	0.089	-4442.612	316.6324
SumVC0adj50	.6020859	.4028306	1.49	0.135	-.1874477	1.391619
SumAdj5Lin2	-.0141596	.2806904	-0.05	0.960	-.5643027	.5359834
SumAlbSal~L1	-.006895	.0385302	-0.18	0.858	-.0824128	.0686228
TotDL0	-.0292904	.0424421	-0.69	0.490	-.1124754	.0538947
SumDLSale~12	.0515997	.1867705	0.28	0.782	-.3144639	.4176632
MSTotHits	.0035708	.0021938	1.63	0.104	-.0007289	.0078705
YTTotHits	-.0020449	.0026351	-0.78	0.438	-.0072096	.0031197
_cons	-7814.144	69313.69	-0.11	0.910	-143666.5	128038.2
SumVC0adj50						
AvgVC0	34.95476	8.94163	3.91	0.000	17.42949	52.48003
NumSupUS0	1081.404	449.2414	2.41	0.016	200.9073	1961.901
NumForeign0	764.9611	623.3653	1.23	0.220	-456.8124	1986.735
TotAlb0	.019604	.0508986	0.39	0.700	-.0801555	.1193634
TotDL0	-.0090175	.0238601	-0.38	0.705	-.0557825	.0377474
MSTotHits	.0014128	.0012107	1.17	0.243	-.0009601	.0037858
YTTotHits	-.0025816	.00144	-1.79	0.073	-.005404	.0002409
SumInvRan~t0	26.67339	867.3308	0.03	0.975	-1673.264	1726.61
SumAdj5Lin2	.1759525	.1092143	1.61	0.107	-.0381036	.3900087
SumAlbSal~L1	.0319316	.0159802	2.00	0.046	.0006109	.0632522
_cons	-35544.21	43154.75	-0.82	0.410	-120126	49037.54

Table C.8: 3SLS Results: Entire Sample, Regression 4 (part 2)

MSTotHits						
MSSDvHits	10.41088	.9788514	10.64	0.000	8.492364	12.32939
SumInvRan~t0	436600.5	68814.34	6.34	0.000	301726.8	571474.1
TotAlb0	6.270381	5.477662	1.14	0.252	-4.465639	17.0064
SumAlbSal~L1	-.1335439	1.909915	-0.07	0.944	-3.876909	3.609821
SumVC0adj50	-19.31693	22.22463	-0.87	0.385	-62.87639	24.24254
SumAdj5Lin2	13.0998	11.42957	1.15	0.252	-9.301739	35.50135
_cons	-9660152	3741297	-2.58	0.010	-1.70e+07	-2327344
YTTotHits						
YTSDvHits	9.30669	.696921	13.35	0.000	7.94075	10.67263
SumInvRan~t0	297600.6	42738	6.96	0.000	213835.6	381365.5
TotAlb0	-1.152102	3.436488	-0.34	0.737	-7.887494	5.58329
SumAlbSal~L1	-.0006318	1.188729	-0.00	1.000	-2.330498	2.329234
SumVC0adj50	-8.483064	14.33641	-0.59	0.554	-36.58191	19.61578
SumAdj5Lin2	18.542	7.513422	2.47	0.014	3.815966	33.26804
_cons	-5152596	2451144	-2.10	0.036	-9956749	-348442.5
SumInvRan~t0						
NumSongsBB0	14.41651	.9364394	15.40	0.000	12.58113	16.2519
AvgWksBB0	1.83079	.1930646	9.48	0.000	1.45239	2.209189
SumIRNSqrtL3	-.0610432	.0482909	-1.26	0.206	-.1556917	.0336053
SumVC0adj50	.0000212	.0000153	1.39	0.166	-8.77e-06	.0000511
SumAdj5Lin2	7.73e-07	9.33e-06	0.08	0.934	-.0000175	.0000191
TotAlb0	-3.87e-06	3.54e-06	-1.09	0.275	-.0000108	3.08e-06
SumAlbSal~L1	-9.55e-07	1.28e-06	-0.75	0.455	-3.46e-06	1.55e-06
TotDL0	3.36e-06	1.54e-06	2.18	0.029	3.37e-07	6.39e-06
SumDLSale~12	-7.80e-06	7.05e-06	-1.11	0.269	-.0000216	6.02e-06
MSTotHits	9.83e-08	7.58e-08	1.30	0.195	-5.03e-08	2.47e-07
YTTotHits	-1.07e-07	9.00e-08	-1.19	0.234	-2.83e-07	6.93e-08
_cons	-37.48362	3.391442	-11.05	0.000	-44.13072	-30.83652

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: AvgAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 AvgVC0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.9: 3SLS Results: Rock Subsample, Regression 1 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	60	10	381224.6	0.7788	201.81	0.0000
SumVC0adj50	60	10	137200.2	0.8788	448.97	0.0000
MSTotHits	60	6	1.05e+07	0.8290	299.78	0.0000
YTTotHits	60	6	1.06e+07	0.8142	288.54	0.0000
SumInvRan~t0	60	11	7.343248	0.9077	616.31	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
NumAlb0	779282.4	124426.1	6.26	0.000	535411.7	1023153
SumInvRan~t0	1337.751	3629.023	0.37	0.712	-5775.004	8450.505
SumIRNSqrtL3	-3461.462	2013.728	-1.72	0.086	-7408.296	485.3725
SumVC0adj50	-1.152924	.4479254	-2.57	0.010	-2.030842	-.2750064
SumAdj5Lin2	.9665694	.2990289	3.23	0.001	.3804835	1.552655
SumAlbSal~L1	.1404171	.0507871	2.76	0.006	.0408762	.239958
TotDL0	.1021588	.0717598	1.42	0.155	-.0384878	.2428055
SumDLSale~12	-.3785487	.2517907	-1.50	0.133	-.8720493	.114952
MSTotHits	.0106062	.0041758	2.54	0.011	.0024218	.0187905
YTTotHits	-.0142514	.0038526	-3.70	0.000	-.0218023	-.0067004
_cons	134776.5	88873.66	1.52	0.129	-39412.71	308965.6
SumVC0adj50						
NumHdlUS0	2838.723	429.1822	6.61	0.000	1997.542	3679.905
NumSupUS0	-882.9961	388.0118	-2.28	0.023	-1643.485	-122.507
NumForeign0	-77.86487	469.7492	-0.17	0.868	-998.5565	842.8267
TotAlb0	.0660024	.0491319	1.34	0.179	-.0302944	.1622992
TotDL0	-.0353927	.0225108	-1.57	0.116	-.0795131	.0087278
MSTotHits	.0031258	.0013868	2.25	0.024	.0004076	.0058439
YTTotHits	.0003289	.0013849	0.24	0.812	-.0023854	.0030432
SumInvRan~t0	2746.751	962.0632	2.86	0.004	861.1413	4632.36
SumAdj5Lin2	.1183582	.0820559	1.44	0.149	-.0424684	.2791847
SumAlbSal~L1	.058712	.0134957	4.35	0.000	.032261	.085163
_cons	-164148.3	43879.2	-3.74	0.000	-250150	-78146.66

Table C.10: 3SLS Results: Rock Subsample, Regression 1 (part 2)

MSTotHits						
MSSDvHits	6.930247	.5287756	13.11	0.000	5.893866	7.966629
SumInvRan~t0	117353.1	78318.9	1.50	0.134	-36149.1	270855.3
TotAlb0	4.226577	3.229183	1.31	0.191	-2.102506	10.55566
SumAlbSal~L1	-2.772593	1.012143	-2.74	0.006	-4.756356	-.7888299
SumVC0adj50	25.37197	9.109533	2.79	0.005	7.517611	43.22633
SumAdj5Lin2	5.928583	6.203561	0.96	0.339	-6.230173	18.08734
_cons	-460200.7	2251112	-0.20	0.838	-4872299	3951898
YTTotHits						
YTSDvHits	7.907953	.5402588	14.64	0.000	6.849066	8.966841
SumInvRan~t0	174204.2	81850.19	2.13	0.033	13780.81	334627.6
TotAlb0	6.425322	3.06668	2.10	0.036	.4147391	12.43591
SumAlbSal~L1	-1.490591	1.006407	-1.48	0.139	-3.463113	.4819308
SumVC0adj50	7.954022	9.255984	0.86	0.390	-10.18737	26.09542
SumAdj5Lin2	4.216839	6.297274	0.67	0.503	-8.125591	16.55927
_cons	-1559592	2263905	-0.69	0.491	-5996764	2877579
SumInvRan~t0						
NumSongsBB0	9.323567	1.013786	9.20	0.000	7.336582	11.31055
AvgWksBB0	1.544005	.1284493	12.02	0.000	1.292249	1.795761
SumIRNSqrtL3	.065004	.0360318	1.80	0.071	-.0056171	.1356251
SumVC0adj50	.0000272	7.21e-06	3.78	0.000	.0000131	.0000413
SumAdj5Lin2	-.0000134	5.95e-06	-2.26	0.024	-.0000251	-1.77e-06
TotAlb0	-1.80e-06	2.89e-06	-0.62	0.534	-7.47e-06	3.87e-06
SumAlbSal~L1	-5.49e-07	9.27e-07	-0.59	0.554	-2.37e-06	1.27e-06
TotDL0	4.18e-06	1.36e-06	3.08	0.002	1.52e-06	6.85e-06
SumDLSale~12	3.78e-06	4.40e-06	0.86	0.391	-4.85e-06	.0000124
MSTotHits	-9.76e-08	8.10e-08	-1.21	0.228	-2.56e-07	6.11e-08
YTTotHits	-2.55e-08	6.78e-08	-0.38	0.707	-1.58e-07	1.07e-07
_cons	-22.1407	2.670122	-8.29	0.000	-27.37404	-16.90736

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0
Exogenous variables: NumAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 NumHdlUS0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.11: 3SLS Results: Rock Subsample, Regression 2 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	60	10	324782.6	0.8394	317.69	0.0000
SumVC0adj50	60	10	131039.6	0.8895	485.49	0.0000
MSTotHits	60	6	1.06e+07	0.8258	296.52	0.0000
YTTotHits	60	6	1.04e+07	0.8200	297.12	0.0000
SumInvRan~t0	60	11	7.381519	0.9067	621.65	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
AvgAlb0	.968467	.1099836	8.81	0.000	.7529031	1.184031
SumInvRan~t0	2690.859	3217.491	0.84	0.403	-3615.307	8997.026
SumIRNSqrtL3	-846.4929	1728.063	-0.49	0.624	-4233.434	2540.448
SumVC0adj50	-.7743479	.3878573	-2.00	0.046	-1.534534	-.0141615
SumAdj5Lin2	.1882173	.2889131	0.65	0.515	-.378042	.7544766
SumAlbSal~L1	.0561332	.0429258	1.31	0.191	-.0279998	.1402663
TotDL0	.0048611	.0627718	0.08	0.938	-.1181694	.1278916
SumDLSale~12	.1469881	.2224305	0.66	0.509	-.2889676	.5829439
MSTotHits	.0062032	.0037021	1.68	0.094	-.0010528	.0134591
YTTotHits	-.0058058	.0031683	-1.83	0.067	-.0120155	.0004039
_cons	129513.7	75523.09	1.71	0.086	-18508.8	277536.3
SumVC0adj50						
NumHdlUS0	2622.155	409.0859	6.41	0.000	1820.362	3423.949
NumSupUS0	-881.4616	358.6384	-2.46	0.014	-1584.38	-178.5431
NumForeign0	106.3806	430.1414	0.25	0.805	-736.681	949.4422
TotAlb0	-.0142144	.0405844	-0.35	0.726	-.0937584	.0653297
TotDL0	-.0232697	.0217966	-1.07	0.286	-.0659903	.0194509
MSTotHits	.0042682	.0013087	3.26	0.001	.0017031	.0068333
YTTotHits	-.0007968	.0013393	-0.59	0.552	-.0034217	.0018281
SumInvRan~t0	2738.696	919.2563	2.98	0.003	936.9863	4540.405
SumAdj5Lin2	.1766103	.0775977	2.28	0.023	.0245217	.328699
SumAlbSal~L1	.0630864	.0127855	4.93	0.000	.0380273	.0881454
_cons	-147481.9	41041.49	-3.59	0.000	-227921.8	-67042.09
MSTotHits						
MSSDvHits	6.648698	.5132982	12.95	0.000	5.642652	7.654744
SumInvRan~t0	122689.2	79331.05	1.55	0.122	-32796.85	278175.2
TotAlb0	5.516348	2.787642	1.98	0.048	.05267	10.98003
SumAlbSal~L1	-2.931602	1.017624	-2.88	0.004	-4.926109	-.9370963
SumVC0adj50	25.57748	9.611766	2.66	0.008	6.738765	44.41619
SumAdj5Lin2	5.198416	6.296015	0.83	0.409	-7.141547	17.53838
_cons	-265854.1	2233384	-0.12	0.905	-4643207	4111498

Table C.12: 3SLS Results: Rock Subsample, Regression 2 (part 2)

YTTotHits						
YTSDvHits	7.82059	.5232604	14.95	0.000	6.795019	8.846162
SumInvRan~t0	189208.7	81926.6	2.31	0.021	28635.47	349781.8
TotAlb0	3.362526	2.643472	1.27	0.203	-1.818585	8.543636
SumAlbSal~L1	-1.131792	.9993944	-1.13	0.257	-3.090569	.8269849
SumVC0adj50	5.74579	9.52987	0.60	0.547	-12.93241	24.42399
SumAdj5Lin2	6.580276	6.306161	1.04	0.297	-5.779573	18.94013
_cons	-920207.9	2217994	-0.41	0.678	-5267397	3426981
SumInvRan~t0						
NumSongsBB0	8.907906	1.05811	8.42	0.000	6.834048	10.98176
AvgWksBB0	1.578017	.1282336	12.31	0.000	1.326684	1.829351
SumIRNSqrtL3	.0693571	.0338389	2.05	0.040	.003034	.1356803
SumVC0adj50	.0000297	8.06e-06	3.69	0.000	.0000139	.0000455
SumAdj5Lin2	-.0000153	5.89e-06	-2.60	0.009	-.0000269	-3.77e-06
TotAlb0	2.62e-07	2.52e-06	0.10	0.917	-4.67e-06	5.19e-06
SumAlbSal~L1	-8.32e-07	9.58e-07	-0.87	0.385	-2.71e-06	1.05e-06
TotDL0	4.01e-06	1.30e-06	3.08	0.002	1.46e-06	6.56e-06
SumDLSale~12	3.58e-06	4.23e-06	0.85	0.397	-4.71e-06	.0000119
MSTotHits	-1.38e-07	8.55e-08	-1.61	0.107	-3.05e-07	2.99e-08
YTTotHits	1.90e-09	7.16e-08	0.03	0.979	-1.38e-07	1.42e-07
_cons	-22.15294	2.689832	-8.24	0.000	-27.42492	-16.88097

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: AvgAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 NumHdlUS0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.13: 3SLS Results: Rock Subsample, Regression 3 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	60	10	542783.9	0.5515	139.50	0.0000
SumVC0adj50	60	10	204133.2	0.7318	213.92	0.0000
MSTotHits	60	6	1.19e+07	0.7796	266.94	0.0000
YTTotHits	60	6	1.24e+07	0.7470	276.52	0.0000
SumInvRan~t0	60	11	7.847171	0.8945	543.54	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
NumAlb0	906378.7	272423.6	3.33	0.001	372438.3	1440319
SumInvRan~t0	10066.16	7199.326	1.40	0.162	-4044.264	24176.58
SumIRNSqrtL3	-3981.645	2278.765	-1.75	0.081	-8447.942	484.6517
SumVC0adj50	-3.324419	1.85312	-1.79	0.073	-6.956468	.3076305
SumAdj5Lin2	1.400453	.4318839	3.24	0.001	.5539759	2.24693
SumAlbSal~L1	.281464	.1328665	2.12	0.034	.0210504	.5418775
TotDL0	-.0452241	.1310663	-0.35	0.730	-.3021094	.2116612
SumDLSale~12	-.3364451	.3136304	-1.07	0.283	-.9511493	.2782591
MSTotHits	.0206998	.0079223	2.61	0.009	.0051723	.0362273
YTTotHits	-.0181527	.0066457	-2.73	0.006	-.0311781	-.0051272
_cons	99121.11	107537.1	0.92	0.357	-111647.8	309890
SumVC0adj50						
AvgVC0	-21.94512	11.70806	-1.87	0.061	-44.89249	1.002244
NumSupUS0	306.6208	568.1639	0.54	0.589	-806.9599	1420.201
NumForeign0	-74.9369	640.5533	-0.12	0.907	-1330.398	1180.524
TotAlb0	.1584379	.0844409	1.88	0.061	-.0070632	.3239391
TotDL0	-.0651089	.0307392	-2.12	0.034	-.1253566	-.0048613
MSTotHits	.0014148	.0020149	0.70	0.483	-.0025343	.005364
YTTotHits	-.0004066	.0019873	-0.20	0.838	-.0043017	.0034884
SumInvRan~t0	5729.679	1333.843	4.30	0.000	3115.395	8343.963
SumAdj5Lin2	.2766776	.1078433	2.57	0.010	.0653085	.4880466
SumAlbSal~L1	.0536803	.0184018	2.92	0.004	.0176135	.0897471
_cons	-3549.016	54657.45	-0.06	0.948	-110675.7	103577.6
MSTotHits						
MSSDvHits	6.658981	.5315071	12.53	0.000	5.617247	7.700716
SumInvRan~t0	251893.9	100319.2	2.51	0.012	55271.84	448516
TotAlb0	10.98946	3.579403	3.07	0.002	3.973954	18.00496
SumAlbSal~L1	-2.128023	1.175527	-1.81	0.070	-4.432013	.1759677
SumVC0adj50	-4.221255	16.92202	-0.25	0.803	-37.3878	28.94529
SumAdj5Lin2	11.66904	7.428345	1.57	0.116	-2.890247	26.22833
_cons	-125408.4	2419447	-0.05	0.959	-4867437	4616620

Table C.14: 3SLS Results: Rock Subsample, Regression 3 (part 2)

YTTotHits						
YTSDvHits	7.666569	.5445097	14.08	0.000	6.599349	8.733788
SumInvRan~t0	291639.3	103938.5	2.81	0.005	87923.58	495355
TotAlb0	10.4207	3.508332	2.97	0.003	3.544493	17.2969
SumAlbSal~L1	-.7373412	1.153096	-0.64	0.523	-2.997368	1.522686
SumVC0adj50	-18.31955	16.99088	-1.08	0.281	-51.62107	14.98197
SumAdj5Lin2	10.59955	7.468897	1.42	0.156	-4.039214	25.23832
_cons	-986979.8	2412870	-0.41	0.683	-5716119	3742159
SumInvRan~t0						
NumSongsBB0	10.88067	1.663743	6.54	0.000	7.619799	14.14155
AvgWksBB0	1.720692	.1654541	10.40	0.000	1.396408	2.044976
SumIRNSqrtL3	.0623802	.0407039	1.53	0.125	-.0173979	.1421583
SumVC0adj50	3.78e-06	.0000207	0.18	0.855	-.0000368	.0000444
SumAdj5Lin2	-6.97e-06	6.40e-06	-1.09	0.276	-.0000195	5.58e-06
TotAlb0	4.24e-07	3.59e-06	0.12	0.906	-6.60e-06	7.45e-06
SumAlbSal~L1	4.11e-07	1.49e-06	0.28	0.782	-2.51e-06	3.33e-06
TotDL0	3.61e-06	1.78e-06	2.03	0.043	1.18e-07	7.09e-06
SumDLSale~12	1.82e-06	5.75e-06	0.32	0.751	-9.45e-06	.0000131
MSTotHits	-5.43e-08	1.00e-07	-0.54	0.588	-2.50e-07	1.42e-07
YTTotHits	-4.57e-08	7.57e-08	-0.60	0.546	-1.94e-07	1.03e-07
_cons	-25.55252	3.502701	-7.30	0.000	-32.41769	-18.68735

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: NumAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 AvgVC0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Table C.15: 3SLS Results: Rock Subsample, Regression 4 (part 1)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
TotAlb0	60	10	670994.7	0.3146	118.71	0.0000
SumVC0adj50	60	10	171616.1	0.8104	265.81	0.0000
MSTotHits	60	6	1.07e+07	0.8228	289.26	0.0000
YTTotHits	60	6	1.06e+07	0.8132	290.71	0.0000
SumInvRan~t0	60	11	7.201195	0.9112	550.65	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TotAlb0						
AvgAlb0	.9054708	.1639205	5.52	0.000	.5841926	1.226749
SumInvRan~t0	-14050.03	6677.662	-2.10	0.035	-27138	-962.0487
SumIRNSqrtL3	-1779.944	1847.71	-0.96	0.335	-5401.39	1841.501
SumVC0adj50	3.0013	1.188981	2.52	0.012	.6709402	5.331659
SumAdj5Lin2	-.7344037	.4632873	-1.59	0.113	-1.64243	.1736226
SumAlbSal~L1	-.1475219	.092525	-1.59	0.111	-.3288676	.0338237
TotDL0	.196337	.1115482	1.76	0.078	-.0222936	.4149675
SumDLSale~12	.0813112	.2306911	0.35	0.724	-.3708351	.5334575
MSTotHits	-.0076044	.0071125	-1.07	0.285	-.0215445	.0063358
YTTotHits	.0025347	.0055001	0.46	0.645	-.0082452	.0133147
_cons	108374	122722.7	0.88	0.377	-132158	348906.1
SumVC0adj50						
AvgVC0	24.98443	7.852961	3.18	0.001	9.592904	40.37595
NumSupUS0	80.60485	340.4981	0.24	0.813	-586.7592	747.9689
NumForeign0	88.08605	383.0612	0.23	0.818	-662.7	838.8721
TotAlb0	-.0453359	.0596962	-0.76	0.448	-.1623382	.0716664
TotDL0	-.0508058	.0286005	-1.78	0.076	-.1068617	.0052501
MSTotHits	.0045335	.001721	2.63	0.008	.0011603	.0079066
YTTotHits	-.0025438	.0017391	-1.46	0.144	-.0059525	.0008649
SumInvRan~t0	3536.138	1180.513	3.00	0.003	1222.374	5849.901
SumAdj5Lin2	.1832043	.0965828	1.90	0.058	-.0060945	.3725031
SumAlbSal~L1	.0485929	.0164918	2.95	0.003	.0162696	.0809162
_cons	-31114.44	44126.7	-0.71	0.481	-117601.2	55372.31

Table C.16: 3SLS Results: Rock Subsample, Regression 4 (part 2)

MSTotHits						
MSSDvHits	6.509704	.5245941	12.41	0.000	5.481518	7.537889
SumInvRan~t0	137989.9	98284.86	1.40	0.160	-54644.92	330624.7
TotAlb0	7.904263	2.712265	2.91	0.004	2.588321	13.2202
SumAlbSal~L1	-3.031909	1.162074	-2.61	0.009	-5.309532	-.7542848
SumVC0adj50	22.52482	17.12027	1.32	0.188	-11.03029	56.07993
SumAdj5Lin2	4.924174	7.419078	0.66	0.507	-9.616952	19.4653
_cons	-340542.5	2265250	-0.15	0.881	-4780351	4099266
YTTotHits						
YTSDvHits	7.728482	.5318309	14.53	0.000	6.686112	8.770851
SumInvRan~t0	215301.9	101667.8	2.12	0.034	16036.7	414567.2
TotAlb0	5.402754	2.662434	2.03	0.042	.1844795	10.62103
SumAlbSal~L1	-1.081895	1.123873	-0.96	0.336	-3.284647	1.120856
SumVC0adj50	-.007644	16.76821	-0.00	1.000	-32.87274	32.85745
SumAdj5Lin2	7.438218	7.376981	1.01	0.313	-7.020398	21.89684
_cons	-956517.8	2274379	-0.42	0.674	-5414219	3501184
SumInvRan~t0						
NumSongsBB0	10.55412	2.319194	4.55	0.000	6.008582	15.09966
AvgWksBB0	1.568177	.1718933	9.12	0.000	1.231272	1.905081
SumIRNSqrtL3	.0430531	.0340993	1.26	0.207	-.0237802	.1098865
SumVC0adj50	.0000178	.0000292	0.61	0.543	-.0000394	.0000749
SumAdj5Lin2	-9.80e-06	7.68e-06	-1.28	0.202	-.0000249	5.26e-06
TotAlb0	-1.42e-06	2.33e-06	-0.61	0.543	-5.98e-06	3.15e-06
SumAlbSal~L1	-1.35e-07	2.05e-06	-0.07	0.947	-4.15e-06	3.88e-06
TotDL0	4.08e-06	1.81e-06	2.26	0.024	5.44e-07	7.62e-06
SumDLSale~12	2.74e-06	5.03e-06	0.54	0.586	-7.12e-06	.0000126
MSTotHits	-7.46e-08	1.47e-07	-0.51	0.611	-3.62e-07	2.13e-07
YTTotHits	-3.24e-08	9.53e-08	-0.34	0.734	-2.19e-07	1.54e-07
_cons	-23.87362	4.204905	-5.68	0.000	-32.11509	-15.63216

Endogenous variables: TotAlb0 SumVC0adj50 MSTotHits YTTotHits SumInvRankNormSqrt0

Exogenous variables: AvgAlb0 SumIRNSqrtL3 SumAdj5Lin2 SumAlbSalesL1 TotDL0 SumDLSalesPd12 AvgVC0 NumSupUS0 NumForeign0 MSSDvHits YTSDvHits NumSongsBB0 AvgWksBB0

Bibliography

Alexander, P. J. (1994). “Entry barriers, release behavior, and multi-product firms in the music recording industry.” Review of Industrial Organization 9(1): 85-98.

Alexander, P. J. (1994). “New technology and market structure: Evidence from the music recording industry.” Journal of Cultural Economics 18(2): 113-123.

Bhattacharjee, S., R. Gopal, et al. (2006). ”Whatever happened to payola? An empirical analysis of online music sharing.” Decision Support Systems 42(1): 104-120.

Black, G. C., M. A. Fox, et al. (2007). “Concert tour success in North America: An examination of the top 100 tours from 1997 to 2005.” Popular Music and Society 30(2): 149-172.

Bourreau, M. “An Economist’s Guide to Digital Music: A Comment on Peitz and Waelbroeck.” CESifo Economic Studies 51(2-3): 429-33.

Bradlow, E. T. and P. S. Fader (2001). “A Bayesian lifetime model for the ”Hot 100” Billboard songs.” Journal of the American Statistical Association 96(454): 368-381.

Butler, S. (2005). Kazaa Under Scrutiny. Billboard.

Connolly, M. and A. Krueger (2005). Rockonomics: The Economics of Popular Music, Princeton University, Department of Economics, Industrial Relations Section., Working Papers: 499.pdf.

Courty, P. “An Economic Guide to Ticket Pricing in the Entertainment Industry.” Recherches Economiques de Louvain 66(2): 167-92.

Earl, P. E. (2001). “Simon’s travel theorem and the demand for live music.” Journal of Economic Psychology 22(3): 335-358.

Gayer, A. and O. Shy "Copyright Enforcement in the Digital Era."CESifo Economic Studies 51(2-3): 477-89.

Gayer, A. and O. Shy (2006). "Publishers, Artists, and Copyright Enforcement." Information Economics and Policy 18(4): 374-84.

Gopal, R. D., S. Bhattacharjee, et al. (2006). "Do Artists Benefit from Online Music Sharing?" Journal of Business 79(3): 1503-33.

Grassi, I. (2007). The Music Market in the Age of Download, Fondazione Eni Enrico Mattei, Working Papers: 2007.80.

Krueger, A. B. (2005). "The economics of real superstars: The market for rock concerts in the material world." Journal of Labor Economics 23(1): 1-30.

Mixon, F. G. and R. W. Ressler (2000). "A Note on Elasticity and Price Dispersions in the Music Recording Industry." Review of Industrial Organization 17(4): 465-470.

Moreau, N. C. F. (2004). Towards A New Business Model for the Music Industry: Accomodating Piracy through Ancillary Products, Laboratoire dconomtrie Conservatoire National des Arts et Mtiers.

Moreau, N. C. F. (2005). The Music Industry in the Digital Era: Towards New Business Frontiers?, Laboratoire dEconomtrie, Conservatoire National des Arts et Mtiers.

Mortimer, J. H. and A. Sorensen (2005). "Supply Responses to Digital Distribution: Recorded Music and Live Performances." Manuscript. Harvard University.

Oberholzer-Gee, F. and K. Strumpf (2007). "The Effect of File Sharing on Record Sales: An Empirical Analysis." Journal of Political Economy 115(1): 1-42.

Peitz, M. and P. Waelbroeck (2004). An Economist's Guide to Digital Music, CESifo GmbH, CESifo Working Paper Series: CESifo Working Paper No. 1333.

Peitz, M. and P. Waelbroeck (2006). "Why the Music Industry May Gain from Free Downloading–The Role of Sampling." International Journal of Industrial Organization 24(5): 907-13.

Richtel, M. (2001). Judge Orders Napster to Police Trading; Judge Orders Napster to End Unauthorized File Trading. New York Times (1857-Current file). New York, N.Y.: C1.

Rob, R. and J. Waldfogel (2006). “Piracy on the High C’s: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students.” Journal of Law and Economics 49(1): 29-62.

Acknowledgements

First and foremost, I would like to thank and acknowledge my parents. Without their love and support none of this would have been possible. I would also like to sincerely thank the Reed College Economics Department for supporting a project that others doubted. Without the Simpson Fund grant, which they provided, it would not have been possible to gather the data that laid the foundation for this thesis. I would like to thank my advisors, Lon Peters and Joe Roberts, not only for their guidance, but for their support and unwavering patience throughout the year. A crucial acknowledgement goes to my programmers, Noah Pepper and Cooper Francis, who worked tirelessly and beyond their paid hours in order to gather the raw data for this thesis, despite the pressures of their own theses. There is no way to sufficiently thank Noah for the role he played in this thesis; his help was invaluable. He served not only as a programmer, but as the primary sounding board for my ideas. Though neither of us can remember how exactly we became involved in this excessively time-consuming project, nor what we were thinking, it is clear that the empirical aspect of this thesis would not have been conceived without him, nor would it have been possible. I would also like to sincerely thank Cooper, who continued to battle against the more difficult data sources long past his anticipated date of completion, despite my own increasing frustration and lack of appreciation. He provided the most unique data analyzed in this thesis. Finally, I would like to thank my friends for feeding me, telling me when I needed to sleep, and for generally taking care of me.