Universität St.Gallen

Is Software Piracy a Middle Class Crime? Investigating the Inequality-Piracy Channel

Justina A.V. Fischer, Antonio Rodríguez Andrés

August 2005 Discussion paper no. 2005-18

Department of Economics University of St. Gallen

Editor:	Prof. Jörg Baumberger University of St. Gallen Department of Economics Bodanstr. 1 CH-9000 St. Gallen Phone ++41 71 224 22 41 Fax ++41 71 224 28 85 Email joerg.baumberger@unisg.ch
Publisher:	Department of Economics University of St. Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 22 98
Electronic Publication:	http://www.vwa.unisg.ch

Is Software Piracy a Middle Class Crime?

Investigating the inequality-piracy channel *

JUSTINA A. V. FISCHER
University of St. Gallen, SIAW-HSG

ANTONIO RODRÍGUEZ ANDRÉS
University of Southern Denmark, Department of Economics

Corresponding author: Justina A. V. Fischer
Swiss Institute for International Economics and Applied
Econonomic Research (SIAW)
University of St. Gallen
Bodanstrasse 8, CH-9000 St. Gallen, Switzerland
Justina.Fischer@unisg.ch

* We thank Hakan Holm, Markus Froelich, F.M. Scherer and participants at the II Workshop on the Law and Economics of Intellectual Property and Information Technology (LIUC, Italy) for comments and suggestions. All errors are our own.

Abstract

This paper uses a sample of 71 countries in a cross-country context to empirically analyze the relationship between income distribution and software piracy rates. It measures income inequality by the Gini coefficient and alternatively by quintile shares. This analysis remedies previous econometric studies by controlling for a wide range of factors that potentially influence national piracy rates and employing an instrumental variables approach. Results indicate that income inequality is negatively associated with piracy rates but also that the impact of various income classes on piracy rates may depend on the geographic region where a country is located. Moreover, the model predicts an inverted U-shaped relationship between piracy and per capita income and reveals an apparent inverse relationship between individualism and software piracy. In addition, the results seem robust to the inclusion of additional covariants often employed in predicting piracy rates and the occurrence of property crime.

Keywords

Intellectual property rights, software piracy, income inequality, economics of crime

JEL Classification

K42, K11, D3

1. Introduction

Economically, the piracy of intellectual property is equivalent to the theft of a tangible good. Piracy can take several forms depending on the type of intellectual or information good[1] and the technological access to it (Watt, 2001). One of the most troubling areas today is certainly the piracy of computer programs. According to the International Planning and Research Corporation (IPRC, 2003), worldwide, in 2002 pirated software constituted 39 percent of all PC business software applications, which translates into a $13.1 billion loss in revenue dollars.[2] Indeed, the emergence of new technologies that enhance replication and diffusion of illegal copies (e.g., CD and DVD burners and mechanisms for digital transfer over the Internet) poses a new threat for business software publishers because these technologies provide the opportunity for intellectual property law violation on a wider scale than ever before (Gallegos, 1999; Gopal and Sanders, 2000; Moorehouse 2001). Therefore, a better understanding of the phenomenon of piracy may be crucial to curbing the software piracy problem.

Most extant economic studies on the causes of piracy focus primarily on the role of factors such as economic development, legal framework, institutions, culture, and intellectual property rights protection, thus giving little attention to the role of income inequality (Marron and Steel, 2000; Husted, 2000; Depken and Simmons, 2004; Shadlen et al., 2005; Rodríguez, 2005b; van Kranenburg and Hogenbirk, 2005). In general, the findings of these papers support the notion that economic and legal factors significantly influence piracy rates. Nonetheless, the formal literature on the impact of income inequality on piracy is scant because only a few studies deal with the effects of income distribution on software piracy (Husted, 2000; Rodríguez, 2005a). Moreover, not only do these studies vary significantly in terms of the number of countries included, the period of analysis, the measure of income distribution, and the set of additional explanatory variables included in the models, but they rely on cross-sectional estimation methods. Indeed, those few studies that do analyze income inequality as an explanatory variable for piracy test its significance using ordinary least squares (OLS) regressions. Among these, Husted (2000) shows that income inequality, measured by the top 10 percent, does exert a negative effect on piracy after variables such as economic development and cultural characteristics are controlled for. In a more recent cross-country analysis, Rodríguez (2005a) finds that income inequality as measured by the Gini index lowers piracy rates. However, both papers omit a number of relevant variables that, if correlated with other explanatory variables, would render the parameter estimates biased and inconsistent. Further, in any study of the relationship between inequality and software piracy, it is unclear which quantile(s) of the income distribution matter(s). In addition, the mechanisms that Husted (2000) proposes as possible avenues through which inequality may impact software piracy rates appear tentative and need more elaborate explanation. Finally, neither Husted (2000) nor Rodríguez (2005a) employs instrumental variable techniques to overcome measurement errors in the income inequality variables.

The contribution of this present assessment of the impact of income distribution on the occurrence of piracy is fourfold. First, theoretically, the paper tries to shed light on the possible arguments of why inequality may influence piracy. Second, in the empirical analysis, using a cross section of international data at country level it attempts to evaluate which part of the income distribution matters most in determining software piracy rates by considering several measures of income inequality. Third, methodologically, the paper explicitly addresses the possibility of measurement error of income inequality by using an instrumental variables estimation approach. Finally, the empirical analysis also tests the importance of other socioeconomic factors suggested by the economics of crime literature that have been neglected in previous empirical studies of piracy.

For the full sample of countries, we find significant effects of several inequality measures on piracy. In fact, the two stage least squares (2SLS) estimates for the coefficients of inequality measures are larger than the OLS estimates, which might suggest that attenuation bias due to measurement error is present in previous studies. We find evidence that inequality impacts piracy negatively, with the lowest and the middle income classes exerting a positive effect and the highest income class exerting a negative one. Moreover, the magnitude of the effect of income inequality measures is greater than that for the effect of per capita income. Our findings also corroborate recent empirical research showing that higher per capita income is associated with lower levels of piracy. In addition, we detect an inverted U-shaped relationship between piracy and national income. We also find that higher levels of individualism are associated with lower levels of piracy. Moreover, splitting up the sample by geographical region, our data reveals different directions of impact for the various income inequality measures.

In the next Section, we present the theoretical arguments why inequality may impact piracy rates. In Sections 3 and 4, we describe the econometric model and the data used in the empirical analysis. Empirical findings and robustness tests are then presented in Section 5. We conclude with some policy implications.

2. Inequality and piracy: theoretical context and hypotheses

Even though several empirical studies have attempted to assess the effect of income inequality on software piracy rates across countries, in general, the economic and societal mechanisms through which inequality may impact piracy rates are not well understood. In fact, an economic rationale for this relationship can be derived from previous studies. Based on this earlier research, it is assumed that income inequality has an inverse relationship with piracy rates. One argument is that the more income equality, or the larger the middle class, the greater the share of illegal copying (Husted, 2000). However, this assumption raises some obvious questions, including why the middle class is expected to be more prone to piracy than other income classes. Indeed, even if the size of middle class is a correct and justifiable predictor, the question remains how the middle class should be defined.[3] Answering all these questions is beyond the scope of this paper, but we will nonetheless provide a thorough discussion by developing arguments on why inequality may influence piracy rates. However, as there is no a clear theoretical framework within which to interpret the effect of income inequality on piracy, we will draw on other strands of economic literature such as that on criminal behavior and personal trust.

According to the economic model of crime (Becker, 1968), rational individuals choose whether or not to engage in illegal activities by comparing the expected returns from these activities to the potential returns from working legally. Thus, the expected benefit from illegal activities could be proxied by mean income as a measure of society's wealth. The cost of committing a crime, however, increases with the potential legal income of the criminal as opportunity cost. Specifically, in societies with high income inequality, the gap between societal mean income and individual income may be large, making those on the lower rungs of the income distribution more prone to commit crime. Findings in prior empirical literature are consistent with this argument (Ehrlich, 1973; Gould et al., 2002; Machin and Meghir, 2000).[4] From a sociological viewpoint, an increase in inequality could lead to what is called an "envy effect" of the less affluent, which reduces some individuals' moral thresholds and exercises a positive impact on crime (Fajnzylber et al., 2002). Based on these arguments, we should see a direct positive relationship between inequality and piracy.

A similar argument is that equal income distributions in many countries are the result of redistribution policies that represent fairness opinions among the citizens. These policies are based partially on the principle of the state's right to claim part of citizens' earnings and property rights. Thus, the decision to pirate may be perceived as a redistribution measure from the rich to the poor through violation of the intellectual property rights law by the latter. It is plausible that in countries whose citizens embrace the principles of redistribution, the population is more prone to find the pirating of legal software for redistributive purposes morally acceptable.

In addition, illegal copying may be based on interpersonal trust. First, those who either copy from or share software with others must trust that the software contains no viruses. Second, individuals giving illegal copies to others must trust these persons not to report them to the police. Zak and Knack (2001) argue that equality in personal income generates personal trust. Thus, in more equal societies, we might expect higher levels of trust that lead to more software piracy.

Predictions can also be made not only with respect to the overall income distribution, but also with respect to the various income classes. In this paper, we distinguish a lower, a middle and an upper income class. In general, we assume the lowest income earners not to be an important determinant of piracy because, in the time period under investigation, poor people may not have been able to afford a computer. Therefore, the lowest income class influences neither the demand nor the supply side of this crime. Rather, following Husted (2000), we would expect the middle income households to positively influence piracy rates because these people can afford a PC but not necessarily the legal acquisition of all the software they want and need; thus, this income group constitutes the demand side and (potentially) the supply side of piracy. The highest income class is expected to have a piracy dampening impact because potential opportunity costs, particularly reputational costs, exceed the gain from having saved a few hundred dollars by acquiring an illegal copy.[5] Based on both the previous arguments and these hypotheses on the sizes of the various income classes, we would expect a higher Gini coefficient to have a crime dampening influence; that is, as the share of highest income earners in total income rises beyond its desired income share of equal income distribution[6], so economic inequality within society rises; while, in general, as the share of the middle class increases, income inequality is lowered.[7]

3. Model and estimation method

3.1. Dependent variable

To study the effect of inequality on piracy rates, we specify a regression equation using estimated piracy rate as the dependent variable (*PR*). Piracy rate is defined as the estimated difference between software programs installed and software applications legally licensed.

3.2. Explanatory variables

As regards the independent variables, as hypothesized in Section 2, income inequality constitutes the main variable of interest in our understanding of why national piracy rates differ across countries, unlike most of the extant studies which have not even considered this determinant. In contrast to the few previous contributions which include income inequality, we employ and test several measures of income distribution, particularly three measures of income shares, distinguishing a low, middle and upper income class, and a measure of the overall income inequality in society.

Because illegal copying may respond to differences in national income, the natural log of per capita GNP for 1970-1995 is also included as a control variable (LGNP) to capture the level of economic development in any particular country. The rationale for its inclusion is that it serves as a proxy for the strictness of protection of intellectual property rights (IPR). Ginarte and Park (1997), Maskus and Penubarti (1995), and Rapp and Rozek (1990) find that countries with high income have stronger patent protection. Further, IPR in less developed countries tend to be "shorter, less comprehensive, and much less vigorously enforced" (Richardson and Gaisford, 1996, p. S376). In their recent article, Shadlen et al. (2005) note a relationship between national income and intellectual property protection that may be thought of in terms of the demand and supply of intellectual property protection. On the demand side, as nations becomes wealthier, local software producers devote more resources to innovative activities and are more likely to demand that national governments increase IPR protection. On the supply side, the degree of software protection may also depend on financial capacity: countries with a higher level of GNP per capita are able to provide stronger protection simply because they can afford to (Ostergard, 2000; Varian, 1998). IPR protection involves large fixed costs – for example, judicial courts and policing – therefore, richer countries are more likely than poorer nations to provide it (Kanwar, 2002; Varian, 1998). In turn, good policies themselves might foster growth of per capita income (Acemoglu et al., 2004). For this reason, the inclusion of income in the regressions could control for other potential determinants of piracy such as government quality.

For all these reasons, we expect a negative correlation between GNP and software piracy. Indeed, previous empirical research (e.g., Gopal and Sanders, 1998, 2000; Holm, 2003; Shin et al., 2004; Husted, 2000; Marron and Steel, 2000; Ronkainen and Guerrero-Cusumano, 2001; Rodríguez, 2005a) corroborates the negative impact of national income on piracy.

However, we take into account that the relationship between the rate of piracy and GNP per capita may well be nonlinear, with piracy first increasing then later declining as the level of per capita income increases. We conjecture that demand for property protection may be lower than the demand for pirated software when the country is in economic transition; in this case, increasing wealth in society might lead to a growing number of prosperous households using pirated software while the government sector is still insufficiently affluent or unwilling to provide IPR. A further rise in GNP might then be accompanied by the development of a domestic software industry or growing political pressure of trading partners leading to better protection of intellectual property, as explained before. To shed some light on this issue, the square term of the log of the real GNP per capita ($LGNP^2$) is included in the regression model based on the expectation that the coefficient on income term will be positive, while its quadratic term will be negative.[8]

Cultural norms may equally influence the formation of individual behavior, which might in turn impact the occurrence of software piracy. Individualism and collectivism are two of the most important cultural dimensions emerging from many cultural studies. According to Hofstede (1997) in collectivistic societies, group pressure is exerted on individuals to share their resources with group members.[9] As Gopal and Sanders (1997) point out, piracy is a group activity: pirated software is distributed among friends, coworkers, and family members. Consequently, we hypothesize that more individualistic societies will tend to experience less software piracy. Indeed, research evidence on piracy does show a negatively significant effect of individualism on piracy (e.g., Husted, 2000; Marron and Steel, 2000; Shin et al., 2004). Therefore, our empirical analysis uses a measure of societal of individualism (*IDV*) to test whether such cultural values do matter in an explanation of cross-national variations in piracy rates.

Whereas these three variables of GNP, income inequality, and individualism constitute the baseline economic model of piracy, two other key explanatory institutional determinants are introduced to check its robustness: (a) the degree of protection of intellectual property rights and (b) institutional quality. Since legal tendencies towards or against protection of IPR may have an impact on piracy in that they affect its opportunity or execution costs, a measure of IPR protection (*PRIGHT*) is included as a representation of the strength of different national IPR regimes. We expect that patent protection will have a crime lowering effect on piracy. Moreover, because the literature on piracy also emphasizes the role of institutional quality on this relationship, we include an according measure, the rule of law (*LAW*), in the regression model. The underlying rationale for its inclusion is that efficient law enforcement may raise the probability or severity of punishment, leading in turn to a decline in the expected economic gains from any illegal activity (for further arguments, see Holm, 2003). In fact, two recent articles (Holm, 2003; Rodríguez, 2005a) do note a strong negative relationship between the rule of law and the share of illegal software.

3.3. Model specification and methodology

Integration of the all above ideas yields the following reduced form model

$$LPR_i = \beta_0 + \beta_1 * LGNP_i + \beta_2 * LGNP_i^2 + \beta_3 * IDV_i + \beta_4 * INEQ_i + \beta_5 * PRIGHT_i + \beta_6 * LAW_i + \beta^* \sum_k Z_i + \varepsilon_i$$

where LPR_i is the natural logarithm of the piracy rate in country i, *LGNP* is the natural logarithm of the GNP per capita, *IDV* is the individualism index, *INEQ* represents our different measures of income inequality, *PRIGHT* is an index of patent protection, and *LAW* is a measure of institutional quality. The vector Z_i includes a set of k additional explanatory variables that are likely to influence piracy rates and probably also linked to income distribution. These remaining additional variables are introduced in detail in the Section on robustness analysis (Section 6). The β are unknown parameters to be estimated, and ε_i is the usual random error term. The focus of this paper is on this model, to be estimated using both ordinary least squares (*OLS*) and two stage least squares (2SLS).

To account for measurement errors and statistical outliers in the dependent and independent variables, we use averages over a longer time span, even though doing so results in a substantial loss of degrees of freedom and a single cross-section to be estimated.[10] In addition, in cross-sectional analysis, the regression results are highly likely to be sensitive to the presence of influential observations or outliers (Kennedy, 2003). Therefore, we make extensive use of several regression diagnostics like Studentized residuals and Cook's D to test for the presence of influential cases. However, when our regression diagnostics suggest the presence of highly influential cases, we rerun the analysis dropping those cases. If the pattern of the results is substantially changed, we present our results with these cases removed.

In this model, there is a risk of endogeneity bias stemming from a reverse causality between software piracy and legal protection of intellectual property rights. That is, high piracy rates may indicate economic dependency on pirated copies of software, in which case policy authority may not strengthen the laws. On the other hand, a high share of illegal software may affect private business or firms, who may then lobby for stronger laws. But it is also reasonable to expect that the degree of endogeneity will not be severe if it is argued that policy maker response to changes in piracy may not be instantaneous. In this case, one simple way to ameliorate the problem is to use values of the IPR protection variable at the beginning of the observation period. In addition, the IPR protection variable may also be subject to measurement error because of the likelihood of gaps between measure and actual protection. Nonetheless, this likelihood may not be so important because countries with strong statutory laws tend also to be those that strictly execute the laws (Park, 2001). [11] Endogeneity might also be present with some of the remaining economic and socio-demographic control variables, such e.g. the national income or the share of PC users in the population. In these cases the simultaneity might be mitigated by choosing a time span for the explanatory variables that is by far larger then the one for the dependent variable (see Table A5 of the Appendix).

The empirical analysis is carried out in four steps: (a) a simple regression model on the level of economic development, individualism, and income inequality measures; (b) examination of the core specification for geographic subregions; (c) integration of the level of IPR legal protection and institutional quality as additional predictors; and (d) control for a number of other variables likely to be linked to piracy and income inequality. The next section discusses the data to be used to test these specifications.

4. Data

In this examination we use an initial sample of 71 countries from Europe, Africa, Asia, and Latin America, a much larger sample than those in other studies on the link between inequality and piracy.[12] However, we exclude from the empirical analysis countries with any missing information. Since we are combining a large number of datasets, we have different numbers of observations for different variables and model specifications, with some regressions covering as few as 43 countries. The data also come from different decades, but some of the data are from the 1990s. The list of countries included in the study is provided in Table A1 of the Appendix.

4.1. Primary variables

The data on the variables used in the empirical analysis come from diverse sources. Data for the dependent variable, the average piracy rate for 1994–2002, are obtained from the International Planning Research Corporation's annual report for the year 2003 (IPRC, 2003).[13] As stated before, piracy rate is defined as the estimated difference between software programs installed and software applications legally licensed. This variable is reported as percentages, ranging from 0 percent (no piracy) to 100 percent (all software installed is pirated). Admittedly, figures on piracy are likely to be underestimated because a large number of software applications are sold without the computer hardware (Husted, 2000; Traphagan and Griffith, 1998; van Kranenburg and Hogenbirk, 2005). On the other hand, they might also be biased upwards as the publishers are business corporations which have an interest in a rigorous software protection or simply because they naively assume that all pirated software would have otherwise been acquired legally (Gayer and Shy, 2003). Nonetheless, this data source is the most reliable for cross-country comparisons and is commonly employed in the empirical research (e.g., Depken and Simmons, 2004; Husted, 2000; Marron and Steel, 2000; Rodríguez, 2005a, b; van Kranenburg and Hogenbirk, 2005). As shown in Table A1 of the Appendix, average piracy rates vary considerably across countries. In some countries, like Russia and Vietnam, almost all software is pirated, while in other countries, like Denmark and Norway, the proportion of illegal software is below 25 percent. Table A3 of the Appendix reports the average piracy rates for subgroups by geographic region. The highest piracy rate is in Eastern Europe, followed by the Middle East and Latin America. The lowest piracy rate is in North America (the U.S. and Canada).

As a measure of income inequality, we use Gini coefficients[14] (*GINI*) and quintile shares (*Q1-Q5*) based on the income distribution data gathered by Dollar and Kraay (2002) from four sources: the United Nations-WIDER Income Inequality Database, Deininger and Squire (1996), Chen and Ravallion (1997), and Lundberg and Squire (2003). The latter dataset includes observations covering 137 countries for the period 1960–1999. We use average values of the measures of economic inequality for the period 1960-1999 for countries with at least one observation for that period; these measures, however, are not expected to change much within countries over time. For empirical analysis, the lower income class is represented by the lowest quintile, the middle income class by the sum of the three middle quintiles, and the highest income class by the fifth quintile. For all countries, the share of the lowest quintile is well below its share in an equal distribution (< 0.2), while the share of the highest income class is far above the value of 0.2 (minimum: 0.32). Only the share of the middle income class comes close to the value of equal distribution of 0.6 in 25 percent of the countries, but never reaches this level (see Table A2 of the Appendix). In our sample, the Gini coefficient is highly correlated with all three quintile share measures (see Table A4 of the Appendix).

As regards the remaining explanatory variables of the baseline model, data on GNP per capita for 1970–95 are extracted from World Development Indicators (WDI), as calculated by La Porta et al. (1999). The cultural dimension, proxied by the individualism index (*IDV*), is derived from an initial survey developed and conducted by Hofstede (1997, 2001) of over 72,000 IBM employees over the period 1967–1973 in 40 countries. In a subsequently constructed estimate for additional countries, Hofstede assigns each country a value between 0 and 100, with a higher value corresponding to a higher level of individualism. Because data on this cultural attribute are unavailable for only a very few cases, we use the average of the individualism score for bordering countries. As instruments for the inequality variables, we employ the average life expectancy at birth and average birth rate per 1,000 people from 1970 to 1995 as provided by the World Bank.

4.2. Other variables

The variables used in the robustness analysis are all from the World Bank World Development Indicators, with the exception of the proxy variable for the degree of institutional quality, the rule of law (*LAW*), which comes from Kauffman et al. (2004). This index, ranges from -2.5 (weakest institution) to 2.5 (strongest institution). We employ the index value for the year 1996. [15]

We also use the intellectual property rights or patent index (*PRIGHT*) for 1995 from Ginarte and Park (1997) as a proxy for the degree of protection of intellectual property rights. Even though this index is not a specific measure of software protection at the country level (Rodríguez, 2005b), it captures a wider range of features regarding the variability of patent protection across countries than a simpler dummy variable approach (see Maskus, 2000, for a detailed discussion). This index covers 110 countries and is based on an evaluation of current national patent laws.[16] It ranges from zero to five, with higher values representing stronger patent rights protection.

Definitions and sources for all variables used in the paper are provided in Table A5, and their descriptive statistics in Table A2 of the Appendix. The results for the Pearson correlation coefficients also computed for the main variables employed in this study indicate that multicollinearity does not appear to be a problem (see Table A4 of the Appendix).

5. Results

5.1. Baseline model

Table 1 displays the results for the OLS and 2SLS estimations using the baseline model of all countries in the sample. All regression models pass a RESET test (Ramsey, 1969). A correction of standard errors for heteroscedasticity before and after instrumentation, where necessary, was carried out.[17] For all specifications, the hypothesis of normality of residuals cannot be rejected according to the Bera Jarque test statistics (results not reported). Columns 1, 3, 5, 7, 9, 11, and 13 give OLS estimates of the core specification. With few exceptions, the control variables reveal the expected signs and are highly significant. The set of explanatory variables together with the constant term explains around 80 percent of the variation in piracy rates, which indicates a good fit for our models.

As conjectured, there appears to be a nonlinear relation between income per capita and piracy. The explanatory power of the models increases considerably after the inclusion of the GNP squared term. All coefficient on the GNP variables are significant, and, in addition, the null hypothesis that the coefficients are jointly equal to zero is rejected for all specifications at the 1 percent level (not reported). The estimation results indicate that illegal copying tends to increase in national income initially, but once nations reach some threshold value of GNP per capita, a decline in piracy rates is observed, probably caused by improvements in the protection of intellectual property rights. This relation is similar to the well-known environmental Kuznet’s curve (EKC) between per capita income and pollution (Grossman and Krueger, 1995). Further, these results seem consistent with Rodriguez's (2005b) finding of an inverted U-shaped relationship between per capita income and piracy for a sample of 23

European countries.[18] Beyond the maximum point of the GNP curve, the level of economic development is negatively linked to piracy in all regression models. This finding is equally consistent with the results in Marron and Steel (2000), Depken and Simmons (2004), and Rodríguez (2005a), who tested and showed a linear negative relation. In addition, the individualism index *(IDV)* is inversely related to the dependent variable. Similar findings are reported by Husted (2000), Marron and Steel (2000), Moores (2003), and Depken and Simmons (2004), among others. As previously stated, the explanatory assumption is that in more individualistic countries, a person feels less group pressure to share software with peers, and thus more individualistic countries tend to experience lower levels of piracy.

Insert Table 1 about here

As regards the variables of interest in the OLS regression outcomes – namely, the measures of income distribution – the significance and the positive sign of the coefficient of the income share of the first quintile (*Q1*), reported in column 1, contradict our hypothesis that poor persons do not influence piracy rates.[19] The higher the income share in total income of the first 20 percent of all income earners, the more software is pirated. Since the actual share of the first quintile is below the share of equal distribution of 20 percent for all countries, we can equally state that the more unequally income is distributed in society with respect to the first quintile, the less software piracy occurs. Moreover, all coefficients of the three different single middle quintiles, given in columns 3, 5 and 7, are positive, but only the lower two are significant at the 1 percent level, whereas that of the fourth quintile is rendered insignificant. In general, this result provides support for the hypothesis that the middle classes are more likely to be engaged in intellectual theft because they have the means to commit it and the least to lose. Results for the aggregated middle class share (*Q2* + *Q3* + *Q4*) usually employed in these types of studies are reported in column 9. Its coefficient is strongly significant and shows the expected positive sign, and is therefore also in line with our initial hypothesis. The richness of the middle income class is positively associated with software piracy, but for all countries its share in total income is below the equal share of 60 percent; stated differently, a more unequal income distribution with respect to the middle income class leads to less piracy, as already observed for the lowest quintile. As shown in column 11, the share of the fifth and highest quintile (*Q5*) has a negative and strongly significant effect on piracy (at the 1 percent level). This negative coefficient could potentially be interpreted to mean that the richest in society can afford and prefer to buy all software legally and that this behavior becomes more likely as their total income share rises. In this case, given that income is unequally distributed with respect to the fifth quintile in all countries in this sample (given a share far above 20 percent), we can state that a higher inequality in income distribution leads to lower rates of intellectual theft. As illustrated in column 13, the overall impact of the degree of income inequality, the coefficient for *GINI* is negative and significant at the 1 percent level, suggesting that such inequality exerts a beneficial impact on piracy because the more unequal the income distribution in society, the less piracy occurs. This result corroborates our findings for the single quintiles. The OLS estimate of income inequality (0.006) is similar in magnitude to that obtained in previous cross-sectional studies (Husted 2000; Rodríguez, 2005a).[20] Overall, most of these empirical results conform with our hypotheses on the influence of income distribution on piracy rates.

Columns 2, 4, 6, 8, 10, 12 and 14 report the results for the baseline model when 2SLS are estimated to correct for the possible measurement error of inequality measures. The 2SLS estimates are also statistically significant but often substantially larger in magnitude than their corresponding OLS estimates.[21] Particularly the coefficients on the middle quintiles are strongly affected through instrumentation. Hence, measurement errors in the income distribution variables tend to bias the coefficients downwards so that their true impact is underestimated in an OLS regression. In contrast, the 2SLS estimates of the coefficients for the remaining explanatory variables are barely affected by the instrumentation of the income inequality variable. The observed directions of the various income inequality measures are the same as in the OLS regressions, but significance levels appear to be affected; particularly, the share of the fourth quintile (*Q4*) in total income, which, as shown by its increased significance (at the 5 percent level), exerts now the predicted piracy raising impact.

Given that the 2SLS approach yields unbiased coefficients of the inequality measures, a brief look at their relative sizes might be interesting. In general, marginal impacts of inequality measures are up to 10 times bigger then the ones observed for the GNP variables. The largest marginal impact on piracy rates in absolute terms is exerted by the lowest quintile (estimate of 2.671), followed by the middle quintiles (1.185), and then by the highest quintile (-0.834). According to these results, a redistribution of income from the lowest to the highest quintile would be most efficient as it would result into the biggest reduction of piracy rates (about -3.50 percentage points). A redistribution from the middle quintiles to the highest quintile, however, would lower piracy rates by only about -2.02 percentage points. Both policies would cause an increase in overall income inequality. On the other hand, a redistributive policy aiming at a more equal income distribution would unambiguously lead to an increase in piracy rates. These conclusions are supported by the negative sign of the coefficient on the *GINI* estimate.

These findings for our baseline model affect our approach for the remaining analyses in the following ways. First, from now on we report only the results from the lowest quintile, the aggregated middle income shares, the upper quintile, and the Gini coefficient, thereby foregoing separate examination of the single second, third, and fourth quintiles. Second, we also base the remaining analyses on the 2SLS results. The F-test of excluded instruments of our first stage regression indicates that the chosen instruments are strong predictors of income inequality measures.

5.2. Analysis of the baseline model on regional data

The results reported thus far are suggestive but are subject to distortion arising from the inability to adequately control for country-specific factors. Therefore, we extend the analysis to examine the impact of income distribution measures for various geographical regions in the world with the aim of providing a check on the baseline model results (see Table 1). Regional unobserved determinants might include factors such as cultural or geographic (climate, temperature) characteristics. The regions, which are classified according to the regional groups defined by the World Bank,[22] were selected based on the availability of a sufficient number of observations. In general, any lack of significance is probably caused by too small a sample size or too low a variation in the sample. In addition, the instruments in such a small sample do not always perform well for similar reasons. In such case, only the signs of the coefficients can be interpreted.

5.2.1. Latin American and Caribbean countries

Table 2 reports the results obtained for the sample that includes only 17 Latin American and Caribbean countries. As in all the regression models, the coefficients on income distribution measures are insignificant. The GNP variables are jointly significant at the 1 percent levels, and, in general, do have the predicted signs. Contrary to expectations, individualism exerts a positive but quantitatively negligible impact, which might be due to the cultural similarities of the countries in that sample. As regards the signs for our variables of interest, we observe a piracy dampening impact of the first and fifth quintiles, but a piracy rate increasing impact of the middle quintiles. In this subsample, the different income classes influence software piracy as suggested in Section 2. Thus, a rising share in total income of the highest or lowest income classes is associated with lower piracy rates. Obviously, in these countries with a relatively medium level of per capita income, only wealthier people of the middle income class are potential offenders (and suppliers of illegal software). The estimates for the quintiles are consistent with the negative value of *GINI*. In this subsample, redistributive policies in favor of the lowest or highest income classes would lead to a decline in piracy rates. As the size of the coefficients indicate, redistribution from the middle quintiles to the lowest quintile would be the most efficient policy compared to a redistribution to the highest quintile (estimates of -3.341 vs. -0.494).

Insert Table 2 about here

5.2.2. West European and North American countries

Table 3 displays the results for the sample that includes West European and North American countries, which leaves 13 countries after exclusion of outliers. First, there seems to be a nonlinear relationship between income and piracy, as an F-test on the joint significance of the GNP variables allows rejection of the hypothesis of no influence (except regression 3.2). The signs of the coefficients indicate that software piracy decreases in national income as long as GNP is below a certain threshold. In addition, despite its negative marginal effect, the individualism variable fails to exert the predicted significant impact, most likely because of the similarity in culture in this region. Interestingly, the signs of the coefficients for both the middle and the low income shares indicate a piracy dampening influence, whereas the sign of the one for the highest share is piracy enhancing. Thus, a higher relative income of the lower and middle class translates into a higher share of more legally bought software, whereas a higher relative income of the highest class leads to a greater demand for pirated software. The reason for this unexpected finding might be that the richest 20 percent in Europe are relatively poor compared to the income of the richest 20 percent of the world. The estimation results suggest that in the wealthiest and most industrialized countries of the world, which are contained in this sample, software piracy appears to be an upper class crime. The positive estimate of the Gini coefficient for these countries corroborates this result: the higher the proportion of the lowest and middle income shares relative to the highest income shares, the less piracy occurs. For these countries, in order to most efficiently prevent piracy a redistribution of income should occur from the wealthiest class particularly to the middle income earners as the magnitude of the coefficients suggests; a redistribution to the lowest income class instead would impact piracy rates at a substantially lower level (estimates of -10.767 vs. -2.432).

Insert Table 3 about here

5.2.3. European and Asian countries

Table 4 presents the results obtained for Eastern European and Central Asian countries (10 observations). Here, the GNP variables do not appear to have a statistically significant impact. In addition, the hypothesis that all coefficients on the GNP variables are jointly equal to zero cannot be rejected in all regressions. Moreover, the coefficient on *IDV* is negative, as predicted, but not significant at any conventional level. In this sample, the lowest quintile and middle quintiles exert a software piracy increasing influence, whereas the highest quintile appears to lower piracy offences; the estimates for the quintiles mirrors the findings for the full sample in Table 1. The coefficient on *GINI* is negative, although not significant. Again, a redistributive policy to prevent piracy should focus on lowering income for the lower or middle classes; the first policy would be more efficient as indicated by the size of the quintile estimates (estimates of 1.769 vs. 1.141).

Insert Table 4 about here

Table 5 outlines the results for the sample that includes South Asia, East Asia and the Pacific (12 countries).[23] As in the full sample, in this subsample, a nonlinear pattern is detected between national income and piracy, and individualism exerts the predicted significant piracy dampening impact. It should be noted that Australia and New Zealand, as well as China and Vietnam, are part of this sample, so that the variation in the individualism variable is sufficiently strong. In this sample, the income shares have the expected signs: the first quintile exerts a software piracy lowering impact, as does the upper quintile, whereas the middle income classes affect piracy rates positively. Hence, in the South Asian, East Asian and Pacific regions, which include mainly OECD countries and the Asian tiger states, intellectual theft does appear to be a middle class crime. The positive coefficient of the GINI supports our earlier findings for the quintiles. In such case, a redistributive policy should distribute income from the middle class (estimate of 0.824) to either the lower class (-2.430) or the upper class (-0.055), with the first policy obviously being the more efficient one.

Insert Table 5 about here

Taken together, the results from the specific geographic regions illustrate that changing the countries in a sample leads to different results for the impact of income classes on software piracy. Software piracy appears to be an exclusively middle class crime in Latin America, the Caribbean, East Asian and the Pacific regions, a crime of both middle and lower class in the Central Asian and Eastern European regions, but an upper class crime in Western Europe and North America. It is plausible to assume that the influence of income classes on software piracy rates depends strongly on the wealth of the countries that form the regional subsample.

6. Robustness Analysis

6.1. Controlling for institutional and legal factors

Up to this point, we have only considered the effects of economic and social determinants on piracy such as economic inequality, the level of economic development, and individualism. Earlier findings in this field are based on such models considering these factors only. However, some more recent empirical research suggests that other – particularly institutional – factors may have a direct impact on intellectual theft equally. Since such factors might in turn be correlated with some of the determinants of our baseline model, we have to take them into account. For this reason, we estimate a complete model of software piracy by including two such factors: a measure of institutional quality (*LAW*) and intellectual property rights protection (*PRIGHT*). Both these variables can be viewed as crime deterrents (see discussion in Section 3.2). Table 6 reports the regression results for the above baseline model with these new variables added. Regressions in columns 6.1 to 6.4 and 6.5 to 6.8 report the results when only one institutional determinant is added, and columns 6.9 to 6.12 give the findings when both are included simultaneously.

Insert Table 6 about here

The first regression model confirms the hypothesis that stronger intellectual property rights protection may lower piracy rates (columns 6.1 to 6.4) (at the 1 percent level), a result that is in line with previous research like that of Rodríguez (2005b) who emphasizes that legal factors proxied by an index of patent right protection matter for piracy. Interestingly, the inclusion of this deterrent does not cause a drop in the significance of the GNP variables compared to the baseline model (Table 1), as both remain significant at the 1 and 5 percent levels. We would have expected such a drop in significance because countries with a higher per capita income also tend to protect property rights more rigorously than do poorer countries (see Table A4 of the Appendix [rho = 0.79]). In this context, the persistence of the nonlinear relationship between income and piracy after this measure of institutional quality is controlled for is particularly noteworthy.

In regressions 6.5 to 6.8, the estimate of the impact of the effect of institutional quality on piracy rates is negative and statistically significant at the 1 percent level. As stated in the theory section, the stricter the rule of law, the lower the level of piracy should be, the underlying rationale being that efficient execution of laws may raise the probability and/or severity of punishment, resulting in a decline of the expected economic gains from software piracy. This finding is also obtained in other empirical studies (Holm, 2003; Rodríguez, 2005a). We should note the considerable difference in size of coefficients compared to the ones of the index of patent rights which suggests that well-functioning legal institution as a general deterrent perform considerably better than the particular protection of software through patents.[24] In explanation, as the patent right index is based on the analysis national laws only, it does not capture the rigorousness of their enforcement. Again, as the correlation matrix indicates, there is a strong positive link between the rule of law variable and the GNP per capita variable (see Table A4 in the Appendix, [rho = 0.86]). Therefore, it is surprising that the GNP variables do not appear less significant than in the baseline model, with significances at the 1 and 5 percent levels.

As a final step, the institutional quality and patent rights index are added simultaneously to the baseline model. Again, the coefficient on the rule of law variable is negative and statistically significant at least at the 1 percent level. In addition, consistent with the results obtained in regressions 6.1 to 6.4, the intellectual property rights variable enters negatively, but with significances at the 5 percent level considerably below the ones observed in the first specification. One possible explanation is that the causal effect of *PRIGHT* variable is being usurped by our institutional quality variable *LAW*. Interestingly, the magnitude of the coefficients on both variables are practically unaffected when we drop one of our measures of institutional quality. Similarly, the regression outcomes yield robust results for the nonlinear relationship between the level of economic development and piracy presented in Table 1, with comparable significance levels.

As regards the inequality measures, their impact is robust to the inclusion of either new institutional deterrent variable. In all specifications, our hypothesis on the impact of middle and high income shares on piracy rates is corroborated: the middle income is associated with rising piracy rates (regressions 6.2, 6.6, and 6.10), and a higher share of the fifth quintile leads to lower intellectual theft rates (regressions 6.3, 6.7, and 6.11). Again, the impact of the lowest quintile does not appear to be consistent with our a priori expectations (regressions 6.1, 6.5 and 6.9). Likewise, the coefficient on *GINI* is always negative and statistically significant (regressions 6.4, 6.8, and 6.12). As regards the size of the coefficients of the inequality measures, again the ranking in absolute sizes is that obtained in Table 1. The inclusion of *LAW*, however, raises the size of the coefficients on all inequality measures considerably. The correlation matrix in the Appendix (Table A4) shows that, indeed, more equal societies tend to be the ones with a better legal institutions. Thus, omission of the *LAW* variable leads to an underestimation of the true impact of income distribution in society. This might suggest that inequality plays a direct role in explaining the degree of intellectual property rights protection and not just through its effects on legal institutions.

In all regression models, our measure of individualism (*IDV*) has the expected negative sign and is statistically significant. This result is particularly interesting as we might expect the state to take over the role of prosecutor and defender of societal rules in individualistic societies in which the control of the group/clan over individual actions is looser. If this were the case, there should be a positive correlation between the degree of individualism and the development of governing state structures. Indeed, the correlation between the two institutional variables and the individualism measure exceeds the value of 0.67 for each (see Table A4 in the Appendix). Finally, given that we achieve respectable F-tests on the excluded instruments in the first stage of the regressions and pass standard overidentification chi-square tests, the instruments appear to be valid.

In sum, the impact of the income inequality measures appears to be robust to the inclusion of additional institutional control variables such as the rule of law and the protection of intellectual property rights. The significance of the key variables enhances our confidence in the initial results displayed in Table 1. Clearly, our results indicate that because variables measuring the quality of the legal institutions act as deterrent factors for piracy and thus appear to be decisive determinants, they should not be neglected in any future analysis of cross-national piracy rates. As a result, we will – from here on – always include these institutional deterrents in our baseline model.

6.2. Potentially omitted determinants of piracy rates and inequality

As a final check, we test the sensitivity of the inequality measures to the inclusion of further variables which might be correlated with income inequality (always also controlling for the two institutional factors *LAW* and *PRIGHT*). In addition, this exercise serves as test whether the model estimated in Section 6.1 is complete or suffers from a potential omitted variable bias.

These additional regressors include legal origin, particularly roots in English Common Law, French Commercial Code, Communist/Socialist Laws, German Commercial Code, or Scandinavian Commercial Code, educational attainment proxied by the secondary and tertiary school enrollment rates, Internet and personal computers penetration rates in the population[25], total unemployment and male youth unemployment rates, trade openness measured as the share of exports and imports in GDP, R&D and information and technology expenditures as percentages of GDP, urbanization that measures the percentage of urban population in the total population, the percentage of the Protestant, Catholic, and Muslim populations in 1980, ethnolinguistic fractionalization in the country, and, finally, the proportion of males aged 15–19, 20–29, and 30–39 years in the population. A description of the source and the time span covered of these determinants can be found in Table A5 of the Appendix. The inclusion of these additional controls is based on a large body of empirical literature on piracy (e.g., Marron and Steel, 2000, van Kranenburg and Hogenbirk, 2005) as well as crime (e.g., Ehrlich, 1973; for a review, see Eide, 2000) [26]. Some of these variables proxy the part of the population most likely to be prone to commit property crimes, the demand for illegal software, the awareness of legal stipulations in the population, cultural factors, or, finally, the self-interest of the government in intellectual property protection. The results of this exercise are presented in Table 7.

Overall, as Table 7 shows, the directions of the impact of the income inequality measures remain robust to the inclusion of any of these additional controls. In most of the cases also the size of the coefficients on the distribution measures are comparable to those displayed in Table 6. In three specifications, however, significance levels are negatively affected, specifically when the legal origin, the information and technology expenditure, or the share of young male population are included in the regressions. In explanation, the historical roots of the legal system might serve as a proxy for preferences for a specific income distribution which was shaped by the same (colonial) history. Consequently, countries with an English legal tradition might accept higher income inequality and actually have a more unequal income distribution whereas countries with a socialist tradition favor a more equal income distribution. Second, in the regression that includes the communication and information technology expenditures variable, our expectation is that higher information and technology expenditures might be associated with lower income inequality, if technological progress is skilled biased. It should also be noted that the number of observations in this specification is by about one third smaller than the number of observations for which the baseline model was estimated (see Table 1). Third, the share of young males in the population might be a sociodemographic factor linked to income inequality because young persons tend to have an income below the average income in society. Thus, a growing share of young principal earners will generate an increase in income inequality.

Insert Table 7 about here

Among the additional explanatory variables used in this robustness test, only a few exert a significant impact on software piracy.[27] Information and communication technology expenditures appear to be strongly negatively correlated with piracy rates at the 1 and 5 percent levels. Since in the same regression model, the institutional quality and intellectual property rights protection index as well as income inequality measures are insignificant, we can interpret this finding to mean that this variable captures partly the effects for well-developed legislative and executive institutions as well as income distribution. Openness, defined as the sum of shares of exports and imports in GDP, accounts for the extent to which a country is linked internationally. It can be conjectured that the stronger this trade link, the better property rights should be protected. The estimation outcomes, however, do indicate that in all regression models more openness in a country is associated with higher piracy rates (at the 5 and 10 percent levels). This finding is somewhat inconsistent with previous empirical studies of piracy (e.g., Shadlen et al., 2005; van Kranenburg and Hogenbirk, 2005). The reason might be either an endogeneity problem or the inclusion of variables measuring institutional quality and a potentially resulting multicollinearity.[28] In addition, the Protestant, Catholic, and Muslim population shares are negatively correlated with piracy rates, although only the coefficient on Protestants is statistically significant at the 5 percent level. Obviously, religious confessions seem to capture some of the effects of the economic development and individualism variables, which loose significance in this specification. The remaining additional determinants tested are not significant at conventional significance levels in any regression model.[29]

7. Conclusions

This paper examines the relationship between income inequality and software piracy rates using a rich dataset on income distribution measures whose averages cover a longer time span than most statistical methodologies have allowed to date. Obviously, this study has limitations; most particularly, the basing of the empirical analysis on a cross-sectional estimation method that is caused by data availability. Nevertheless, we found that income inequality seemingly has a negative and significant effect on piracy rates which supports previous empirical studies that potentially suffer from methodological shortcomings (e.g., Husted, 2000). Most particularly, an increasing share of the highest quintile appears to be crime lowering, whereas the ones of the remaining quintiles seem to increase piracy rates. The regression results also reveal that national income, rule of law, the degree of protection of intellectual property rights, and individualism are important determinants in any explanation of piracy rate variations at a cross-country level. Moreover, these results are robust to the inclusion of a wide range of additional control variables often employed in predicting piracy rates and the occurrence of property crime.

Most interestingly, the impact of income distribution may differ across subsamples defined by geographic region. In particular, the Gini index is observed to impact piracy rates both negatively and positively depending on the sample considered. To shed light on this issue, we also investigate the influence of the various income quintiles on the occurrence of software piracy. Software piracy appears to be a middle class crime in the full sample, and in all regional subsamples except the West European and North American geographic region. In contrast, a richer middle class leads to less piracy in Western Europe and North America. The upper class also influences piracy rates in different ways: specifically, and contrary to prediction, it appears to be piracy enhancing in Western Europe and Northern America, but piracy lowering in the rest of the world. Finally, also the impact of the lowest income class differs between regions because it appears to be piracy enhancing only in the economically more developed region of Eastern Europe and Central Asia.

This study also has some policy implications. Our results indicate that piracy may respond to economic differences across countries and that many other noneconomic factors are also important. Thus, initiatives to fight piracy may interfere with the cultural and institutional values of a nation. Therefore, preventive policies against piracy need to be adapted to a nation's cultural and institutional values. In addition, economic policies should emphasize the role of income distribution, as well as which parts of the income distribution matter in reducing levels of piracy. Changing income distribution, however, can only be achieved in the long-run and might be opposed by many societal groups. Moreover, in some regions the redistributions necessary to prevent piracy might be unfavorable from other perspectives, such as distributional fairness considerations. Short-term successes in fighting piracy are most likely to be achieved through the strengthening of legal institutions, which, in turn might also enhance economic growth and prosperity (see e.g. Acemoglu et al. 2004).

Appendix

Countries included in the sample**:** Argentina, Australia, Austria, Bahrain, Bolivia, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Croatia, Cyprus, Czech. Rep, Denmark, Dom. Rep, Ecuador Egypt, El Salvador, Estonia, Finland, France, Germany, Greece, Guatemala, Honduras, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kenya, Korea. Rep, Kuwait, Lebanon, Lithuania, Malaysia, Malta, Mauritius, Mexico, Morocco, Netherlands, New Zealand, Nicaragua, Nigeria, Norway, Oman, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Reunion, Romania, Russian Federation, Saudi Arabia, Singapore, South Africa, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, UAE, United Kingdom, Ukraine, Uruguay, United States, Venezuela, Vietnam, Zimbabwe.

Insert Tables A1 through A6 here

References

Acemoglu, D., Johnson, S., Robinson, J.A. 2001. The colonial origins of comparative development: an empirical investigation. American Economic Review, 91, 1369–1401.

Acemoglu, D., Johnson, S., Robinson, J.A. 2004. Institutions as the fundamental cause of long-run growth, CEPR Discussion Papers No. 4458.

Becker, G.S. 1968. Crime and punishment: an economic approach. Journal of Political Economy, 76, 169–217.

Bourguignon, F. 1998. Crime as a social cost of poverty and inequality: a review focusing on developing countries. 1999/2000 WDR Background paper; Washington, D.C.

Chen, S., Ravallion, M. 1997. What can new survey data tell us about recent changes in distribution and poverty?. World Bank Economic Review, 11, 357–382.

Cohen, L.E., Land, K.C. 1987. Age structure and crime: symmetry versus asymmetry and the projection of crime rates through the 1990s. American Sociological Review, 52, 170 – 183.

Deininger, K., Squire, L. 1996. A new data set measuring economic inequality. World Bank Economic Review, 10, 565–591.

Depken, C.A., Simmons, L.C. 2004. Social construct and the propensity for software piracy. Applied Economics Letters, 11, 97–100.

Dollar, D., Kraay, A. 2002. Growth is good for the poor. Journal of Economic Growth, 7, 195–225.

Doyle, J.M., Ahmed, E., Horn, R.N. 1999. The effects of labor markets and income inequality on crime: evidence from panel data. Southern Economic Journal, 65, 717–738.

Ehrlich, I. 1973. Participation in illegitimate activities: a theoretical and empirical investigation. Journal of Political Economy, 81, 521–565.

Eide, E. 2000. Economics of criminal behavior, in: Bouckaert, B., de Geest, G. (Eds.), Encyclopedia of Law and Economics, Vol. V. Elgar, Cheltenham, 345–389.

Fajnzylber, P., Lederman, D., Loyaza, N. 2002. What causes violent crime?. European Economic Review, 46, 1323–1357.

Feldman, Y., Nadler, J. 2005. The effect of law on perceived consensus about file sharing norms, Bar-Ilan University, Faculty of Law, unpublished manuscript.

Freeman, R.B. 1999. The economics of crime, in: Ashenfelter, O.C. Card, D. (Eds.), Handbook of Labor Economics, Vol. 3C. Elsevier, Amsterdam, chapter 52, 3529–3571.

Gallegos, F. 1999. Software piracy: some facts, figures, and issues. Information Systems Security, 8, 34–56.

Gayer, A., Shy, O. 2003. Copyright protection and hardware taxation. Information Economics & Policy, 15, 467–483.

Ginarte, J.C., Park, W.G. 1997. Determinants of patent rights: a cross-national study. Research Policy, 26, 283–301.

Glaeser, E.L, Sacerdote, B. 1999. Why is there more crime in cities?, Journal of Political Economy 107, part 2, S225 – S258.

Gopal, R.D. 2000. Global software piracy: you can't get blood out of a turnip. Communications of the ACM, 43, 82–89.

Gopal, R.D., Sanders, G.L. 1998. International software piracy: analysis of key issues and impacts. Information Systems Research, 9, 380–397.

Gopal, R.D., Sanders, G.L. 1997. Preventive and deterrent controls for software piracy. Journal of Management Information Systems, 13, 29–47.

Gould, E.D., Weinberg, B.A., Mustard, D.B. 2002. Crime rates and local labor market opportunities in the United States: 1979–1997. Review of Economics and Statistics, 84, 45–61.

Grossman, G.M., Krueger, A.B. 1995. Economic growth and the environment. Quarterly Journal of Economics, 110, 353–377.

Hofstede, G. 1983. Dimensions of national culture in fifty countries and three regions, in: Deregowski, J.B., Dziurawiec, S., Annis, R.C. (Eds.), Explorations in Cross-Cultural Psychology, Swets and Zeitlinger, Lisse, The Netherlands, 335–355.

Hofstede, G. 1997. Cultures and Organizations: Software of the Mind (rev. ed.). McGraw Hill, New York

Hofstede, G. 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations (2nd ed.). Sage Publications, Thousand Oaks, CA.

Holm, H.J. 2003. Can economic theory explain piracy behavior?. Topics in Economic Analysis & Policy, 3, article 5, 1–15.

Husted, B.W. 2000. The impact of national culture on software piracy. Journal of Business Ethics, 26, 197–211.

International Planning Research Corporation. 2003. Eighth Annual BSA Global Software Piracy Study. Trends in Software Piracy 1994–2002. Washington, DC.

Kanwar, S. 2002. Does intellectual property lead to intellectual property protection?. Working Paper No. 108, Center for Development Economics. Delhi School of Economics.

Kaufmann, D., Kraay, A., Mastruzzi, M. 2004. Governance matters III: governance indicators for 1996, 1998, 2000, and 2002. The World Bank Economic Review, 18, 253-287.

Kennedy, P. 2003. A Guide to Econometrics (5th ed.). Massachusetts Institute of Technology Press, Boston.

La Porta, R., Lopez-de-Silanes, F. Shleifer, A., Vishny, R. 1999. The quality of government. Journal of Law, Economics and Organization, 15, 222–279.

Lundberg, M., Squire, L. 2003. The simultaneous evolution of growth and inequality. Economic Journal, 113, 326–344.

Machin, S., Meghir, C. 2004. Crime and economic incentives. Journal of Human Resources, 39, 958–979.

Marron, D.B., Steel, D.G. 2000. Which countries protect intellectual property?. The case of software piracy. Economic Inquiry, 38, 159–174.

Maskus, K.E., Penubarti, M. 1995. How trade-related are intellectual property rights?. Journal of International Economics, 39, 227–248.

Maskus, K.E. 2000. The global effects of intellectual property rights: measuring what cannot be seen, in Maskus, K.E.(Ed.), Intellectual Property Rights in the Global Economy, chapter 4, 87–142, Institute for International Economics, Washington, DC.

Moorehouse, J.C. 2001. Property rights, technology, and Internet distribution. Journal of Technology Transfer, 26, 351–361.

Moores, T.T. 2003. The effect of national culture and economic wealth on global software piracy rates. Communications of the ACM, 46, 207–215.

Osorio, C.A. 2002. A contribution to the understanding of illegal copying of software: empirical and analytical evidence against conventional wisdom. Working Paper, Center for Technology, Policy, and Industrial Development (CTPID), MIT Massachusetts. Available through http://hdl.handle.net/1721.1/1479.

Ostergard, R.L. 2000. The measurement of intellectual property rights protection. Journal of International Business Studies, 31, 349–360.

Park, W. 2001. Intellectual property and patent regimes, in: Gwartney J., Lawson, R. (Eds.), Economic Freedom of the World, Fraser Institute, Vancouver, 101–117

Ramsey, J.B. 1969. Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society. Series B, 31, 350–371.

Rapp, R.T., Rozek, R.P. 1990. Benefits and costs of intellectual property protection in developing countries. Journal of World Trade, 24, 75–102.

Richardson, R.S., Gaisford, J.D. 1996. North-south disputes over the protection of intellectual property. Canadian Journal of Economics, 29, S376–S381.

Rodríguez, A. 2005 a. Software piracy and income inequality. Forthcoming in: Applied Economics Letters.

Rodríguez, A. 2005 b. The relationship between copyright software protection and piracy: evidence from Europe. Forthcoming in: European Journal of Law and Economics.

Ronkainen, I.A., Guerrero-Cusumano, J.-L. 2001. Correlates of intellectual property violation. Multinational Business Review, 9, 59–65.

Shadlen, K.C., Schrank, A., Kurtz, M.J. 2005. The political economy of intellectual property protection: the case of software. International Studies Quarterly, 49, 45–71.

Shin, S.K., Gopal, R.D., Sanders, G.L., Whinston, A.B. 2004. Global software piracy revisited. Communications of the ACM, 47, 103–107.

Traphagan, M., Griffith, A.1998. Software piracy and global competitiveness: report on global software piracy. International Review of Law, Computers & Technology, 12, 431–451.

Van Kranenburg, H., Hogenbirk, A 2005. Multimedia, entertainment, and business software copyright piracy: a cross-national study. Journal of Media Economics, 18, 109–129.

Varian, H.R. 1998. Markets for information goods. Working Paper, University of California, Berkeley. Available through http://www.sims.berkeley.edu/~hal/Papers/japan/japan.pdf.

Watt, R. 2001. Prices in cultural markets: a theoretical introduction for beginners. Working paper. Available through http://www.serci.org/documents/pricing.pdf.

Witte, A.D., Tauchen, H. 1994. Work and crime: an exploration using panel data. Public Finance, 49, 155–167.

World Bank. 2003. World Development Indicators. Washington, DC.

Zak, P.J., Knack, S. 2001. Trust and growth. Economic Journal, 470, 295–321.

Table 1

Empirical results for OLS and 2SLS estimates using the full sample

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS	(11) OLS	(12) 2SLS	(13) OLS	(14) 2SLS
LGNP	0.273*	0.280*	0.301**	0.318*	0.298**	0.334*	0.278*	0.366*	0.302**	0.336*	0.299**	0.320*	0.290**	0.306*
	(1.97)	(1.72)	(2.10)	(1.91)	(2.07)	(1.94)	(1.88)	(1.88)	(2.08)	(1.96)	(2.09)	(1.91)	(2.05)	(1.84)
$(LGNP)^2$	-0.027***	-0.027**	-0.029***	-0.030***	-0.029***	-0.032***	-0.028***	-0.035***	-0.030***	-0.032**	-0.029***	-0.030***	-0.029***	-0.029***
	(2.96)	(2.55)	(3.10)	(2.78)	(3.08)	(2.81)	(2.83)	(2.71)	(3.08)	(2.84)	(3.09)	(2.79)	(3.06)	(2.72)
IDV	-0.005***	-0.006***	-0.005***	-0.006***	-0.005***	-0.006***	-0.005***	-0.006***	-0.005***	-0.006***	-0.005***	-0.006***	-0.005***	-0.006***
	(4.48)	(4.83)	(4.45)	(4.90)	(4.35)	(4.83)	(4.01)	(4.44)	(4.36)	(4.87)	(4.43)	(4.90)	(4.47)	(4.89)
Q1	2.474***	2.671**												
	(2.85)	(2.22)												
Q2 / Q3 / Q4			2.185***	2.588**	2.090***	3.177**	1.863	5.930**						
			(3.07)	(2.39)	(2.83)	(2.44)	(1.56)	(2.44)						
Q2-Q4									0.816***	1.185**				
									(2.91)	(2.50)				
Q5											-0.657***	-0.834**		
											(3.08)	(2.46)		
GINI													-0.006***	-0.007**
													(2.98)	(2.42)
Constant	3.701***	3.658***	3.556***	3.456***	3.498***	3.232***	3.472***	2.385**	3.415***	3.143***	4.109***	4.125***	4.057***	4.053***
	(6.86)	(5.73)	(6.30)	(5.22)	(6.08)	(4.54)	(5.32)	(2.47)	(5.78)	(4.36)	(8.01)	(6.47)	(8.00)	(6.42)
# obs.	67	66	67	66	67	66	67	66	67	66	67	66	67	66
R^2 (centered)	0.8262	0.8282	0.8237	0.8255	0.8178	0.8153	0.8066	0.7701	0.8190	0.8174	0.8224	0.8232	0.8237	0.8250
F(GNP vars)	42.37	29.43	47.98	32.62	48.78	31.35	37.22	24.10	48.14	31.82	48.06	32.59	48.81	32.65
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Partial R^2 (1st)		0.4461		0.4819		0.4175		0.2770		0.4466		0.4719		0.4680
F (1st)		24.17		27.90		21.51		11.49		24.21		26.81		26.39
(p-value)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Sargan over-id test chi^2 (p-value)		3.393 (0.066)		2.429 (0.119)		1.685 (0.194)		0.115 (0.734)		1.474 (0.225)		1.968 (0.161)		2.267 (0.132)

Note: Dependent variable: log of crime rate (1994-2002). Absolute value of t-statistics in parentheses; robust t-statistics in the OLS regressions. A Pagan Hall test does not reject homoscedastic errors in the 2SLS regressions. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. Lebanon, Russia, United Kingdom, and the Ukraine are removed from the analysis because of influential cases. Results are robust to outliers and influential observations. In the 2SLS, income inequality measures are instrumented with the average life expectancy and birth rate for 1970–95.

Table 2

Subsample of Latin America and the Caribbean region (2SLS estimates)

	(2.1)	(2.2)	(2.3)	(2.4)
LGNP	-0.032	0.155	0.104	0.039
	(0.03)	(0.18)	(0.12)	(0.04)
$(LGNP)^2$	-0.011	-0.025	-0.021	-0.016
	(0.14)	(0.42)	(0.35)	(0.25)
IDV	0.002	0.002	0.002	0.002
	(0.67)	(0.98)	(0.91)	(0.79)
Q1	-3.341			
	(0.37)			
Q2-Q4		0.769		
		(0.52)		
Q5			-0.494	
			(0.36)	
GINI				- 0.002
				(0.12)
Constant	5.173	4.092	4.852	4.879
	(1.20)	(1.23)	(1.64)	(1.53)
# observations	17	17	17	17
R^2 (centered)	0.4846	0.7558	0.7528	0.7176
F (GNP vars)	7.14	12.38	12.07	10.78
(p-value)	(0.009)	(0.001)	(0.001)	(0.002)
Partial R^2 (1st)	0.0895	0.1950	0.1533	0.1292
Sargan over-id test chi^2	0.328	0.707	0.895	0.929
(p-value)	(0.567)	(0.401)	(0.344)	(0.335)
F (1st)	0.54	1.33	1.00	0.82
(p-value)	(0.597)	(0.303)	(0.400)	(0.467)

Note: Dependent variable: log of piracy rate (94-02). Absolute value of t-statistics in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. Results are robust to outliers and influential observations. Income inequality measures are instrumented with the average life expectancy and birth rate for 1970–95. A Pagan Hall test does not reject homoscedastic errors.

Table 3

Subsample of Western Europe and Northern America (2SLS estimates)

	(3.1)	(3.2)	(3.3)	(3.4)
LGNP	-4.003	-1.412	-3.504	-4.487
	(1.41)	(0.24)	(1.31)	(1.77)
$(LGNP)^2$	0.195	0.058	0.169	0.219
	(1.32)	(0.19)	(1.21)	(1.66)
IDV	-0.005	-0.014	-0.007	-0.007
	(1.85)	(1.06)	(1.57)	(1.62)
Q1	-2.432			
	(0.58)			
Q2-Q4		-10.767		
		(0.83)		
Q5			2.727	
			(1.03)	
GINI				0.034
				(1.38)
Constant	24.627	18.620	21.185	25.834*
	(1.81)	(0.74)	(1.66)	(2.17)
# observations	13	13	13	13
R^2 (centered)	0.6469	-0.2592	0.6681	0.7180
F (GNP vars)	5.00	1.42	5.40	6.87
(p-value)	(0.039)	(0.298)	(0.033)	(0.018)
Partial R^2 (1st)	0.6929	0.0817	0.3449	0.1960
Sargan over-id test chi^2	5.446	0.573	4.658	4.424
(p-value)	(0.020)	(0.449)	(0.031)	(0.035)
F (1st)	7.90	0.31	1.84	0.85
(p-value)	(0.016)	(0.742)	(0.228)	(0.466)

Note: Dependent variable: log of piracy rate (94-02). Absolute value of t-statistics in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. Results are robust to outliers and influential observations. Income inequality measures are instrumented with the average life expectancy and birth rate for 1970–95. A Pagan Hall test does not reject homoscedastic errors.

Table 4

Subsample of Eastern Europe and Central Asia (2SLS estimates)

	(4.1)	(4.2)	(4.3)	(4.4)
LGNP	-6.300	-6.423	-6.413	-6.256
	(1.12)	(1.16)	(1.15)	(1.11)
$(LGNP)^2$	0.377	0.385	0.384	0.373
	(1.08)	(1.11)	(1.11)	(1.07)
IDV	-0.006	-0.006	-0.006	-0.006
	(0.65)	(0.73)	(0.71)	(0.71)
Q1	1.769			
	(0.34)			
Q2-Q4		1.141		
		(0.62)		
Q5			-0.751	
			(0.55)	
GINI				-0.006
				(0.49)
Constant	30.448	30.525	31.384	30.716
	(1.36)	(1.40)	(1.41)	(1.37)
# observations	10	10	10	10
R^2 (centered)	0.4397	0.4461	0.4440	0.4297
F (GNP vars)	1.20	1.57	1.49	1.35
(p-value)	(0.374)	(0.296)	(0.311)	(0.340)
Partial R^2 (1st)	0.9241	0.9352	0.9395	0.8988
Sargan over-id test chi^2	2.224	1.726	1.878	1.938
(p-value)	(0.136)	(0.189)	(0.171)	(0.164)
F (1st)	24.36	28.86	31.06	17.77
(p-value)	(0.006)	(0.004)	(0.004)	(0.010)

Note: Dependent variable: log of piracy rate (94-02). Absolute value of t-statistics in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. Results are robust to outliers and influential observations. Income inequality measures are instrumented with the average life expectancy and birth rate for 1970–95. A Pagan Hall test does not reject homoscedastic errors.

Table 5

Subsample of South Asia, East Asia and the Pacific region (2SLS estimates)

	(5.1)	(5.2)	(5.3)	(5.4)
LGNP	-0.122	0.324	0.147	-0.022
	(0.36)	(0.59)	(0.31)	(0.05)
$(LGNP)^2$	-0.002	-0.030	-0.019	-0.007
	(0.07)	(0.83)	(0.61)	(0.27)
IDV	-0.009***	-0.010***	-0.009***	-0.009***
	(8.88)	(4.52)	(6.08)	(8.18)
Q1	-2.430			
	(0.87)			
Q2-Q4		0.824		
		(0.40)		
Q5			-0.055	
			(0.05)	
GINI				0.003
				(0.40)
Constant	5.623***	3.436	4.478***	4.939***
	(3.91)	(1.21)	(3.68)	(3.92)
# observations	12	12	12	12
R^2 (centered)	0.9856	0.9736	0.9814	0.9846
F (GNP vars)	29.83	19.80	26.45	33.75
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
Partial R^2 (1st)	0.5605	0.1653	0.2168	0.3689
Sargan over-id test chi^2	4.004	2.609	4.098	4.690
(p-value)	(0.045)	(0.106)	(0.043)	(0.030)
F (1st)	3.83	0.59	0.83	1.75
(p-value)	(0.09)	(0.582)	(0.481)	(0.251)

Note: Dependent variable: log of piracy rate (94-02). Absolute value of t-statistics in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. Results are robust to outliers and influential observations. Income inequality measures are instrumented with are the average life expectancy and birth rate for 1970–95. A Pagan Hall test does not reject the homoscedastic errors.

Table 6

Empirical results for 2SLS estimates institutional determinants included†

	(6.1)	(6.2)	(6.3)	(6.4)	(6.5)	(6.6)	(6.7)	(6.8)	(6.9)	(6.10)	(6.11)	(6.12)
LGNP	0.537***	0.600***	0.584***	0.550***	0.399**	0.473***	0.452**	0.415**	0.447**	0.510***	0.493***	0.458**
	(2.89)	(2.96)	(2.96)	(2.88)	(2.09)	(2.45)	(2.39)	(2.22)	(2.45)	(2.78)	(2.72)	(2.54)
$(LGNP)^2$	-0.042***	-0.047***	-0.046***	-0.044***	-0.030**	-0.036***	-0.034***	-0.032**	-0.031**	-0.036***	-0.035***	-0.033***
	(3.49)	(3.51)	(3.53)	(3.48)	(2.41)	(2.84)	(2.76)	(2.62)	(2.62)	(3.02)	(2.95)	(2.80)
IDV	-0.004**	-0.004**	-0.004**	-0.004**	-0.005***	-0.005***	-0.005***	-0.005***	-0.004**	-0.004***	-0.004***	-0.004***
	(2.51)	(2.47)	(2.51)	(2.53)	(3.37)	(3.51)	(3.53)	(3.50)	(2.66)	(2.78)	(2.79)	(2.76)
Q1	3.400**				4.424***				3.932***			
	(2.21)				(3.02)				(2.77)			
Q2-Q4		1.228***				1.756***				1.549***		
		(2.69)				(3.27)				(2.98)		
Q5			-0.913**				-1.268***				-1.121***	
			(2.57)				(3.26)				(2.97)	
GINI				-0.008**				-0.010***				-0.009***
				(2.43)				(3.22)				(2.92)
PRIGHT	-0.084***	-0.084***	-0.083***	-0.080***					-0.079**	-0.077**	-0.077**	-0.074**
	(3.76)	(3.67)	(3.72)	(3.41)					(2.13)	(2.08)	(2.11)	(2.00)
LAW					-0.120***	-0.144***	-0.138***	-0.126***	-0.119***	-0.140***	-0.134***	-0.125***
					(3.14)	(3.72)	(3.62)	(3.37)	(3.28)	(3.81)	(3.72)	(3.48)
Constant	2.704***	2.185**	3.228***	3.229***	2.861***	2.106**	3.581***	3.553***	2.771***	2.112**	3.412***	3.395***
	(3.58)	(2.50)	(4.84)	(4.80)	(3.87)	(2.62)	(4.89)	(4.89)	(3.94)	(2.77)	(4.88)	(4.86)
# observations	63	63	63	63	63	63	63	63	63	63	63	63
R^2 (centered)	0.7886	0.7832	0.7872	0.7909	0.7931	0.7931	0.7981	0.8006	0.8164	0.8176	0.8212	0.8212
Partial R^2 (1st)	0.4006	0.4081	0.4296	0.4311	0.4144	0.4098	0.4348	0.4454	0.4050	0.3989	0.4241	0.4326
Sargan over-id test chi^2 /												
Hansen J statistic chi^2	0.154	0.041	0.004	0.014	2.782	1.046	1.477	1.902	1.813	0.541	0.835	1.173
(p-value)	(0.695)	(0.841)	(0.951)	(0.904)	(0.095)	(0.306)	(0.224)	(0.168)	(0.178)	(0.462)	(0.361)	(0.279)
Pagan Hall test chi^2	Robust	Robust	Robust	Robust	8.885	9.468	9.739	10.100	9.082	10.475	10.520	11.015
(p-value)	std. errors	std. errors	std. errors	std. errors	(0.1802)	(0.149)	(0.136)	(0.121)	(0.2468)	(0.163)	(0.161)	(0.138)
F (1st)	13.15	17.30	17.00	16.74	19.81	19.44	21.54	22.49	18.72	18.25	20.25	20.97
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: †Dependent variable: log of piracy rate (94–02). Absolute value of t-statistics in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. Results are robust to outliers or influential observations. Income inequality measures are instrumented with the average life expectancy and birth rate for 1970–95.

Table 7

Sensitivity tests for the full sample (2SLS estimates)†

Additional variables(s)	Income Share Q1 (1)	Income Share Q2-Q4 (2)	Income Share Q5 (3)	GINI (4)
Legal origin (German, French, English, Socialist)	4.252	3.019*	-2.237	-0.019
	(0.90)	(1.74)	(1.64)	(1.54)
	n= 63	*n= 63*	*n= 63*	*n= 63*
Secondary and Tertiary school enrollment (%)	4.290**	1.746**	-1.252**	-0.010**
	(2.51)	(2.64)	(2.64)	(2.59)
	n= 63	*n= 63*	*n= 63*	*n= 63*
PCs (%)	3.819**	1.438***	-1.054***	-0.009***
	(2.64)	(2.90)	(2.87)	(2.81)
	n= 61	*n= 61*	*n= 61*	*n= 61*
Internet users (%)	3.857***	1.480***	-1.079***	-0.009**
	(2.66)	(2.91)	(2.88)	(2.83)
	n= 63	*n= 63*	*n= 63*	*n= 63*
Male youth Unemployment (%)	3.675***	1.479***	-1.066***	-0.009***
	(2.71)	(2.79)	(2.80)	(2.80)
	n= 56	*n= 56*	*n= 56*	*n= 56*
Total unemployment (%)	3.953**	1.523***	-1.109***	-0.010***
	(2.64)	(2.83)	(2.82)	(2.76)
	n= 61	*n= 61*	*n= 61*	*n= 61*
Openness (%)	4.954***	1.975***	-1.424***	-0.012***
	(3.56)	(3.75)	(3.78)	(3.82)
	n= 63	*n= 63*	*n= 63*	*n= 63*
Research and Development expenditure (% GDP)	4.411***	1.701***	-1.235***	-0.010***
	(2.86)	(2.91)	(2.94)	(2.95)
	n= 55	*n= 55*	*n= 55*	*n= 55*
Information and technology expenditure (% GDP)	-2.056	0.806	-0.589	-0.005
	(1.19)	(1.45)	(1.39)	(1.32)
	n= 43	*n= 43*	*n= 43*	*n= 43*
Urbanization (%)	3.824***	1.554***	-1.118***	-0.009***
	(2.69)	(2.91)	(2.89)	(2.84)
	n= 63	*n= 63*	*n= 63*	*n= 63*
% Muslim, Catholic or Protestant	3.280**	1.311**	-0.946**	-0.008**
	(2.34)	(2.64)	(2.59)	(2.54)
	n= 63	*n= 63*	*n= 63*	*n= 63*
Ethnolinguistic fractionalization	3.741***	1.496***	-1.080***	-0.009***
	(2.75)	(3.09)	(3.06)	(2.96)
	n= 58	*n=58*	*n=58*	*n= 58*
Share of young male, aged 15-19, 20-29, 30-39 years	3.563	2.071	-1.476	-0.011
	(0.78)	(1.29)	(1.21)	(1.13)
	n= 63	*n= 63*	*n= 63*	*n= 63*

Note: †Dependent variable: log of piracy rates (94–02) Specifications (1), (2), (3), and (4) include the *LGNP* and $(LGNP)^2$ for 1970-1995, individualism (*IDV*), institutional quality (*LAW*), and patent rights index (*PRIGHT*) and a constant (not shown). Absolute values of t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The number of observations is in italics. Results are robust to outliers and influential observations. Income inequality measures are instrumented with the average life expectancy and birth rate for 1970–95. A Pagan Hall test does not reject homoscedastic errors.

Appendix tables

Table A1

Average piracy rates by country, 1994–2002

Country	Piracy rate (%)	Country	Piracy rate (%)	Country	Piracy rate (%)
Argentina	66.89	Hungary	58.78	Philippines	77.67
Australia	32.55	India	70.44	Poland	62.89
Austria	39.00	Indonesia	92.00	Portugal	49.67
Bolivia	85.11	Ireland	56.89	Puerto Rico	52.67
Brazil	63.22	Israel	54.00	Romania	82.78
Bulgaria	85.55	Italy	50.55	Russia	90.44
Canada	40.78	Japan	40.55	Singapore	53.44
Chile	56.78	Jordan	77.11	Slovakia	52.67
China	94.33	Kenya	73.67	Slovenia	75.78
Colombia	60.89	Latvia	58.50	South Africa	48.00
Costa Rica	74.44	Lebanon	86.89	Spain	59.11
Czech. Rep	49.55	Malaysia	73.00	Sweden	40.67
Denmark	33.11	Mauritius	76.11	Switzerland	36.89
Dominican Republic	74.67	Mexico	62.44	Taiwan	59.22
Ecuador	73.67	Morocco	69.78	Thailand	81.00
Egypt	74.11	Netherlands	48.00	Turkey	76.55
El Salvador	85.00	New Zealand	32.55	UK	30.78
Finland	36.11	Nicaragua	83.55	Ukraine	91.44
France	44.89	Nigeria	72.66	Uruguay	72.67
Germany	34.22	Norway	42.78	US	25.89
Greece	73.55	Pakistan	86.89	Venezuela	63.00
Guatemala	82.11	Panama	68.55	Vietnam	97.44
Honduras	76.89	Paraguay	83.67	Zimbabwe	65.00
Hong Kong	59.55	Peru	68.67		

Note: This table lists only those countries with no missing observations for estimating the baseline model.

Table A2

Summary statistics of variables

Variable	Obs	Mean	Std. Dev	Min	Max
Baseline model					
LPR	71	4.11	0.319	3.25	4.58
GINI	71	39.43	10.08	20.55	60.8
LGNP	71	7.83	1.24	5.31	10.52
IDV	71	41.52	22.97	6	91
Q1	71	0.062	0.022	0.025	0.113
Q2-Q4	71	0.476	0.068	0.328	0.564
Q5	71	0.461	0.086	0.329	0.648
Institutional measures					
LAW	70	0.51	0.94	-1.14	2.05
PRIGHT	66	3.00	0.86	0.92	4.86
Variables for robustness check					
Legal origin (British)	70	0.24	0.43	0	1
Legal origin (French)	70	0.46	0.50	0	1
Legal origin (Socialist)	70	0.17	0.38	0	1
Legal origin (German)	70	0.07	0.26	0	1
Male youth unemployment rate	61	15.99	9.26	4.27	51.45
Total unemployment rate	67	8.56	4.23	2.29	20.82
Secondary school enrollment rate	70	74.72	28.04	24.25	127.83
Tertiary school enrollment rate	70	28.46	17.45	1.63	82.73
% Urbanization	70	62.67	19.62	18.87	100
Information and technology expenditures (% GDP)	47	4.96	1.97	1.25	9.21
% Computers	68	1.02	1.69	0.00013	6.50
% Internet users	70	2.97	3.78	0.031	16.57
Openness (% GDP)	70	71.52	48.99	17.33	328.58
Ethnolinguistic fractionalization	65	24.29	24.42	0.25	85,67
Research and Development expenditure (% GDP)	59	1.04	0.87	0.09	3.44
% male 15–19 years	70	4.56	0.88	3.13	6.06
% male 20–29 years	70	8.25	0.79	6.91	9.84
% male 30–39 years	70	7.03	0.95	4.96	9.93
% Protestant	71	12.43	22.55	0	97.8
% Catholic	71	44.03	40.06	0	96.9
% Muslim	71	10.69	25.41	0	99.4
Instrumental variables					
Life expectancy	70	67.72	7.12	46.55	76.77
Birth rate	70	24.40	10.80	10.76	47.35

Table A3

Average piracy rate by region (1994–2002)

Region	Mean
Western Europe	39.89
Eastern Europe	74.67
North America	26.89
Latin America and Caribbean	64.22
Asia and Pacific	55.00
Middle East	67.40
Africa	61.00

Note: Regions are as defined by the World Bank (2003).

Table A4

Correlation matrix for key variables (N= 66)

	LGNP	IDV	LAW	PRIGHT	Q1	Q2-Q4	Q5	GINI
LGNP	1							
IDV	0.7306	1						
LAW	0.8607	0.6675	1					
PRIGHT	0.7916	0.7049	0.7072	1				
Q1	0.1790	0.3510	0.1960	0.1462	1			
Q2-Q4	0.5279	0.6059	0.5454	0.4413	0.8011	1		
Q5	-0.4618	-0.5722	-0.6283	-0.3851	-0.8808	-0.9890	1	
GINI	-0.4276	-0.5506	-0.4274	-0.3466	-0.9142	-0.9548	0.9812	1

Table A5

Definitions of variables and sources

Variable	Time span	Source
Software piracy rate (%)	1994–2002	International Planning Research Corporation (IPRC, 2003)
Gross national product (per capita in current US$)	1970–1995	La Porta et al. (1999)
Gini index, and income shares (quintiles)	1960–1995	Dollar and Kraay (2002)
Individualism index	1967–1973	Hofstede (1997, 2000)
Patent rights index	1995	Ginarte and Park (1997)
Rule of law	1996	Kaufman, Kraay, and Mastruzzi (2004) Available through www.worldbank.org/ wbi/governance/govdata2002
Secondary school enrollment, gross (%)	1985–1999	World Development Indicators database (World Bank, 2003)
Tertiary school enrollment, gross (%)	1985–1999	World Development Indicators database (World Bank, 2003)
Urbanization (the share of the total population living in urban areas)	1985–1999	World Development Indicators database (World Bank, 2003)
Male youth unemployment rate (% of male labor force between 15 and 24 years)	985–1999	World Development Indicators database (World Bank, 2003)
Unemployment rate (% of total labor force)	1985–1999	World Development Indicators database (World Bank, 2003)
Openness (ratio of imports plus exports to GDP)	1985–1999	World Development Indicators database (World Bank, 2003)
Internet users (the share of people with access to the worldwide network in total population)	1985–1999	World Development Indicators database (World Bank, 2003)
Personal computers (the share of computers in total population)	1985–1999	World Development Indicators database (World Bank, 2003)
Legal origin (British, Socialist, French, German, and Scandinavian)	1960	La Porta et al. (1999)
Ethnolinguistic fractionalization (in %)	1960	La Porta et al. (1999)
Information and technology expenditures (information and technology expenditures as a percentage of GDP)	1985–1999	World Development Indicators database (World Bank, 2003)
Research and development expenditures (research and development expenditures as a percentage of GDP)	1985–1999	World Development Indicators database (World Bank, 2003)
Share of male 15–19 years (% of population)	1985–2000	Available through http://esa.un.org/unpp/
Share of male 20–29 years (% of population)	1985–2000	Available through http://esa.un.org/unpp/
Share of male 30–39 years (% of population)	1985–2000	Available through http://esa.un.org/unpp/
Share of Protestants, Catholics, Muslims (% of population)	1980	La Porta et al. (1999)
Life expectancy (the number of years that a newborn infant is expected to live at birth)	1970–1995	World Development Indicators database (World Bank, 2003)
Birth rate (the number live births occurring during the year, per 1,000 population)	1970–1995	World Development Indicators database (World Bank, 2003)

Notes

1. Varian (1998) defines an information good as anything that can be digitized; therefore, books, records, magazines, and software fall into this category.
2. However, the latter number should be treated with caution as it is based on the assumption that each pirated copy transaction represents the loss of a legitimate sale. This argument is incorrect because pirated copies are always transacted at a lower price than legitimate copies.
3. The answer to this question determines which proxy variable would be the most proper measure for the middle class (the Gini index, the income share of the top X percent, or other distribution measures).
4. For reviews of the literature on crime and inequality, see Bourguignon (1998), Eide (2000), and Freeman (1999).
5. Experiments have shown that reputational costs ('shaming') have the strongest piracy deterring impact, while formal sanctions decrease acceptability of crime to much lower extent, and the knowledge of the existence of some dubious copyright law alone does not prevent people from committing this felony at all (Feldman and Nadler, 2005). An alternative explanation could be that members of the higher income class have a more diversified demand for differentiated software products, which leads to higher search costs of illegal copies.
6. If income is equally distributed, x percent of the income earners in society possess x percent of the total income earned in the economy. In such a case, the Lorenz curve has a diagonal shape.
7. However, as the absolute income level of an income class changes in the national income of a country, these general predictions might not hold for rich and poor countries likewise, particularly because relative prices for tradable goods do not vary sufficiently in national GDP levels due to potential cross-country arbitrage possibilities (re-imports etc.). Based on these arguments, both software and hardware will be more expensive in purchasing power parity terms in poorer countries than in wealthier countries.
8. In case of a linear relation the coefficient on the quadratic term should be rendered insignificant.
9. Hofstede (1983, p. 336) defines individualism as “a preference for a loosely knit social framework… in which individuals are supposed to take care of themselves and their immediate families only”. Collectivism, in turn, is a "preference for tightly knit social framework in which individuals can expect their relatives, clan, or other in-group to look after them in exchange for unquestioning loyalty”.
10. Measurement error in the dependent variable typically causes inefficiency in the regression analysis. So measurement error in the piracy rate variable makes the standard errors in the coefficients on the explanatory variables large and loose statistical significance. If the measurement error in the dependent variables is systematically related to one or more of the explanatory variables, OLS estimates will be biased. Taking the natural log of the dependent variable, however, lets the bias move into the error term if measurement error is systematic and persistent.
11. A similar argument applies for the rule of law variable, which is also measured at the beginning of the observation period.
12. For instance, Husted (2000) includes only 30 countries.
13. Even though piracy rate data are available for more than 70 countries for this period, we only consider observations from countries for which individual piracy rates are available. Thus, we net out the merged observations from Belgium and Luxembourg.
14. The Gini index ranges from zero (perfect equality) to one (perfect inequality).

15. The advantage of the rule of law measure used in this study is that it is available for a larger sample of countries than are other potential measures of institutional quality like the risk of expropriation data compiled by Political Risk Services (Acemoglu et al., 2001)
16. This measure is obtained as the unweighted sum of five different categories of patent laws: (i) extent of coverage, (ii) provisions for loss of protection, (iii) membership in international patent treaties, (iv) an enforcement mechanism, and (v) duration of protection. Each of these categories is coded between zero and one.
17. In such cases, a Huber/White/sandwich variance estimator is applied.
18. Maskus (2000) also finds a quadratic relationship between the level of economic development proxied by the GNP per capita and the strength of property rights proxied by the patent rights index.
19. One possible explanation is that the first quintile already includes too many households that are wealthy enough to afford a PC, particularly because averages over a long time span are used. A more accurate measure might be the first decile. Estimation of the same equation using only data for 1995 renders an insignificant coefficient for the first quintile, which is in line with our prediction.
20. Husted's (2000) and Rodriguez's (2005a) estimated coefficients for GINI are around 0.006.
21. Using averages over 5 years or even values of only a single year an instrumentation of the inequality measures leads to an even more sizeable differences between its 2SLS and OLS estimates. This outcome shows that using averages over a long time span, here from 1960 until 1999, may mitigate the (potential) measurement error.
22. Inclusion of regional dummies in the full sample removes all variation in the GNP and inequality variables and was therefore not considered useful by the authors.
23. This region comprises the World Bank regions of South Asia (SA) and East Asia and the Pacific (EAP). The majority of the countries in this sample belongs to the second region and only two nations to the first.
24. Computer programs could be protected against piracy by copyrights or patents. Copyright law might be the more important source of protection because patent law usually applies only when the software is particularly designed to run a patented machine.
25. Personal Computers are defined as self-contained computers designed to be used b a single individual only. Internet users are people with access to the worldwide network. Definitions and data are from the International Telecommunication Union, World Telecommunication Development Report and database.
26. For the impact of unemployment on crime, see Witte and Tauchen (1994) or Doyle et al. (1999); see Cohen and Land (1987) for which gender and age groups are prone to commit property crime, and see Glaeser and Sacerdote (1999) for why more crime occurs in agglomerations.
27. Complete estimation results are not reported, but available upon request.
28. An investigation into the data has not brought about strong evidence for a multicollinearity between the explanatory variables. Simple correlation coefficients between the other regressors and the openness measure never exceed 0.25, and regressing openness on the remaining predictors of piracy leads to an adjusted R^2 of just 0.22. Highly significant coefficients are the ones on *IDV* and *LAW* (at the 1 percent levels).
29. The insignificance observed for the share of computers or Internet users in the population might well be a result of a potential endogeneity (Osorio, 2002): It is very likely that at least the promulgation of PCs is caused by greater availability of pirated software.