Illegal File Sharing & The Film Industry

Daren Zhang
Faculty Advisor: Professor Brian D. Wright

University of California, Berkeley
Department of Economics
December, 2015

Abstract

This paper investigates the effects of illegal file-sharing activity on worldwide box office revenues. The study uses measures such as software piracy rate, intellectual property protection index, number of BitTorrent trackers, and file-sharing client download rate as measures for illegal file-sharing activity. There were two separate datasets and studies conducted for this paper. The first is a panel data of 111 countries over a span of 5 years which offers only software piracy rates as a variable of interest. The second dataset is a much more detailed cross-sectional dataset within a single year and includes all four variables of interest as well as a number of control variables. The results indicate that illegal file-sharing activity does not have a significant impact on total box office returns.

Acknowledgement

I would like to thank my faculty advisor, Professor Brian Wright for his invaluable time and advice throughout this writing process. I would also like to thank Professor Glenn Woroch, whose teachings and encouragement was essential for starting this paper in the first place. Lastly, I would like to thank Dr. Alex Kigerl, who generously provided his original dataset, and to the advisors of The Data Lab and D-Lab, who helped me collect and clean the necessary data. This paper would not have been possible without all the guidance and support, thank you.

I. Introduction

Industry Overview

There are two industries that this paper focuses on: the film industry and the illegal file-sharing "industry". The film industry is very well known; it is a multi-billion-dollar industry that has had a long history.[1] Major studios headquartered in Hollywood, California, produce the vast majority of top-grossing blockbuster films, and these are the movies that will be used in this study. They are consumed worldwide in many different countries, and these Hollywood produced movies usually enjoy a very high demand in international countries.

Perhaps the lesser-known "industry" is the illegal file-sharing industry. The suppliers upload illegal copies of movies and the consumers download these movies using a technology called "torrenting". The upload and download process is done on two-way platforms called file-sharing sites. These file-sharing sites are essentially the major "firms" within this industry, facilitating content upload and distribution. These sites are organized by a small private group of owners or administrators and with the exception of top sites, they are usually not very profitable[2]. There are an endless number of sites that host a file-sharing platform. However, the vast majority of traffic and illegal downloads are from the top sites. Historically there has been one dominant site, The Piratebay[3]. There is significant turnover among leading sites, and when one site's website goes down for a few days or even weeks, nearly all the traffic shifts to other sites. However, there is nearly no decrease in total consumption, and it's distributed among all the remaining firms, with the large sites being the biggest gainers. Also, when a major site such as The Piratebay comes back from its temporary absence, it usually is able to come back and regain the position as the top site very quickly[4]. This was evident when The Piratebay was shut

---

[1] In 2015, the total global box office revenue was $39.1 Billion, see http://www.statista.com/statistics/259987

[2] Digital Citizens Alliance and MediaLink estimates that top "pirate" sites generate around an annual revenue of $200 million. See "Good Money Gone Bad" report (2015).

[3] See http://www.ebizmba.com/articles/torrent-websites.

[4] See http://www.qdtricks.org/top-best-torrent-sites/. After the temporary shut down in 2014, The Piratebay lost all of its users, but quickly came all the way back to being the top site.

down temporarily when Swedish police raided its servers on December, 2014[5]. Users seamlessly transitioned to other sites, with the second largest site, Kickass Torrent, being the biggest gainer.

These file sharing sites use BitTorrent's file-sharing technology, which allows for fast downloads. BitTorrent's torrent technology itself and sharing platforms are completely legal. However, it's the use of such sharing technology and platforms to share content that are protected through copyright laws that make sites illegal. This is when file-sharing sites traditionally come into major conflicts with the film industry. For most governments, the current view is that any file-sharing site that hosts a platform of pirated goods available for anyone to download is illegal. For example, in 2012 the U.S. government shut down one of the most popular file-sharing websites, MegaUpload.com, because it deemed its design features and operational model as evidence showing criminal intent and venture.[6] Sites are constantly being raided, and a lot of governments, such as the U.K. and France, openly banned such sites. There is some evidence that when a government makes an effort to ban such sites, there is generally a slight decrease in piracy activity.[7] There is also evidence that in countries where content is made available legally with services such as Netflix, there is a general decrease in piracy activity.[8]

Discussion of Relevant Literature and Economic Theory

There have been many papers written about the effects of illegal file sharing on movie box office revenue. Some studies attempt to show that illegal file sharing has a significant impact on box office revenue, but many of these studies either do not have much statistical power or are based on dubious assumptions inflating expected revenues of movies. For example, De Vany and Walls (2007) estimated that piracy has a substantial effect on box office revenue

[5] http://venturebeat.com/2015/02/05/pirate-bay-rises-from-the-ashes-again-a-stats-eye-view/

[6] see Department of Justice Indictment: MegaUpload.com Indictment, Wall Street Journal. http://www.wsj.com/articles/SB10001424052970204616504577171180266957116

[7] See http://www.lifehacker.com.au/2015/06/why-blocking-torrent-sites-wont-curb-piracy/

[8] See http://www.fool.com/investing/general/2015/04/26/how-hbo-and-netflix-inc-are-fighting-piracy.aspx

with a median loss of $41.70 million.[9] That estimate, however, is based on a study of a single widely released movie and has questionable assumptions behind the theory of its model. For example, the paper empirically models the rate of decline in weekly revenues in relation to the number of pirated prints available for download on the internet. However, move revenues naturally decline after opening week in theaters and number of pirated copies of a movie naturally increases over time. Just by establishing this correlation does not prove any causal effect between illegal file sharing activity and decline in movie revenues. Other studies such as Strumpf (2014) show that illegal file sharing does not have a significant impact on box office revenue.[10] Strumpf (2014), however, is a domestic study using data such as domestic box office revenues, and proxies such as HSX stock prices. A concern with this is that the U.S. has been historically one of the least piratical countries, enjoys the largest GDP in the world, and is home to Hollywood and the studios that produce all the blockbuster films. [11] Other studies such as Danaher and Waldfogel (2010) do use international data and estimated that international box office returns were at least 7% lower than they would have been in the absence of pre-released piracy[12]. However, Danaher and Waldfogel (2010) studied data from 2003-2006, during which studios did not have a uniform strategy for releasing movies abroad. Movies were always released first in a few countries then later in other countries, sometimes even months later. With digital technology improved over the years and a growing global market for movie piracy, major studios have since adopted the strategy of opening on the same date around the world[13].

---

[9] A.S. DeVany and D.W. Walls. "Estimating the effects of movie piracy on box-office revenue." *Rev Ind Organ* 30 (2007): 291-301

[10] Koleman Strumpf. "Using Markets to Measure the Impact of File Sharing on Movie Revenues." *Preliminary*, University of Kansas School of Business, 2014

[11] See "Spotting the Pirates." The Economist. August 20, 2011. Accessed April 22, 2015. http://www.economist.com/node/21526299.

[12] B. Danaher and J. Waldfogel, "Reel piracy: The effect of online film piracy on box office sales." Technical report, Working Paper, 2010.

[13] Mickey Ferri, "Rent, Buy, or Pirate: Consumer Preferences in the Movie Industry." PhD diss., University of Chicago, 2013.

This paper seeks to overcome some of the weaknesses of the aforementioned papers by using data that includes foreign box office revenues during a time period after which major studios adapted a strategy to counter piracy. The rest of the paper is organized as follows. Section II describes the methods and data used, Section III describes the empirical results, and Section IV gives a conclusive summary of the paper and its limitations.

II. Method and Data Description

Method Overview

This paper examines the effects on worldwide box office revenue using two datasets. The first dataset is a panel data of 111 countries over a span of 5 years. The second dataset is a much more detailed cross-sectional dataset within a single year that includes a number of control variables. The method of the analysis is to use measures such as software piracy rates, number of BitTorrent trackers, intellectual property protection index, and file sharing client download rates as a proxy for illegal file sharing activity. I will first take a general approach using the panel data and see if software piracy rates have any effect on total box office revenue while taking into consideration movie, time, and country fixed effects. I will then perform a more detailed analysis using the cross sectional dataset and examine what are the actual factors that drive the differences in box office revenue between countries with different piracy rates.

Data Summary

The first dataset consists of box office revenues of the top worldwide grossing movies as well as software piracy rates of 111 countries from 2005-2009. The software piracy rates were taken from the annual Global Software Piracy Study done by the Business Software Alliance from the years 2005-2009. The worldwide box office revenues are collected from Box Office Mojo.[14] I collected the total gross revenue broken down by country of the top three worldwide grossing movies from the years 2005-2009. The number of countries each movie was released in ranges from around 46 to 80, so this was not a balanced panel. A description of the variables used in the first dataset is listed below:

Total Box Office Revenue (Totalgross) - The main dependent variable is the total gross revenue of the top three worldwide grossing movies in the years 2005-2009, broken down by country. The gross revenue accounts only for box office revenue generated by the movies, and does not include DVD sales. The number of countries each movie was released in ranges from 46

[14] www.boxofficemojo.com

to 81. This data is collected from Box Office Mojo. Totalgross is measured in millions of dollars.

Software Piracy Rate (SoftRate) – Software piracy rate is an estimated percentage of all illegally installed software on business PCs. It is calculated by taking the total number of illegal installations divided by the count of all software installed on business computers per country. The data is collected from the Business Software Alliance report (2005, 2006, 2007, 2008, 2009), and includes data for around 116 countries. SoftRate is measured in percentage points going from 0-100.

The second dataset is used to perform a cross-sectional study of the effects of illegal file sharing on worldwide box office revenues. This dataset includes data from 107 countries and the top 20 worldwide grossing movies in 2011. This dataset is mostly based on an original dataset generously provided by Dr. Alex C. Kigerl from Washington State University. The original dataset was used in a study to predict software piracy rates in different countries.[15] A description of each variable for the second dataset is listed below[16]:

Dependent Variables

Total Box Office Revenue (Totalgross) - The main dependent variable is the total gross revenue of the top 20 worldwide grossing movies in 2011, broken down by country. The gross revenue accounts only for box office revenue generated by the movies, and does not include DVD sales. The number of countries each movie was released in ranges from 50 to 81. This data is also collected from the website www.boxofficemojocom. Totalgross is measured in millions of dollars.

Variables of Interest

Software Piracy Rate (SoftRate) – Software piracy rate is an estimated percentage of all illegally installed software on business PCs. It is calculated by taking the total number of illegal installations divided by the count of all software installed on business computers per country.

[15] Alex Kigerl, "Infringing Nations: Predicting Software Piracy Rates, BitTorrent Tracker Hosting, and P2P File Sharing Client Downloads Between Countries." International Journal of Cyber Criminology 7, no. 1 (2013): 62-80.

[16] See Alex Kigerl (2013) for a more detailed explanation of certain variables

The data are collected from the Business Software Alliance report (2011), and includes data for 116 countries. SoftRate is measured in percentage points going from 0-100.

Intellectual Property Protection Index (IntPropIndex) – The intellectual property protection index is a measure based on survey data from 139 countries involving 13,607 respondents. This measure was taken from the Global Competitiveness Report (Schwab, 2010)[17]. Respondents include leaders from international, public, and private organizations who were asked to rate their country's level of IP protection such as anti-counterfeiting measures. The index measure is on a scale of 1 to 7, with higher values indicating a higher level of protection.

Number of BitTorrent Trackers (Trackers) – BitTorrent networks are managed by servers called trackers. Users of the network connect to these trackers in order to begin downloading a chosen file. BitTorrent trackers can be hosted anywhere in the world where there is an internet connection. Users may download BitTorrent tracker lists to add to their file-sharing client in order to facilitate faster and more efficient downloads. The data were taken from multiple BitTorrent indexes and file locker websites[18]. Trackers are measured as an absolute count variable, not as a rate.

File Sharing Client Downloads per Internet User (DownRate) – DownRate is calculated by taking the total number of downloads of file sharing clients and dividing it by the number of internet users per country. The total number of downloads consists of the number of downloads of the top four most frequently downloaded clients in 2010. The data are taken from SourceForge.net. SourceForge is a free online software repository where file-sharing clients such as BitTorrent are available for download. The top four clients are Ares Galaxy, Vuze, Shareza, and a Mule. There were over 82 million downloads in 251 countries and sovereign states in 2010

[17] Schwab, K. "The Global Competitiveness Report 2010-2011." World Economic Forum, 2010
[18] See Alex Kigerl (2013) for a detailed explanation of how the data was collected.

among the four file-sharing clients. DownRate is measured in percentage points going from 0-100.

Control Variables

Gini Coefficient (Gini) – The Gini coefficient represents the income inequality in a nation and traditionally ranges from 0 to 1. A minimum score of 0 indicates perfect income equality while higher scores represent higher inequality. In this paper the measure is multiplied by 100, for range of 0-100 consistent with other variables. The data are collected from The World Bank (2009) website and includes data on 171 countries from the years 1992 to 2009.

GDP (GDP) – The measure of GDP is from the International Monetary Fund website (World Economic Outlook Database, 2011). It includes data on 184 countries in 2010 and is measured in billions of dollars.

GDP Per Capita (GDPCapita) – The measure of GDP per capita also comes from the International Monetary Fund website (World Economic Outlook Database, 2011). It is simply the GDP divided by the population size. It includes data on 184 countries in 2010 and is measured in thousands of dollars.

Average Years of Schooling (MeanSchool) – Average years of schooling is the mean number of years of formal schooling (primary, secondary, postsecondary) for adults age 25 and older. It is meant to serve as a measure of a nation's educational level. The data are collected from the United Nations Development Program (2011) and includes 187 countries.

Life Expectancy (LifeExp) – Life expectancy is a measure of general health levels of a country, and is a measure of the number of years newborns are expected to live assuming prevailing patterns of mortality rates remain constant throughout the child's life. The data are also collected from the United Nations Development Program (2011) and includes 187 countries.

Unemployment Rate (Unemployment) – Unemployment rate measures the percentage of people who are unemployed when compared to the total number of people in the labor force. The data is taken from the World Factbook website. The data represent unemployment rates in 2008 and includes 197 nations. Unemployment rate is measured in percentage points going from 0-100.

Discussion on Variables of Interest

Table 1.1 Correlation Between Variables of Interest

	SoftRate	IntPropIndex	Trackers	DownRate
SoftRate	1.0000			
IntPropIndex	-0.8304	1.0000		
Trackers	-0.3473	0.3069	1.0000	
DownRate	-0.2375	0.1070	0.0394	1.0000

The four variables of interest are meant to jointly serve as a proxy for representing a general picture of illegal file-sharing activity in different countries. There are several concerns when using these variables as proxies. But overall, the variables provide a general picture of a country's illegal file-sharing activity.

Software piracy rate is one of the most commonly used proxies for measuring illegal file-sharing activity in countries. Data on the percentage of pirated software, number of pirated downloads, as well as estimated cost of pirated materials are readily available. The nature of these pirated materials, however, is mostly software for business and differs from those of entertainment value such as movies, music, ebooks, etc.

The intellectual property protection index is survey data based on leaders of a country giving their ratings for their country's level of intellectual property protection. The exact method for evaluating each country is unclear and may vary. Looking at Table 1.1 we can also see that there is an extremely high correlation between this index and software piracy rates, which suggests that software piracy rates may be a factor of how the index is measured.

The number of BitTorrent trackers is important for file-sharing purposes, however it is not a measure of infringing traffic. Users can reside anywhere in the world when using such servers. It is important to note that this measure does not give us any data on where the pirates reside and the level of pirate traffic in a country, but rather where the facilitators of file-sharing host their tracking servers. This measure is intended to capture piracy facilitating countries, rather than pirate residing countries[19].

File-sharing client download rate measures the number of file sharing client downloads per internet user. Although it is legal to use the BitTorrent technology and P2P networks can be used for non-illegal file-sharing activities, it is estimated that the majority of these clients downloaded will be used for illegal activities and the majority of P2P traffic is in fact infringing[20]. If we assume that each person only downloads one file-sharing client, this measure could be used as a rough estimate of the percentage of internet users who are engaging in illegal file-sharing activity online. Throughout this paper, I will also refer to this measure as the illegal download rate. One concern here is that a measure of file-sharing client download rate is not a direct measure of how much illegal file-sharing activity is present. Some people use clients to download a song or two while others may use it to download hundreds of movies.

Discussion on Control Variables

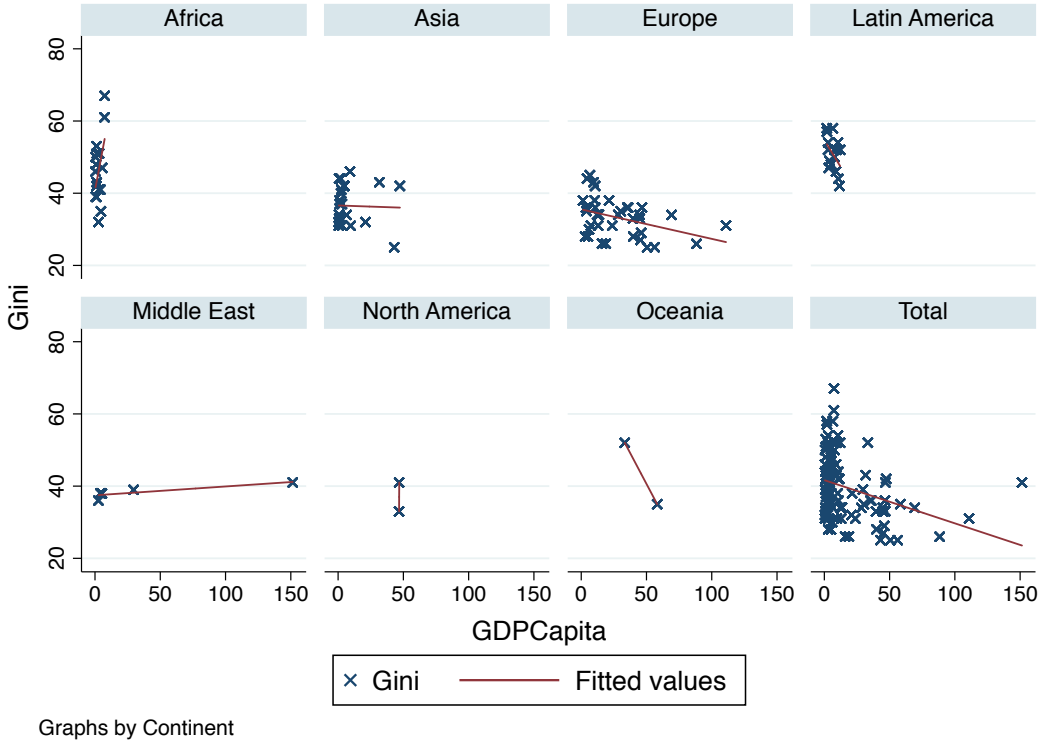
The control variables are meant to be measures that could be correlated with the variables of interest and could have an impact on total box office revenues. These variables are also meant to serve as a way to compare countries and distinguish developing and low-income countries from developed and high-income countries. Developing and low-income countries will tend to have a lower GDP, average years of schooling, and life expectancy while having a higher unemployment rate. Theories behind the Gini coefficient is more inconsistent as it measures income inequality and varies greatly due to other potential factors such as political structure,

[19] Alex Kigerl, "Infringing Nations: Predicting Software Piracy Rates, BitTorrent Tracker Hosting, and P2P File Sharing Client Downloads Between Countries." International Journal of Cyber Criminology 7, no. 1 (2013): 62-80.

[20] Envisional (Janurary, 2011). Technical report: An estimate of infringing use of the internet. Technical report.

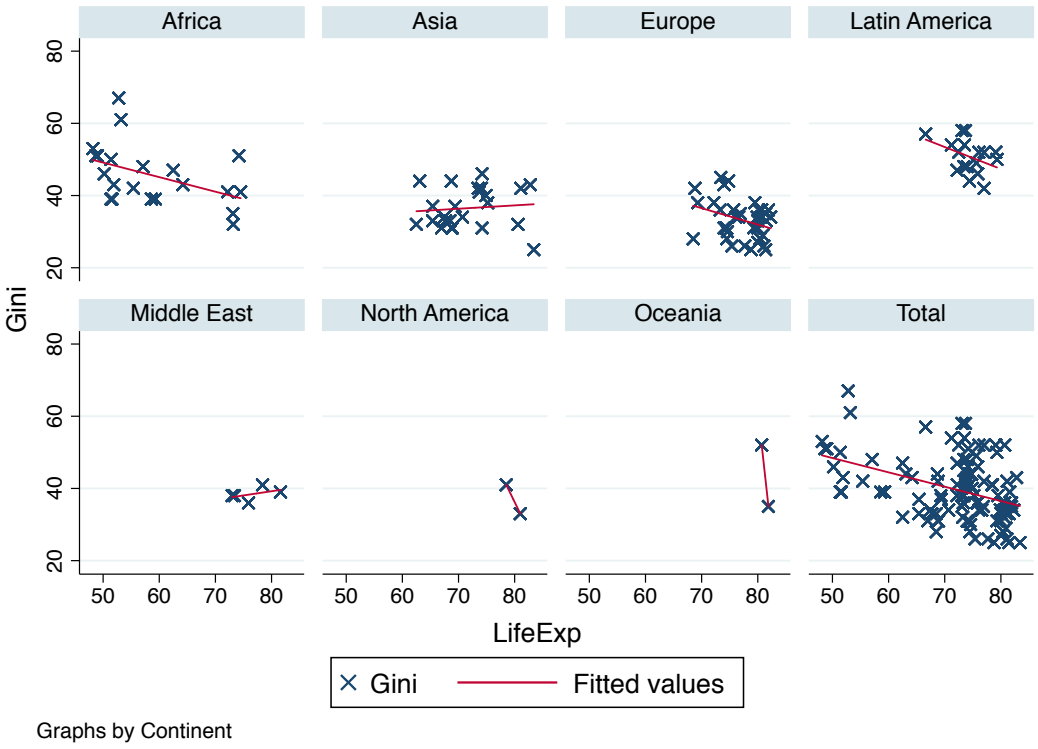
tax rate, and social mobility, regardless of whether a country is considered a developing or developed country. In general, however, it can be argued that developing and low-income countries have a higher Gini coefficient. This is evident when we see Graph 1.1 and Graph 1.2 below, which plots each country's Gini coefficient against its GDP per capita and life expectancy. I used GDP per capita and life expectancy as proxies because they are generally considered to be a great measure for distinguishing between developing countries and developed countries[21].

Graph 1.1



[21] Maddison, Angus. "A Comparison of Levels of GDP Per Capita in Developed and Developing Countries, 1700-1980." Journal of Economic History 43, (1983): pp 27-41.

Graph 1.2



Looking at the graph results we can see two trends that support the notion that developing and low-income countries have higher Gini coefficients. First we can see that in continents such as Africa and South America, where there is a higher concentration of developing countries versus a continent such as Europe, the countries tend to have higher Gini coefficients. Secondly, within each continent and looking at the entire world as a whole, we can see that there is a clear negative relationship between Gini coefficient with GDP per capita and life expectancy. For more evidence using other measures, see Table 1.3 in the Appendix.

III. Empirical Results

Empirical Model

The objective of this paper is to use proxies to measure the impact of illegal file sharing on total box office revenue. The models used for both datasets are simple linear models with a variable of interest and a number of control variables and fixed effects. The model used for the first panel data study is written as,

$$\mathrm{T}_{itm} = \beta_0 + \beta_1 \mathrm{P}_{it} + \mathrm{T}_t + \mathrm{C}_i + \mathrm{M}_m + \mu_{itm}$$

where T_{itm} is the total box office gross of movie m at country i and time t. P_{it} is the software piracy rate of country i at time t. C_i is country fixed effects, T_t is time fixed effects, M_m is movie fixed effects, and μ_{itm} is the error term. β_0 is the constant term, and one term from each fixed effect is dropped to avoid perfect multicollinearity.

The model used for the second cross-sectional data study is written as,

$$\mathrm{T}_{im} = \beta_0 + \beta_1 \nu_i + \gamma_1 \omega_{1i} + \gamma_2 \omega_{2i} + \gamma_3 \omega_{3i} \ldots + \mathrm{M}_m + \mu_{im}$$

where T_{im} is the total box office gross of movie m at country i. ν_i is the variable of interest measured in country i, which represent the software piracy rate, intellectual property protection index, number of BitTorrent trackers, or file sharing client download rate. ω_{1i}, ω_{2i}, ω_{3i}, etc. are the control variables measured in country i, which represent the Gini coefficient, GDP, average years of schooling, life expectancy, and unemployment rate. Again, M_m is movie fixed effects and μ_{im} is the error term. β_0 is the constant term, and one term from the movie fixed effect is dropped to avoid perfect multicollinearity.

Panel Data Regression Results

Table 2.2

Totalgross Regression with Fixed Effects

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV
SoftRate	-0.335***	-0.335***	-0.613	-0.583
	(0.0556)	(0.0550)	(0.753)	(0.733)
Constant	28.38***	28.38***	48.90	50.71
	(3.346)	(3.673)	(56.78)	(55.78)
Observations	583	583	583	583
R-squared	0.082	0.111	0.649	0.663
Adjusted R-squared	0.0808	0.103	0.604	0.618
Year FE		Yes	Yes	Yes
Country FE			Yes	Yes
Movie FE				Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Using the first panel dataset, we can see that in Model I, where I did not account for any fixed effects and simply did a panel pooled regression of Totalgross on the variable SoftRate, the variable was highly significant and negative. It seems to show that as software piracy rate increases by 1%, it would on average lead to a $335,000 decrease in box office revenue. The R^2 and adjusted-R^2 in this Model I are pretty low, with both just slightly above 0.08. The results did not change much in Model II, when I added year fixed effects. This is not surprising, because although total gross revenue for top movies tends to increase every year, it does not vary too much between the years.

However, when I added country fixed-effects, the results changed dramatically. Although Software piracy rate seems to have a larger effect with its coefficient dropping down to -0.613, it is no longer statistically significant. The R^2 and adjusted R^2 also jumped up to around 0.649 and 0.604 respectively. Similar changes occur when adding movie fixed effects. The coefficient for SoftRate went up a little to -0.583 and remained insignificant. The R^2 and adjusted R^2 also went up a bit more, to 0.663 and 0.618 respectively.

The results of the panel data regression support the idea that illegal file-sharing activity does not have a significant impact on box office revenue. On the surface level, it seems that

software piracy rate may have a significant and negative effect on worldwide box office revenue. But when we consider differences between countries and movies, that effect completely goes away. In order to further examine why that effect became statistically insignificant, we will look into the results of the cross-sectional dataset for a more detailed explanation of what factors are actually influencing box office revenue.

Relationship Between Variables of Interest and Control Variables

Using the second cross-sectional dataset, I will first start off by examining the relationship between illegal file-sharing activity and the control variables. The initial hypothesis is that illegal file-sharing activity will tend to be higher in developing and lower-income countries. The method of this study is to first test the relationship between illegal file-sharing activity and the measures that define a lower-income country, in order to ultimately test whether illegal file-sharing activity has a significant impact on movie box office revenue or if other correlated factors that define developing and low-income countries such as GDP, unemployment, life expectancy, etc. are the actual driving force behind lower box office returns in these countries.

Table 3.1 Correlation between variables of interest and control variables

	SoftRate	IntPropIndex	Trackers	DownRate
Gini	**0.292**	**-0.2923**	**-0.1121**	**0.2394**
GDP	**-0.345**	**0.2748**	**0.8084**	**0.0015**
MeanSchool	**-0.6446**	**0.5446**	**0.2961**	**0.0956**
LifeExp	**-0.6358**	**0.5027**	**0.1897**	**0.3426**
Unemployment	**0.3883**	**-0.3229**	**-0.1358**	**-0.2188**

Looking at the table of the correlations between the variables of interest and the control variables gives us a good general picture of the relationship between illegal file sharing and measures that define developing and low-income countries. As mentioned earlier in the discussion of control variables, developing and low-income countries tend to have higher Gini coefficients and unemployment rates as well as lower GDP, average years of school, and life expectancy.

For software piracy rates, it has a negative correlation with GDP, average years of

schooling, and life expectancy. It has a positive correlation with the Gini coefficient and unemployment rate. This is consistent with the notion that software piracy rates tend to increase as we move from developed and high-income countries to developing and low-income countries.

For the intellectual property protection index, it has the completely opposite signs when compared to software piracy rates. It has a negative correlation with the Gini coefficient and unemployment rate, and it has a positive correlation with GDP, average years of schooling, and life expectancy. This is also consistent with the previous notion, as we expect countries with higher intellectual property protections level to have lower illegal file sharing activity.

For the number of BitTorrent trackers, it has a negative correlation with the Gini coefficient and unemployment rate, and it has a positive correlation with GDP, average years of schooling, and life expectancy. This result is interesting, as it means that more facilitating trackers are located in developed countries. It is not surprising, however, as BitTorrent trackers are a measure of piracy facilitating countries and not piracy residing countries. Furthermore, the BitTorrent technology and many file-sharing websites were founded in developed countries in the first place.

The only variable that is not quite consistent with all the control variables is file-sharing client download rate. It has a positive correlation with all the control variables except unemployment. Its positive correlation with GDP, average years of schooling, and life expectancy combined with its negative correlation with unemployment suggests that developed and high-income countries have a higher illegal download rate. However, it also has a positive correlation with the Gini coefficient, which suggests the opposite and that developing and low-income countries have a higher illegal download rate.

In order to get a more detailed picture of the relationships, I ran regressions of the four variables of interest on the control variables. This will give a better insight on some of the relationships, especially for file-sharing client download rate. Looking at the tables 3.2-3.5 below, we can see that most of the control variables are highly significant.

Table 3.2

Software Piracy Rate Regression on Control Variables

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V
Gini	0.757***	0.687***	0.184***	-0.00949	-0.00842
	(0.0607)	(0.0586)	(0.0475)	(0.0488)	(0.0489)
GDP		-0.00448***	-0.00289***	-0.00251***	-0.00252***
		(0.000270)	(0.000232)	(0.000188)	(0.000188)
MeanSchool			-4.943***	-3.271***	-3.276***
			(0.137)	(0.184)	(0.184)
LifeExp				-0.924***	-0.951***
				(0.0489)	(0.0578)
Unemployment					-0.0323
					(0.0216)
Constant	33.38***	38.68***	99.78***	159.6***	161.9***
	(2.532)	(2.460)	(2.458)	(3.477)	(4.347)
Observations	2,140	2,140	2,140	2,140	2,140
R-squared	0.085	0.189	0.461	0.535	0.535
Adjusted R-squared	0.0848	0.188	0.460	0.534	0.534
F test model	155.6	251.5	830.9	766.4	672.5
P-value of F model	0	0	0	0	0

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.3

Intellectual Property Protection Index Regression on Control Variables

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V
Gini	-0.0387***	-0.0359***	-0.0149***	-0.00857***	-0.00855***
	(0.00309)	(0.00304)	(0.00266)	(0.00278)	(0.00279)
GDP		0.000179***	0.000112***	9.97e-05***	9.96e-05***
		(1.27e-05)	(1.17e-05)	(9.94e-06)	(9.95e-06)
MeanSchool			0.207***	0.153***	0.152***
			(0.00686)	(0.00921)	(0.00917)
LifeExp				0.0301***	0.0296***
				(0.00290)	(0.00343)
Unemployment					-0.000576
					(0.00169)
Constant	5.219***	5.007***	2.449***	0.504**	0.546**
	(0.128)	(0.127)	(0.131)	(0.202)	(0.241)
Observations	2,140	2,140	2,140	2,140	2,140
R-squared	0.085	0.149	0.331	0.361	0.361
Adjusted R-squared	0.0850	0.148	0.330	0.360	0.360
F test model	157.2	194.8	535.5	462.5	371.6
P-value of F model	0	0	0	0	0

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.4

BitTorrent Trackers Regression on Control Variables

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V
Gini	-0.445***	-0.179***	-0.0238	-0.101**	-0.102**
	(0.0414)	(0.0465)	(0.0416)	(0.0415)	(0.0416)
GDP		0.0171***	0.0166***	0.0167***	0.0167***
		(0.000860)	(0.000870)	(0.000853)	(0.000853)
MeanSchool			1.522***	2.186***	2.194***
			(0.158)	(0.159)	(0.159)
LifeExp				-0.367***	-0.328***
				(0.0329)	(0.0357)
Unemployment					0.0468***
					(0.0158)
Constant	26.61***	6.386***	-12.42***	11.33***	7.938***
	(1.943)	(2.078)	(1.997)	(2.705)	(2.886)
Observations	2,140	2,140	2,140	2,140	2,140
R-squared	0.013	0.656	0.667	0.672	0.672
Adjusted R-squared	0.0121	0.655	0.666	0.671	0.671
F test model	115.4	206.2	191.5	159.4	131.1
P-value of F model	0	0	0	0	0

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5

Illegal Download Rate Regression on Control Variables

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V
Gini	0.147***	0.148***	0.192***	0.269***	0.269***
	(0.0152)	(0.0154)	(0.0167)	(0.0141)	(0.0141)
GDP		7.09e-05	-6.87e-05	-0.000219***	-0.000223***
		(4.93e-05)	(4.84e-05)	(4.00e-05)	(4.01e-05)
MeanSchool			0.434***	-0.225***	-0.228***
			(0.0378)	(0.0463)	(0.0460)
LifeExp				0.364***	0.350***
				(0.0137)	(0.0149)
Unemployment					-0.0173**
					(0.00697)
Constant	-0.611	-0.695	-6.057***	-29.61***	-28.36***
	(0.561)	(0.573)	(0.799)	(1.264)	(1.308)
Observations	2,140	2,140	2,140	2,140	2,140
R-squared	0.057	0.058	0.095	0.300	0.301
Adjusted R-squared	0.0569	0.0569	0.0939	0.299	0.300
F test model	93.11	46.61	62.01	199.4	160.8
P-value of F model	0	0	0	0	0

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The regressions confirm the correlation tables and show that software piracy rates are higher in developing and low-income countries while intellectual property protection index and number of BitTorrent trackers are higher in developed and high-income countries. For illegal download rate, it is again a complicated situation and it is not clear whether it is higher in developing or developed countries. I will now briefly go over each variable of interest.

For software piracy rate, some important changes occur when more control variables were included in the regression. First off, the coefficients for the Gini and Unemployment now have a negative sign, but are not statistically significant. GDP, MeanSchool, and LifeExp all have a negative effect on software piracy rate and are highly statistically significant. Model V also has a high F-stat and a p-value of essentially zero, which means that we reject the null hypothesis that all the variables have zero effect on software piracy rates. Overall, the regression results support the notion that software piracy rates are higher in developing and low-income countries.

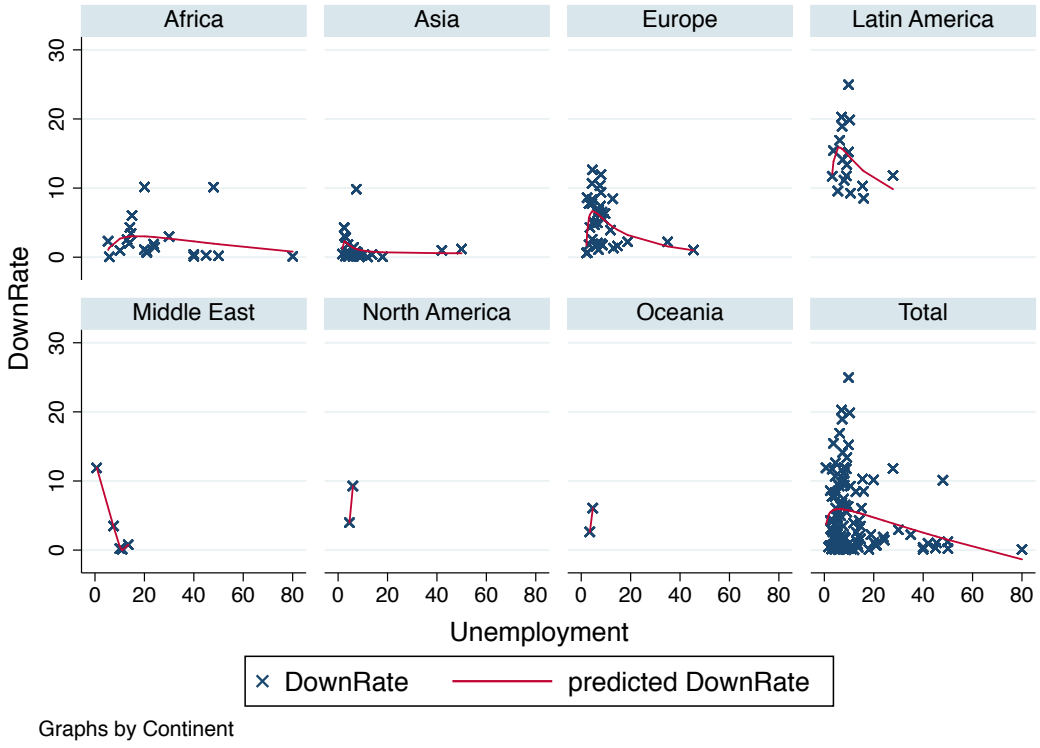
For intellectual property protection index, all the control variables are highly statistically significant with the exception of Unemployment. The variable Gini has a negative effect while GDP, MeanSchool, and LifeExp have positive effects. This result supports the notion that intellectual property protection is higher in developed and high-income countries.

For the number of BitTorrent trackers, almost all the control variables are highly statistically significant and the variable Gini is significant at the 5% level. While previously a simple correlation table showed that the number of BitTorrent trackers are higher in developed and high-income countries, when we look at the regression results of Model V some inconsistencies occur. LifeExp has a negative effect while unemployment has a positive effect. This would support the opposite. The results show that there is a more complicated story here behind why certain countries have higher number of BitTorrent trackers. There are many factors that are not included in this regression that can be influencing where a BitTorrent tracker may be hosted, such as digital infrastructure, political situation of a country, culture/attitude towards file-sharing services, etc. But overall the results combined with the

correlation table support the notion that developed and high-income countries in general have more BitTorrent trackers.

For the file-sharing client download rate, all the variables are highly significant with the exception of Unemployment, which is significant at the 5% level. Again, the regression results show a more complicated story. Gini has a positive effect while GDP and MeanSchool have negative effects. These variables support the notion that developing and low-income countries have higher illegal download rates. On the other hand, LifeExp has a positive effect and Unemployment has a negative effect. These variables support the notion that illegal download rate is higher in developed and high-income country. A more detailed examination of the relationship between DownRate, Unemployment, and LifeExp can provide some possible explanations for this inconsistency.

Graph 3.1



A scatterplot of DownRate and Unemployment overlaid with a fractional-polynomial prediction plot and separated by continent (Graph 3.1) provides some interesting insight into the relationship. It seems that among most continents such as Africa, Asia, Europe, and Latin

America there is a sharp increase in illegal download rate as unemployment rate increases at lower levels under 10 percentage points. However, as unemployment rates increases at levels above 10 percentage points there is a gradual decrease in illegal download rate and this is especially affected by a number of outlier countries that have extremely high unemployment rates (over 50) and very low illegal download rates. Some possible explanation for this is that in general as we move from developed and high-income countries to developing and low-income countries, illegal download rate increases. However, after a certain point, countries with extremely high unemployment rates are starting to see a decrease in illegal download rate because their citizens are so affected by poor social welfare statuses that they are possibly more concerned with surviving and getting a job instead of trying to download and watch movies. Chen and Yeh (2010), a study on software piracy, gives some support to this notion as it notes that "a higher unemployment rate indicates that the unemployed people have less demand for software applications, resulting in a lower probability for software piracy".[22]

Table 3.6

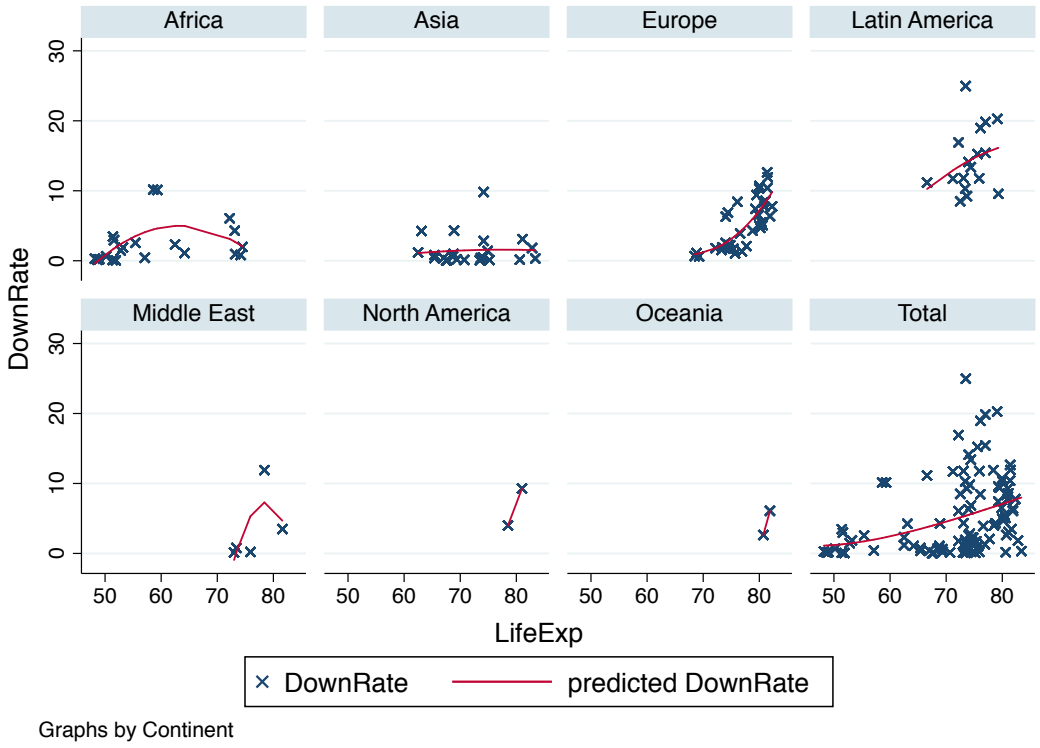
Percentiles	LifeExp		
1%	48.7	Min	48.2
5%	51.4	Max	83.4
10%	55.4	Obs	2140
25%	68.7		
50%	74.1	Mean	71.99159
		Std. Dev.	8.932997
75%	79.1		
90%	81.1	Variance	79.79843
95%	81.6	Skewness	-1.196376
99%	82.8	Kurtosis	3.646858

Similar inconsistencies exist for LifeExp, and a possible explanation could be that it is a measure that is not a good indicator of technological advances for mid to high-level income and developed countries but rather it is much better at capturing the advances in basic medical conditions and technology in poor countries that allows its population to survive and live a

[22] Chen, C., C. Chen, and C. Yeh. "Determinants of Software Piracy: Evidence From Far East Countries." *Journal of International and Global Economic Studies* 3(2), (2010): 53-662

relatively healthy life. This can be seen in Table 3.6 above, where we can see that the spread of LifeExp of countries below the 50th percentile is huge while the spread of the countries above the 50th percentile is relatively small. It can be argued that a country must have a certain level of life expectancy as an indicator of whether it has enough infrastructures available such as internet and computers that allows people to download file-sharing clients and enjoy movies. So life expectancy would have a positive effect on illegal download rate. As we move up in life expectancy, however, it should start to have less and less of a positive effect on illegal download rate, and at a certain point may even start to have a negative effect on illegal download rate. Similar to Unemployment, I tested this theory and further examined the relationship by creating a scatterplot of DownRate and LifeExp overlaid with a fractional-polynomial prediction plot, separated by continent (Graph 3.2 below).

Graph 3.2



This trend is confirmed in poor continents such as Africa and the Middle East. The decreasing effect of life expectancy is somewhat showing in Latin America. In richer continents such as Europe and North America, however, it seems that as life expectancy increases, illegal download rate increases. One possible theory could be that in these developed and high-income countries in Europe and North America there is consequently a higher demand for movies. For example, the countries that have higher life expectancies have better healthcare and social services, which would mean that their citizens have less incentive to work hard, thus spending more time watching movies. Good universal healthcare systems and good social services are traits that define countries in the European continent with the highest life expectancies such as France, Iceland, Sweden, and Switzerland. By looking at the scatterplot we can definitely see that this is a complicated relationship and varies a lot between continents. Despite some inconsistencies, the results combined with the correlation table are enough to show that developing and low-income countries in general have higher illegal download rates.

Cross-sectional Data Regression Results

I will now discuss regression results of the cross-sectional data. After establishing a strong relationship between illegal file-sharing activity and developing and low-income countries, I test whether illegal file-sharing activity has a significant impact on movie box office revenues or if other correlated measures could be the significant factors that actually lower box office returns in low-income countries that have high illegal file-sharing activities. Tables 4.1 - 4.4 below show the regression results of Totalgross on the four variables of interest.

Table 4.1

Totalgross Regression on Software Piracy Rates

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V	(6) Model VI	(7) Model VII
SoftRate	-0.123***	-0.142***	-0.0559**	-0.0260	0.00803	0.0126	0.00816
	(0.0292)	(0.0306)	(0.0263)	(0.0315)	(0.0354)	(0.0366)	(0.0364)
Gini		0.0772*	0.163***	0.188***	0.176***	0.174***	0.174***
		(0.0416)	(0.0243)	(0.0225)	(0.0216)	(0.0215)	(0.0224)
GDP			0.00818***	0.00829***	0.00819***	0.00819***	0.00809***
			(0.00104)	(0.00105)	(0.00105)	(0.00105)	(0.00100)
MeanSchool				0.499***	0.373**	0.335**	0.336**
				(0.172)	(0.165)	(0.165)	(0.165)
LifeExp					0.191***	0.185***	0.187***
					(0.0502)	(0.0495)	(0.0527)
Unemployment						-0.128	-0.139*
						(0.0776)	(0.0763)
Constant	12.98***	10.97***	-2.391	-9.781***	-24.31***	-22.75***	-24.26***
	(1.559)	(2.010)	(1.734)	(3.338)	(5.730)	(5.444)	(5.705)
Observations	937	937	937	937	937	937	937
R-squared	0.029	0.031	0.447	0.450	0.452	0.453	0.511
Adjusted R-squared	0.0281	0.0289	0.446	0.448	0.449	0.449	0.498
Movie FE							Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.2

Totalgross Regression on Intellectual Property Protection Index

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V	(6) Model VI	(7) Model VII
IntPropIndex	2.132***	2.537***	1.073***	0.793***	0.553*	0.465	0.452
	(0.315)	(0.323)	(0.246)	(0.271)	(0.312)	(0.358)	(0.352)
Gini		0.104**	0.178***	0.212***	0.202***	0.199***	0.196***
		(0.0481)	(0.0340)	(0.0289)	(0.0310)	(0.0315)	(0.0316)
GDP			0.00814***	0.00822***	0.00816***	0.00816***	0.00806***
			(0.00102)	(0.00102)	(0.00104)	(0.00104)	(0.000988)
MeanSchool				0.469***	0.344**	0.311**	0.322**
				(0.161)	(0.144)	(0.148)	(0.148)
LifeExp					0.109	0.108	0.120*
					(0.0662)	(0.0662)	(0.0660)
Unemployment						-0.0930	-0.106
						(0.0818)	(0.0809)
Constant	-1.935*	-7.569***	-10.14***	-14.89***	-20.60***	-19.12***	-21.48***
	(1.135)	(2.501)	(1.751)	(1.701)	(3.849)	(4.030)	(4.143)

Observations	937	937	937	937	937	937	937
R-squared	0.036	0.039	0.449	0.452	0.453	0.453	0.512
Adjusted R-squared	0.0347	0.0370	0.448	0.450	0.450	0.450	0.498
Movie FE							Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.3

Totalgross Regression on BitTorrent Trackers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Model I	Model II	Model III	Model IV	Model V	Model VI	Model VII
Trackers	0.119***	0.119***	0.0147	0.00331	0.00593	0.00665	0.00781
	(0.0224)	(0.0235)	(0.0213)	(0.0224)	(0.0227)	(0.0225)	(0.0204)
Gini		0.000401	0.113***	0.178***	0.180***	0.180***	0.178***
		(0.0477)	(0.0342)	(0.0262)	(0.0256)	(0.0256)	(0.0258)
GDP			0.00822***	0.00834***	0.00815***	0.00814***	0.00803***
			(0.00107)	(0.00108)	(0.00113)	(0.00113)	(0.00107)
MeanSchool				0.628***	0.331*	0.281*	0.287*
				(0.159)	(0.171)	(0.170)	(0.170)
LifeExp					0.178***	0.164***	0.175***
					(0.0607)	(0.0628)	(0.0628)
Unemployment						-0.124*	-0.138**
						(0.0669)	(0.0665)
Constant	5.688***	5.673***	-3.433**	-11.99***	-22.72***	-20.30***	-22.69***
	(0.423)	(2.007)	(1.440)	(1.644)	(3.697)	(4.093)	(4.206)
Observations	937	937	937	937	937	937	937
R-squared	0.045	0.045	0.444	0.449	0.452	0.453	0.511
Adjusted R-squared	0.0440	0.0429	0.442	0.447	0.449	0.449	0.498
Movie FE							Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.4

Totalgross Regression on Illegal Download Rate

VARIABLES	(1) Model I	(2) Model II	(3) Model III	(4) Model IV	(5) Model V	(6) Model VI	(7) Model VII
DownRate	0.135	0.221**	0.120*	0.0956	0.0530	0.0741	0.0804
	(0.0854)	(0.0934)	(0.0727)	(0.0719)	(0.0716)	(0.0685)	(0.0653)
Gini		-0.133***	0.0741*	0.148***	0.163***	0.157***	0.152***
		(0.0494)	(0.0422)	(0.0354)	(0.0341)	(0.0332)	(0.0330)
GDP			0.00828***	0.00833***	0.00820***	0.00819***	0.00809***
			(0.00102)	(0.00102)	(0.00103)	(0.00103)	(0.000979)
MeanSchool				0.603***	0.373***	0.325**	0.337**
				(0.138)	(0.139)	(0.147)	(0.147)
LifeExp					0.152***	0.125***	0.133***
					(0.0508)	(0.0483)	(0.0496)
Unemployment						-0.147**	-0.163***
						(0.0621)	(0.0624)
Constant	5.975***	10.45***	-2.644*	-11.22***	-20.77***	-17.15***	-19.32***
	(0.870)	(1.890)	(1.598)	(1.652)	(3.405)	(3.191)	(3.499)
Observations	937	937	937	937	937	937	937
R-squared	0.003	0.009	0.445	0.451	0.452	0.453	0.512
Adjusted R-squared	0.00207	0.00734	0.443	0.448	0.449	0.450	0.498
Movie FE							Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

For all four variables of interest, the regression tables show a similar trend. In general, looking at the first couple of models, illegal file-sharing activity seems to have a significant impact on movie revenues. However, as I add in more control variables and move up the models, the significance goes away. This result combined with the fact that there is an established a correlation between the variables of interest and control variables signifies that in the first couple of models there was a lot of omitted variable bias. This is evident when examining the change in the coefficients from Model I to Model VII. Looking at Model VII, which includes all of the control variables and accounts for movie fixed-effects, it tells a compelling story for all four variables of interest. The coefficient for SoftRate in Table 4.1 is negative from Model I to Model IV, but eventually turns positive and insignificant. The coefficient for IntPropIndex is positive throughout, but decreases dramatically. Trackers in Table 4.3 has a positive coefficient throughout and also decreases dramatically. Also, the positive sign would seem to indicate that

pirate-facilitating countries tend to have higher box office revenues. The surprising result is Downrate in Table 4.4, which has a positive coefficient throughout. This could be due to some network effects. Specifically, high illegal file-sharing activities lead to more awareness of movies and more consumption, which is a positive network externality[23]. This theory is further supported by the positive coefficients of Trackers and SoftRate in Model VII. It is important to note, however, that the reverse could be true, where countries that have a high demand for movies are the ones that have high illegal file-sharing rates. This positive network externality theory or reverse causality could possible explain some of the inconsistencies earlier in this paper when I discussed the relationship between illegal download rate and control variables, to try and establish the trend that developing and low-income countries have higher illegal download rates. In the end, however, all four variable of interest are not statistically significant, which shows that in general, illegal file-sharing activity does not have any impact on total box office revenues. It is also important to note that for all four variables of interest, as I added in more control variables, the standard error did not increase dramatically. In fact, for illegal download rate it actually decreased. This shows that although there is an established a relationship between the variables of interest and control variables, there isn't a significant imperfect multicollinearity problem here.

On the other hand, almost all the control variables are statistically significant and the vast majority is highly significant. The only exception is Unemployment for the regression on intellectual property protection index (see Table 4.2). The control variables Gini, GDP, MeanSchool, and LifeExp all have a positive effect while the variable Unemployment has a negative effect. This all makes sense, as a higher GDP, MeanSchool, and LifeExp will tend to signify a more developed and high-income country, which would lead to higher box office returns. Unemployment obviously would have a negative effect on box office returns, as when people don't have jobs they won't have money to spend on movie tickets. The results suggest that these control variables, and not illegal file-sharing activity, are the true drivers behind box

[23] Smith, Michael. "Piracy or Promotion? The Impact of Broadband Internet Penetration on DVD Sales." Information Economics and Policy 22(4), (2010): 289-298

office revenues.

The one surprising trend here is that the variable Gini has a positive effect, and this contradicts a lot of the underlying theories so far in this paper. First off, when I run a regression of Totalgross on Gini, it shows that the Gini coefficient has a negative effect on total box office revenue (see Table 4.5 in Appendix). Secondly, a positive effect on total box office revenue for Gini would seem to indicate that developing and low-income countries would have higher box office revenues. There are some theories that may explain this inconsistency. It could be argued that since the Gini coefficient measures the level of income inequality in a country, when accounting for all the other control variables such as GDP, LifeExp, MeanSchool, and Unemployment, Gini serves as a way of describing the political atmosphere of a country. For example, a more socialist country would have lower income inequality and a lower Gini coefficient. Socialist/communist governments might tend to discourage consumption of western Hollywood movies, which would lower total box office returns. The entire top twenty worldwide grossing movie list used in this dataset are movies produced by western studios in Hollywood. There is some evidence that supports this notion. For example, in socialist countries such as France there are domestic context regulations that limit the umber of Hollywood movies shown in theaters there.[24] Hollywood movies' market share runs 80%-90% in Latin America countries, but in France it is below 50% and is around 60%-75% over Western Europe.[25]

[24] See French Film Quotas and Cultural Protectionism, http://www1.american.edu/ted/frenchtv.htm

[25] Hopewell, John. "Hollywood Stymied as Europe Sticks With Its Limits on Film and TV." *Variety.* The Variety, June 14, 2013.

IV. Discussion and Conclusion

Discussion of Validity and Limitations

One of the most obvious concerns regarding the validity of this paper is the use of the software piracy rates, intellectual property protection index, number of BitTorrent trackers, and file-sharing client download rates as proxies for illegal file-sharing activity. Although in the data section of this paper I outlined many reasons why taken as a whole the four variables of interest provide a good indication of the nature of illegal file-sharing activity in a country, it is still not a direct source of data. There are a number of concerns with each variable of interest, and none of the variables directly measure the volume of movies downloaded illegally in a country and there are no ways to empirically test how strongly these variables relate to illegal file-sharing activity using the data of this paper. Other concerns include the numerous surprising and largely unexplained outcome, such as the relationship between illegal download rate, unemployment rate, and life expectancy or the Gini coefficient having a strong positive effect on total box office revenues. Although there were attempts to explain such inconsistencies, it remains largely unsolved with the limited data available in this paper.

The main limitation of this paper goes back to data. One way to drastically improve this paper is by getting direct data on illegally downloaded movies, including measures such as volume, location downloaded, date downloaded, and date uploaded. The control variables could also be improved, such as including data for each country on the number of movie theaters, number of movie theater seats, number of movie distributors, number of tickets sold, and ticket price to name a few. Ideally I would also like to obtain data in more years for the cross-sectional data, especially for the years 2005-2009 so it matches the general panel data set.

This paper establishes that for global box office revenues, illegal file-sharing activity has no significant impact on how much box office revenue is received from a certain country. However, it does not show that within each country, whether fluctuations in illegal file-sharing activity affects local box office revenues. Each country is unique, with many different factors not included in this paper's data that can potentially influence box office revenues and illegal file-

sharing activity. This paper also uses the top grossing movies world wide which are all produced by Hollywood studios, and does not show whether illegal file-sharing activity has an impact on domestic film productions. Box office revenues do not include DVD sales, which may be significantly impacted by piracy.

Conclusion and Summary

This paper explores the effect of illegal file-sharing activity on movie box office revenues worldwide. To provide a holistic view, this paper includes data not only on domestic revenues in the U.S. but also every country in the world that played one of the top grossing movies listed in the dataset. The approach is to use measures such as software piracy rates, number of BitTorrent trackers, intellectual protection index and file-sharing client download rates as a proxy for illegal file-sharing activity. There were two separate datasets and studies conducted for this paper. The first is a panel data of 111 countries over a span of 5 years which offers only software piracy rates as a variable of interest. The second dataset is a much more detailed cross-sectional dataset within a single year and includes all four variables of interest as well as a number of control variables.

With the exception for control variables such as life expectancy and unemployment rate, a consistent finding is that illegal file-sharing activity tends to be higher in developing and low-income countries than in developed and high-income countries. Estimates of the panel study show that on the surface level it may seem that illegal file-sharing activity has a significant and negative effect on movie revenues, but when country and movie fixed effects are included, that effect completely goes away and is no longer significant. Estimates of the cross-sectional study confirm this notion by showing all four variables of interest are not significant and have no effect on box office revenues. Furthermore, although the variables are insignificant, the coefficients completely changed as we added in control variables and illegal file-sharing activity seems to have a mild positive effect on box office revenues. Variables such as the Gini coefficient, GDP, life expectancy, average years of schooling, and unemployment, on the other hand, are very significant and have a huge impact on box office revenues.

References

Chen, C., C. Chen, and C. Yeh. "Determinants of Software Piracy: Evidence From Far East Countries." *Journal of International and Global Economic Studies* 3(2), (2010): 53-662

Danaher, B. and J. Waldfogel. "Reel piracy: The effect of online film piracy on box office sales." Technical report, Working Paper, 2010.

DeVany, A.S. and D.W. Walls. "Estimating the effects of movie piracy on box-office revenue." *Rev Ind Organ* 30, (2007): 291-301.

Hopewell, John. "Hollywood Stymied as Europe Sticks With Its Limits on Film and TV." *Variety.* The Variety, June 14, 2013.

Jolibert, Alain and Gary Baumgartner. "Toward a Definition of the Consumerist Segment in France". Journal of Consumer Research 8.1 (1901): 114-117

Kigerl, Alex. "Infringing Nations: Predicting Software Piracy Rates, BitTorrent Tracker Hosting, and P2P File Sharing Client Downloads Between Countries." International Journal of Cyber Criminology 7, no. 1 (2013): 62-80.

Ferri, Mickey. "Rent, Buy, or Pirate: Consumer Preferences in the Movie Industry." PhD diss., University of Chicago, 2013.

Maddison, Angus. "A Comparison of Levels of GDP Per Capita in Developed and Developing Countries, 1700-1980." Journal of Economic History 43, (1983): pp 27-41.

Schwab, K. "The Global Competitiveness Report 2010-2011." World Economic Forum, 2010.

Smith, Michael. "Piracy or Promotion? The Impact of Broadband Internet Penetration on DVD Sales." Information Economics and Policy 22(4), (2010): 289-298

"Spotting the Pirates." The Economist. August 20, 2011. Accessed April 22, 2015. http://www.economist.com/node/21526299.

Strumpf, Koleman. "Using Markets to Measure the Impact of File Sharing on Movie Revenues." *Preliminary*, University of Kansas School of Business, 2014.

Appendix

Table 1.3 Correlation between Gini and Control Variables

	Gini
Gini	**1**
LifeExp	**-0.4033**
GDPCapita	**-0.3272**
MeanSchool	**-0.3562**
Unemployment	**0.2656**

Note: As expected, Gini has a negative correlation with LifeExp, GDPCapita, and MeanSchool. It has a positive correlation with Unemployment. All are consistent with the notion that the Gini coefficient is higher in developing and low-income countries.

Table 4.5

Totalgross Regression on the Gini Coefficient

	(1)	(2)
VARIABLES	Model I	Model II
Gini	-0.0761*	-0.0806*
	(0.0461)	(0.0454)
Constant	9.780***	7.757***
	(1.930)	(1.986)
Observations	937	937
R-squared	0.002	0.069
Adjusted R-squared	0.00143	0.0492
Movie FE		Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1