



Infringing Nations: Predicting Software Piracy Rates, BitTorrent Tracker Hosting, and P2P File Sharing Client Downloads Between Countries

Alex C. Kigerl¹

Washington State University, United States of America

Abstract

This study sought to investigate the predictors of digital piracy at the national level. The bulk of previous research on this subject has relied almost exclusively on measures of piracy taken from reports created by copyright industry representatives, which may not be objective sources. For this research, two new measures of piracy related activity in addition to the usual software piracy rate and software piracy cost measures were used. The number of BitTorrent tracking servers and the number of peer-to-peer file sharing client downloads per country were measured. It was determined that these new measures tended to have predictors that were different than the standard software piracy rates. Additionally, it appeared that measuring piracy as a rate relative to legal purchases had the opposite effect than when measuring piracy in absolute terms (such as the absolute number of BitTorrent trackers and absolute dollar amount lost due to piracy). Smaller, poorer, and less technologically developed countries had higher piracy rates, but lower absolute piracy activity. An absolute measure of piracy may be more appropriate, as it likely reflects larger costs to copyright stakeholders, and therefore policy ought to focus on wealthier nations, not poorer ones, when it comes to targeting pirating behavior.

Keywords: Digital piracy, P2P, BitTorrent, Copyright infringement, Cyber crime.

Introduction

Copyright protects the expression of ideas by conferring certain exclusive rights on the creator for a period of time. Copyright is infringed and piracy occurs when a person exercises one of those exclusive rights without the copyright holder's authorization (Clough, 2010). Since the advent of P2P file sharing networks, distribution of copyrighted works on such networks has increased (OECD, 2008). Much of this traffic is illicit, with 86.4% of P2P file sharing network traffic estimated to be infringing (Envisional, 2011). The expansion of file sharing networks is estimated to have harmed legitimate sales of copyrighted content to some extent (Stevens et al., 2005).

¹PhD Candidate and Research Assistant, Department of Criminal Justice and Criminology, Washington State University, 1401 NE Merman Dr Pullman WA 99163, United States of America. Email: alex@kigerl.com

Given the worldwide reach of the internet and the fact that the internet does not have national borders, piracy is a crime which can cross many jurisdictions around the globe. Existing research has focused on the predictors of piracy at the national level. Studies of national piracy rates have found that wealthier nations that have higher income per capita with stronger legal protections and law enforcement against copyright infringement have lower piracy rates (Andrés, 2006a; Andrés, 2006b; Bagchi et al., 2006; Chen et al., 2010; Depken et al., 2004; Ki, Chang et al., 2006; Kovačić, 2007; Kranenburg et al., 2005; Marron et al., 2000). It is suggested that poorer nations are less able to afford legal purchases of copyrighted goods and have cultures more conducive to copyright infringement.

However, the vast majority of this research relies on measures of piracy almost exclusively created by large copyright industries and lobby organizations. The most common variable representing piracy used in the literature is that of software piracy rates. This measure is intended to reflect the percentage of software installed on business PCs that was illegally acquired. The measure is released in reports written by the Business Software Alliance (BSA) (2011), and the process by which the measure is estimated is not entirely transparent.

The creation of the measure relies heavily on estimates of the number of software units installed on each business machine, how much of that software was illegally acquired, and what the retail costs of each software unit would have been were each to have been legally purchased (Goldman, 2003; Png, 2010). The number of estimates involved in the measure makes its accuracy seem unclear.

This study attempts to test the impact that economic, legal, technological, and other predictors have on piracy related rates and outcomes at the national level. In addition to the usual outcomes of software piracy rates and estimated piracy costs released by the BSA, this study seeks to incorporate two other variables that are intended to reflect piracy related activity among nations. The first is a measure of the count of BitTorrent trackers hosted in each country. BitTorrent trackers are servers that help connect users to each other in a P2P network allowing decentralized and optimized downloading of files. The number of file sharing client downloads per internet user will also be utilized. In order to distribute files on a P2P network, users must first download and install a P2P client. The number of these client downloads are thought to relate to actual piracy activity.

It is intended that these additional measures might further help determine what predicts digital piracy at the national level. Neither of these two measures was created by the content industry, so they might be different in terms of their relation to national characteristics. It is to be tested whether these measures might have similar or different predictive validity than prior piracy estimates.

Literature review

Aggregate level studies on the predictors of piracy activity tend to focus on samples at the national or state level. Typical predictors often found to be associated with piracy and piracy rates include economic, technological, legal, and cultural variables. A majority of the literature attempts to measure piracy as a rate, most often software piracy specifically. Piracy here is conceptualized as the percent of all intellectual goods of a certain category (software, music) of which are illegally acquired, relative to legal copies held by individuals or businesses.

Economic Predictors of Piracy

By far the most common and often associated predictor of national piracy rates is the wealth or income of a nation. In the case of software piracy rates (pirated software relative to legally owned software installed by businesses), Gross domestic product (GDP) per capita is typically found to be significantly and inversely related at the national level (Bagchi et al., 2006; Chen et al., 2010; Depken et al., 2004; Kovačić, 2007; Marron et al., 2000). Even if not significant, the effect is still negative (Andrés, 2006a; Andrés, 2006b). That is, nations with lower GDP per capita tend to have the highest software piracy rates.

Wealth measured as gross national income (GNI) per capita also produces the same negative sign relating to software piracy (Robertson et al., 2008; Yang et al., 2009). Bezmen and colleagues (2005) found similar results at the state level within the United States where those states with lower gross state product (GSP) per capita had higher estimated software piracy rates.

The relationship between wealth and software piracy may be more complicated than a simple linear association, as some investigations have yielded an inverted U shape of GDP per capita on piracy rates (Andrés, 2006b; Fischer et al., 2005). It is possible that GDP increases piracy rates initially, as the nation becomes more technologically advanced; until a certain peak where the nation's intellectual property protections become better enforced.

Other economic predictors aside from wealth have also been used to attempt to predict national piracy. While poorer nations tend to have higher piracy rates, nations with fewer economic hardships may also have higher piracy rates, controlling for wealth. Specifically, the Gini coefficient, which measures income inequality within a nation, has been found to be associated with a decrease in software piracy rates (Andrés, 2006a; Moores, 2008). That is, more inequality is associated with lower piracy rates. Music piracy rates, conversely, are associated with more inequality (Ki et al., 2006). Unemployment rates appear to also decrease piracy at the national level (Chen et al., 2010) but are not related at the state level within the USA (Bezmen et al., 2005).

Poorer nations may have higher piracy rates due to greater acceptance of piracy in such nations, in addition to some hostilities towards wealthier nations which tend to be home to the copyright owners in the first place from which intellectual property is acquired and distributed (Mattelart, 2009). It may also be that piracy itself can make an economy poorer, as it is suggested to be involved in subverting tax revenues and may harm legitimate sales (Desierto et al., 2010).

Technological Predictors of Piracy

Piracy and counterfeiting of intellectual goods depends on information technology for the copying and illegal distribution involved in such crimes. The technological nature of piracy naturally implicates measures of IT development and exposure at the national level when considering predictors of piracy. In the case of software piracy rates, illegal installations depend on business PCs being available, not to mention internet connectivity for distribution or at least networks within the country for purchasing (cheaper) counterfeit software CDs. However, it may also be that poorer nations (which tend to have higher piracy rates) have less easy access to legal technology and therefore resort to seeking illegal versions of a given product, suggesting less technologically advanced nations might have higher rates of piracy.

Technological variables can be measured in a number of different ways. Yang, Sonmez, Bosworth, and Fryxell (2009) attempted to measure each nation's overall technological development in the form of its information and communications technology (ICT). A nation's ICT development was found to decrease rates of software piracy, and it is argued that ICT infrastructure raises the demand for authentic software.

Bagchi, Kirs, and Cherveney (2006) also measured IT infrastructure, capturing it as the prevalence of both PCs and telephones available. Internet growth rate was also measured as the number of ISPs within a given nation. Piracy rates were available for three different years; 1996, 2001, and 2003. However, both IT infrastructure and the number of ISPs were mixed in terms of their impact on piracy rates. When these indicators were significant, they were inversely related with piracy outcomes.

A final study utilizing software piracy rates measured internet users per capita, and along similar lines as the other related literature, nations with higher software piracy tended to have fewer internet users (Robertson et al., 2008). However, Kranenburg and colleagues (2005) relied on estimated losses due to piracy per industry, not piracy rates, which may change the nature of the relationship between this measure and technology indicators. It was discovered that a nation's density of TVs was positively associated with reported motion picture industry losses due to piracy. PC density did not relate to any form of industry loss.

It has been suggested that at the individual level, those who engage in piracy may have more technological skill (Hinduja, 2003). However, at the national level, when using piracy as a rate, there appears to be more piracy activity in nations with poorer access to technology, even after controlling for that nation's wealth and other economic variables. Yet capturing piracy as a rate is not the only means to measure this type of crime.

Legal Predictors of Piracy

Given the illegal nature of digital piracy, copyright protections and other legal measures to combat this form of crime ought to be related. Stronger legal protections tend to be associated with lower piracy of all types. The causal ordering from which the two relate may not be clear, however. The question becomes whether laws deter piracy or whether the laws are simply better enforced in wealthier nations (which tend to have lower piracy rates).

Proxy measures of piracy related laws are often used. Marron and Steel (2000) found an index indicating the strength of a nation's economic institutions (strength of contract law, efficiency of bureaucracy, etc.) had a negative impact on software piracy rates. Andrés (2006a) also found an inverse relationship between software piracy and the chosen legal proxy variable; in this case the efficiency of the judicial system. Lastly, Kovačić, (2007) discovered a negative relationship on piracy for an index measuring perceptions on the effectiveness of the judiciary, incidence of crime, and the enforceability of contracts.

Proxy variables such as contract law and bureaucratic efficiency may only closely reflect the intellectual property protections in place that the variables attempt to measure. More direct measures include an index representing a nation's ratification of international treaties involving intellectual property (IP) laws (Andrés, 2006b), which are similarly found to be associated with lower software piracy. In the case of music piracy rates, a scale from the Economics of Freedom of the World annual reports representing a nation's rated IP protection is also associated with lower piracy. Even when piracy is measured in estimated financial losses due to file sharing and counterfeiting, the legal impact is mostly negative,

measured by national membership in international copyright conventions and treaties (Kranenburg et al., 2005).

The majority of the findings indicate nations with higher piracy tend to have weaker laws preventing such illegal activity. Andrés (2006a) suggests stronger institutions will increase the potential costs of engaging in violation of copyright law, making piracy less attractive. There is evidence that those who engage in piracy have a more rational thinking process that motivates them to infringe copyright at the individual level (Chen, Shang, & Lin, 2008; Wolfe, Higgins, & Marcum, 2007; Yoon, 2011); so deterrence strategies may be effective.

Additionally, knowledge of the laws on copyright (Goles, 2007), as well as prediction of legal consequences for such infringement (Chiou et al., 2005; Kwong et al., 2002; Limayem et al., 2004; Tan, 2002), are associated with lower intents to infringe copyright. However, there is skepticism that enhancements to the laws on legal copyright protections have much of an impact on piracy within the given country (Kilpatrick-Lee, 2005). Copyright protections may deter pirates, or copyright protections may additionally be stronger in nations with cultures less tolerant of piracy.

Educational Predictors of Piracy

Measures of the average educational level of citizens within a given country are mixed in terms of their relation to local piracy rates when controlling for other potential predictors. Marron and Steel (2000) measured education as the average total years spent in formal schooling for citizens aged 25 and over. Average years of schooling was found to be associated with lower software piracy rates. However, Andrés (2006a) found no significant relationship between the same measure and software piracy.

Educational attainment can also be measured as a literacy rate. However, Depken and Simmons (2004) found that it is mixed in terms of predicting software piracy. An educational index combining both average years of schooling and literacy rates fares little better, as it has been found to be unrelated to music piracy rates, at least (Ki, Chang, Khang, 2006). It appears education is a weak predictor of piracy rates once other related factors are taken into account.

Measures of Piracy Outcomes

The majority of the research reviewed here uses software piracy rates as the preferred measure of piracy outcome at the national level. Piracy rates are the percent of software units installed by businesses that were illegally acquired. The measure is taken from reports released by the Business Software Alliance (BSA), and relies heavily on estimates about each country's number of business PCs, software use, and illegal software installations. The estimation of such rates is a three step process that requires (1) determining how much PC software was deployed during a given year, (2) how much was paid for or otherwise legally acquired during the year, and (3) subtracting one from the other to get the amount of unlicensed software (Business Software Alliance, 2011).

The reliability and validity of these estimates have been questioned (Goldman, 2003). These piracy rate numbers are estimated by taking the number of computers shipped to a given country, estimating why those computers were purchased and estimating the number of business software programs that would have been licensed based on that nation's technological maturity. The number of legitimate sales is estimated via confidential data reported by various BSA member companies. It has been suggested that

there are too many estimates and not enough transparency in this data gathering process (ibid).

The measure's creation also relies on average expectations for business software load (amount of installed business software); yet these averages were based on research solely in the United States and it is unclear how they were adjusted to estimate the software load in other countries (Png, 2010). For this measure, the BSA had contracted the service of a consultant company, the International Planning and Research Corporation (IPRC).

However, in 2003, the BSA switched to the International Data Corporation (IDC) as their primary consultant for this measure, which used a slightly different methodology for estimating piracy rates. The IDC's estimates come from consumer and business user surveys in 15 countries to estimate software load, which are then adjusted for the remaining countries in the sample based on estimation. How the estimates are adjusted for other countries is not disclosed by the IDC, although it is suggested that national wealth (GDP, etc.) was involved in the estimation process (ibid). Many studies have found that income is the single most important influence on software piracy rates, and it may be due to the means by which these rates are estimated in the first place.

Regardless of the veracity of such measures, it is still important to consider other measures of piracy at the national level. For instance, measures released by the BSA only capture piracy activity of businesses; and in addition, it is a rate based on legitimate software. Different results may be gleaned from other measures, hopefully reflecting piracy activities of the general public. It may also be found that absolute piracy, not simply piracy as a rate, may have different predictive implications.

It is generally found that wealthier nations have lower software piracy rates. In developing countries, software piracy loss is estimated to be greater relative to that nation's GDP, as compared to developed and wealthier nations. However, the absolute loss due to piracy is higher in industrialized countries, even though it is lower relative to its GDP (Ding et al., 2009).

Alternative measures of piracy, such as absolute losses or piracy activity, may be equally important in terms of policy considerations. While lower income nations may have higher piracy rates, their fiscal impact on copyright owners is possibly much less than the harms of illegal copyright infringement carried out in richer nations. Thus, fighting piracy in higher income nations may yield better rewards than targeting developing nations. It is the purpose of this research to address these considerations.

Methods

Sample

The population from which the sample was drawn from includes all sovereign countries in the world. The estimates for the total number of countries in the world vary depending on how the number is estimated (Rosenberg, 2010). The count of countries can range from between 193 and 200. The Bureau of Intelligence and Research (2009) estimates the count to be 194.

The data used comes from multiple sources, and not all data sources had complete information on all nations. For any missing cases for a given country, a list wise deletion was conducted, resulting in 107 nations total in the sample for which there were no missing data points for all variables used in the study. This included list wise deletions across each of the four dependent variables as well. This was done to make the results from the four models based on these dependent variables comparable. Each subsequent

regression analysis would derive from the same 107 countries so as to compare the different measures of piracy with one another. Together, there are 2 countries from North America, 17 from Latin America, 36 from Europe, 5 from the Middle East, 24 from Asia, 21 from Africa, and 2 from Oceania.

Independent Variables

The independent variables selected are intended to represent each nation's economic, technological, legal, educational, and health related aspects, as well as country size. It should be noted that there is not a measure of culture in the dataset (individualism, etc.). This was deliberate as all sources of cultural measures located did not have data on a large enough sample of countries ($n < 100$). Including them would result in suffering a hit to sample size.

GDP per Capita (GDPPC): The measure of GDP comes from the International Monetary Fund website (World Economic Outlook Database, 2011). GDP represents a nation's wealth in 2010 in billions of dollars, and includes data on 184 countries. GDP was converted into a per capita measure based on population size, described below. The natural log of GDP per capita will be used to correct for positive skew.

Gini Coefficient (Gini): The Gini coefficient represents the income inequality in a nation. Higher scores represent higher inequality. A minimum score of 0 indicates complete equality (everyone has identical wealth), whereas a maximum score of 100 suggests complete inequality (one person has 100% of all income). The measure comes from The World Bank (2009) website. The variable includes data on 171 countries, and represents the Gini coefficient per country between the years 1992 and 2009. The measure was log transformed to correct for positive skew.

Unemployment Rate (Unemployment): Unemployment rates were taken from the World Factbook website (CIA World Factbook, 2008). The measure represents unemployment rates in 2008 and includes 197 nations. The log of unemployment rates will be used in subsequent analyses.

Percentage of Internet Users (Net Users): The Internet World Statistics website provided both national population size and percent of the population who are internet users (Internet World Stats, 2010). The estimates are from June 30, 2010 and include data on 216 countries.

Intellectual Property Protection (IP Protection): Protective IP laws were taken from the Global Competitiveness Report (Schwab, 2010). The measure is based off of survey data conducted on 139 countries involving 13,607 respondents (about 98 respondents per country). Leaders from international, public, and private organizations were surveyed for the report. Respondents were asked to rate their country's level of IP protection, such as anti-counterfeiting measures on a scale of 1 to 7. The inverse of the scale multiplied by -1 was used for subsequent analysis to correct for skew.

Average Years of Schooling (Yrs School): The mean number of years of formal schooling (primary, secondary, postsecondary) for adults age 25 and older is intended to measure the nation's educational level. There is complete data on 187 countries taken from the United Nations Development Program (2011). The variable is squared to adjust for negative skew.

Life Expectancy (Life Expect): To measure general health, life expectancy at birth will be included in the models used. The variable represents the number of years newborns are expected to live assuming prevailing patterns of mortality rates at the time of birth remain

the same throughout the child's life. The data were acquired from the United Nations Development Program (2011), and includes 187 countries.

Population Size (Population): Population size was acquired from the same source as the variable of internet users per capita (see above), and has data on 216 nations. Population size is intended as a control variable to account for the differing sizes of each respective country. The logarithm of this measure was computed to adjust for positive skew.

Dependent Variables

There are four dependent variables, each of which will be used in a separate regression model that includes all the aforementioned independent variables. The first two DVs (software piracy rate and value of unlicensed software) both come from the Business Software Alliance, which the majority of previous literature relies on. The second two DVs include the count of BitTorrent trackers and the number of file sharing client downloads per internet user. These two measures do not come from the BSA, or any large copyright stakeholder lobby.

Software Piracy Rate (Software Piracy): Software piracy is an estimate of the percent of all software installed on business PCs that were illegally acquired. It is calculated as the total number of illegal installations divided by the count of all software installed on business computers per country. The measure was taken from a report by the Business Software Alliance (2011), and includes rates for 116 countries. Software piracy rate was cubed to correct for negative skew.

Commercial Value of Unlicensed Software (Piracy Cost): The value of unlicensed (pirated) software is also taken from the Business Software Alliance report (2011). The variable represents the value of all pirated software installed during a given year as if it had been sold in the market. Unlicensed software is calculated as the number of unlicensed software units multiplied by the average software unit price.

Count of BitTorrent Trackers (Trackers): BitTorrent is a decentralized P2P protocol used for file sharing. BitTorrent networks are managed by servers called trackers, which users of the network must connect to in order to begin downloading a chosen file (Cuevas et al., 2010). Unlike previous P2P networks, BitTorrent trackers are less centralized and tracking servers can be hosted anywhere in the world where there is an internet connection. There also exist many BitTorrent tracker lists users may download to add to their chosen file sharing client; the intended purpose of which is to facilitate faster and more efficient downloads.

Eight BitTorrent tracker lists were downloaded for this research from multiple BitTorrent indexes and file locker websites. The lists included the URL of each tracker, the total of which was 2,476 trackers from the 8 lists. A script written in PERL was executed to automatically run DNS lookups of each URL to retrieve its IP address. At the time the script was run, 576 trackers could not be resolved on January 3, 2012. Of the IP addresses acquired, 998 were unique. The addresses were geolocated by country of origin on January 3, 2012 (via the following service: <http://software77.net/geo-ip/multi-lookup>). Six IP addresses were reserved and could not be geolocated, totaling 992 unique trackers from 51 nations that had at least one tracker. All remaining countries not included were assigned a zero indicating no trackers present. Trackers are an absolute count variable, and not measured as a rate.

It should be mentioned that the number of BitTorrent trackers is not a direct measure of piracy activity. While BitTorrent trackers are necessary to use the networks for

infringing purposes, the number of trackers does not reflect infringing traffic, but rather reflects the number of piracy facilitating servers and where they are hosted. Additionally, the actual pirates (or legitimate users) themselves can reside anywhere in the world when using such servers to facilitate BitTorrent file distribution. The trackers are not intended to represent where the pirates themselves reside, but rather where the facilitators of those pirates decide to host tracking servers. This measure is intended to capture piracy facilitating countries, rather than pirate residing countries.

File Sharing Client Downloads per Internet User (Download Rate): The number of downloads of four file sharing clients available from SourceForge.net in 2010 were used for this variable. SourceForge is a free online software repository for developers to maintain and share open source software that can be downloaded. File sharing clients (such as that of BitTorrent) are available for download at this website. A search for “p2p” on the SourceForge website was run on December 16, 2011 and the first four results were used for subsequent analysis (all file sharing clients). Top results are the most frequently downloaded (tens to hundreds of thousands of downloads per week). The four clients were Ares Galaxy, Vuze, Shareaza, and a Mule.

SourceForge has a statistical reporting form that users can utilize to compute download statistics for each software application present on the site. For each of the four clients, the total number of downloads for the year 2010 per country were computed. There were over 82 million downloads in 251 countries and sovereign states in 2010 among the four file sharing clients. The number of downloads per country will be used as a rate based on the number of internet users in each nation (taken from the same source as the internet users per capita variable described above).

It is the intent of this research that this measure captures some of the variation in infringing behavior, as P2P clients are often necessary to infringe copyright via distributional networks. It is expected that many of these P2P clients that were downloaded were subsequently used for infringing purposes. While P2P networks can easily be used for non-illegal activities, such as sharing non-copyrighted works, much of the evidence monitoring these types of networks indicates the majority of P2P traffic is in fact infringing (Envisional, 2011).

Results

Bivariate Analysis

Table 1 presents a Spearman correlation matrix of the eight independent and four dependent variables. Of the four outcome variables measuring piracy, software piracy rate (Software Piracy) is inversely related to the remaining three piracy measures: Piracy Cost, Trackers, and Download Rate, although Piracy Cost is not significant. Software Piracy is in fact negatively correlated (even if not significant) with every other variable except Gini and Unemployment. Also note that GDPPC is, as expected, negatively associated with Software Piracy, but is positively related to the remaining three piracy measures without controlling for anything.

Table 1. Spearman Correlation Matrix of Outcome and Predictor Variables (n = 107)

	1	2	3	4	5	6	7	8	9	10	11	12
1. Software Piracy	1.0											
2. Piracy Cost	-.14	1.0										
3. Trackers	-.55***	.41***	1.0									
4. Download Rate	-.42***	-.06	.14	1.0								
5. GDPPC	-.86***	.19*	.56***	.45***	1.0							
6. Gini	.25**	-.1	-.3**	.13	-.3**	1.0						
7. Unemployment	.41***	-.26**	-.53***	-.17	.5***	.25**	1.0					
8. Net Users	-.82***	.16	.62***	.28**	.89***	-.4***	-.48***	1.0				
9. IP Protection	-.75***	.12	.44***	.17	.73***	-.28**	-.39***	.64***	1.0			
10. Yrs School	-.63***	.03	.61***	.17	.74***	-.46***	-.43***	.78***	.53***	1.0		
11. Life Expect	-.76***	.18	.5***	.44***	.84***	-.36***	-.47***	.84***	.63***	.66***	1.0	
12. Population	-.07	.58***	.22*	-.09	-.08	.01	-.12	-.06	-.05	-.19	-.04	1.0

* $p < .05$, ** $p < .01$, *** $p < .001$

Multivariate Predictors of Software Piracy Rates

Table 2 contains the results of the OLS regression used to assess software piracy rates. The eight predictor variables included in the model jointly and uniquely explain a significant amount of the variance in software piracy ($R^2 = .80$, $F(8, 98) = 50.54$, $p < .0001$). The size of the R^2 yielded by the model was curiously higher than expected, so separate OLS models were run for each of the independent variables alone to determine the specific variables inflating the variance explained. By far the biggest predictor was the measure of GDP per capita. Alone in the model, it explained 73% of the variance in software piracy rates. GDP is a consistent predictor of piracy rates in prior literature. However, how the BSA calculates piracy rates is not completely transparent, and there are some suggestions that their estimates rely on measures of GDP itself (Png, 2010).

However, the original model formulation is left intact for the purposes of this study. The mean variance inflation factor for the predictors used in the model is 3.24, suggesting multicollinearity is not problematic. As is expected, GDP per capita (GDPPC) has the strongest impact on software piracy rates ($B = -.51$, $p < .001$), suggesting businesses in wealthier nations are less likely to install counterfeit software relative to legitimate software installed. The remaining predictors in the model are all significant and negative, except for average years of schooling (Yrs School) and life expectancy (Life Expect).

Table 2. OLS Regression of Software Piracy Rates (n = 107)

Variable	Unstandardized	Standardized	SE
GDPPC	-.092***	-.506	.022
Gini	-.13*	-.11	.063
Unemployment	-.036*	-.117	.017
Net Users	-.004**	-.384	.001
IP Protection	-.522*	-.176	.198
Yrs School	.0002	.034	.0005
Life Expect	-.0007	-.024	.002
Population	-.027**	-.159	.008
Intercept	2.169***		.343
Model diagnostics			
F(8, 98)		50.54***	
R ²		80.49%	

* $p < .05$, ** $p < .01$, *** $p < .001$

Consistent with previous literature, intellectual property protections (IP Protection) appear to be associated with a decrease in software piracy rates ($B = -.18, p = .01$). Also consistent with prior research, a higher percentage of the population who are internet users (Net Users) is associated with less software piracy ($B = -.38, p = .001$). Lastly, higher income inequality (Gini) ($B = -.11, p = .043$) and unemployment ($B = -.12, p = .041$) predict lower software piracy rates.

Notice that income inequality and unemployment are negatively associated with software piracy rates in the regression model, but they were positively and significantly associated with software piracy rates in correlation matrix in Table 1. This indicates a suppressive effect of a third variable. Sequential elimination of predictors from the regression model (not shown) reveal that once GDP per capita is removed from the model, unemployment and income inequality become positive again. This suggests that economic turmoil generally decreases software piracy rates, but that when two countries have equal amounts of wealth, the nation with higher economic turmoil (unemployment, inequality) will actually have higher software piracy rates.

Multivariate Predictors of the Cost of Unlicensed Software

An OLS model using Piracy Cost as an outcome (not shown) failed the Cook-Weisberg test for heteroscedasticity ($\chi^2 = 32.58, p < .0001$). Significant heteroscedasticity can indicate specification error or the omission of modeling certain curvilinear relationships present in the data. A scatter plot matrix (not shown) revealed a noticeable curvilinear relationship between life expectancy (Life Expect) and internet users per capita (Net Users). It was suspected that this might indicate an interaction, so the model was rerun with a Life Expect by Net Users interaction included (not shown). However, while this did decrease heteroscedasticity, it did not do so sufficiently ($\chi^2 = 21.7, p < .0001$), although the interaction was significant.

Table 3. WLS Regression of Software Piracy Costs ($n = 107$)

Variable	Unstandardized	Standardized	SE
Life Expect x Net Users	.274*	.122	.117
GDPPC	.375**	.207	.114
Gini	-.104	-.015	.248
Unemployment	-.117	-.052	.094
Net Users	.008	.004	.183
IP Protection	-.874	-.024	1.518
Yrs School	-.002	-.06	.002
Life Expect	.093	.02	.187
Population	.912***	1.026	.034
Intercept	-12.777***		1.99
Model diagnostics			
$F(8, 97)$		254.36***	
R^2		95.94%	

* $p < .05$, ** $p < .01$, *** $p < .001$

Weighted least squares (WLS) regression was used to correct for non-constant variance in the residuals. A plot of the residuals against GDPPC and Net Users (not shown)

revealed these two predictors to be problematic, so they were weighted using the inverse of the variance for Piracy Cost. A subsequent plot of the residuals against the fitted values of the model indicated heteroscedasticity was substantially reduced (Christensen, 2002).

The results of the WLS model can be seen in Table 3. The model is significant ($R^2 = .96$, $F(9, 97) = 254.36$, $p < .0001$), suggesting at least one coefficient is greater than zero. To investigate why the coefficient of determination was so high at above 90%, separate models were run for each individual variable alone. By far the biggest culprit was population size, explaining 55.49% of the variation. In the full model, the only significant predictors of Piracy Cost include the interaction term between Net Users and Life Expect ($B = .12$, $p = .022$), GDP per capita (GDPPC) ($B = .21$, $p = .001$), and of course population size ($B = 1.03$, $p < .001$). Specifically, nations with higher GDPPC have more costly amounts of pirated software. This is in the opposite direction of software piracy rates. Also, larger nations with more citizens accrue more costs in terms of pirated software.

The significant interaction can be interpreted to mean that for nations with a higher number of internet users (one standard deviation above the mean on Net Users), life expectancy can be found to increase piracy costs. However, for nations with a low number of internet users (one standard deviation below the mean), life expectancy is found to decrease piracy costs.

Multivariate Predictors of BitTorrent Tracker Count

Negative binomial regression was used to predict Trackers since count data tends to fit the negative binomial distribution best as opposed to a poisson distribution (Lawless, 1987). The model can be found in Table 4. A likelihood ratio test of the overdispersion parameter, alpha, indicated significant overdispersion ($\chi^2 = 600.27$, $p < .001$). This suggests negative binomial is more appropriate than a poisson regression model, as negative binomial models assume overdispersion (the variance exceeds the mean on average) (Gardner et al., 1995).

Table 4. Negative Binomial Regression of BitTorrent Trackers ($n = 107$)

Variable	Unstandardized	Standardized	SE
Life Expect x Net Users	-1.194*	.326	.584
GDPPC	-.021	.97	.354
Gini	-.748	.848	1.184
Unemployment	-1.069**	.406	.346
Net Users	1.803**	6.07	.622
IP Protection	4.067	1.429	4.521
Yrs School	.028***	3.12	.007
Life Expect	-.516	.597	.592
Population	.727***	3.067	.109
Intercept	-7.793		.5671
Model diagnostics			
Log likelihood		-175.19	
Likelihood-ratio test		111.81***	
McFadden's R^2		24.19%	

* $p < .05$, ** $p < .01$, *** $p < .001$

The likelihood-ratio test of the model in Table 4 indicates that at least one coefficient in the model is significantly different from zero ($\chi^2 = 111.81, p < .0001$). Higher unemployment suggests fewer number of trackers in a given country ($B = .41, p = .002$). As might be expected, more internet users per capita (Net Users) is associated with more trackers ($B = 6.07, p = .004$), probably because internet users are needed to set up P2P trackers in the first place. Higher levels of education (Yrs School) indicate more trackers as well ($B = 3.2, p < .001$); perhaps because tracker management requires a certain level of technical skill. Larger countries represented by higher population sizes are also associated with more absolute numbers of trackers ($B = 3.07, p < .001$). More people within a country probably means more people to set up tracking servers, as this was not a rate variable.

The interaction between life expectancy (Life Expect) and the number of internet users (Net Users) on Trackers was also included in the model, which was significant ($B = .33, p = .041$). The interaction indicates that in nations with higher concentrations of internet users, life expectancy decreases tracker count. In areas low in internet users, life expectancy can be shown to increase tracker counts. This finding is in the opposite direction of that found in the interaction in the previous model analyzing Piracy Cost.

Multivariate Predictors of File Sharing Client Downloads per Internet User

Client download rate (Download Rate) is based on the count of downloads per country. All nations in the sample had at least one or more downloads of the four file sharing clients in 2010. This means that Download Rate has no zero values in it. Therefore, zero truncated negative binomial regression is required, as there are no zero values in the model to estimate (Long et al., 2006).

Table 5. Zero Truncated Negative Binomial Regression of P2P Client Download Rate ($n = 107$)

Variable	Unstandardized	z	SE
Life Expect x Net Users	.577**	2.99	.193
GDPPC	.66**	2.81	.235
Gini	1.234*	2.24	.55
Unemployment	.152	1.03	.148
Net Users	-.734**	-2.75	.267
IP Protection	-4.575*	-2.43	1.88
Yrs School	-.009*	-1.98	.004
Life Expect	1.298***	4.25	.305
Population	-.083	-1.1	.075
Intercept	-13.603***	-4.88	2.79
Model diagnostics			
Log likelihood		-1375.94	
Likelihood-ratio test		53.45***	
McFadden's R^2		1.91%	

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 5 presents the results of the zero truncated negative binomial regression of Download Rate on the predictors. The model uses the number of internet users as the exposure term, effectively turning the number of downloads (which is the outcome variable) into a rate (Frome, 1983). The likelihood-ratio test of the model in Table 5 suggests that at least one coefficient in the model is significantly different from zero ($\chi^2 = 53.45, p < .0001$).

All predictors in the model except for population size and unemployment are significant. Wealthier nations (GDPPC) ($b = .66, p = .005$) with higher income inequality (Gini) ($b = 1.23, p = .025$), and longer life expectancy (Life Expect) ($b = 1.3, p < .001$), have higher file sharing client downloads per internet user. Countries with more internet users (Net Users) ($b = .73, p = .006$), higher intellectual property protections (IP Protection) ($b = -4.57, p = .015$), and higher levels of education (Yrs School) ($b = -.01, p = .048$), had lower file sharing downloads. The negative relationship between both Net Users and IP Protection on Download Rate mirror that found for the measure of software piracy rates.

It should be noted that Net Users was positively associated with Download Rate in the correlation matrix presented in Table 1 ($r = .28, p = .003$). However, once Life Expect is controlled for in this negative binomial model, the relationship becomes negative. Where two nations have the same life expectancy, the country with more internet users will download fewer file sharing clients.

The interaction between Life Expect and Net Users significantly relates to downloads. Specifically, nations with fewer internet users per capita have a weak positive impact of life expectancy on download rates. High internet use countries have a stronger positive impact of life expectancy on download rates. Additionally, looking at the interaction a different way, in nations with higher life expectancy, the number of internet users has little impact on downloads. Yet in low life expectancy nations, internet use is negative in its impact on download rates.

Discussion and Conclusion

Many of the predictors of software piracy rates in this analysis have also been found to predict the alternate measures of piracy chosen for this study. However, the direction in which the selected variables predict piracy outcomes is not consistently the same across the different ways in which piracy was measured. In many instances, it appears that piracy costs, BitTorrent tracker hosting, and file sharing client download rates have some similar effects in terms of their prediction.

GDP per capita, for instance, positively predicts piracy costs and download rates, yet consistent with previous research is inversely related to software piracy rates (while BitTorrent trackers are not significant). GDP has traditionally been the largest predictor, in terms of effect size, of software piracy rates, and that was supported here. For the remaining three models, however, GDP tended to be highly predictive, but was not the largest of the predictors chosen in terms of its impact on piracy outcomes.

This research also uncovered some suppressive and interaction effects in the data that were not the original intent of this investigation. Three significant interactions between life expectancy and internet users per capita on piracy outcomes were found. The interaction term was not significant for software piracy rates (not shown in table). The interpretation and meaning behind the different interactions is unclear, as each of the three had slightly different interactive effects on piracy.

In high internet use countries, life expectancy increases the costs of piracy and P2P client downloads, but decreases the number of BitTorrent trackers. In nations with low internet connectivity, life expectancy decreases the costs of piracy, increases the count of trackers, and has a very weak positive impact on file sharing downloads per capita. This could suggest that the three separate measures of piracy are too different from each other. They may measure different aspects of piracy (rate vs. absolute values), or some may not reflect piracies well as others (and be spuriously related to other national characteristics). The different measures of piracy can tell us different things.

The previous literature has relied mostly on one or two different measures of piracy. Most research on piracy at the national and state level to date has so far utilized almost exclusively piracy rates and estimated fiscal losses due to piracy reported by copyright owner industries. This study sought to incorporate additional measures that were intended to capture illegal piracy activity independent of the usual copyright industry's influence.

While the differing piracy measures related to the selected variables in different ways, there appear to be some patterns in the data. Software piracy rates tended to be inversely related to predictors such as population size, GDP, and the number of internet users per capita. That is, bigger, richer, more technologically advanced nations had lower business piracy rates. However, piracy measured in absolute terms tended to be positively associated with these variables. The absolute number of BitTorrent trackers was positively associated with population size and internet access, whereas the absolute fiscal amount pirated positively related to GDP per capita. While bigger, richer, more advanced countries have lower piracy rates, they tend to have much higher absolute piracy activity.

The question becomes which measure of piracy activity is of more use to stakeholders. Absolute piracy is probably a better measure of the fiscal and legal impact piracy has on businesses and stakeholders. If one country has a higher piracy rate, but is a smaller nation, that country is not affecting businesses as much as larger nations with higher absolute piracy activity. Piracy relative to legal software installations tells us less of the overall impact piracy can have on business if those countries are poorer and smaller to begin with. The total costs and activity of piracy are probably better predictors of which nations are more costly to stakeholders.

Wealthier nations' pirate more, it appears, but they likely also contribute more to legal purchases. Richer nations likely benefit copyright stakeholders more in this way in absolute terms, but they are also similarly infringing more of those same goods in absolute terms. Richer and larger countries buy more copyrighted goods, and should be of interest to stakeholders in the case of sales, but also pirate more of those goods, and should similarly be equally of interest to stakeholders in terms of reducing infringement.

It therefore may be advisable to target bigger, more advanced nations, assuming a choice has to be made between the two. Richer and more technologically developed countries are likely the bigger criminal culprits when it comes to the costs imposed on the copyright industry that can potentially be turned around with policy efforts. There may be a bigger yield on investments if the industry focuses on these types of nations, instead of poorer ones. While infringing acquisitions of copyrighted goods are less relative to legally purchased goods in such countries, the absolute value of such infringed goods is higher. Considering this, there ought to be a greater return on investments for focusing on such nations. Wealthier countries higher in GDP should be a greater concern to copyright owners.

However, the new measures of piracy chosen are not without limitations. Neither of these variables are direct measures of piracy. They are indirect, instead measuring file sharing related activity or file sharing facilitating structures of a given country. That file sharing activity may or may not be used for piracy. While the majority of traffic on file sharing networks is infringing (Envisional, 2011), not all of it is; and it is even less clear how the number of trackers or client downloads relates to actual illegal downloads. Illegal downloads were not measured. BitTorrent trackers are legal to host, and it is similarly legal to download and install file sharing clients. It is how those networks are used that we can distinguish between infringing and non-infringing activity.

Additionally, the number of trackers only facilitates others who wish to download content from the internet. It may not reflect which country is actually doing the downloading. One country may host many trackers, but users from other countries may be the ones using those foreign trackers to conduct illegal file sharing. Lastly, client download rate does not necessarily reflect the majority of client downloads. Most client downloads likely do not come from SourceForge. Although there were millions of downloads at this particular website, these downloaders may be systematically different from downloaders elsewhere.

However, it is clear that different results can be gleaned depending on the intended measure of piracy used. Most research has measured piracy as a rate based on total copyrighted goods owned. Many authors use the words piracy and piracy rates interchangeably, although piracy as a rate is not the sole means to measure this form of crime. Instead of piracy relative to legal activity, this study sought to also measure absolute piracy, in the form of software piracy costs and BitTorrent tracker counts.

Whether to focus on piracy as a rate or not can determine different policy recommendations. Poorer nations tend to have higher piracy activity relative to legal activity (at least among businesses), whereas piracy in general can indicate that wealthier nations cost and contribute more to piracy. In terms of which countries ought to be a bigger concern to copyright stakeholders, there may be a bigger return on investments for focusing on wealthier nations, as they may cost copyright owners more in absolute dollars.

References

- Andrés, A. R. (2006a). Software piracy and income inequality. *Applied Economics Letters*, 13, 101-105.
- Andrés, A. R. (2006b). The relationship between copyright software protection and piracy: Evidence from Europe. *European Journal of Law and Economics*, 21(1), 29-51.
- Bagchi, K., Kirs, P., & Cervený, R. (2006). Global software piracy: Can economic factors alone explain the trend? *Communications of the ACM*, 49(6), 70-75.
- Bezmen, T. L., & Depken, C. A. (2006). Influences on software piracy: Evidence from the various United States. *Economics Letters*, 90, 356-361.
- Bureau of Intelligence and Research (July 29, 2009). Independent states in the world. U.S. Department of State. Retrieved November 13, 2010 from <http://www.state.gov/s/inr/rls/4250.htm>
- Business Software Alliance (May, 2011). *Eighth annual BSA global software 2010 piracy study*.
- Chen, C., Chen, C., & Yeh, C. (2010). Determinants of software piracy: Evidence from far East countries. *Journal of International and Global Economic Studies*, 3(2), 53-62.

- Chen, Y. C., Shang, R. A., & Lin, A. K. (2008). The intention to download music files in a P2P environment: Consumption value, fashion, and ethical decision perspectives. *Elsevier*, 411-422.
- Chiou, J. S., Huang, C. Y., & Lee, H. H. (2005). The antecedents of music piracy attitudes and intentions. *Journal of Business Ethics*, 57, 161-174.
- Christensen, R. (2002). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer: New York.
- CIA World Factbook (2008). Retrieved February 13, 2011 from http://www.photius.com/rankings/economy/unemployment_rate_2008_1.html
- Clough, J. (2010). *Principles of Cybercrime*, Cambridge University Press: Cambridge.
- Copyright Law of the United States. 17 U.S.C. Sec. 506 (October, 2009)
- Cuevas, R., Kryczka, M., Cuevas, A., Kaune, S., Guerrero, C., & Rejaie, R. (2010). Is content publishing in bittorrent altruistic or profit-driven? Co-NEXT '10 Proceedings of the 6th International Conference. ACM New York, NY, USA. ISBN: 978-1-4503-0448-1.
- Depken, C. A., & Simmons, L. C. (2004). Social construct and the propensity for software piracy. *Applied Economics Letters*, 11, 97-100.
- Desierto, D. A., Lazaro, K. A. D., Cruz, K. T. G. (2010). Is it worth taxing pirated products? The case of optical media discs in the Philippines. *Asian Economic Papers*, 9(2), 79-112.
- Ding, C. G., & Liu, N. (2008). Productivity changes of Asian economies by taking into account software piracy. *Economic Inquiry*, 47(1), 135-145.
- Envisional (January, 2011). Technical report: An estimate of infringing use of the internet. Technical report.
- Fischer, J. A. V., & Andrés, A. R. (August, 2005). Is software piracy a middle class crime? Investigating the inequality-piracy channel. Discussion paper no. 2005-18, Department of Economics, University of St. Gallen.
- Frome, E. L. (September, 1983). The analysis of rates using poisson regression models. *Biometrics*, 39, 665-674.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Quantitative Methods in Psychology*, 118(3), 392-404.
- Goldman, E. (2003). A road to no warez: The No Electronic Theft Act and criminal copyright infringement. *Oregon Law Review*, Summer 2003.
- Goles, T., Jayatilaka, B., George, B., Parsons, L., Chambers, V., Taylor, D. & Brune, R. (2007). Softlifting: Exploring determinants of attitude. *Journal of Business Ethics*, 77, 481-499.
- Hinduja, S. (2003). Trends and patterns among online software pirates. *Ethics and Information Technology*, 5, 49-61.
- Internet World Stats (2010). Retrieved December 2, 2010 from <http://www.internetworldstats.com>
- Ki, E. J., Chang, B. H., & Khang, H. (2006). Exploring influential factors on music piracy across countries. *Journal of Communication*, 56, 406-426.
- Kilpatrick-Lee, D. L. (2005). Criminal copyright law: Preventing a clear danger to the U.S. economy or clearly preventing the original purpose of copyright law? *University of Baltimore Intellectual Property Law Journal*, Fall, 2005.

- Kovačić, Z.J. (2007). Determinants of worldwide software piracy. Paper presented at the InSITE conference Ljubljana, Slovenia.
- Kranenburg, H. V., & Hogenbirk, A. (2005). Multimedia, entertainment, and business software copyright piracy: a cross national study. *Journal of Media Economics*, 18(2), 109-129.
- Kwong, T. C. H. & Lee, M. K. O. (2002). Behavioral intention model for the exchange mode internet music piracy. Proceedings of the 35th Hawaii International Conference on Systems Sciences.
- Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics*, 15(3), 209-225.
- Limayem, M., Khalifa, M., & Chin, W. W. (2004). Factors motivating software piracy: A longitudinal study. *IEEE Transactions on Engineering Management*, 51(4).
- Long, J. S., & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata* (Second Edition). College Station, TX: Stata Press.
- Marron, D. B., & Steel, D. G. (2000). Which countries protect intellectual property? The case of software piracy. *Economic Inquiry*, 38(2), 159-174.
- Mattelart, T. (2009). Audio-visual piracy: Towards a study of the underground networks of cultural globalization. *Global Media and Communication*, 5(3), 308-326.
- Moores, T. T. (2008). An analysis of the impact of economic wealth and national culture on the rise and fall of software piracy rates. *Journal of Business Ethics*, 81, 39-51.
- No Electronic Theft Act (1997). Pub. L. No. 105-147, 111 Stat. 2678.
- OECD (2008). *Information Technology Outlook*. Organization for Economic Cooperation and Development.
- Png I. P. L. (2010) On the Reliability of Software Piracy Statistics. *Electronic Commerce Research and Applications*, 9(5), 365-373.
- Robertson, C. J., Gilley, K. M., Crittenden, V., & Crittenden, W. F. (2008). An analysis of the predictors of software piracy within Latin America. *Journal of Business Research*, 61(6), 651-656.
- Rosenberg, M. (June 14, 2010). The number of countries in the world. About.com. Retrieved on November 13, 2010 from <http://geography.about.com/cs/countries/a/numbercountries.htm>
- Schwab, K. (2010). *The global competitiveness report 2010-2011*. World Economic Forum.
- Stevens, L. K., & Sessions, D. N. (2005). An empirical investigation into the effect of music downloading on the consumer expenditure of recorded music: A time series approach. *Journal of Consumer Policy*, 28, 311-324.
- Tan, B. (2002). Understanding consumer ethical decision making with respect to purchase of pirated software. *Journal of Consumer Marketing*, 19(2), 96-111.
- The World Bank (2009). Retrieved December 31, 2011 from <http://data.worldbank.org/indicator/SI.POV.GINI>
- United Nations Development Program (2011). *Human development report 2011: Sustainability and equity: A better future for all*.
- Wolfe, S. E., Higgins, G. E., & Marcum, C. D. (2007). Deterrence and digital piracy: A preliminary examination of the role of viruses. *Social Science Computer Review*, 26(3), 317-333.

- World Economic Outlook Database (2011). International Monetary Fund. Retrieved December 27, 2011 from <http://www.imf.org/external/pubs/ft/weo/2011/02/weodata/index.aspx>
- Yang, D., Sonmez, M., Bosworth, D., & Fryxell, G. (2009). Global software piracy: Searching for further explanations. *Journal of Business Ethics*, 87, 269-283.
- Yar, M. (2008). The rhetorics and myths of anti-piracy campaigns: Criminalization, moral pedagogy and capitalist property relations in the classroom. *New Media and Society*, 10(4), 605-623.
- Yoon, C. (2011). Theory of planned behavior and ethics theory in digital piracy: An integrated model. *Journal of Business Ethics*, 100, 405-417.