University Honors Program
University of South Florida
St. Petersburg, Florida

CERTIFICATE OF APPROVAL

Honors Thesis

This is to certify that the Honors Thesis of

Jesse S. Daw

Has been approved by the Examining Committee
on April 29, 2014
as satisfying the thesis requirement
of the University Honors Program

Examining Committee:

S Dieringer

Thesis Director: Suzanne Dieringer, M.A. in Economics
Visiting Professor, College of Business

Nicole Stowell

Thesis Committee Member: Nicole Stowell, JD
Professor, College of Business

Looting on the Digital Sea: An Economic Analysis of Music Piracy

By

Jesse S. Daw

A thesis submitted in partial fulfillment
Of the requirements of the
University Honors Program
University of South Florida, St. Petersburg

April 29, 2014

Thesis Director: Suzanne Dieringer, M.A. in Economics
Visiting Professor, College of Business

Looting on the Digital Sea:

An Economic Analysis of Music Piracy

Jesse Daw

University of South Florida St. Petersburg

Abstract

The development of the internet has brought about a profound shift in the music industry for both record labels and the artists they represent. This shift has come as a result of modern capabilities for music to be hosted online and available to be downloaded by consumers over the internet. These innovations have allowed record labels and independent musicians to sell their music through a new and convenient medium, and services provided by companies such as ITunes and Amazon have introduced more options for consumers purchasing music. However, the ability to access music over the internet has created a new type of criminal who can easily steal or "pirate" music that has been openly shared by others on online file sharing networks.

The purpose of this study is to identify variables that effect consumer demand for illegally downloaded music. In pursuit of this goal, a logit regression model will be used with maximum likelihood estimation to test multiple variables that are expected to affect demand. The variables that will be tested are: age, gender, grade point average, race, annual income bracket, use of streaming services, convenience of illegal downloads, perceived moral obligations, price of music relative to income, and low perceived risk of legal ramifications. By using a logit model to determine how these variables affect the likelihood that consumers will download music illegally, it will become clear how each variable weighs into the consumer decision making process and affects their demand for illegal music downloads.

Looting on the Digital Sea:

An Economic Analysis of Music Piracy

Introduction

The music industry has been forever changed by the development of the modern internet. The internet has ushered in a new and convenient means of acquiring songs through online stores such as ITunes and Amazon which allow consumers to purchase and gain immediate access to the music they want. These online stores have grown in popularity and are "replacing record stores as sources to purchase music" (McCubbin, 2012, p. 327). Music purchased in online stores often comes in the form of MP3 files and these files have replaced the Compact Disk as the most popular medium for music with CD sales having "decreased approximately thirteen percent from their peak in 2002" (McCubbin, 2012, p.327). However, other online services have also developed which allow for the free downloading of music that has been shared by others over the internet. The downloading of copyrighted music over the internet without payment has been dubbed "music piracy" and has become quite common. The illegal sharing and downloading of music is commonly done through Peer to Peer networks which allow for the quick uploading and downloading of music files (Tyler, 2013, p. 2102). Using Peer to Peer networks in order to illegally download music in violation of that music's copyright became a popular practice largely due to the creation of the Napster Peer to Peer network in 1999 (Tyler, 2013, 2102) (Clouse, 2003, p. 109). In the years since Napster was still in operation, many other Peer to Peer Networks have developed with the same general function of providing users with the ability of easily sharing and acquiring copyrighted files (Tyler, 2013, p. 2107).

With the rise of music piracy, record companies have made many attempts and experimented with multiple strategies to decrease music piracy rates. Record companies were previously successful in shutting down many early Peer to Peer networks through litigation in large lawsuits (Tyler, 2003, p. 2107). One example of this is the case *A&M Records v. Napster, Inc.* This case successfully shut down the Napster file sharing network after Napster was found "liable for contributory copyright infringement because Napster provided the material structure to accomplish infringement through its site architecture and knew that users of its site were primarily using it to infringe" (McDonald, 2011, 569). In *MGM Studios Inc. v. Grokster, Ltd.*, The Grokster file sharing network was also shut down when the Supreme Court found the website to be liable for contributory copyright infringement (McDonald, 2011, 569-570). These successes were short-lived however; as the framework for Peer to Peer services was adapted and new services were developed. As a result, lawsuits became much more expensive and harder for copyright owners to win (Tyler, 2003, p. 2107-2108). Record labels and, more specifically, the Recording Industry Association of America also initiated over 30,000 lawsuits against individuals that were participating in music piracy. Record labels and the Recording Industry Association of America abandoned this enforcement option due to the expenses of the lawsuits, their lack of effectiveness at deterring potential music pirates, and the resulting backlash of many music consumers (Tyler, 2003, p. 2108). In addition to bringing lawsuits against individuals who were participating in the illegal downloading of copyrighted music, the Recording Industry Association of America also used threats of legal consequences in attempts to deter potential downloaders (Chiou, Cheng, & Huang, 2011, p. 182-183). Record companies have also attempted to protect their copyrighted music by having internet accounts banned for individuals found pirating music and attempting to protect files from being shared and downloaded using

Digital Rights Management; although consumers of music have found ways to share and download music despite these efforts (Tyler, 2013, p. 2108-2109). Record companies have since backed-off many of these efforts due not only to the high costs, but also because of the consumer backlash that they have caused (Tyler, 2013, p. 2109).

Before the internet became a common medium for procuring music, the music industry operated under a different framework than it does today. In the past, record companies would invest large sums of money in artists and those artists would be paid royalties for each of their albums sold. These royalties paid to the artists would be taken from the money remaining from each album sold after the record company's costs had been recovered (McCubbin, 2012, p. 327). This royalty system remains in effect and has been adapted for music purchased online. To see how online options for purchasing music have drastically affected the way revenue is distributed from album sales, these charts contain a breakdown of album sale distribution for a the sale of a physical CD, and for an album purchased online:

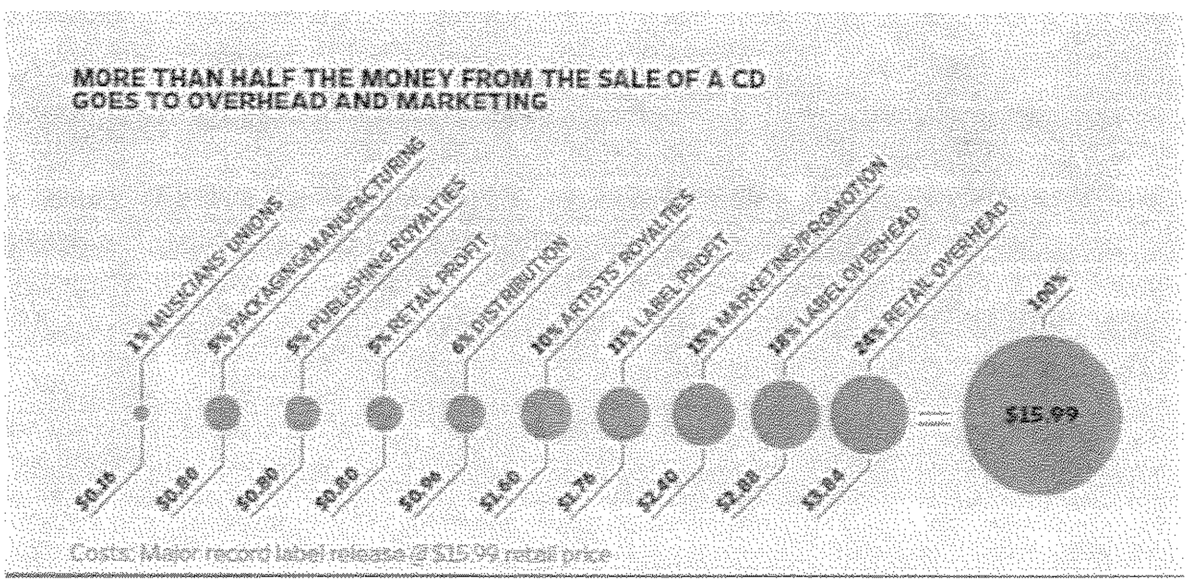

Figure 1. (McCubbin, 2012, p. 328)

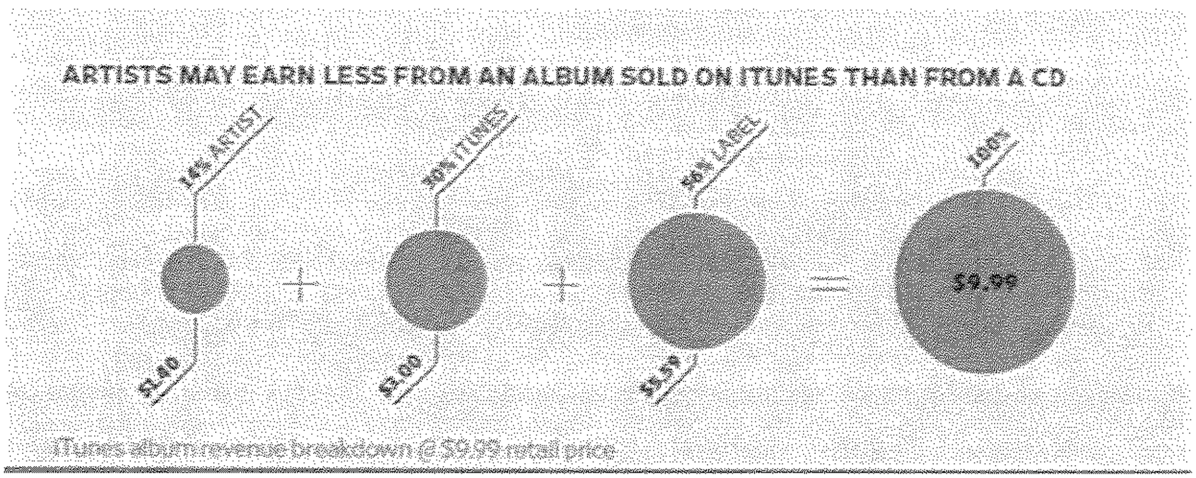

Figure 2. (McCubbin, 2012, p. 328)

Music piracy is an interesting case in which music, a good, can be purchased from a genuine source or simply downloaded for free, albeit illegally. The nature of music as a product effectively make it characteristic of a "quasi-public good in that once the good is provided to some consumers, it is very difficult to preclude other consumers from consuming it" (Gopal, Bhattacharjee, & Sanders, 2006, p. 1504). This quasi-good status of music as a product can cause a free rider effect and "undermine market efficiencies" (Gopal et al, 2006, p. 1504). The ability to share and download music for free brings this issue to a much larger scale where anyone with an internet connection--as opposed to just acquaintances of the music sharer--can obtain the music they are searching for.

This study will not address the question of how music piracy has affected the music industry as a whole, but will instead identify different variables that affect consumer demand for illegally obtained music. The variables that will be tested in this study are: age, gender, grade point average, race, annual income bracket, use of streaming services, convenience of illegal downloads, perceived moral obligations, price of music relative to income, and low perceived risk of legal ramifications. To determine the effects of these variables upon consumer demand, maximum likelihood estimation will be used with a logit regression model.

Literature Review

There is a wealth of literature available relating to the issue of music piracy and how it has affected the music industry. With the rise in popularity of illegally downloading music, the effects of downloading on the recording industry have become quite profound. Illegal downloading of music has become so popular in fact, that some studies estimate that 95% of all music downloaded over the internet is obtained illegally (Tyler, 2013, p. 2102). The practice of music piracy affects both the record labels as well as the music artists themselves. One article which discusses the effects of illegal downloading on the record industry states that as a result of rising piracy rates, "record labels have suffered from consistently decreasing annual revenues [since the 1990's]" (Tyler, 2013, 2102). The article goes on to discuss the actual losses in revenue showing that these losses represent a significant amount of revenue with music sales decreasing by more than half—14.6 billion down to 6.3 billion—from 1999 to 2009" (Tyler, 2013, p. 2102). Another article describes estimates putting the "total commercial value of music piracy" at almost 40 billion dollars as of 2008 (Sheehan, Tsao, & Pokrywczynski, 2012, p. 309). This same article also references a study done by the Institute of Policy Innovation in 2007 which found that the illegal downloading of music was costing the United States economy over 12 billion dollars each year and that it had led to the loss of over 70,000 jobs (Sheehan et al, 2012, p. 302). Research has further shown that online music piracy has caused a 2.7 billion dollar loss worldwide in earnings of music industry personnel (Chiou et al, 2011, p. 182).

There is also information in the existing literature on the effects of music piracy on music artists. One article describes how advancements in modern music production technology have led to a decreased reliance on record companies to front these costs. The article goes on to describe the medium the internet has created for artists to distribute and share their music with large audiences at no cost (McCubbin, 2012, p. 329). This article also explains that music artists generally earn larger revenues from touring and performing live than from actual music sales with the assertion that, "with lower distribution and publicity costs in the Digital Age, artists can spend their resources on touring and increase their revenues through live performance, despite a decline in hard-copy sales" (McCubbin, 2012, p. 329-330). The low percentage of album sales that artists receive as royalties for actual album sales can be seen in figures 1 and 2 presented in the introduction. Due to the lower costs for record companies in the case of digitally distributing music as opposed to the higher costs—such as packaging and shipping—associated with physical album distribution, the article argues that low royalties are no longer justified. In making this argument the article claims that the "low royalties for artists are not justified due to labels no longer [needing] to recover the high distribution costs associated with the CD-based model" (McCubbin, 2012, p. 331).

Research has also shown some characteristics commonly shared by individuals who choose to participate in music piracy. One article states that a large number of college students commonly download music illegally and that "two-thirds of college students who download music do not care whether the music is copyrighted" (Sheehan et al, 2012, p. 309). This article goes on to say that college students were already estimated to be pirating over 1.3 billion songs annually in 2006 (Sheehan et al, 2012, p. 309). Another research study found that young consumers often do not have any "serious moral or ethical concerns" associated with copyright infringement (Gilleti, 2012, 25). This study went on to explain how consumers may justify the illegal downloading of music due to their belief that "information on the internet should be free and accessible to all" (Giletti, 2012, 25).

One possible alternative to downloading music illegally comes in the form of music streaming services. These services allow consumers to listen to music hosted on the internet and either listen to commercial breaks between songs, or pay a subscription fee (Thomes, 2011, 1-2). A study found that an increase in legal risk of illegal music downloader causes an initial increase in demand for free streaming services which results in more advertisements and an eventual overall lower demand for the services (Thomes, 2011, 32-33).

Information is also available on how certain variables affect piracy rates. One study surveyed consumers and used multiple qualitative models in order to find how different consumer responses to artists and their music affected their decisions on whether to legally or illegally obtain music. The results of this study found "an empirical link between individuals' responses to a musical artist (but not to the music itself) and their decision to acquire that music legally or illegally" (Ouellet, 2007, p. 116-117). Another study found that a high probability of getting caught and being punished as perceived by students in the United States and in Taiwan decreased their likelihood of pirating music (Chiou et al, 2011, p. 191-192). Finally, a research study conducted by Theodore Giletti found no significant relationship between the price of mp3 files and "[consumers'] attitudes toward illegal downloading" (2011, 26).

In order to take a new approach to the subject of music piracy, this study does not seek to debate the moral implications of piracy; nor the economic effects it has on the recording industry and individual artists. Instead, this study seeks to estimate consumer demand for illegally downloaded music, and how that demand is affected by multiple variables.

Variables

The variables tested in this study were used in order to seek a deeper economic understanding of what incentives affect participants when they are choosing whether or not to download music illegally. The dependent variable in the study is a binary value with a one indicating that the consumer downloads music illegally, and a zero indicating that the participant has never downloaded music illegally. The independent variables tested in the study are as follows:

There were expectations about the effects of some of these variables before data was collected and analyzed. The variables that were expected to have certain results were: age, annual income bracket, use of streaming services, utility from the convenience of illegal downloads, price of music relative to income, utility from fulfilling perceived moral obligations, and preference for risk aversion regarding the law.

Table 1: Variables

Independent Variable	**Description**
Age	Age of participant at the time they completed the survey
Gender	Male or female
GPA	The college grade point average of the participant.
Race	Race of the participant categorized as: white, black, Hispanic, Asian, or Other.
Annual Income Bracket	The Annual Income of the survey participant with possible brackets of: Under $10,000; $10,001-$15,000; $15,001-$20,000; $20,001-$25,000; and Over $25,000
Use of Streaming Services	Whether or not the consumer uses internet music streaming services such as Pandora or Spotify
Perceived Convenience of illegal music downloads	Whether or not participants have a preference for illegal music downloads because they perceive them to be more convenient.
Price of music relative to consumer income	Whether or not participants believe music prices are too high or that they cannot afford all of the music they wish to acquire
Perceived moral obligations	Whether or not participants believe it is morally wrong to download music illegally
Consumer Preference for risk aversion regarding legal repercussions of downloading music illegally.	Whether or not participants believe the risk of getting caught is high and that the legal penalties for illegal downloading are severe.

It was expected that younger participants would be more likely to download music illegally than older participants. Individuals in low income brackets were also expected to be more likely to download music illegally than individuals in higher income brackets. The use of streaming services was expected to decrease the likelihood that a consumer would download music illegally due to the possibility of music streaming services acting as a substitute for actually owning digital copies of the music. Participants who believe that illegally downloading music is more convenient than purchasing it were expected to be more likely to download music illegally; these participants would have a higher demand for the services they believe are the most convenient. Participants were also expected to be more likely to download music illegally if they believed music prices were high relative to their personal income, or that they should not have to pay for music that so many others already download for free. Finally, consumers with a low preference for risk aversion were expected to have a higher demand for illegal music downloads; It was, however, believed that this variable would have a major impact due to the expectation that many participants would perceive the chances of being caught at all to be minimal.

Methods

Prior to gathering any of the data that was used for this research, the study was submitted to an Institution Review Board (IRB) to seek approval for all methods used in the data collection process. IRB approval was granted for this study after it was determined that the study procedures presented no more than minimal risk to participants and that all data collected would be properly handled and secured according to the appropriate protocols.

The collection of data for this study was accomplished through the anonymous surveying of 197 individuals. Survey participants were primarily undergraduate students at the University of South Florida St. Petersburg campus. College students were chosen for the sample population of this study due in part to the findings of other research which has shown that college students are responsible for a large portion of illegal music downloading (Sheehan et al, 2012, p. 309). Surveys were administered in the classroom to students taking University of South Florida St. Petersburg College of Business Courses. All participants were informed that the surveys would be entirely anonymous and that no identifying information would be collected that could link any of the participants to the survey they had completed. Additionally, an informed consent document was provided with each survey outlining the research study and the participants' part in it. A signed consent document was not collected because it would have been able to link participants back to their survey responses.

The objective of the surveys was to first identify common characteristics about the participants. Questions were asked to determine the age, gender, race, GPA, and income bracket of those who participated in the survey. The next objective of the survey was to discover the online shopping, downloading, and music streaming habits of participants. Lastly, participants were asked if they had ever downloaded any music illegally. Follow up questions were asked related to the reasons why consumers either chose to download music illegally, or chose not to.

Responses from survey participants were entered into a data spreadsheet with each possible response being assigned a number. Once all data was entered, a new spreadsheet was created and data that was deemed irrelevant due to the participant responding that they had never downloaded music was removed. After the removal of this data, 171 observations remained to be analyzed. Descriptive information about this data including percentage breakdowns of survey responses was compiled and can be viewed in Table 2 in the results section. The updated spreadsheet data was then imported to the statistical analysis software STATA.

Before running the statistical tests on the data, the dependent variable had to be created from the responses in the surveys. In order to create this variable, responses to questions 20 and 21 in the survey were used. Participants who responded that they had downloaded music, and that greater than zero percent of the music they downloaded had been acquired illegally, were assigned a 1 for the dependent variable in the data sheet (indicating they have illegally downloaded music). Participants who responded that they had downloaded music, but that zero percent of that music was downloaded illegally, were assigned a value of zero for the dependent variable in the data sheet (indicating that they have not downloaded any music illegally).

Seven of the independent variables used in this study were asked about directly on the survey. Age of the participant was requested in question one, gender in question two, grade point average in question five, race in question seven, annual income bracket in question eight, and use of streaming services in question sixteen. Additionally, question 23a was used to determine whether participants believed it was morally wrong to download music illegally.

The remaining three independent variables were created from survey responses to multiple questions. To create the independent variable of convenience, the answers to survey questions 22a and 22f were grouped together; one of these questions asked if the participant chooses to download music illegally because they believe it is more convenient, and the other asked if the consumer chooses to download music illegally because they can do so conveniently on their phone. The variable measuring the effect of music prices relative to income was created by grouping together the answers to questions 22b, 22e, and 22f from the survey; these questions asked the participants if they believe music prices are too high, if they thought they would not be able to afford purchasing all of the music they want, and if they believed they should not have to pay for music that so many others are already downloading for free. Finally, the variable measuring the effect of perceived risk regarding possible legal ramifications was created by grouping together the answers to questions 22g and 22h from the survey; these questions asked participants if they were afraid of being caught, and if they perceived the legal penalties of being caught downloading music illegally as being severe.

Because the dependent variable in this study has only two possible values-the participant downloads music illegally, or the participant does not-a linear regression using Ordinary Least Squares (OLS) would not be appropriate for analyzing the data. In order to properly analyze the data, maximum likelihood estimation (MLS) was used with a logit regression model. The logit regression model is defined as:

$$P[Y = 1/X] = F(X'\beta) = F(\beta_0 + \beta_1 R_i + \beta_2 Z_i + \varepsilon_i)$$

$$\text{where } F(Z) = \frac{e^Z}{(1+e^Z)}$$

This model is used to find the probability (P) of participants downloading music illegally given values for variables R and Z.

Four separate logit models were analyzed and used for this study. Each of the models tested for the independent variables of age, gender, grade point average, race, annual income bracket, and use of music streaming services. These descriptive variables related to participants in the study are represented by *Z* in the logit function above. Additionally, each model also tested for one of the independent variables related to consumer incentives to download music illegally or purchase it legitimately. This seventh tested variable was thus different in each unique model and is represented by R in the above logit function.

In order to interpret the implications of the results from the logit regression models, the marginal effects of the probability that a consumer will download music illegally were included for each independent variable. Asterisks were used in the logit model tables in order to identify data that was found to be statistically significant at an alpha value of five percent. A percentage value to indicate the amount of data that was correctly classified in each model was also included in tables 3.1 and 3.2.

Results

Preliminary tests on the data collected from the 197 surveys found that 118 participants admitted they had downloaded music illegally while only 53 stated that they had downloaded music, but not illegally. An additional 26 participants claimed to have never downloaded music and were removed from the dataset. This shows that in the sample for this study, 31% of participants purchase all of their music legitimately, while 69% of participants download at least a portion of the music they acquire illegally. Additional descriptive statistics about the data are summarized in table 2.

Relevant data from the four logit models is presented in the tables 3.1 and 3.2. The tables give the logit estimates and the marginal effects of each variable tested. Asterisks in the data table signify results found to be statistically significant at an alpha value of five percent.

Table 2 Descriptive Statistics		
Variable	**Mean**	**Range**
Age (Q1)	22	18-52
GPA (Q5)	3.18	2.00-4.00
Gender (Q2)	**Percent**	
Male	50.6%	
Female	49.4%	
Race (Q7)		
White	72.9%	
Black	7.7%	
Hispanic	10.3%	
Asian	5.8%	
Other	3.2%	
Income (Q8)		
Under $10,000	50.0%	
$10,001-$15,000	15.9%	
$15,001-$20,000	8.5%	
$20,001-$25,000	6.7%	
Over $25,000	18.9%	
Use Streaming Services (Q16)	90.6%	
Believe illegal downloads are more convenient (Q22af)	52.6%	
Believe music price is high relative to income (Q22bef)	59.6%	
Believe there is a moral obligation to pay for music (Q23a)	24.6%	
Not concerned about legal risk (Q22gh)	27.5%	
number of observations = 171		

Table 3.1 Data From Logit Model				
Variable	Model 1		Model 2	
	Logit Estimates	Marginal Effect	Logit Estimates	Marginal Effects
Age	-.130* (.053)	-2.1	-.153* (.057)	-2.6
Gender	-.522 (.452)	-8.6	-.528 (.461)	-9.0
Grade Point Average	.281 (.551)	4.6	.454 (.559)	7.7
Race	.201 (.218)	3.3	-.178 (.211)	-3.0
Annual income Bracket	.088 (.173)	1.5	.243 (.184)	4.1
Use of Streaming Services	-.507 (.820)	-7.4	-.509 (.768)	-7.6
Convenience of illegal downloads	2.377* (.517)	40.1		
Price of Music relative to consumer income			2.370* (.493)	45.0
Perceived moral obligations				
Perceived low risk of legal ramifications				
Correctly Classified	77.78%		80.56%	

Table 3.2 Data From Logit Model (Continued)

Variable	Model 3		Model 4	
	Logit Estimates	Marginal Effect	Logit Estimates	Marginal Effects
Age	-.117* (.054)	-2.2	-.101 (.055)	-1.9
Gender	-.383 (.414)	-7.2	-.199 (.410)	-3.7
Grade Point Average	.186 (.507)	3.5	.127 (.506)	2.4
Race	.020 (.192)	0.4	.004 (.193)	0.1
Annual income Bracket	-.062 (.156)	-1.2	-.028 (.155)	-0.5
Use of Streaming Services	-.283 (.786)	-5.0	-.281 (.745)	-4.9
Convenience of illegal downloads				
Price of Music relative to consumer income				
Perceived moral obligations	-1.393* (.449)	-30.0		
Perceived low risk of legal ramifications			1.359 (.590)	21.7
Correctly Classified	75.69%		73.61%	

Certain variables tested in this study were not found to have statistically significant results in any of the four logit models at an alpha value of five percent. Gender, grade point average, and race were not found to be statistically significant variables for determining the likelihood that a participant downloads music illegally. The marginal effect for the income variable was expected to be negative—indicating a decrease in likelihood of downloading music illegally for participants in higher income brackets—but was actually found to be positive (see tables 3.1 and 3.2). However, the results for this income variable were found to not be statistically significant. Participants' use of streaming services was expected to decrease their likelihood of downloading music illegally and this was supported by the negative marginal effect of this variable shown in all four tables; although, these results were also found to not be statistically significant and are therefore not conclusive.

The age of participants in this study was found to be a statistically significant predictor for illegal music downloads in logit models one, two, and three. The age range of participants in this study was between 18 and 52 (see table 2). As expected, the marginal effect of this variable was negative indicating a decreased likelihood of illegally downloading music for older participants. The marginal effect values for this variable were found to be -2.1%, -2.6%, and -2.2% in tables one, two, and three respectively. These values show that for each additional year of age, participants were this percent less likely to download music illegally. This is also consistent with the findings of a previous study which showed that younger consumers have a more favorable view of music piracy than older consumers (Giletti, 2012, 2).

The first logit regression model tested the convenience of illegal music downloads. Convenience was found to be statistically significant and the model shows that this variable has a marginal effect of 40.1%. This result indicates that consumers who view illegal music downloads as being more convenient than purchasing the music legally were 40.1% more likely to illegally download music. This result is consistent with the expectation that convenience would increase demand and that participants would choose to download music illegally if they believed it was more convenient than purchasing the music. The large marginal effect of this variable shows that consumers value convenience and that convenience is a major factor when consumers are deciding how they will acquire music.

Price of music relative to consumer income was tested in the second logit model and was also found to be a statistically significant variable at an alpha value of five percent. The marginal effect of this variable was found to be 45% indicating that consumers who considered music prices to be high relative to their income were 45% more likely to download music illegally rather than purchase it. This result reinforces the expectations about the effects of this variable, and the large marginal effect--the largest of any variable in the study--reveals that price is a major factor for consumers when they are deciding whether to purchase music or download it illegally. The results found for this variable differ from the results of the Theodore Giletti study which discovered no relationship between the price of music and consumers' "attitude towards illegal downloading" (2012, 26).

Perceived moral obligations was tested in the third logit model and emerged as a statistically significant predictor of the likelihood that a consumer will download music illegally. Significance of this variable was established at an alpha value of five percent. The marginal effect of this variable was found to be -30% which indicates that consumers who believe it is morally wrong to download music without paying for it were 30% less likely to pirate music. This finding is consistent with the expectation that consumers who believe they have a moral obligation to pay for music will have a lower demand for illegal music downloads.

Low perceived risk with regards to the law was tested in the fourth logit model and was found to be a statistically significant predictor of the likelihood that consumers will choose to download music illegally. This variable was found to be statistically significant at an alpha value of 5% and the marginal effect of this preference for risk aversion was found to be 21.7%. This marginal effect means that participants who have a low preference for risk aversion and do not fear the legal ramifications that could result from pirating music were 21.7% more likely to download music illegally. This variable was not expected to be a major factor in consumers' decision whether or not to download music illegally due to the expectation that most participants would not believe there was a significant chance of getting caught. However, only 27.5% of respondents were found to be unafraid of the legal risks (see table 2). This percent and the marginal effect of 21.7% show that participants did in fact give consideration to the law and had a higher demand for illegal music downloads if they had a low preference for risk aversion.

The results found in this study are not without their limitations. Some of the variables in the models were not found to have statistically significant results. This lack of significance for some of the variables may be a result of the lack of variation in the sample surveyed for this study. Additionally, the variables that were constant across the four models had notable differences in their coefficients and in their marginal effects. Because these variables did not change in any of the models, the coefficient and marginal effect values should have been similar in each model. Further statistical tests will be required in order to verify the validity of the results of this study and determine the reasons for these discrepancies.

Conclusion

This Study sought to identify variables that have an effect on consumer demand for illegal music downloads. Multiple variables were tested using maximum likelihood estimation with a logit model and the results of these tests are summarized in tables 3.1 and 3.2.

Of the ten variables tested, five were found to have statistically significant results. Older participants were found to be less likely to download music illegally than younger survey participants. Convenience was found to be an important factor influencing consumer demand for illegal music downloads; participants who believed illegal downloads to be more convenient were more likely to download music illegally. Consumers also had a higher demand for illegal music downloads and were more likely to pirate music if they believed the price of music to be high relative to their personal income, or if they were not risk averse and were not afraid of possible legal repercussions. Finally, participants were found to have a lower demand for illegal music downloads if they believed they had moral obligations to pay for the music they acquire.

Many additional statistical tests must still be run to determine the validity of the results of this study and explain the lack of significance found in the results for certain variables. More tests will also be necessary in order to expand upon the findings and make further assertions about how consumer demand for illegal music downloads is affected by each variable. Future studies may benefit by choosing a larger and more diverse sample population. Additionally, future research studies could work to better determine the relationship between music streaming services and consumer demand for illegal music downloads due to the statistically insignificant results for the variable in this study. Finally, future research may be needed in order to identify even more variables that effect consumer demand and their decision whether or not to pirate music.

References

Chiou, J., Cheng, H., & Huang, C. (2011). The Effects of Artist Adoration and Perceived Risk of Getting Caught on Attitude and Intention to Pirate Music in the United States and Taiwan. *Ethics & Behavior*, *21*(3), 182-196. doi:10.1080/10508422.2011.570163

Clouse, L. (2003). Virtual Border Customs: Prevention of International Online Music Piracy within the Ever-Evolving Technological Landscape. *Valparaiso University Law Review*, *38*(1), 109-164.

Giletti, T. (2012). *Why Pay if it's free? Streaming, downloading, and digital music consumption in the "iTunes era"* (Masters Dissertation). Retrieved from MEDIA@LSE Electronic MSc dissertation series

Gopal, R. D., Bhattacharjee, S., & Sanders, G. (2006). Do Artists Benefit from Online Music Sharing? *Journal Of Business*, *79*(3), 1503-1533.

McCubbin, M. (2012). The Aftermath of Aftermath: The Impact of Digital Music Distribution on the Recording Industry. *University Of New Hampshire Law Review*, *10*(2), 323-343.

McDonald, V. R. (2011). Stirring the waters: whether the Pirate Bay case and the Thomas-Rasset case will impact file sharing and piracy in Sweden and the United States. *Transnational Law & Contemporary Problems*, (2), 569-570.

Ouellet, J. (2007). The Purchase Versus Illegal Download of Music by Consumers: The Influence of Consumer Response towards the Artist and Music. *Canadian Journal Of Administrative Sciences (Canadian Journal Of Administrative Sciences)*, *24*(2), 116-119. doi:10.1002/CJAS.16

Sheehan, B., Tsao, J., & Pokrywczynski, J. (2012). Stop the Music!. *Journal Of Advertising Research, 52*(3), 309-321. doi:10.2501/JAR-52-3-309-321 S

Thomes, T. P. (2011). An economic analysis of online streaming: How the music industry can generate revenues from cloud computing. *ZEW Discussion Papers, No. 11-039 [rev.]*

Tyler, N. S. (2013). Music Piracy and Diminishing Revenues: How Compulsory Licensing for Interactive Webcasters can Lead the Recording Industry Back to Prominence. *University Of Pennsylvania Law Review, 161*(7), 2101-2150.

Appendix A: Survey used for data collection.

INFORMED CONSENT TO PARTICIPATE IN RESEARCH

Information to Consider Before Taking Part in this Research Study
IRB Study # Pro00016573

The University of South Florida St. Petersburg encourages undergraduate students to engage in research to further their learning experience and benefit the community. To do this, we need the help of volunteers willing to participate in research studies. This form will provide information on this study.

We are asking you to participate in a research study titled: Looting on the Digital Sea, An Economic Analysis of Music Piracy.

The person in charge of this research study is Jesse Daw, This person is known as the Principal Investigator. This research will be done by collecting your responses to an anonymous survey.

PURPOSE OF THE STUDY

The purpose of this study is to estimate consumer demand for music under copyright downloaded without payment. You are being asked because college students make up a large portion of the music downloading population.

STUDY PROCEDURES

If you take part in this study you will be asked to complete an anonymous survey related to the practice of downloading music and the alternatives to downloading music.

VOLUNTARY PARTICIPATION/WITHDRAWAL

You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study. You are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study. Your decision whether or not to take part in this research will not affect your course grade or standing at the University in any way.

ALTERNATIVES

You have the alternative to choose not to participate in this research.

BENEFITS

The findings of this research may help encourage better services for providing music to consumers in the future.

RISKS OR DISCOMFORT

This research is considered to be minimal risk. That means that the risks associated with this study are the same as what you face every day. There are no known additional risks to those who take part in this study.

COMPENSATION

We will not pay you for the time you volunteer while being in this study.

CONFIDENTIALITY

All Surveys will be entirely anonymous and there will be no identifying information linking your responses to your identity. All surveys will be kept in a locked file cabinet and data will be stored in digital format on a computer with a password known only to the study staff (Collected data may be stored for five years after submission of the final report to the IRB.). However, certain people may require access to the anonymous data . The only people who will be allowed to see the study records are:

- The research team, including the Principal Investigator, the Advising Professor, and all other research staff.
- Certain government and university people who need to know more about the study. For example, individuals who provide oversight on this study may need to look at the records. This is done to make sure that we are doing the study in the right way. They also need to make sure that we are protecting your rights and your safety. These include:
    - The University of South Florida Institutional Review Board (IRB) and the staff that work for the IRB. Other individuals who work for USF that provide other kinds of oversight may also need to look at the study records.
    - The Department of Health and Human Services (DHHS).

Additionally, the study data, including your data, are not protected from disclosure to legal authorities if those authorities request them (Although all data gathered in the surveys will be anonymous).

We may publish what we learn from this study. If we do, all data will remain completely anonymous and there will be nothing in the study that would let anyone know who you are.

The University of South Florida St. Petersburg encourages undergraduate students to engage in research to further their learning experience and benefit the community. Please complete the short survey below as part of an undergraduate research project.

<u>All surveys are anonymous.</u> Please do NOT include your U# or name on the survey. Your participation is strictly voluntary and confidential. Your participation will not affect your class grade or status in anyway.

Thank you for supporting student research!

1. Age__________
2. Male__________ Female __________
3. Freshman__________ Sophomore__________ Junior__________ Senior __________
4. Intended major____________________
5. GPA __________
6. Nationality ____________________
7. Race ____________________
8. Annual Income:

   Under $10,000__________ $10,001-15,000__________ $15,001-20,000 __________

   $20,001-25,000__________ Over $25,000__________
9. Do you own a computer?

   Yes__________ No __________
10. Do you have internet service available at your home/residence?

    Yes__________ No __________
11. Where do you use the internet <u>most often</u>?
    a) Home __________
    b) School __________
    c) Friends' house__________
    d) Work __________
    e) Wifi hotspot e.g. coffee shop, restaurant __________
    f) Other ____________________
12. Do you shop on the internet or have you ever made any kind of purchase online (*Amazon, eBay, retail store, etc.*)?

    Yes__________ No __________
13. If yes, how many times in the past 30 days?

    0__________ 1-5__________ 6-10 __________ more than 10__________

14. Have you ever downloaded a movie off the internet?

Yes__________ No __________

15. If yes,

a. How many times in the past 30 days?

0 __________ 1-5__________ 6-10 __________ more than 10__________

b. What percentage of movies have you downloaded that you were supposed to pay for but didn't?

0%__________ 1-25%__________ 26-50%__________

51-75% __________ 76-99% __________ 100% __________

16. Do you use any music streaming services?

Yes __________ No __________

17. If yes,

a. How many times in the past 30 days?

0 __________ 1-5__________ 6-10 __________ more than 10__________

b. Select which music streaming service do you use the most.

Spotify __________ Xbox Music __________

Pandora __________ Grooveshark__________

Songza__________ Soundcloud__________

Last.fm__________ Other ________________________________

18. Have you ever purchased music through an online service (Amazon, iTunes, etc.)?

Yes __________ No __________

19. If yes,

a. How many times in the past 30 days?

0 __________ 1-5__________ 6-10 __________ more than 10__________

b. What (average) price did you pay per song? __________

20. Have you ever downloaded music over the internet?

Yes __________ No __________

21. If yes,

a. How many times in the past 30 days?

0 _______ 1-5_______ 6-10 _______ more than 10_____

b. What percentage of music do you download that you were supposed to pay for but didn't?

0% __________ *(skip question 22)* 1-25% __________ 26-50% __________

51-75% _______ 76-99% __________ 100% __________ *(skip question 23)*

22. For what reason(s) have downloaded music that you were supposed to pay for but didn't?
Select all that apply.

It's more convenient than purchasing the music__________

I refuse to pay the price because it's not worth it to me __________

The artist is successful enough that it won't hurt him/her__________

I do not believe it is wrong to download it __________

I could not afford to purchase all of the music I would want to download__________

I can conveniently download music through an app on my phone__________

I do not believe there is a high risk of getting caught__________

I do not believe the consequences are very severe if I do get caught__________

I do not want to support the record company __________

Why should I pay when everyone else gets it for free? __________

Other reason ____________________

23. For what reason(s) have you purchased music off the internet rather than downloading it without paying?
Select all that apply.

It's the right thing to do ___________

I'm afraid I would get caught __________

The legal penalties are too severe to risk getting caught__________

I want to support artists whose music I enjoy__________

I want to support the record company __________

I'm concerned about causing harm to my computer, e.g. virus __________

I would feel guilty toward the artist__________

I would feel guilty toward the record label__________

There is no need to download music since there are many music streaming services available _______

Other reason ____________________

Appendix B: Data Output from STATA

```
. global xlist2 q1 q2 q5 q7 q8 q16 q22bej

. global xlist3 q1 q2 q5 q7 q8 q16 q22gh

. global xlist4 q1 q2 q5 q7 q8 q16 q23a

.
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
obs	171	96.30409	56.21284	1	195
q1	169	22.34911	4.656282	18	52
q2	170	1.494118	.5014424	1	2
q5	161	3.179752	.4257155	2	4
q7	155	1.587097	1.091806	1	5
q8	164	2.286585	1.577313	1	5
q16	171	.9064327	.292081	0	1
q20c	171	.6900585	.4638274	0	1
q22af	171	.5263158	.5007734	0	1
q22bej	171	.5964912	.4920419	0	1
q22gh	171	.2748538	.4477517	0	1
q23a	171	.245614	.4317148	0	1

```
.
. tabulate $ylist
```

Q20C	Freq.	Percent	Cum.
0	53	30.99	30.99
1	118	69.01	100.00
Total	171	100.00	

```
.
. * Logit model
. logit $ylist $xlist1

Iteration 0:   log likelihood =  -85.081283
Iteration 1:   log likelihood =  -66.577941
Iteration 2:   log likelihood =  -65.290721
Iteration 3:   log likelihood =  -65.280387
Iteration 4:   log likelihood =  -65.280383

Logistic regression                               Number of obs   =        144
                                                  LR chi2(    7)      =      39.60
                                                  Prob > chi2     =     0.0000
Log likelihood =  -65.280383                       Pseudo R2       =     0.2327
```

q20c	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
q1	-.1298596	.053292	-2.44	0.015	-.2343099	-.0254092
q2	-.5220746	.4520543	-1.15	0.248	-1.408085	.3639356
q5	.2805447	.551422	0.51	0.611	-.8002225	1.361312
q7	.2013367	.2176029	0.93	0.355	-.2251572	.6278306
q8	.0884006	.1727369	0.51	0.609	-.2501575	.4269586
q16	-.5067152	.819746	-0.62	0.536	-2.113388	1.099957
q22af	2.376726	.5165174	4.60	0.000	1.364371	3.389082
_cons	2.813726	2.59345	1.08	0.278	-2.269343	7.896795

```
26 . test $xlist1

 ( 1)  [q20c]q1 = 0
 ( 2)  [q20c]q2 = 0
 ( 3)  [q20c]q5 = 0
 ( 4)  [q20c]q7 = 0
 ( 5)  [q20c]q8 = 0
 ( 6)  [q20c]q16 = 0
 ( 7)  [q20c]q22af = 0

           chi2(  7) =   27.77
         Prob > chi2 =    0.0002

27 . logit $ylist $xlist2

Iteration 0:   log likelihood =  -85.081283
Iteration 1:   log likelihood =  -65.926509
Iteration 2:   log likelihood =  -64.852084
Iteration 3:   log likelihood =  -64.846014
Iteration 4:   log likelihood =  -64.846013

Logistic regression                               Number of obs   =        144
                                                  LR chi2(    7)      =      40.47
                                                  Prob > chi2     =     0.0000
Log likelihood =  -64.846013                       Pseudo R2       =     0.2378
```

q20c	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
q1	-.1529456	.0570564	-2.68	0.007	-.264774	-.0411171
q2	-.528012	.4609869	-1.15	0.252	-1.43153	.3755056
q5	.4543634	.5592767	0.81	0.417	-.6417988	1.550526
q7	-.1778955	.2111625	-0.84	0.400	-.5917664	.2359754
q8	.2431307	.1843705	1.32	0.187	-.1182289	.6044903
q16	-.5091062	.7675301	-0.66	0.507	-2.013438	.9952252
q22bej	2.370222	.4934234	4.80	0.000	1.40313	3.337314
_cons	2.736627	2.463178	1.11	0.267	-2.091113	7.564368

```
28 . test $xlist2

 ( 1)  [q20c]q1 = 0
 ( 2)  [q20c]q2 = 0
 ( 3)  [q20c]q5 = 0
 ( 4)  [q20c]q7 = 0
 ( 5)  [q20c]q8 = 0
 ( 6)  [q20c]q16 = 0
 ( 7)  [q20c]q22bej = 0

           chi2(  7) =   29.91
         Prob > chi2 =    0.0001

29 . logit $ylist $xlist3

Iteration 0:   log likelihood =  -85.081283
Iteration 1:   log likelihood =  -75.911177
Iteration 2:   log likelihood =  -75.499815
Iteration 3:   log likelihood =  -75.496939
Iteration 4:   log likelihood =  -75.496939

Logistic regression                               Number of obs   =        144
                                                  LR chi2(    7)      =      19.17
                                                  Prob > chi2     =     0.0077
Log likelihood =  -75.496939                       Pseudo R2       =     0.1126
```

q20c	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
q1	-.1013137	.0546975	-1.85	0.064	-.2085188	.0058914
q2	-.1991281	.4096345	-0.49	0.627	-1.001997	.6037408
q5	.1274893	.5059911	0.25	0.801	-.864235	1.119214
q7	.0043171	.1926544	0.02	0.982	-.3732785	.3819127
q8	-.0282565	.154746	-0.18	0.855	-.3315532	.2750401
q16	-.2807442	.7452978	-0.38	0.706	-1.741501	1.180013
q22gh	1.359204	.589602	2.31	0.021	.2036053	2.514803
_cons	3.182647	2.380757	1.34	0.181	-1.483551	7.848844

```
30 . test $xlist3

 ( 1)  [q20c]q1 = 0
 ( 2)  [q20c]q2 = 0
 ( 3)  [q20c]q5 = 0
 ( 4)  [q20c]q7 = 0
 ( 5)  [q20c]q8 = 0
 ( 6)  [q20c]q16 = 0
 ( 7)  [q20c]q22gh = 0

           chi2(  7) =   14.73
         Prob > chi2 =    0.0396

31 . logit $ylist $xlist4

Iteration 0:   log likelihood = -85.081283
Iteration 1:   log likelihood = -74.095996
Iteration 2:   log likelihood = -73.874265
Iteration 3:   log likelihood = -73.873928
Iteration 4:   log likelihood = -73.873928

Logistic regression                               Number of obs   =        144
                                                  LR chi2(    7)      =      22.41
                                                  Prob > chi2     =     0.0022
Log likelihood =  -73.873928                       Pseudo R2       =     0.1317
```

q20c	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
q1	-.1170811	.0538332	-2.17	0.030	-.2225921	-.01157
q2	-.3825011	.4140361	-0.92	0.356	-1.193997	.4289947
q5	.1857743	.5078598	0.37	0.715	-.8096126	1.181161
q7	.0200155	.1919763	0.10	0.917	-.3562511	.3962821
q8	-.0615264	.1564837	-0.39	0.694	-.3682288	.2451761
q16	-.2826219	.7857071	-0.36	0.719	-1.82258	1.257336
q23a	-1.392558	.4492119	-3.10	0.002	-2.272997	-.5121188
_cons	4.369233	2.417493	1.81	0.071	-.3689664	9.107433

```
32 . test $xlist4

 ( 1)  [q20c]q1 = 0
 ( 2)  [q20c]q2 = 0
 ( 3)  [q20c]q5 = 0
 ( 4)  [q20c]q7 = 0
 ( 5)  [q20c]q8 = 0
 ( 6)  [q20c]q16 = 0
 ( 7)  [q20c]q23a = 0

           chi2(  7) =   18.46
         Prob > chi2 =    0.0100
```

```
33 .
34 . * Marginal effects (at the mean and average marginal effect)
35 . quietly logit $ylist $xlist1

36 . mfx, dydx at(mean)

Marginal effects after logit
      y  = Pr(q20c) (predict)
         =  .79142144
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
q1	-.0214364	.00885	-2.42	0.015	-.038783	-.00409	22.3403
q2	-.0861807	.07446	-1.16	0.247	-.232128	.059766	1.5
q5	.0463105	.09097	0.51	0.611	-.131997	.224618	3.18396
q7	.0332354	.03583	0.93	0.354	-.03699	.103461	1.56944
q8	.0145926	.02846	0.51	0.608	-.041184	.07037	2.27778
q16*	-.0737068	.10355	-0.71	0.477	-.276658	.129245	.909722
q22af*	.4012162	.07351	5.46	0.000	.257145	.545288	.527778

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
37 . mfx, dydx

Marginal effects after logit
      y  = Pr(q20c) (predict)
         =  .79142144
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
q1	-.0214364	.00885	-2.42	0.015	-.038783	-.00409	22.3403
q2	-.0861807	.07446	-1.16	0.247	-.232128	.059766	1.5
q5	.0463105	.09097	0.51	0.611	-.131997	.224618	3.18396
q7	.0332354	.03583	0.93	0.354	-.03699	.103461	1.56944
q8	.0145926	.02846	0.51	0.608	-.041184	.07037	2.27778
q16*	-.0737068	.10355	-0.71	0.477	-.276658	.129245	.909722
q22af*	.4012162	.07351	5.46	0.000	.257145	.545288	.527778

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
38 .
39 . quietly logit $ylist $xlist2

40 . mfx, dydx at(mean)

Marginal effects after logit
      y  = Pr(q20c) (predict)
         =  .7830859
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
q1	-.0259797	.00973	-2.67	0.008	-.045048	-.006912	22.3403
q2	-.0896894	.07739	-1.16	0.246	-.241369	.061991	1.5
q5	.0771793	.09419	0.82	0.413	-.107439	.261797	3.18396
q7	-.0302178	.03569	-0.85	0.397	-.100175	.03974	1.56944
q8	.0412988	.03103	1.33	0.183	-.019519	.102116	2.27778
q16*	-.0763823	.1009	-0.76	0.449	-.274151	.121386	.909722
q22bej*	.4495621	.085	5.29	0.000	.28296	.616164	.631944

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
41 . mfx, dydx

   Marginal effects after logit
         y  = Pr(q20c) (predict)
            =    .7830859
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
q1	-.0259797	.00973	-2.67	0.008	-.045048	-.006912	22.3403
q2	-.0896894	.07739	-1.16	0.246	-.241369	.061991	1.5
q5	.0771793	.09419	0.82	0.413	-.107439	.261797	3.18396
q7	-.0302178	.03569	-0.85	0.397	-.100175	.03974	1.56944
q8	.0412988	.03103	1.33	0.183	-.019519	.102116	2.27778
q16*	-.0763823	.1009	-0.76	0.449	-.274151	.121386	.909722
q22bej*	.4495621	.085	5.29	0.000	.28296	.616164	.631944

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
42 .
43 . quietly logit $ylist $xlist3

44 . mfx, dydx at(mean)

   Marginal effects after logit
         y  = Pr(q20c) (predict)
            =   .75387484
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
q1	-.0187985	.01031	-1.82	0.068	-.039001	.001404	22.3403
q2	-.0369477	.07601	-0.49	0.627	-.185928	.112032	1.5
q5	.0236553	.09386	0.25	0.801	-.160304	.207615	3.18396
q7	.000801	.03575	0.02	0.982	-.06926	.070862	1.56944
q8	-.0052429	.0287	-0.18	0.855	-.061489	.051003	2.27778
q16*	-.0490103	.12186	-0.40	0.688	-.287847	.189827	.909722
q22gh*	.2170461	.07397	2.93	0.003	.072059	.362033	.298611

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
45 . mfx, dydx

   Marginal effects after logit
         y  = Pr(q20c) (predict)
            =   .75387484
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
q1	-.0187985	.01031	-1.82	0.068	-.039001	.001404	22.3403
q2	-.0369477	.07601	-0.49	0.627	-.185928	.112032	1.5
q5	.0236553	.09386	0.25	0.801	-.160304	.207615	3.18396
q7	.000801	.03575	0.02	0.982	-.06926	.070862	1.56944
q8	-.0052429	.0287	-0.18	0.855	-.061489	.051003	2.27778
q16*	-.0490103	.12186	-0.40	0.688	-.287847	.189827	.909722
q22gh*	.2170461	.07397	2.93	0.003	.072059	.362033	.298611

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
46 .
47 . quietly logit $ylist $xlist4

48 . mfx, dydx at(mean)

   Marginal effects after logit
         y  = Pr(q20c) (predict)
            =  .74770136
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      q1 |  -.0220866      .01026   -2.15   0.031   -.04219 -.001983   22.3403
      q2 |  -.0721565      .07785   -0.93   0.354  -.224733  .080419       1.5
      q5 |   .0350452      .09576    0.37   0.714  -.152633  .222723   3.18396
      q7 |   .0037758      .03622    0.10   0.917  -.067221  .074773   1.56944
      q8 |  -.0116066      .02947   -0.39   0.694   -.06936  .046147   2.27778
    q16* |  -.0502067      .13079   -0.38   0.701  -.306546  .206133   .909722
   q23a* |  -.2998181      .10101   -2.97   0.003  -.497792 -.101844   .229167
------------------------------------------------------------------------------
   (*) dy/dx is for discrete change of dummy variable from 0 to 1

49 . mfx, dydx

   Marginal effects after logit
         y  = Pr(q20c) (predict)
            =  .74770136
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      q1 |  -.0220866      .01026   -2.15   0.031   -.04219 -.001983   22.3403
      q2 |  -.0721565      .07785   -0.93   0.354  -.224733  .080419       1.5
      q5 |   .0350452      .09576    0.37   0.714  -.152633  .222723   3.18396
      q7 |   .0037758      .03622    0.10   0.917  -.067221  .074773   1.56944
      q8 |  -.0116066      .02947   -0.39   0.694   -.06936  .046147   2.27778
    q16* |  -.0502067      .13079   -0.38   0.701  -.306546  .206133   .909722
   q23a* |  -.2998181      .10101   -2.97   0.003  -.497792 -.101844   .229167
------------------------------------------------------------------------------
   (*) dy/dx is for discrete change of dummy variable from 0 to 1

50 .
51 . * Predicted probabilities
52 . quietly logit $ylist $xlist1

53 . predict plogit1, pr
   (27 missing values generated)

54 .
55 . quietly logit $ylist $xlist2

56 . predict plogit2, pr
   (27 missing values generated)

57 .
58 . quietly logit $ylist $xlist3
```

```
59 . predict plogit3, pr
   (27 missing values generated)

60 .
61 . quietly logit $ylist $xlist4

62 . predict plogit4, pr
   (27 missing values generated)

63 .
64 . summarize $ylist plogit1 plogit2 plogit3 plogit4
```

Variable	Obs	Mean	Std. Dev.	Min	Max
q20c	171	.6900585	.4638274	0	1
plogit1	144	.7222222	.2302374	.1547588	.9758505
plogit2	144	.7222222	.2368131	.1542459	.9743905
plogit3	144	.7222222	.1610829	.0920014	.9443599
plogit4	144	.7222222	.1806881	.1120657	.9146928

```
65 .
66 . * Post test, fitness of the model and percent correctly predicted values
67 . quietly logit $ylist $xlist1

68 . estat classification
```

Logistic model for q20c

Classified	True D	True ~D	Total
+	90	18	108
-	14	22	36
Total	104	40	144

Classified + if predicted Pr(D) >= .5
True D defined as q20c != 0

Sensitivity	Pr(+| D)	86.54%
Specificity	Pr(-|~D)	55.00%
Positive predictive value	Pr(D| +)	83.33%
Negative predictive value	Pr(~D| -)	61.11%
False + rate for true ~D	Pr(+|~D)	45.00%
False - rate for true D	Pr(-| D)	13.46%
False + rate for classified +	Pr(~D| +)	16.67%
False - rate for classified -	Pr(D| -)	38.89%
Correctly classified		77.78%

```
69 . estat gof
```

Logistic model for q20c, goodness-of-fit test

```
       number of observations =       144
 number of covariate patterns =       141
             Pearson chi2( 133) =     143.66
                   Prob > chi2 =       0.2489
```

70 . lroc

Logistic model for q20c

number of observations = 144
area under ROC curve = 0.8221

71 .
72 . quietly logit $ylist $xlist2

73 . estat classification

Logistic model for q20c

Classified	True D	True ~D	Total
+	93	17	110
-	11	23	34
Total	104	40	144

Classified + if predicted Pr(D) >= .5
True D defined as q20c != 0

Sensitivity	Pr(+| D)	89.42%
Specificity	Pr(-|~D)	57.50%
Positive predictive value	Pr(D| +)	84.55%
Negative predictive value	Pr(~D| -)	67.65%
False + rate for true ~D	Pr(+|~D)	42.50%
False - rate for true D	Pr(-| D)	10.58%
False + rate for classified +	Pr(~D| +)	15.45%
False - rate for classified -	Pr(D| -)	32.35%
Correctly classified		80.56%

74 . estat gof

Logistic model for q20c, goodness-of-fit test

number of observations = 144
number of covariate patterns = 139
Pearson chi2(131) = 139.49
Prob > chi2 = 0.2897

75 . lroc

Logistic model for q20c

number of observations = 144
area under ROC curve = 0.8186

76 .

```
77 . quietly logit $ylist $xlist3

78 . estat classification

   Logistic model for q20c

              -------- True --------
   Classified |         D            ~D  |      Total
   -----------+--------------------------+-----------
        +     |        99            33  |        132
        -     |         5             7  |         12
   -----------+--------------------------+-----------
      Total   |       104            40  |        144

   Classified + if predicted Pr(D) >= .5
   True D defined as q20c != 0
   --------------------------------------------------
   Sensitivity                     Pr( +| D)   95.19%
   Specificity                     Pr( -|~D)   17.50%
   Positive predictive value       Pr( D| +)   75.00%
   Negative predictive value       Pr(~D| -)   58.33%
   --------------------------------------------------
   False + rate for true ~D        Pr( +|~D)   82.50%
   False - rate for true D         Pr( -| D)    4.81%
   False + rate for classified +   Pr(~D| +)   25.00%
   False - rate for classified -   Pr( D| -)   41.67%
   --------------------------------------------------
   Correctly classified                        73.61%
   --------------------------------------------------

79 . test $xlist3

    ( 1)  [q20c]q1 = 0
    ( 2)  [q20c]q2 = 0
    ( 3)  [q20c]q5 = 0
    ( 4)  [q20c]q7 = 0
    ( 5)  [q20c]q8 = 0
    ( 6)  [q20c]q16 = 0
    ( 7)  [q20c]q22gh = 0

              chi2(  7) =   14.73
            Prob > chi2 =    0.0396

80 . estat gof

   Logistic model for q20c, goodness-of-fit test

          number of observations =       144
    number of covariate patterns =       140
               Pearson chi2( 132) =      139.09
                     Prob > chi2 =        0.3192

81 . lroc

   Logistic model for q20c

   number of observations =        144
   area under ROC curve   =     0.7249
```

```
82 .
83 . quietly logit $ylist $xlist4

84 . estat classification

   Logistic model for q20c

               -------- True --------
   Classified |        D           ~D  |      Total
   -----------+------------------------+-----------
        +     |       98           29  |        127
        -     |        6           11  |         17
   -----------+------------------------+-----------
      Total   |      104           40  |        144

   Classified + if predicted Pr(D) >= .5
   True D defined as q20c != 0
   --------------------------------------------------
   Sensitivity                     Pr( +| D)   94.23%
   Specificity                     Pr( -|~D)   27.50%
   Positive predictive value       Pr( D| +)   77.17%
   Negative predictive value       Pr(~D| -)   64.71%
   --------------------------------------------------
   False + rate for true ~D        Pr( +|~D)   72.50%
   False - rate for true D         Pr( -| D)    5.77%
   False + rate for classified +   Pr(~D| +)   22.83%
   False - rate for classified -   Pr( D| -)   35.29%
   --------------------------------------------------
   Correctly classified                        75.69%
   --------------------------------------------------

85 . test $xlist4

    ( 1)  [q20c]q1 = 0
    ( 2)  [q20c]q2 = 0
    ( 3)  [q20c]q5 = 0
    ( 4)  [q20c]q7 = 0
    ( 5)  [q20c]q8 = 0
    ( 6)  [q20c]q16 = 0
    ( 7)  [q20c]q23a = 0

             chi2(  7) =   18.46
           Prob > chi2 =    0.0100

86 . estat gof

   Logistic model for q20c, goodness-of-fit test

          number of observations =       144
    number of covariate patterns =       139
               Pearson chi2( 131) =      141.89
                      Prob > chi2 =        0.2432

87 . lroc

   Logistic model for q20c

   number of observations =      144
   area under ROC curve   =   0.7393

88 .
89 .
90 .
   end of do-file

91 .
```