Technological Development and Software Piracy[1]

Francisco Martínez-Sánchez[2]
University of Murcia

Andrés Romeu[3]
University of Murcia

Abstract

In this paper, we analyze the differences in piracy rates from one country to another. Like previous papers on the topic, we find that more developed countries have lower incentives for pirating. Unlike previous papers, we find that the piracy rate is positively correlated with the tax burden rate but negatively correlated with the domestic market size and exports over GDP. We also separate the impacts of education and R&D on piracy, and find two effects with opposite signs. Moreover, we find that those countries with smaller, more efficient bureaucracies are likely to protect intellectual property more effectively. Finally, we show that the spread of access to the Internet is negatively correlated with the software piracy rate.

Keywords: piracy rate; tax burden rate; domestic market size; education; R&D ; quality bureaucracies; intellectual property; Internet
JEL Classification: K42; L86; O3; O57

---

1 We acknowledge financial support from the Spanish *Ministerio de Economía y Competitividad* under projects ECO2010-19830 and ECO2013-45698.

2 Departmento de Métodos Cuantitativos para la Economía y la Empresa, University of Murcia, 30100 Murcia, Spain. E-mail: fms@um.es

3 Departamento de Fundamentos del Análisis Económico, University of Murcia, 30100 Murcia, Spain. E-mail: aromeu@um.es

1. Introduction

The software piracy rate measures the ratio of the value of unlicensed softwareto the total market value of software. In this paper we set out to find the causes for the differences in the software piracy rate from one country to another.[4] This kind of analysis is very important given the negative effects of piracy on economic growth (Andrés and Goel, 2012) and industry profits (Gomes et al., 2013). We contribute to the existing literature by using the most recent available piracy data from the latest Global Software Survey provided by the Business Software Alliance (BSA) for 111 countries and 7 years, and by building a panel with data from the Global Competitiveness Index data platform (CGI) of the World Economic Forum (WEF), that both provides its own data and collects data from other platforms such as the World Bank or UNESCO among others. Using these data, we build a model to assess the explanatory power of different factors that have been proposed in the relevant literature using model-selection techniques.

Software piracy rates across countries have been widely analysed ineconomic literature.[5] Thus, we know that one of the main factors that explain the differences in software piracy rates from one country to another is the national income per capita: countries with lower incomes per capita have higher piracy rates (Marron and Steel, 2000; Husted, 2000). This result has been confirmed using data from the United States as in Bezmen and Depken II (2006) and from 23 European countries as in Andrés (2006b), among others.

4 In this paper we do not distinguish between commercial piracy, which takes place when some firms illegally reproduce and sell copies of original software, and end-user copying, which refers to non-commercial copying by final consumers. However, this distinction is drawn in theoretical papers. For papers on commercial piracy see Banerjee (2003), Martínez-Sánchez (2010, 2013) and López-Cuñat and Martínez-Sánchez (2015), and for end-user copying see Bae and Choi (2006) and Martínez-Sánchez (2011, 2012, 2014).

5 For surveys of the theoretical literature see Belleflamme and Peitz (2012, 2014) and Peitz and Waelbroeck (2006), but for surveys of the empirical literature see Dejean (2009) and Waldfogel (2012).

Furthermore, national income is related to piracy rates, and not only in per capita terms. Andrés (2006a) finds that inequality has a negative impact on piracy.

The strength of institutions is also a major factor in explaining differences between countries in piracy. From Marron and Steel (2000) we know that piracy rates are lower in countries that have strong institutions able to enforce contracts and protect property from misappropriation. Shadlen et al. (2005) empirically show that countries with small, efficient, high-quality bureaucracies are likely to protect intellectual property more effectively.[6] Furthermore, a bigger shadow economy leads to higher piracy rates as shown in Goel and Nelson (2012), and Andrés and Asongu (2013) find that controlling corruption is the most effective tool for fighting piracy in Africa. Like these authors, we also find that our measures of high-quality bureaucracies are negatively correlated with piracy.

Bezmen and Depken II (2006) find a significant, negative relationship between state tax burdens and software piracy rates. They suggest that a higher tax burden might mean that the government has more resources to enforce property rights, thus resulting in lower piracy rates. However, it is well known that there is a strong relationship between tax burden and development. Our results confirm that more developed countries do indeed have lower piracy rates but unlike the said authors we find that the piracy rate is positively correlated with the tax burden rate, which we attribute to a price effect of sales taxes on incentives to engage in piracy. Furthermore, we also find that after controlling for domestic market size and exports over GDP the piracy rate is negatively correlated with these factors. Therefore it is reasonable to think that

6 However, these results are not robust to alternative specifications of the model Shadlen et al. (2005).

incentives for software piracy are smaller in large, open economies with competitive pricing.

In the literature the results concerned with the significance of education for the piracy rate are controversial. For instance, Marron and Steel (2000) and Andrés (2006b) find that a weak negative correlation between education and piracy while Depken II and Simmons (2004) and Shadlen et al. (2005) find a strong negative effect .[7] These ambiguous results cold be the result of two opposing effects, as suggested by Scalise (1997). On the one hand, education facilitates imitation and replication, which increases the rate of piracy. On the other hand, education also increases the demand for intellectual protection, which in turn decreases the piracy rate. We control for enrollment rates indifferent levels of education and find that the signs of the coefficients for higher education and lower levels differ.

The relevant literature also contains ambiguous results regarding the effect of R&D on software piracy rates. Marron and Steel (2000) and Andrés (2006b) show that countries with higher levels of R&D have low piracy rates, although the correlation is not significant. On the other hand, Shadlen et al. (2005) empirically find that the effect of R&D is positive and robust, except when they include a linear time-trend indicator. In this last case, the coefficient is still positive though not significant.[8] We again find that different measures of R&D have opposing effects: the number of scientists and engineers and

7 To measure national education levels, Marron and Steel (2000) use the average number of years of schooling among people aged 25 or older in 1990, Andrés (2006b) uses secondary school enrollment, Depken II and Simmons (2004) use the literacy rate as a proxy of education and Shadlen et al. (2005) uses a combined primary, secondary, and tertiary gross enrollment ratio for 1997.

8 To measure national levels of R&D, Marron and Steel (2000) and Andrés (2006b) use R&D expenditures as a percentage of GDP, while Shadlen et al. (2005) use the number of scientists and technicians working in R&D per 1,000 inhabitants (average 1990–1996).

technology absorption at firm-level positively affect piracy rates while R&D spending and other R&D measures show less piracy.

Marron and Steel (2000) show that intellectual property receives greater protection in developed countries, so low-income countries have higher piracy rates. In the same way, Andrés (2006b) finds that copyright protection of software is one of the most determinant factors of low software piracy rates. However, we find that the enforcement of ownership rights is not significant.

In recentyears the effect of wider access to Internet on legitimate sales has received increasing attention from scholars and the results are not entirely clear. On the one hand, Peitz and Waelbroeck (2004) suggests that Internet piracy played a significant role in the decline in music sales during the early days of file-sharing networks. On the other hand, Oberholzer-Gee and Strumpf (2007) and McKenzie (2009) empirically find that file sharing has no effect on sales in the music industry. A recent result by Goel and Nelson (2012) points in the direction of the latter. They empirically find that the spread of the Internet has positive spillovers in terms of lowering piracy. This probably happens because greater Internet availability makes tracking piracy easier. Our results support this hypothesis since Internet availability measured as broadband access penetration has a clearly negative impact on piracy rates.

Along the same lines asour work but focusing on the software piracy losses, Gomes et al. (2013) show that the number of patents by residents has a positive effect while the effect of R&D is negative. They also show that more spending on education increases losses from piracy but, at the same time, more years of schooling have the opposite effect. Finally, they find that access to the Internet diminishes losses while the share of Internet broadband subscriptions has no effect.

The rest of the paper is organized as follow. Section 2 describes the data. Section 3 explains the methodology used. Section 4 analyzes the empirical results. Section 5 concludes.

2. Data

We obtain the variable PINDEX, defined as the ratio of the value of unlicensed software to the total commercial value of software, from the Business Software Alliance (BSA). We consider data from 111 countries over the period 2006 to 2013, except 2012, for which the BSA provides no information. In the empirical analysis of the determinants of PINDEX we obtain data from other sources in order to build a panel for 2006-2013.

Our main source of data is the Global Competitiveness Index data platform (CGI) of the World Economic Forum (WEF). This platform collects data from several other sources, in particular the World Bank, the International Monetary Fund (IMF), UNESCO, the World Trade Organization (WTO) and the Executive Opinion Survey carried out by the WEF.

The GDP per capita at purchasing power parity (PPP current international $) is obtained from the International Monetary Fund (IMF). The variables "domestic market size" (DMS) and "foreign market size" (FMS) are constructed by the WEF. DMS represents the sum of the gross domestic product valued at purchasing power parity plus the total value of imports of goods and services, minus the total value of exports of goods and services, while FMS represents the value of exports of goods and services. Both variables are normalized on a 1-to-7 ranking scale.[9]

9 For more details see p. 51 of *The Global Competitiveness Report 2013–2014* published by the World Economic Forum. Available at http:// www. weforum. Org / issues / global-competitiveness

The Executive Opinion Survey elaborated by the WEF aims to capture the opinion of business leaders in 148 countries. From this survey we use the following variables: protection of property rights (PPRIGHT), protection of intellectual property (IPPROC), wastefulness of government spending (WGS), burden of government regulation (BGR), business cost of crime and violence (BCC), organized crime (OGC), quality of math and science education (QMS), Internet access in schools (IAS), intensity of local competition (ILC), extent of market dominance (EMM), efficiency of antimonopoly laws (EAM), availability of alternative technologies (ALT), firm-level technology absorption (FTT), foreign investment and technology transfer (FDI), capacity for innovation (CCI), quality of scientific research and institution (QSR), company spending on R&D (CSRD), university-industry collaboration (UICOL), government procurement of advanced technology products (GPTECH) and the availability of scientists and engineers (AVSCI). The following table summarizes the questions posed to the business leaders in the interview:

Table 1. CGI variables and definitions		
Name	**Question**	**Rank (1-7)**
PPRIGHT	*how strong is the protection of property rights, including financial assets?*	1 = extremely weak.
IPPROC	*how strong is the protection of intellectual property , including anti-counterfeiting measures?*	1 = extremely weak.
WGS	*how efficiently does the government spend public revenue?*	1 = extremely inefficient.
BGR	*how burdensome is it for businesses to comply with governmental administrative requirements?*	1 = extremely burdensome.
BCC	*to what extent does the incidence of crime and violence impose costs on businesses?*	1 = to a great extent.
OGC	*to what extent does organized crime (mafia-oriented racketeering, extortion) impose costs on businesses?*	1 = to a great extent.
QMS	*how would you assess the quality of math and science education in schools?*	1 = extremely poor.
QSR	*how would you assess the quality of scientific research institutions?*	1 = extremely poor.
IAS	*how widespread is Internet access in schools?*	1 = nonexistent.
ILC	*how intense is competition in the local markets?*	1 = not intense at all.
EMM	*how would you characterize corporate activity?*	1 = dominated by a few.
EAM	*to what extent does anti-monopoly policy promote competition?*	1 = does not promote competition.
ALT	*to what extent are the latest technologies available?*	1 = not available at all.
FTT		1 = not available at all.
AVSCI	*to what extent are scientists and engineers available?*	1 = not available at all
FDI	*to what extent does foreign direct investment bring new technology into your country?*	1 = not at all.
CCI	*to what extent do companies have the capacity to innovate?*	1 = not at all.
GPTECH	*to what extent do government purchasing decisions foster innovation?*	1 = not at all.
CSRD	*to what extent do companies spend on research and development?*	1 = do not spend on R&D.
UICOL	*to what extent do business and universities collaborate on research and development?*	1 = not collaborate at all.

As mentioned above, the GCI also provides data from other sources. The variables “life expectancy at birth in years” (LEY), “number of procedures required to start a business” (PSB), “number of days to start a business” (DSB) and “total tax rate” (TAX) are obtained from the *Doing Business 2013: Smarter Regulations for Small and Medium-Size Enterprises and the World Development Indicators* (April 2013 edition) provided by the World Bank. TAX measures the amount of taxes and mandatory contributions payable by a business in the second year of operation, expressed as a share of commercial profits. The total amount of taxes is the sum of five different types of taxes and contributions payable after accounting for deductions and exemptions: profit or corporate income tax, employer's social welfare contributions and labor taxes , property taxes, turnover taxes, and other small taxes.[10]

Fixed broadband Internet subscriptions per 100 head of population (FBS) and the percentage of individuals using the Internet (IUI) are obtained from the World Telecommunication/ICT Indicators 2013 (June 2013 edition) of the International Telecommunication Union. Primary education enrollment (PEED), secondary education enrollment (SEED) and higher education enrollment (TEED) are obtained from the UNESCO Institute for Statistics. Finally, the levels of imports of goods and services as a percentage of gross domestic product (MPGDP) and the levels of exports of goods and services as a percentage of gross domestic product (XPGDP) are obtained from the WTO and IMF.

For a statistical description of all variables see Table 2. Notice that we have different numbers of observations for different variables. This is because we are combining different datasets.

10 For more details about the methodology employed and the assumptions made to compute this indicator, please visit http://www.doingbusiness.org/methodologysurveys /

Variable	Description	Units	Source	Mean	SD	VIF
	Table 2: Summary Statistics					
Variable	**Description**	**Units**	**Source**	**Mean**	**SD**	**VIF**
PINDEX	Piracy Index (dependent)	%	BSA	59.23	21.12	
PPRIGHT	Protection of Property Rights	0-7 (best)	EOS	4.48	1.08	2.85
IPPROC	Protection of Intellectual Property	0-7 (best)	EOS	3.77	1.17	2.78
WGS	Wastefulness of Govt. Spending	0-7 (best)	EOS	3.36	0.90	2.04
BGR	Burden of Govt. Regulation	0-7 (best)	EOS	3.28	0.68	1.59
BCC	Business cost of crime and violence	0-7 (best)	EOS	4.60	1.14	2.68
OGC	Organized crime	0-7 (best)	EOS	5.06	1.08	3.12
LEY	Life expectancy	Years	WB	69.77	10.38	1.60
PEED	Primary Education Enrollment	%	UNESCO	90.12	10.42	1.27
SEED	Secondary Education Enrollment	%	UNESCO	78.88	27.98	1.44
TEED	Higher Education Enrollment	%	UNESCO	35.63	25.88	1.44
QMS	Quality of math&sci education	0-7 (best)	EOS	4.02	0.99	2.08
IAS	Internet Access in Schools	0-7 (best)	EOS	3.94	1.35	3.66
ILC	Intensity of local competition	0-7 (best)	EOS	4.85	0.70	1.78
EMM	Extent of market dominance	0-7 (best)	EOS	3.87	0.86	2.74
EAM	Efficiency of anti-monop. laws	0-7 (best)	EOS	4.04	0.85	2.37
PSB	Procedures to start a business	Procedures	WB	8.31	3.59	1.85
DSB	Days to start a business	Days	WB	30.96	27.80	1.78
ALT	Alternative Technologies	0-7 (best)	EOS	4.71	1.08	4.48
FTT	Firm-level technology absorption	0-7 (best)	EOS	4.79	0.77	3.08
FDI	Foreign investment & technology transfer	0-7 (best)	EOS	4.70	0.64	1.66
IUI	Individuals using Internet	%	ITU	32.03	26.89	1.89
FBS	Fixed broadband Internet/100 pop.	%	ITU	8.65	10.48	1.68
XPGDP	Exports over GDP	%	WTO & IMF	46.48	32.51	10.80
MPGDP	Imports over GDP	%	WTO & IMF	50.55	30.00	7.10
DMS	Domestic market size	0-7 (best)	GCI	3.61	1.21	4.70
FMS	Foreign market size	0-7 (best)	GCI	4.23	1.20	4.77
CCI	Capacity for innovation	0-7 (best)	EOS	3.37	0.94	3.32
QSR	Quality of scientific research & institutions	0-7 (best)	EOS	3.87	1.01	3.75
CSRD	Company spending on R&D	0-7 (best)	EOS	3.32	0.91	3.15
UICOL	University-Industry Collaboration	0-7 (best)	EOS	3.54	0.93	3.93
GPTECH	Gov. procurement of adv. technology	0-7 (best)	EOS	3.63	0.65	3.01
AVSCI	Availability of Scientists	0-7 (best)	EOS	4.19	0.81	1.71
TAX	Tax burden	%	WB	46.72	32.83	1.47
GDPCIMF	Gross Domestic Product PPP	Logs.USD	IMF	9.03	1.24	3.50

BSA: Business Software Alliance; EOS: Executive Opinion Survey of the Global Competitiveness Index (GCI) provided by the World Economic Forum (WEF); WB: The World Bank through the GCI provided by the WEF; UNESCO: UNESCO through the GCI provided by the WEF; ITU: International Telecommunication Union through the GCI provided by the WEF; WTO: World Trade Organization through the GCI provided by the WEF; IMF: International Monetary Fund.

3. Methodology

The data consist of a panel of 111 countries from the years 2006-2011 and 2013. We consider a linear specification for the dependent variable y_{it}, the piracy rate for country $i = 1,...,n$ at year $t = 1,...,T$. Assuming that $\mathbf{z}_{it}$ is a vector of potentially relevant regressors, the baseline model is given by

$$y_{it} = \alpha_i + \mathbf{z}_{it}\beta + v_{it} \tag{1}$$

where α_i is a country fixed effect and $\mathbf{v}_{\cdot t}$ is a vector error term with covariance matrix $\mathbf{V}_i$ for $t = 1,...,T$. As usual, the model may be consistently estimated using differences with respect to the cross section mean

$$\tilde{y}_{it} = \beta\, \tilde{z}_{it} + \tilde{v}_{it} \tag{2}$$

where tilde variables now represent the original variables measured in deviations from the cross-section mean. Ordinary Least Squares on the transformed model in (2) yield consistent estimators of β. For the covariance matrix, a robust consistent estimator of cross-section heteroskedasticity is obtained using standard procedures as in White (1980).

It is reasonable to presume that most of the GCI variables should be highly correlated since they are trying to proxy the same dimension: country competitiveness. Therefore it should be considered that the estimation of (1) with all regressors could lead to imprecise estimates that display considerable variance. The last column of Table 2 computes the Variance Inflation Factor (VIF references), i.e. the eigenvalues of the cross-correlation matrix of the regressors for the model in (1). Maximum values are above 10 and almost 50% of the regressors have VIFs higher than 2.5, indicating serious multiple correlation. But multiple collinearity is not the only reason for preferring a smaller model. Indeed, when there is a large number of regressors in (1) the determinants for piracy become more difficult to interpret and the researcher may prefer a sparser model with a smaller subset of regressors that exhibit the strongest effects. Prediction accuracy acctually improves if the true data generating process is sparse and it is possible to identify the set of important variables. For this reason, apart from computing (1) with all regressors, we also compute and report the results of a model selection analysis.

The relevant literature has addressed the problem of model selection and proposed several algorithms such as the Forward Selection (FS) and Backward Elimination (BE) methods (Weisberg, 1980) and All Subsets regression. They implement an automatic procedure to add or discard variables from the specification or, in the case of All Subsets, choose among a potentially large number of alternative specifications taking some predefined measures of goodness of fit such as the Bayesian Information Criterion or others as a guide.[11] Forward Selection and Backward Deletion have the advantage of being faster and simpler than All Subset methods specially when the number of potential regressors is high, but they are known to be greedy methods (Tibshirani, 1996) in that they tend to eliminate too many regressors from the specification that are relevant as explanatory variables.

Two modern approaches to the problem are the Lasso and the Incremental Forward Stagewise Regression (IFS_ε) algorithms. The geometry of both can be interpreted in terms of an algorithm called Least Angle Regression (LARS, Efron et al. 2004) which is itself a modification of the more traditional Forward Stepwise Regression. All these methods have the advantage of being computationally simpler, faster and less greedy than the aforesaid approaches.

Though these methods are initially devised for the standard linear model, Incremental Forward Stagewise Regression is particularly easy to implement in the context of a panel data model with unobserved cross-section heterogeneity since it comprises merely applying the algorithm steps to the model in differences in (2). The IFS_ε algorithm starts with an empty model where all variability around the mean is unexplained and at each iteration the coefficient vector is updated in the direction pointed to by the regressor that shows the highest correlation with the current residual. Then the unexplained residual is updated for a new iteration. The length of the step at each iteration of the IFS_ε algorithm is controlled by a parameter ε which is small enough. Thus, denote by $\tilde{y}$ the nT column vector of observations of the dependent variable in (2) and by $\tilde{z_{(k)}}$ the kth column vector of $nT\times K$ matrix $\tilde{Z}$, the algorithm begins with residual vector $\boldsymbol{u}_0 = \tilde{y}$, $\beta_0 = \mathbf{0}$. At step j the algorithm computes $k(j)$ as the index of the maximum column in $|u'_j\tilde{Z}|$ and computes

$$\begin{aligned}\beta_{j+1}&=\beta_j+\varepsilon\, sgn\left[\boldsymbol{u}_j'\tilde{\boldsymbol{Z}}\right]\boldsymbol{I}_{k_j}\\ \boldsymbol{u}_{j+1}&=\boldsymbol{u}_j-\varepsilon\, sgn\left[\boldsymbol{u}_j'\tilde{\boldsymbol{Z}}\right]\boldsymbol{I}_{j_k}\tilde{z_{(k_j)}}\end{aligned} \tag{3}$$

11 See Royston and Sauerbrei (2008) for a practical description of these methods.

where *sgn* is the sign function and $\boldsymbol{I}_k$ is the index column vector of size K of zeros with a one in the kth position. A step length as big as $\varepsilon=|\boldsymbol{u}(j)'\tilde{\boldsymbol{z}}_{(k_j)}|$ leads to the standard Forward Stepwise Regression, which is known to produce models which are too small. The trick is therefore to choose a smaller ε in order to take smaller steps. Freund, Grigas and Mazumder (2013) show that a dynamic choice of the shrinkage parameter such as

$$\delta_j=\frac{|\boldsymbol{u}_j{}'\tilde{\boldsymbol{z}}_{(k_j)}|}{\|\tilde{\boldsymbol{z}}_{(k_j)}\|^2} \tag{4}$$

leaves the correlation with residuals at step j bounded by a $O(j^{-1/2})$ term.

The algorithm runs until the stop criterion is met. By virtue of (4), the correlation of regressors and residuals decreases at each iteration and the algorithm is stopped when such a correlation lies below a small number chosen as $1\text{x}10^{-4}$ here. The Bayesian Information Criterion (BIC) is calculated each time a new variable enters the specification and the preferred model is chosen as that which maximizes the current BIC.

4. Empirical Analysis

The first columns of Table 3 show the results of an estimation of a Fixed Effects complete specification of (1) including all regressors in Table 2. To improve stability, variables were previously scaled by their standard deviation after cross-section mean subtraction as required for fixed effects estimation. The regressor coefficients reported in Table 3 are presented rescaled to the original units. The model shows a good fit with the adjusted R^2 coefficient of 0.47 with good explanatory power for differences across countries (0.51) and within differences (0.50).

Table 3: Fixed Effects Panel Model

	FE All Vars.		IFS$_\varepsilon$ Regression				
Variable	**Coef.**	**T-stat.**	**Iter.**	**C.res.**	**BIC**	**Coef.**	**T-stat.**
IUI	-0.0124	-1.54	1	0.0075	-8.4199	-0.0114	-1.53
FBS	-0.0746	-2.97	7	0.0069	-8.4495	-0.0918	-4.20
AVSCI	0.9470	2.85	37	0.0044	-8.6138	1.0559	3.60
LEY	-0.1514	-2.5	40	0.0042	-8.6177	-0.1554	-2.89
IAS	0.3384	1.38	44	0.0040	-8.6259	-0.1196	-0.59
PSB	0.0325	0.62	50	0.0037	-8.6426	0.0451	1.02
GDPCIMF	-1.0408	-3.48	69	0.0029	-8.7096	-1.0567	-3.56
TEED	-0.0313	-2.32	71	0.0028	-8.7062	-0.0386	-3.03
DMS	-0.4135	-0.93	94	0.0019	-8.7682	0.0290	0.08
TAX	0.0378	3.01	99	0.0018	-8.7703	0.0466	3.92
BGR	-0.5087	-1.73	108	0.0015	-8.7801	-0.8143	-3.34
QMS	0.5174	1.61	120	0.0012	-8.7923	0.3339	1.09
FTT	1.3359	3.28	127	0.0010	-8.7927	0.6105	1.91
UICOL	-0.6079	-1.53	133	0.0010	-8.7906		
BCC	-0.4204	-1.49	134	0.0009	-8.7812		
EMM	0.2242	0.76	150	0.0008	-8.7900		
SEED	0.0420	2.77	152	0.0007	-8.7814		
ALT	-0.4998	-1.76	154	0.0007	-8.7728		
IPPROC	-0.3765	-1.04	171	0.0006	-8.7778		
CSRD	0.4051	0.98	178	0.0005	-8.7726		
WGS	-0.2754	-1.06	184	0.0005	-8.7662		
GPTECH	-0.2441	-0.79	195	0.0004	-8.7626		
XPGDP	-0.0094	-0.71	197	0.0004	-8.7530		
PPRIGHT	0.0949	0.34					
OGC	0.3100	1.06					
PEED	-0.0065	-0.28					
EAM	-0.3492	-0.98					
DSB	0.0072	0.99					
FDI	-0.2055	-0.65					
MPGDP	0.0076	0.65					
FMS	-0.1139	-0.32					
CCI	0.1480	0.45					
QSR	-0.0005	0					
ILC	0.1406	0.45					

The number of observations is 671 in the FE All Vars model and 679 in the IFS$_\varepsilon$ selected model. The difference is due to missing observations in some variables. The adjusted R^2 are 0.47 and 0.64 respectively. The Between and Within R^2 coefficients are 0.51 and 0.50 in the FE All Vars model and 0.66 and 0.42 in the IFS$_\varepsilon$ selected model.

Apart from FE estimation of (2), the last part of Table 3 displays the results of model selection after IFS_ε estimation of the model. Because variables enter the specification sequentially in IFS_ε the regressors in the first column are shown in the order given by the iteration numbers in column 4. Remember that the algorithm is run until the correlation of regressors and current residuals is small enough, which in our case means that the algorithm was stopped when this correlation was below $4x10^{-4}$ at iteration 197. At that point, 23 out of 34 variables had been included in the specification. From these, we now choose the point at which the model showed the best BIC at -8.7927 with 13 variables selected. The adjusted R^2 coefficient of the selected model is 17 points greater than the all-variables model and the "between" R^2 is 15 points better. The "within" adjustment is however 8 points smaller.

As expected, we find that income per capita is negatively and significantly correlated with the software piracy rate. Note that the coefficient is similar in the complete and selected models, meaning an almost unitary negative elasticity between GDP and piracy. The coefficient is significant and much greater than any other, suggesting that development is the fastest way to reduce piracy. This is confirmed by the coefficient for life expectancy, which is also negative. This means that every extra year of life expectancy reduces piracy by 0.15 points.

Bezmen and Depken II (2006), find that the tax burden and the piracy rate are negatively correlated. They suggest that a higher tax burden might reflect that the government has more resources to enforce property rights, and thus the piracy rate decreases. Moreover governments in developing countries find it more difficult to raise taxes and to devote resources to useful spending. Also, many rich countries have democratic institutions specifically designed to control corruption and government spending and are much more efficient at raising tax revenue from taxpayers. Thus, these variables could be reflecting improvements in the rule of law and government control that may simultaneously reduce impunity and piracy rates. However, we find a positive relationship between tax burden and piracy rates. We interpret our results as reflecting two opposing effects: on the one hand, it is reasonable to think that a higher tax burden means more efficiency and more resources for property rights enforcement, but the PPRIGHT variable which specifically controls for protection of property rights is nevertheless non-significant in all specifications. Therefore, we suspect that the induced effect of the tax burden on piracy rates via property rights enforcement may be weak. On the other hand, higher sales taxes increase the price of legally sold software, sothe incentive to copy or distribute software via shadow markets increases.[12] This role of price in the software market may also explain the negative signs in the "domestic market size" and the "exports over GDP" variables in our specification. The intuition is that larger, more open markets exhibit tougher competition which results in lower average mark-ups (Melitz and Ottaviano, 2008), and in lower prices of the original products so that the incentives to copy original software decrease. However, the effect of the extent of market dominance is not due to the relationship between the price of the original products and the incentives to copy them. In this case, the results suggest that a more monopolistic market lets firms make higher profits, so that they have more resources with which to fight piracy.[13]

Education is an important factor in explaining piracy. As in Depken II and Simmons (2004) and Shadlen et al. (2005) we find a strong correlation between education and piracy but the evidence shows two opposite effects of education as suggested by Scalise (1997). In particular, the quality of math and science education and secondary schooling are positively correlated with the piracy rate, which suggests that they provide the knowledge resources to facilitate the illegal copying of original software, but higher education enrollment rates are negatively correlated with piracy, which suggests that higher education is associated with the demand for intellectual protection.

It is also possible to discriminate between different effects of R&D on piracy, since we use several variables to measure its impact. On the one hand, the availability of scientists and engineers and firm-level technology absorption are strongly and positively correlated with the piracy rate. Thus, these R&D factors could facilitate illegal copying. On the other hand, we find that the availability of the latest technology and university-industry collaboration in R&D are negatively correlated with software piracy, but their effect is weak and irrelevant as shown in the selected model in Table 3. This is also the case for the protection of intellectual property as perceived by respondents, the coefficient for which is negative but non-significant in all specifications. Indeed, tougher legislation on the protection of intellectual property may be due to higher piracy rates, which reduces the significance of this coefficient.

12 Papadopoulos (2003, 2004) empirically shows the negative correlation between the piracy rate and the price of the original products, and Martínez-Sánchez (2010) provides a theoretical explanation of this correlation.

13 Notice that a lower extent of market dominance (EMM) implies that the market is more monopolistic.

As in earlier literature, we show that the strength of institutions is an important factor in reducing piracy. In particular, those countries with governments that require a high number of procedures to start up a business and have burdensome administrative requirements are associated with higher software piracy rates.[14] Those countries where the incidence of crime and violence impose higher costs on businesses also display higher piracy rates. By contrast, we find that those governments that foster innovation and are more efficient in providing goods and services are also more successful in reducing piracy. These results suggest that states with higher-quality bureaucracies are likely to protect intellectual property more effectively, as indicated in Shadlen et al. (2005).

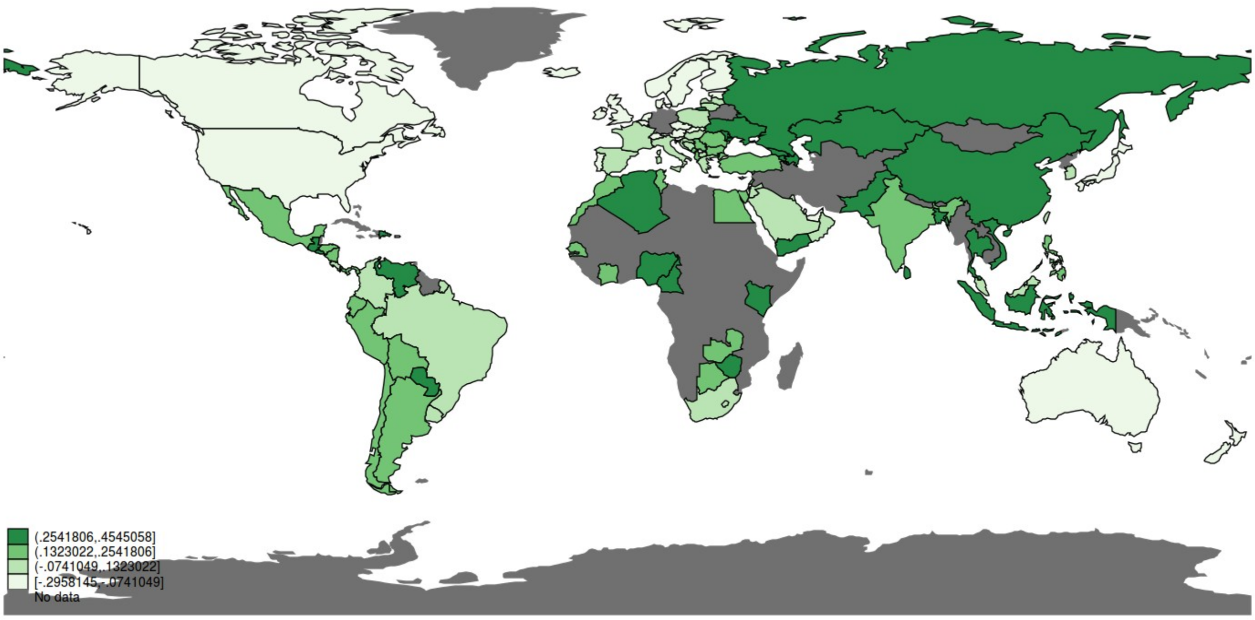
We measure the availability of the Internet in a country using three variables: Internet access in schools, the percentage of individuals using the Internet and the number of fixed broadband Internet subscriptions per 100 head of population. Our results show that the availability of the Internet is negatively correlated with the software piracy rate, especially for the FBS variable in the selected model, which shows that a 1% increase in the fixed broadband penetration rate in the population decreases the piracy rate by 0.07 percentage points. Our results coincide with Goel and Nelson (2012). These authors suggest that greater Internet availability makes tracking piracy easier. This result is consonant with recent empirical results that find that free music streaming (where the consumer does not possess the music but only has access to it) has no significant effect on CD sales but positively affects live music attendance for national and foreign performers (Nguyen, Dejean and Moreau (2014)). Moreover, it is also consistent with those empirical papers that find that file sharing has no effect on sales in the music industry (Oberholzer-Gee and Strumpf (2007) and McKenzie (2009)).[15] However, the explanation in the case of software piracy may differ substantially. As already pointed out, Internet access is in many senses the gateway to freeware or open source games and applications. The incentives for piracy are expected to decrease considerably if a substantial proportion of the things that could only be done via proprietary software before broadband access can now be obtained at lower cost online. Of course, proprietary software still provides functionalities and extensions that are rarely obtained under other types of licensing but this is in some cases a software firm's reply to Internet increasing competition in the lowest segments of demand.

14 Notice that a high value of variable BGR means that complying with governmental administrative requirements is less burdensome for businesses.

15 Another effect of broadband connectedness is that it increased the closedown rate of brick and mortar music stores and reduced their number (Zentner (2008)). Moreover, it is also found that the presence of a university leads to a reduction in the number of music specialty stores in the local area (Zentner (2008)).

The selected model explains more than 65% of the differences across countries but some of those differences remain as black-box residual country effects. Figure 1 plots the levels of α_i for the countries analyzed on a world map. It is apparent at a glance that there is a pattern of regional clustering with levels which are very highin regions such as Asia and Africa, high in South America and lower in Western Countries. (La Porta et al., 2008). These country effects may respond to many control factors which are country specific and which have not been considered previously. In their paper, Goel and Nelson (2009) find that legal tradition and ethnic, linguistic and religious diversity among other institutional country specific factors may have effects on piracy rates.

Figure 1: Fixed country effects.



We look for potential explanatory factors and build a model for the country effects α_i. We consider the three classic legal systems, i.e. Civil Law (also known as Continental European Law), Common Law (in the Anglo-Saxon tradition), and Religious Law (in particular Sharia Law). Countries were then classified according to the source of their legal system into seven different types:[16] Civil Law, Common Law, Civil/Sharia Law, Civil/Common Law, Common/Religious Law, Religious/Sharia Law and Mixed/Unclassified. For country fractionalization we take the Alesina and La Ferrara (2005) religious, linguistic and ethnic diversification indexes. In a recent paper, Ayyagari, Demirguç-Kunt and Maksimovic (2012) use cross-country data to evaluate the effect of

16 The classification was taken from the CIA World Factbook. (https://www.cia.gov/library/publications/the-world-factbook/fields/2100.html)

competing hypotheses such as legal origin, endowments, religion andr ethnic diversification on the determinants of the protection of property rights. Theyfind that ethnic diversification is robust compared to all other factors. Moreover, it is also explained that most of these diversification indexes are highly correlated. For this reason, we include ethnic diversification alone as an indicator of cultural diversity or country fractionalization. The results for this model are shown in Table 4. There is a significantly lower piracy rate in countries with legal systems based on pure Common Law than in any others. The model does not detect significant differences between any legal systems other than Common Law, and the effect of ethnic fractionalization on the unexplained country effects for piracy is significant and positive. This latter result does not coincide with the findings of Goel and Nelson (2009), who find a negative relationship between ethnic diversification and piracy, but they are in line with the findings of Ayyagari, Demirguç-Kunt and Maksimovic (2012), whose results point in the direction of a positive correlation between these two variables. The intuition is as follows: since the relationship between ethnic diversification and property rights protection is negative (Ayyagari, Demirguç-Kunt and Maksimovic, 2012), those countries with higher ethnic diversification protect property rights less and therefore have higher piracy rates.

Table 4. Fixed Country Effects

Number of Observations	105	R-squared		0.1761
F-statistic (7,97)	2.96	Adjusted		0.1166
P-value	0.0074	Root MSE		0.1759
Legal System	*Coefficient*	*Std. Err.*	*T-Stat.*	*P-value*
Civil Law	-0.0514	0.0344	-1.49	0.1380
Common Law	-0.2266	0.0717	-3.16	0.0020
Civil and Sharia	-0.0602	0.4302	-0.14	0.8900
Civil and Common	-0.0331	0.1182	-0.28	0.7820
Common and Religious	0.0059	0.0078	0.75	0.4560
Religious and Sharia	0.0853	0.0797	1.07	0.2870
Mixed/Unclass.	-0.0817	0.2817	-0.29	0.7720
Ethnic Diversity	0.2038	0.0763	-2.67	0.0092

5. Conclusions

In this paper we analyze the differences in piracy rates between countries by collecting data for 111 countries over the period 2006 to 2013, except for the year 2012, from the BSA, CGI data platform provided by the WEF and the IMF. The approach that we use is Incremental Forward Stagewise Regression (IFS_ε).

Like previous papers, we find that as a country develops so the incentives for pirating software decrease. This result is deduced from the negative correlation between the piracy rate and GDP per capita and life expectancy.

Unlike the result obtained by Bezmen and Depken II (2006), we obtain that the piracy rate is positively correlated with the tax burden but negatively correlated with domestic market size and exports over GDP. The intuition behind these results is based on the role of the price of the original software. Lower tax burdens decrease the price of the original products, as do larger markets and more exporter countries, so the price gap between the original and pirated products is lower. Thus, the incentives to copy original software decrease. On the other hand, the effect of the extent of market dominance cannot be attributed to the relationship between the price of the original products and the incentives to copy them. In this case, the results suggest that a more monopolistic market allows firms to make higher profits, which gives them more resources for fighting against piracy.

The factors behind the two opposite effects of education suggested by Scalise (1997) can now be explained. In particular, the quality of math and science education and secondary education are positively correlated with the piracy rate, which suggests that they facilitate the illegal copying of original software. But higher education has a negative relationship, which suggests that it is associated with the demand for intellectual protection.

We find that R&D has two opposite effects on software piracy. On the one hand, the availability of scientists and engineers and firm-level technology absorption could facilitate illegal copying. On the other hand, higher availability of the latest technology, more private spending on R&D and better university-industry collaboration is associated with a higher intellectual property protection by national governments, which obstructs piracy. Moreover, we also find that higher intellectual property protection leads to a lower piracy rate, although this correlation is weak.

Ours results suggest that those countries with higher-quality bureaucracies are likely to protect intellectual property more effectively, as in Shadlen et al. (2005).

Finally, our results show that the availability of Internet access is negatively correlated with the software piracy rate, as found by Goel and Nelson (2012), who suggest that greater Internet access makes tracking piracy easier. This result is consistent with recent empirical results which find that free music streaming has no significant effect on CD sales but positively affects live music attendance for national and foreign artists (Nguyen, Dejean and Moreau (2014)).

References

Alesina, A. and E. La Ferrara. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43 (3), 762–800.

Andrés, A. R., (2006a). Software piracy and income inequality. *Applied Economics Letters*, **13** (2), 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.

Andrés, A. R., and K. R. Goel. (2012). Does software piracy affect economic growth? Evidence across countries. *Journal of Policy Modeling*, 34(2), 284–295.

Andrés, A. R., and S. A. Asongu. (2013). Fighting Software Piracy: Which Governance Tools Matter in Africa?. *Journal of Business Ethics*, 118 (3), 667-682.

Ayyagari M., Demirgüç-Kunt, A. and V. Maksimovic. (2013). What Determines Protection of Property Rights? An Analysis of Direct and Indirect Effects. *Journal of Financial Econometrics*, **11** (4), 610-649.

Bae, S.-H. and J. P. Choi. (2006). A Model of Piracy. *Information Economics and Policy*, 18 (3), 303-320.

Banerjee, D. S. (2003). Software Piracy: a Strategic Analysis and Piracy Instruments. *International Journal of Industrial Organization*, 21 (1), 97-127.

Belleflamme, P. and M. Peitz. (2012). Digital piracy: In M. Peitz and J. Waldfogel (Eds.), *Oxford Handbook of the Digital Economy*, Oxford University Press.

Belleflamme, P. and M. Peitz. (2014). Digital piracy: An Update, *CORE Discussion Paper*.

Bezmen, T. L. and C. A. Depken II. (2006) Influences on software piracy: Evidence from the various United States. *Economics Letters*, 90 (3), 356-361.

Business Software Alliance (BSA). (2014). The Compliance Gap: BSA Global Software Survey. http://globalstudy.bsa.org/2013/index.html

Dejean, S. (2009). What can we learn from empirical studies about piracy?. *CESifo Economic Studies*, 55(2), 326–352.

Depken II, C. A. and L. C. Simmons. (2004). Social Construct and the Propensity for Software Piracy. *Applied Economics Letters,* 11 (2), 97–100.

Efron, B., T. Hastie, I. Johnstone and R. Tibshirani. (2004). Least angle regression. *The Annals of Statistics*, 32 (2), 407–499.

Freund, R., Grigas, P., and Mazumder, R. (2013). Incremental Forward Stagewise Regression: Computational Complexity and Connection to LASSO. ROKS 2013, available at [http: // www.esat. kuleuven. be / sista / ROKS2013 / program.php](http://www.esat.kuleuven.be/sista/ROKS2013/program.php).

Goel, R.K., and M.A. Nelson. (2009). Determinants of Software Piracy: Economics, Institutions and Technology. *The Journal of Technology Transfer*, 34(6), 637-658.

Goel, R. K. and M. A. Nelson. (2012). Shadow economy and international software piracy. *Applied Financial Economics*, 22 (23), 1951-1959.

Gomes, N. D., P. A. Cerqueira and L. Alçada-Almeida. (2013). Determinants of worldwide software piracy losses. Working Papers Estudos do GEMF 2013-19.

Husted, B. W. (2000). The impact of national culture on software piracy. *Journal of Business Ethics*, 26 (3), 197-211.

La Porta, R., F. Lopez-de-Silanes and A. Shleifer. (2008). The Economic Consequences of Legal Origins. *Journal of Economic Literature*, 46 (2), 285-332.

López-Cuñat, J. M. and F. Martínez-Sánchez. (2015). Anti-piracy policy and quality differential in markets for information goods. *European Journal of Law and Economics*, forthcoming.

Marron, D.B. and D.G. Steel. (2000). Which Countries Protect Intellectual Property? The Case of Software Piracy. *Economic Inquiry*, 38 (2), 159-174.

Martínez-Sánchez, F. (2010). Avoiding Commercial Piracy. *Information Economics and Policy*, 22 (4), 398-408.

Martínez-Sánchez, F. (2011). Collusion, Competition and piracy. *Applied Economics Letters*, 18 (11), 1043-1047.

Martínez-Sánchez, F. (2012). Collusion in Software Markets. *Economics Bulletin*, 32 (2), 1345-1352.

Martínez-Sánchez, F. (2013). Time to Market and Impatient Customers. *Bulletin of Economic Research*, 65 (2), 194-202.

Martínez-Sánchez, F. (2014). Learning by Copying a Piratable Good. *Theoretical Economics Letters*, **4**, 787-794.

McKenzie, J. (2009). Illegal Music Downloading and its Impact on Legitimate Sales: Australian Empirical Evidence. *Australian Economic Papers*, 48(4), 296-307.

Melitz, M.J. and G.I.P. Ottaviano. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75 (1). 295-316.

Nguyen, G.D., S. Dejean and F. Moreau. (2014). On the complementarity between online and offline music consumption: the case of free streaming. *Journal of Cultural Economics*, 38 (4), 315-330.

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

Papadopoulos, T. (2003). Determinants of International Sound Recording Piracy. *Economics Bulletin*, 10 (6), 1-9.

Papadopoulos, T. (2004). Pricing and pirate product market formation. *Journal of Product and Brand Management*, 13(1), 56–63.

Peitz, M. and P. Waelbroeck. (2004). The Effect of Internet Piracy on CD Sales – Cross Section Evidence. *Review of Economic Research on Copyright*, (1), 71-79.

Peitz, M. and P. Waelbroeck. (2006). Piracy of Digital Products: A Critical Review of the Theoretical Literature. *Information Economics and Policy*, 18 (4), 449-476.

Royston, P. and W. Sauerbrei. (2008). *Multivariable Model-building.* Wiley Series in Probability and Statistics.

Scalise, C. T. (1997). Natural Intellectual Property Protection Reform. (Doctoral dissertation). University of Chicago, Graduate School of Business.

Shadlen, K. C., A. Schrank and M. J. Kurtz. (2005). The Political Economy of Intellectual Property Protection: The Case of Software. *International Studies Quarterly*, 49 (1), 45–71.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. **58** (1), 267-288.

The Global Competitiveness Report 2013–2014, World Economic Forum. Available at http:// www.weforum.org / issues / global-competitiveness.

Waldfogel, J. (2012). Digital Piracy: Empirics. In M. Peitz and J. Waldfogel (Eds.), *Oxford Handbook of the Digital Economy*, Oxford University Press.

Weisberg, S. (1980). *Applied Linear Regression*. Wiley, New York.

White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.

Zentner, A. (2008). Online sales, Internet use, file sharing, and the decline of retail music specialty stores. *Information Economics and Policy*, 20 (3), 288-300.