CARLOS A. OSORIO[*]

A Contribution to the Understanding of Illegal Copying of Software:

Empirical and analytical evidence against conventional wisdom

Abstract

This paper analyzes the causes of illegal copying and its effects in the software market across 66 countries. By studying the aggregated and joint effects of different variables, the analysis shows that supply constraints in the software market, shortage of after-sale support, and characteristics legal framework are major drivers of illegal copying when controlling for income per capita. It also concludes there is enough evidence to show there is a threshold of illegal copying over which its aggregated effect on the software market is positive, and this is an efficient mechanism for market creation. Thus, allowing illegal copying in some countries and at certain periods of time may be a profit generating decision in the long-term, especially in countries with low-developed software markets and with presence of Open Source software. For other hand, the results provide evidence to understand why proprietary software companies would prefer enforcing their copyrights and intellectual property rights contingently and as result of a rational decision-making process.

JEL Codes: L86, O34, N40

Keywords: illegal copying, piracy, software, network effects, copyrights, intellectual property

I. Introduction

The electronically networked environment creates a scenario of high vulnerability for protecting intellectual property rights and copyright of digital goods. In the case of software, conventional wisdom suggests that less expensive and faster copying devices would foster illegal copying, damaging the software industry, affecting the size of its market, and diminishing companies' incentives to develop new and better products. Empirical and analytical evidence, however, contradicts this belief.

After reviewing the literature about causes and effects of illegal copying of software, this paper studies the evolution of illegal copying of software for 66 countries between 1995 and 2000, and examines three questions. First, it studies different causes of illegal copying. Second, it presents an analysis of the importance of the illegal user base in the interaction between closed and open source software companies. Finally, it examines the effects of illegal copying in the hindering and fostering supply and demand for software.

Carlos A. Osorio is Research Associate at the Information Technologies Group at the Center for International Development at Harvard University, and Visiting Research Scientist at the Media Laboratory at the Massachusetts Institute of Technology. e-mail: carlos_osorio@post.harvard.edu. Copyright © 2002 by Carlos A. Osorio.

Finally, it presents the conclusions and policy implications for each section, and a discussion on future directions for research.

II. Economics of Digital Goods and Illegal Copying of Software

Computer software has the characteristic of a non-rival and quasi non-excludable good. One may prohibit a third person from using it only by not letting him (or her) to access a version of the software. Once access is granted, however, the software can be copied at almost zero cost. By doing so, new users cannot exclude the earlier from using the software -as with a bicycle or a jacket- and, by direct and indirect network effects, the new user adds value to the whole network of users (legal and illegal)[i].

What are the effects of illegal copying of software, commonly known as "piracy", on the overall software market? Why do some software companies enforce their intellectual property rights differently across countries? These are some of many questions surrounding a very difficult and politically charged subject where, sometimes, maintaining a public position is privileged over strategy and available evidence.

This paper does not pretend to address all the related subjects, but provide evidence for future richer analyses. It identifies the major determinants in generating software illegal copying, and the way in which it affects the creation of the software market. Regardless of its ethical value or moral justification, illegal copying of software violates copyright and the legislation in a large number of countries. The existence of this public and business policy problem has several interesting implications ranging from how to protect copyright, to the ways in which such protection would affect the size and value of the market, and different ways of creating legal constraints for software copying, and companies' role on this process.

Novos and Waldman (1984) first analyzed the effects of copyright protection of software -as partially non-excludable goods- as way to minimize social welfare loss generated by illegal copying. One of their findings points that increasing copyright protection would decrease social welfare loss due to underproduction and underutilization of the software. In their approach, however, they assumed the decision to be an illegal user is function of income, and cost of copying relative to the software's license. Their analysis, however, has two shortcomings: it does not consider the dynamics generated by network effects, and it does not account for the fact that, while cost of copying has been decreasing over time, illegal copying has also decreased.

Network effects are important because, in terms of the total user base, the illegal users of software add value to all the users, legal and illegal, and act as agents in fostering the software's diffusion process by word-of-mouth. By this way, they indirectly generate additional positive effect for the software company. Givon, Mahajan and Muller (1997), analyze this dynamic by studying differential illegal copying and brand switching comparing market estimates based on units' sales and user base. Their work analyzes the dynamics between brands using the Bass Model, and concluding that a brand capable of building the larger user base –not necessarily the best product- has a better chance to succeed in the long run. Their analysis, however, considered the rate of illegal copying as constant over time.

From a different perspective, Gayer and Shy (2001a) proposed a model to analyze the loss of sales associated with illegal copying, suggesting that software companies tend to overestimate the losses. This overestimation, they argue, results from assuming that all "pirated" software would have been legal sales if illegal copying becomes impossible. Building over their results, the relevance of supply constraints are underestimated in their effect on the availability of complementary goods or services by overstating the relevance of the intrinsic value of a software. For consumers in need of after-sale support or value-added services, this effect can be as powerful as the availability of the software itself.

Additionally, Chen and Png (2001) suggest that software producers are likely to manage illegal copying more efficiently through price cuts than by enforcement, a line of reasoning close to Varian (2001). Al-Jabri and Adbul-Gader (1997) provided a good example of how illegal copying of software can be affected by peer beliefs about the justifications for copying. Generalizing their work, one would understand illegal copying of software as being constrained or encouraged by social justifications, paradigms, or special circumstances. These findings are related with understanding software as quasi non-excludable goods and that excludability is, as Varian (1998) suggests, to a certain extent function of the legal regime.

Finally, corporate strategy may play a role in both fostering and fighting illegal copying. First, as suggested by Katz and Shapiro (1986), market creation needs a company that is willing to serve as "sponsor" of its product by investing in marketing, after-sale support and service. These activities require local presence regardless of having critical mass. DeLong and Froomkin (2000) present how transparency in the markets is used as market creation strategy by offering trial periods. The "try-before-you-buy" approach is a period in which the user has permission to use software free of charge, after which he must pay a license and register the copy in order to continue doing so. The period is calculated to be large enough to generate lock-in, short enough to give a sense of what the product is really about, without giving too much value. Some people may argue the conceptual difference between these different approaches and illegal copying is subtle (Obviously the company's permission to use software for free makes all the difference).

III. Empirical Evidence on the Causes of Illegal Copying of Software

Summarizing from the previous section, illegal copying is commonly assumed to be function of the software's price, potential market's average income per capita, and marginal cost of copying versus producing the software[ii]. Varian (1998, and 2001) considers excludability, and Al-Jabri and Abdul-Gader (1997) a possible social justification resulting from inadequate supply conditions and cultural characteristics. Additionally, direct and indirect network effects play an important role in explaining the importance of illegal users in the diffusion process, as well as the relationship between software and computer.

After reviewing the literature one may suggest that illegal copying could be caused by different reasons. First, it may be a social response to inadequate supply conditions (i.e. the extent to which software fits local needs -language, metaphor and sophistication-, availability of software after-sale support and complementary products and services). Second, illegal copying can also be explained by social beliefs and rules that, to some extent, justify duplicating an original version of software. These beliefs and rules would be made operative through each community's legal framework structuring, thus, the grounds for allowing different levels of excludability. This would be the case, for instance, in communities with less notion of private property. Third, illegal copying could also result from and affordability problem: some people make illegal copies when the cost of copying is inferior of the software's license. Some of the foundations for the previous hypothesis are found in the work of Varian (2001), Al-Jabri and Abdul-Gader (1997), Katz and Shapiro (1986).

If the previous hypotheses hold, the results would support the argument of Gayer and Shy (2001) that, assuming that all illegal users would be legal if total enforcement is possible, closed source companies overestimate their losses due to illegal copying. Yet more, it would give evidence that software companies might have a direct and indirect role in helping the generation of illegal copying in underdeveloped markets, and incentives for doing so. In terms of business strategies, for instance, some ways of doing it are by undersupplying system compatibility, generating lock-in for users of their product, and by lack of "sponsorship" in some markets [Katz and Shapiro (1986)].

The previous arguments generate four basic hypotheses tat will be tested to examine if each one holds by itself, and they hold together.

The dependent variable is the rate of illegal copying in the software market, which is defined using data from the Business Software Alliance (BSA 2001). BSA defines it as "retail software revenue lost to piracy". By product aggregation, however, one can define a representative piece of software, and assume a one-to-one relationship between a piece of software and a computer[iii] [Mas-Collel, Whinston, and Green (1995)]. Some people may argue that BSA's data is inaccurate and overstate the rate of illegal copying. It is, however, the voice of the industry in the issue, and the most respected source for this estimate.

The first hypothesis, H1, supports that illegal copying or software results from the cost of copying relative to the license's fee, which is defined here as an affordability or income problem. Thus, countries with lower income per capita, would have higher levels of illegal copying. Income per capita in Purchasing Power Parity is used as a proxy variable for testing H1.

Figure 1: Average Illegal Copying rate by countries grouped by legal framework 2000
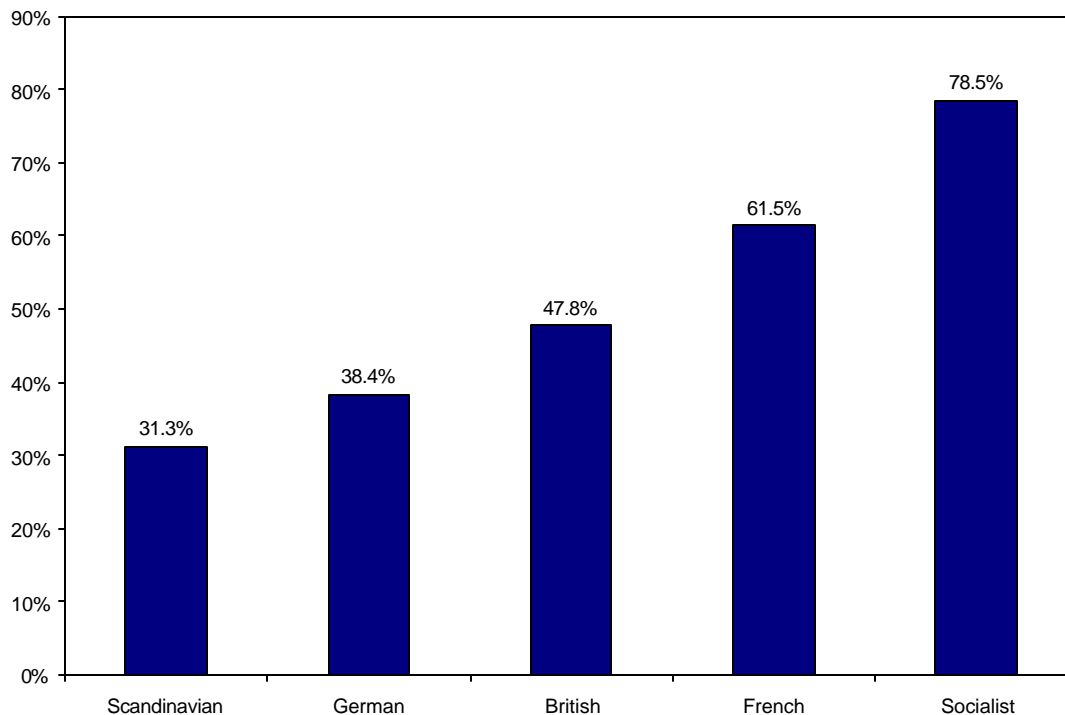
Source: Author's development using BSA (2001), Easterly and Sewadeh (2001), and complemented by author's research.

The second, H2, maintain that illegal copying can respond to social characteristics and justifications towards, or against, protecting intellectual property rights. Assuming each community's legal frameworks capture social embeddedness, a dummy variable of legal frameworks are, thus, included to account for their effect as coding of social beliefs, conducts, and representation of different degrees of rigorousness in allowing illegal copying, protecting copyrights, and explaining excludability. In order to test H2, the analysis includes two options. First, dummies variables according to the legal framework of each country. Second, a new variable –Legal- is defined based on the stringency of each legal system in allowing Illegal

copying as illustrated in Figure 1. The variable takes values 1, 2, 3, 4 and 5 for Scandinavian, German, British, French, and Socialist legal systems respectively. The source of data is Easterly and Sewadeh (2001), completed with research of the author.

The third hypothesis, H3, makes the case that illegal copying of software can be caused by software that fails to fit local needs. The variable used for testing H3 was taken from the Global Information Technology Report, responded by 4601 CEOs from 75 countries. The question asked to grade from 1 to 7 the statement "Software products sold in your country", where 1=need to be highly modified to fit local needs, and 7=fit local needs[iv].

The fourth hypothesis, H4, claim that countries with higher levels of corporate presence, after-sale support and complementary services in the software industry will have lower levels of illegal copying. Two variables from the Global Information Technology Report can be used to test H4. First, a question asking for grading between 1 and 7 regarding "Specialized information technology services are" 1: Not available in the country, 7: available from world-class local institutions". The second variable asks "How many local software and software services companies are competing in domestic markets?" The responses range from 1=none, to 7=a large number, the domestic market is competitive.

Table 1 shows the final results of regressing legal framework, software fitting to local needs, and availability of after-sale service and support against illegal copying, controlling for the effect of income per capita. In the first four regressions, all coefficients are statistically significant at 5% level. Independent and joint hypotheses testing support that the coefficients for the variables associated with H1, H2, H3 and H4 are different from zero. Standard errors are robust to arbitrary heteroskedasticity.

Table 1: Dependent variable: Illegal copying

ln(ill_copying)	Regression 1	Regression 2	Regression 3	Regression 4	Regression 5	Regression 6
ln(GDPpercapita)	-.1067242 (.0439764)	-.1139176 (.0469067)	-.1109888 (.0436576)	-.1242429 (.0439508)	-.1298821 (.048604)	-.1186449 (.0467395)
Legal	.0875975 (.0230182)	..	.0839777 (.0231634)	.1065903 (.0230416)	..	..
leg_brit	..	-.2353027 (.0835377)	..	..	-.2765267 (.0850853)	-.2278336 (.0852571)
leg_fren	..	-.1366632 (.0596089)	..	..	-.1662173 (.0603671)	-.1253074 (.0614664)
leg_soci	..	(dropped)	..	..	(dropped)	(dropped)
leg_germ	..	-.2706424 (.0849913)	..	..	-.3401392 (.0880929)	-.2554127 (.08613)
leg_scan	..	-.3744794 (.1048897)	..	..	-.4491515 (.1113392)	-.3539794 (.1023022)
soft_fitting	-.1009226 (.0417895)	-.0918603 (.0454799)	-.1079618 (.0415013)	-.1238218 (.0443512)	-.1165095 (.0468408)	-.0975915 (.0443225)
IT_services	-.1402486 (.0361633)	-.1398878 (.039278)	-.2031711 (.0506435)	..	..	-.2019932 (.0509779)
Soft_Comp	..	..	.0851195* (.0547561)	-.092331 (.0415828)	-.0932881 (.0441602)	.0841042* (.0566654)
Cons	1.232618 (.3105402)	1.728726 (.2574875)	1.200009 (.3206821)	1.220608 (.3029678)	1.819949 (.2489386)	1.67213 (.2606675)
Observations	66	66	66	66	66	66
R-Squared	0.8108	0.8143	0.8149	0.7869	0.7910	0.8183

Source: The author. All coefficients, but those marked * in reg. 3 and 6, are statistically significant at a 5% confidence level. Robust standard errors are in parenthesis. Please see Exhibit 1 for a key of the variables.

All results in Table 1 point to the same direction, with small variations in the value of the coefficients resulting from the different variables chosen. Regressions 3 and 6 help identify possible multicollinearity between "Availability of Local IT services" (IT_services) and "Competition in the Local Software Market" (Soft_Comp), and find the best proxy to test H4. In both cases, the null hypothesis for Soft_Comp can be rejected at 15% level, but not at 5% level. From the results there is multicollinearity, and Soft_Comp is dropped from the final equation. Thus, regressions 1 and 2 will be used in the analysis, although regressions 4 and 5 also show statistically significant coefficients.

In terms of the validity of H1, the regressions in Table 1 continue to show the most basic relationship between income level and illegal copying: the lower the income, the higher the illegal copying. There is a negative statistically significant relationship between income per capita and illegal copying of software. The statistical significance of ln(GDPpercapita) support the assumption that illegal copying is generated by an affordability problem, and it is used here as a control variable. As previously stated, the idea is to analyze to what extent there are other variables that, considering the effect of income, also help generating illegal copying.

For H2 there is little difference between in the results from using the variable "Legal" or the dummy variables for legal system (leg_brit, leg_fren, leg_soci, leg_germ, and leg_scan). In both cases, all coefficients are statistically significant at a 5% level. Hypothesis testing shows the individual null hypothesis can be rejected, as well as the joint hypothesis. Regression 2 shows the difference in the effect of legal frameworks when controlling for all the other variables, matching the evidence from Figure 1.

H3 proposes that illegal copying of software can also be caused when software fails to fit local needs, which is tested by including soft_fitting into the equations. Regression results show that soft_fitting is statistically significant when controlling for income per capita, as well as when controlling for all the other variables. The concept of "fitting local needs" include the metaphors and paradigms under which the software is designed, its appropriateness to the local culture, needs, and more basic issues such as language. In the early stages of the development of a software market, for instance, most of the available software is in English, which is likely to affect potential users' willingness to pay for a license, in the same way it would affect yours if your software were in Japanese, and you don't read Japanese.

Finally, H4 shows that illegal copying is also likely to be created as a social response to inadequate supply conditions, which was considered in the regressions by including and testing the variables IT_Services and Soft_Comp. After testing, IT_Services is used as proxy for local presence of software companies, related IT services, and competition in the local software market. Regression results show statistically significant coefficients for the proxy variable, when controlling by income per capita as well as when doing it by the other variables. The effect of local availability of IT services and competition in the local software market can be explained by their impact on price, geographical availability, and after-sale support, technical assistance, and complementary software and services.

In terms of pricing, more competition lead to lower prices, and lower prices would lead to a lower rate of illegal copying. Geographical availability is likely to affect illegal copying in a similar way, by lowering the cost of having a copy. After-sale support, technical assistance, and complementary software and services are likely to increasing the value of being a legal user. Additionally, competition in the software market generates versioning, a corporate strategy used to renew software markets, which fight illegal copying by periodically renewing the market with versions that better fit local needs, by self-obsolescence, and undersupplying compatibility between versions[v].

Thus, there is evidence that –controlling for the effect of income per capita- legal framework, software fitting to local needs, and local availability of World-Class IT services and competition in the software market do matter in explaining illegal copying. Joint hypotheses testing show they matter together, and the

reasons for the existence of illegal copying go beyond a balance between the marginal costs of copying and the software's license.

Income per capita results from each country's economic performance, but price can be managed. The legal framework is characteristic of each society and can be, and is being, forced to adjust to better protect copyrights (through international trade policy and politics). Software fitting to local needs and quality of supply conditions, however, are direct results from corporate strategy. These findings have interesting direct consequences.

First, coincidentally, the last two subjects are not issues among most OECD countries, which do not face major problems of supply constraints, or software not responding to market's needs. These facts have helped to craft the current approach towards illegal copying: addressing the problem by forcing legislative changes, increasing the cost of copying devices by taxes, and compensating copyright owners. Evidence shows there is ground to assume the problem can be addressed in more effective ways –mainly in developing markets- and companies have power for doing it.

So the question is: Why is the market "sponsorship" approach not as widely used? A partial answer can be that affecting illegal copying by these means requires decision and important investments that generate results in the medium to long run, which is risky and profitable in case of reaching critical mass. It would be more easy, less risky, and maybe more profitable to force low-income countries to address legislative changes than investing in best products and better serving the markets.

Second, the differences between closed and open source software. Thinking back about 20 years, and looking where Apple and Microsoft were and where they are, one major explanation of the change is that Apple maintained its proprietary operative system "closed" to its computers, while Microsoft "open" approach was to license it to different hardware producers. Similarly, now in terms of the current closed and open source software approaches; the major differences are in the access to source code, possibilities for modifying it and freely improve software's fit to local needs, and lower cost. The open source business model is basically based in support and technical service and not in direct revenues from software licensing. This generates an interesting additional question.

If open source software undermines the market for closed source, then is it possible that (by not enforcing their copyrights in low-developed software markets) closed source software companies are competing with open source by using illegal copying for market creation as a more cost-effective "sponsorship" strategy?

Each customer's attitude towards software is function not only of the intrinsic value of the software, service availability and cost, but also of the size of the network using it, and innovations diffusion process can be affected by supply constraints [Kumar and Swaminathan (2001)]. A positive answer to the latest questions would help to better understand the central subject of this paper.

Furthermore, and considering the role of network effects in diffusion process, these results would illustrate there are enough incentives to allow illegal copying grow, and later enforce copyrights by legislative changes, fines, taxing copying devices, and compensating copyright owners.

IV. Closed and Open Source interaction

Let's suppose there are two incompatible and competing software pieces, which diffusion process can be modeled using the Bass Model[vi]. The first option is open source software, available at no charge, which can be freely shared, copied and modified. The second is a proprietary closed source software available for

a license fee, and which copying and sharing is illegal. This problem can be modeled as an interactive diffusion process of both software as an extension of the work of Givon, Mahajan and Muller (1997). Let be:

N(t) = Cumulative number of PC users at period t.

LUi(t) = Cumulative number of legal users of software i at period t

IUi(t) = Cumulative number of illegal users of software i at period t

Ui(t) = Number of legal users of software i at period t

Wi(t) = Number of illegal users of software i at period t

Zi(t) = Total number of users of software i at period t, Zi(t)=Ui(t)+Wi(t).

Z(t) = Total number of users of both software at period t, Z(t)=Z_{CS}(t)+Z_{OS}(t).

a_i: coefficient of external influence of software i,

b_i: coefficient of imitation of software i,

b'_i: coefficient of "switching" of software i

α_i= Percentage of legal users (copying) of software i.

0< a_i, α_{cs} ,b_i, b'_i<1, with α_{os}=1 because there is no illegal copying of open source software.

The model was extended to allow switching from being legal user from one brand to be illegal user in the other, and vice-versa. Considering these new assumptions, the variation on the cumulative number of legal users of closed source software at period t can be represented by:

$$\frac{dLU_{CS}(t)}{dt} = \left(a_{CS} + \boldsymbol{a}_{CS}b_{CS}\frac{Z_{CS}(t)}{N(t)} \right)(N(t) - Z(t)) + \boldsymbol{a}_{CS}b'_{CS}\frac{Z_{CS}(t)}{N(t)}Z_{OS}(t) \qquad [1]$$

Half of the first term – a_{CS}(N(t)-Z(t)) – represents the adoptions due to external influence (by innovation) or part of the adoption that does not result from word-of-mouth effect. If there is no word-of-mouth, there is no way to obtain an illegal copy of the software and, by definition; these users can only be legal users.

The following term, $\alpha_{CS}b_{CS}$(N(t)-Z(t))Z_{CS}(t)/N(t), represents the "legal" adoption of closed source software due to word-of-mouth between users of the software with non-users of any software (neither open nor closed source). It represents only the adoption that is reflected in the legal user base. Finally, the third term represents the increase in legal users by users that switched from open to closed source software.

For open source software, the variation on the cumulative number of legal users at period t is as follows:

$$\frac{dLU_{OS}(t)}{dt} = \left(a_{OS} + b_{OS}\frac{Z_{OS}(t)}{N(t)} \right)(N(t) - Z(t)) + b'_{OS}\frac{Z_{OS}(t)}{N(t)}Z_{CS}(t) \qquad [2]$$

There is, however, one major difference in this case: there are no "illegal" users of open source. In the same way, the variation on the illegal user base of closed and open source software in period t are represented by equations [3] and [4]:

$$\frac{dIU_{CS}(t)}{dt} = (1-\boldsymbol{a}_{CS})b_{CS}\frac{Z_{CS}(t)}{N(t)}(N(t)-Z_{CS}(t)) + (1-\boldsymbol{a}_{CS})b_{CS}^{'}\frac{Z_{CS}(t)}{N(t)}Z_{OS}(t) \qquad [3]$$

$$\frac{dIU_{OS}(t)}{dt} = 0 \qquad [4]$$

Let's assume both types of software have similar coefficients of external and internal influence, and of switching. Thus, the major difference between equations 1 and 2 relies in the fact that $\alpha_{OS}=0$ and, therefore, the increase in number of legal users due to brand switching is higher for open source. If $0< \alpha_{cs}, b_i, b'_i<1$, then $b'_{OS}>\alpha_{CS}b'_{CS}$, and $b'_{OS}\sim b'_{CS}$.

In order to have an overall increase in closed source users, the company must rely in the effect of word-of-mouth, which is based in the number of users at every period, $Z_{CS}(t)$, including legal and illegal users. Given the assumption of similarity in coefficients, the equality between equations 1 and 2 is achieved when:

$$\boldsymbol{a}_{CS}b_{CS}\frac{Z_{CS}(t)}{N(t)}\big(N(t)-Z(t)\big) + (\boldsymbol{a}_{CS}b_{CS}^{'}-b_{OS}^{'})\frac{Z_{CS}(t)}{N(t)}Z_{OS}(t) = b_{OS}\frac{Z_{OS}(t)}{N(t)}\big(N(t)-Z(t)\big) \quad [5]$$

Reorganizing, it is equivalent to:

$$\big(N(t)-Z(t)\big)\big(\boldsymbol{a}_{CS}b_{CS}Z_{CS}(t)-b_{OS}Z_{OS}(t)\big) + (\boldsymbol{a}_{CS}b_{CS}^{'}-b_{OS}^{'})Z_{CS}(t)Z_{OS}(t) = 0 \qquad [6]$$

Given the previous discussion, the second term of equation 6 is negative; the increase in legal users due to brand switching is greater for open source. The first term, however, needs more attention. In order to beat open source in diffusion, $\alpha_{CS}b_{CS}Z_{CS}(t)$ needs to be larger than $b_{OS}Z_{OS}(t)$ in at least an amount that, multiplied by the non-adopters, $(N(t)-Z(t))$, could equal the difference in brand switching. The number of closed source software users at period t, $Z_{CS}(t)$, equals the number of legal and illegal users, $U_{CS}(t)+W_{CS}(t)$. Additionally, $a_{CS}b_{CS}<b_{OS}$, requiring the closed source company to take advantage of the network effects generated by its illegal user base, $W_{CS}(t)$, as a way to compete with open source. The evidence for this is in equation [3], which explains the variation of the cumulative number of illegal users of closed source software at period t, $dIU_{CS}(t)/dt$.

This analytical evidence not only supports that open source software undermines the market for closed source software, but also that closed source companies need their illegal user base to cope with the diffusion of open source.

From a general perspective, one would suggest that closed source companies could work this out by increasing lock-in and switching costs in order to reverse the previous situation. Even in this case, it would be hard to refute the benefits of taking advantage of the illegal user base. Other option would be trying to reduce the rate of illegal copying for closed source $-(1-\alpha_{CS})-$ that, from the previous section, depends on income (and price), and other variables. So, an additional way to combat open source would be by decreasing costs, customizing software to local needs, and increasing local presence and services availability.

There are two last options. First, not to care about who is and is not a legal user at least until reaching some critical mass. Second, try to generate $(1-\alpha_{CS})=0$, i.e. eliminate the reasons for being an illegal user, by setting a zero price for certain period of time $-$we may assume both software companies may offer similar support services-, or by becoming an open source company.

V. Allowing illegal copying to grow in low-developed software markets: optimal business strategy and the benefits of the illegal user base

The previous section analyzed the interaction between the diffusion of closed and open source software, concluding the illegal user base is valuable because it generates network effects and, thus, attracts new adopters. In terms of the "sponsorship" strategy, proposed by Katz and Shapiro (1986), the decision of being a sponsor explains an important part of the differences between winning and losing a market. However, regardless of the differences between open and closed source software, there is a deeper question about the effects of illegal copying.

In the early stages of the creation of a market, a current reality among many developing countries, this rivalry may or may not exist, and illegal copying of software is rampant, reaching 97% and 94% in countries like Vietnam and China, respectively[vii]. Scenarios like this generates high uncertainty and there are few incentives for being a sponsor and investing in market development, given the high investments, fixed and variable costs needed for creating market presence.

Figure 2: Direct effects of illegal copying in software companies
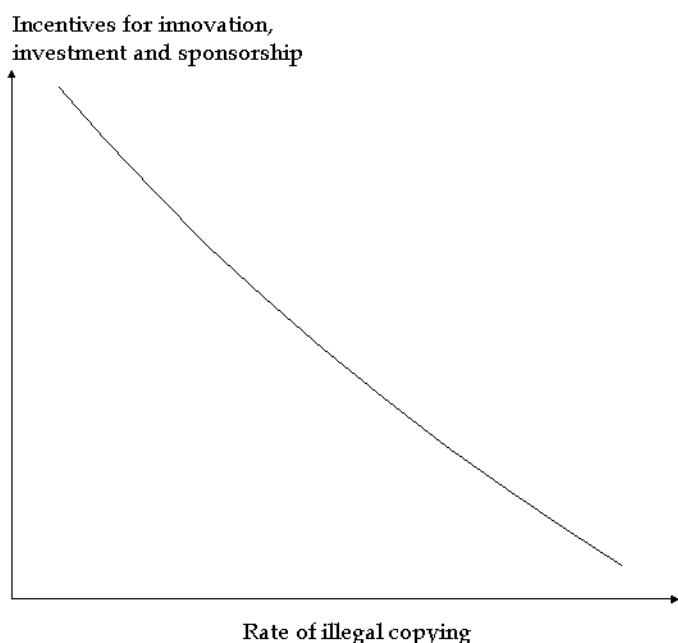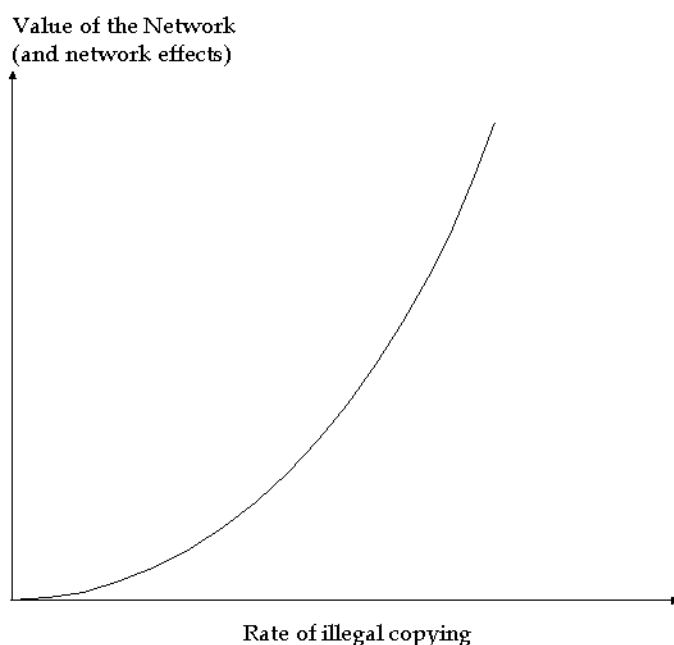

Source: The author.

So, regardless the presence of open source, what are the effects of illegal copying on the software market? The answers for this question vary depending of the level of aggregation, analysis of strategic interaction among software companies, and type of software products. Some issues of special relevance for this, are consideration of general-purpose (e.g. a $120 per-license) versus specific-purpose (e.g. a $10,000 per-license) software, and different level of network effects present across types of applications (word processors, database managing, etc.). In this work, the assumption of aggregation considers there is a representative piece of software that is sold in a global market (66 countries).

Varian (2001) suggests indirect network effects create dual causality between hardware and software, describing it as a "chicken-and-egg" problem. Given the previous assumptions, there is a closed relationship between adoption of the representative software and of personal computers, which from now is assumed as equivalent for potential software market.

Intuitively, one may think that illegal copying of software might have at least two types of effect. First, there would be one created by the costs and risks involved in being a sponsor and being a sponsor in a market with illegal copying. The latest case diminishes the incentives for innovation, market creation, and all the actions and investments needed to be an effective "sponsor". Thus, the first effect of illegal copying is on the supply side, and would be represented by Figure 2. The figure is not at scale, and its shape only represent an increasing negative effect of illegal copying on the software market.

Figure 3: Direct effects of illegal copying in the user base
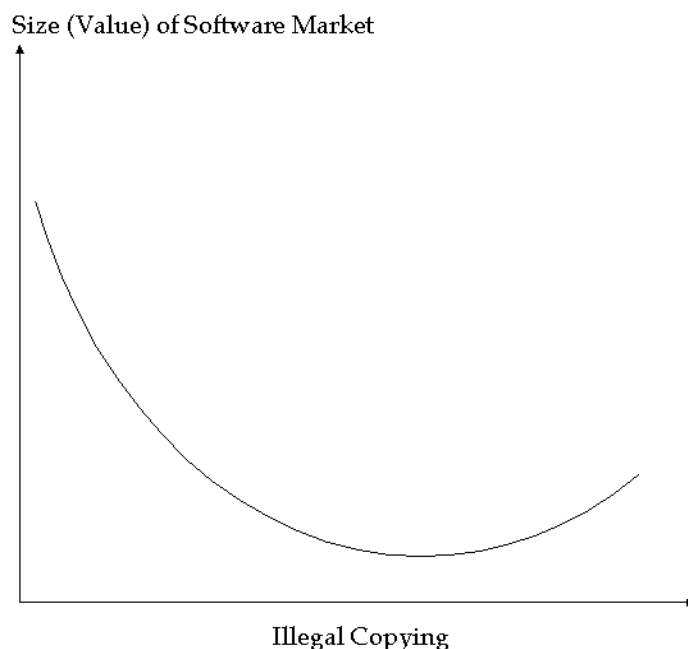

Source: The author.

Software production process, however, exhibits economies of scale and is, therefore, sensitive to the size of the user base. These facts, as Katz and Shapiro (1994) suggests, are the major determinants of the future diffusion of software added to network effects among users. Rogers (1995) defines diffusion as "the process by which an innovation [in this case software] is communicated through certain channels over time, among the members of a social system". About 90% of the adoption results from word-of-mouth, i.e. from the interaction of people in networks. Thus, in cases like the Vietnamese and Chinese, the size of the illegal user base is likely to be important in adding value to the network, as result of network effects. Consequently, the second effect of illegal copying –illustrated on Figure 3- is on the demand side. In this figure, also not scaled, one may assume a quadratic form based on Metcalfe's Law: The value of a network increases with the square of the number of users.

Not only members of the network perceive this value -in this case legal and illegal users-, but also those outside the network and companies that potentially may do business with the network, which includes the owner of the proprietary software. One may argue, therefore, the size of the network has positive effects in three ways. First, it generates lock-in and minimizes the incentives to switch to a network of lower value. Second, it generates word-of-mouth, attracting more potential users. Third, it increases the potential value of the network by signaling to both the supply and demand.

Figure 4: Combined effect of illegal copying in the software market

Source: The author.

Table 2: Dependent variable personal computers, per 100 inhabitants (as proxy for market size)

Software Market (PC)	Regression 7	Regression 8	Regression 9
ln(GDP per capita, PPP)	.1568826 (.0107665)	.0888032 (.0108663)	.0788415 (.0094402)
Illegal Copying	..	-.3019815 (.0267792)	-1.414235 (.0977837)
Illegal Copying ^2	..	..	.869772 (.0743261)
Const	-1.3055 (.0984444)	-.4952331 (.1093001)	-.081646* (.1009334)
Observations	377	377	377
Adj. R-squared	0.6449	0.7209	0.7892

Source: The author All coefficients, but * in regression 9, are statistically significant at 5%. Standard errors are in parenthesis..

Thus, the next step is to study the overall effect of illegal copying on the software market. By simply adding the effects represented in both figures, one may expect a behavior like the represented in Figure 4.

Here, the central idea is the following. The combined effects of illegal copying may be positive in low developed markets where illegal users are responsible for generating most of network value and effects. Additionally, a small market does not generate incentives for "sponsorship".

Thus, the next step is empirically analyzing the effect of illegal copying in the software market. As working assumption, and previously discussed, computer penetration is used as proxy for the size of the software market. BSA (2001) is the source for data on illegal copying. Once again, the regressions are done controlling by income per capita in PPP.

Due to the lack of historical data, only six years for 66 countries, the regression analysis was done using random-effects linear model with Generalized Least Squares, which results are in Table 2. First, regression 7 shows the relationship between income per capita and software market. With statistically significant coefficients, it makes evident the obvious importance of income in the generation of the market. the richer the country, the bigger its relative market.

Illegal copying is then included in a regression (8) against market size, controlling by income per capita. Its coefficient is statistically significant at 5% level, and negative. This can be interpreted as the first of the previously mentioned effects of illegal copying over the supply side: the greater illegal copying, the lower the incentives for investing in innovation, and sponsorship on the market, the larger the effect of supply constraints in the diffusion process[viii].

Finally, regression 9 incorporates the square of illegal copying as a proxy for direct network effects, and as a way to analyze its overall effect on the software market. Always controlling by income per capita, results show that all coefficients but the constant are statistically significant at 5% level. The coefficient of the square of illegal copying is positive, supporting the hypothesis about positive effects.

Thus, by regression 9, the relationship between illegal copying and the software market, controlling by income per capita, may be represented as:

$$SM_{it} = \alpha IC_{it} + \beta IC_{it}^{2} + \phi Y_{it} + u_{i} + e_{it}, \text{ with } \alpha, \beta, \chi, \phi \in R \qquad [7]$$

where u_i represents the random effect, and e_{it} the error. Additionally:

SM_{it}: Software market (penetration of personal computers in the population) for country i, in period t.

IC_{it}: Rate of illegal copying for country i in period t.

Y_{it}: GDP per capita (PPP) for country i in period t (log).

Constant is zero, as dropped by result in regression 9, and α, β and φ are the GLS estimators.

Thus, now the task is to analyze the effect of a change on illegal copying on the evolution of the software market. Given the results, the amount of this change is represented by equation [8]:

$$\frac{dSM_{it}}{dIC_{it}} = \alpha + 2\beta IC_{it} = 0 \Rightarrow IC^{*}_{it} = -\frac{\alpha}{2\beta} \qquad [8]$$

The change is function of illegal copying. α has a negative sign (-1.414235), β is positive (0.869772), and equation [7] is quadratic, which implies it may have a minimum or maximum, which makes of IC^{*}_{it} an

interesting value. $IC^*_{it}= 0.8129$ is a point of slope zero and, by second order conditions (and by $\beta>0$ and $\alpha<0$), is a minimum:

$$\frac{d^2 SM_{it}}{dIC_{it}^2} = 2\boldsymbol{b} > 0 \qquad\qquad [9]$$

The implications of this result are interesting. In terms of its relationship with software market, the rate of illegal copying of 81.29% marks a threshold. By values under it, one may say the overall effect of illegal copying over the market is negative. By values over it, however, the overall effect is positive. This is the behavior represented in Figure 4.

The importance of IC^*_{it} relies in the relevance of network effects. In countries with high rates of illegal copying, the illegal user base is responsible for most of the word-of-mouth, network effects and value of the network. For example, in 2000, Vietnam had a rate of illegal copying of 97%, meaning that only 21,000 of the 700,000 users of the market were legal (Only 1,000 users more than the size of the total legal and illegal market in 1993, when the illegal copying was close to 100%).

From a market creation perspective, having only 21,000 users (all legal) it is very different from also having 679,000 more that –while illegal at the moment- are helping the diffusion of the product, and can be transformed into legal users by different means.

Besides, low incentives to invest in sponsorship and market creation in Vietnam does not come only from the fact of having 97% of illegal copying but, among other things, from dimensioning the market with a unit sales-based perspective[ix]. A small market does not generate the same incentives that knowing there is a market more than 30 times bigger already created. In this case, the network effects generated by illegal users are more important than the negative effect from companies' disincentives to supply the market. The fact this "new market" is of illegal users only makes different the appropriate strategy for capturing it.

Can this problem be solved by the traditional approach of enforcing copyrights? Most certainly it cannot. First, this is unlikely because illegal copying is generated by reasons different than just people wanting to use software for free. Indeed in almost all the analyzed countries, illegal copying has decreased over time. Second, if the legal consequences of being an illegal user get too tough, there is an open source option with switching costs likely to be lower than the legal penalties. Third, if total enforcement were possible and illegal copying suddenly became unfeasible, then the diffusion process would be slowed down, potential market would be smaller, and the diffusion path longer. In other words, if illegal copying is made impossible by enforcement, the size of the network would be smaller.

For values below IC^*_{it}, however, the overall effects of illegal copying of software on market potential respond to conventional wisdom and are negative. This fact is important as well. As the software market grows and the rate of illegal copying decreases, the structure of incentives changes and proprietary software companies would have increasing incentives to address the issue both as individual corporations and as a group.

Corporations have more incentives to better supply and serve the market when it gets larger (the closer to critical mass, the better). Each copyright owner's "sponsorship" strategy is tailored to its product and, from this perspective, makes sense doing it alone. For copyright enforcement, however, makes sense act as a group because influencing legal changes in countries with frameworks less protective of copyright is not particular to any company, but common to all. An example of group initiative is the performed by, among others, the Business Software Alliance[x].

The international dimension of the problem provides evidence for interesting analysis; especially in places where the software market is still in early stages of development. The results allow explaining why

software companies tend to enforce their copyrights differently across countries, which can also be a partial endogenous consequence of not having local presence. This behavior, however, can also result from decision-making oriented towards taking advantage of the effects generated by illegal users in the diffusion process.

VI. Conclusions, policy implications, and future directions for research

While the Internet revolution and availability of new and less expensive storage and copying technologies would have predicted an increase in illegal copying of software, it has decreased between 1995 and 2000 relative to the software market[xi]. This paper argues that illegal copying of software results from pure matters of license price versus cost of copying; suggesting it can also be created by more relevant factors. In terms of its effects, the results propose that not everything is negative, considering the network effects generated by the illegal user base. The first two sections introduce the work, and present a literature review of the aspects most relevant for addressing these issues.

Causes of Illegal Copying

Section III presents the findings of the analysis of causes of illegal copying among 66 nations. By hypothesis testing, it concludes there is, at least, four simultaneous causes of illegal copying.

First, illegal copying of software can be caused by an affordability problem. People duplicate as long as the cost of copying is lower than software's license (both for illegal personal and commercial use). Second, it may be caused by the way in which each community's social perceptions and beliefs about the justifications towards copyrights, which is explained by its legal framework. Third, it can also be generated by the lack of software fitting to local needs -in terms of language, metaphor, etc.- diminishing the value of software to an extent to which potential users are not willing to pay for it. Finally, and with a similar effect, illegal copying can also be generated in undersupplied and poorly competitive markets by low levels and quality of value-added services, and after sales support.

This result has interesting public and business policy implications, highlighting new ways for addressing the illegal copying problem.

The common approach to fight illegal copying –by different types of legal means – can be understood as most OECD countries (home-base of the most important software companies) had well developed markets, none or few supply constraints, and few problems of software fitting to local needs. This has helped to craft the conventional wisdom about illegal copying, understanding the issue as a matter of relative cost, and dealing with it only in its dimension the illegality of people that don't want to pay for the license of to use a service.

In the biggest share of the World software market –developing countries-, however, the other reasons are as important and maybe more relevant than affordability, and First-World ways of dealing with illegal copying are not necessarily the most effective. The results support that total enforcing will not make legal users from all illegal users, less yet in presence of open source software.

Acting in the "best interest" is always contingent to "whose" best interest. One may support the idea that product "sponsorship" would be an ideal and socially responsible way to fight illegal copying. Results in sections IV and V, however, show the value of the illegal user base for proprietary software companies. From this perspective, there is evidence of higher incentives to let illegal copying grow to a certain level and capture rents by legal enforcement, than being a full product "sponsor" on the market.

Open and Closed Source Software Interactions: incentives for using the illegal user base

Section IV analyzes the interaction between closed and open source software companies. By extending the work of Givon, Mahajan and Muller (1997), analytical results show that an optimal strategy for a closed software company is to take advantage of the whole user base, which includes illegal users, and the network effects they generate.

Network effects are especially relevant in software diffusion in two ways. First, they matter as source of word-of-mouth in creating adoption by interaction between current users and non-users (internal effect), and in generating brand switching by interactions between users of closed and open source brands. Second, they matter as the value of a network increase with the square of its number of users.

These findings highlight an important point. Commonly, illegal copying is thought to hinder and threaten the development of the software market, being depicted as "piracy". Section III shown some of its causes, suggesting some possible courses of action. Results in Section IV show that illegal copying also helps the development of closed source software market.

This paradox helps answering -partially- the question about the absence of "sponsorship" of software products as way of fighting illegal copying in the early stages of development of a market. Being a sponsor would generate results similar to fighting illegal copying, but at higher costs. For closed source companies, the corporate operative marginal cost of an extra adopter by illegal means is zero, while it is positive for legal users (and includes all the costs and investments required in market entry).

Thus, as users respond to price, illegal copying is likely to be used as a "second-best sponsorship strategy". Companies decide to do not earn present income (from licenses and service), at no marginal cost (illegal users are the distribution channel), in order to obtain a higher future income (generated by network effects) at similar marginal costs, but in a market closer to critical mass, and taking advantage of economies of scale.

This finding has critical policy implications. Copyright owners claim that intellectual property and copyright infringements are equivalent to costs generated by an illegal activity that should be, somehow, compensated. They try to enforce copyrights in order to collect the estimated "revenue lost to piracy", and by their lobby try to craft national and international policies for facilitating enforcement. As we have shown, an important portion of their legal revenue, would have not been possible without the "help" of illegal users.

Thus, the question now is how to account for this. How should these positive private effects be considered, given they result from an illegal activity? Should they be included as part of the compensation to copyright owners, or not at all? This is especially relevant for the design of compensation policies, such as taxing copying hardware, as a way to redistribute social costs and incentives to both hardware and software companies[xii].

Effects of Illegal Copying

Section V builds on the previous results by, controlling for the effect of income per capita, analyzing the effect of illegal copying on the software market regardless of open source. This is done assuming illegal copying may generate a negative effect on the supply, and positive effect on the demand.

The negative direct effects on supply are assumed to result from lower incentives for companies to serve and invest in developing the market, innovating in new products, and making the investments needed to provide more and better services, which partially supports Novos and Waldman (1984). The

assumption about the positive direct effects on demand is generated partially by the previous results, and supported by Varian (1998 and 2001), and Givon, Mahajan and Muller (1997), among others.

Both legal and illegal users add value to the network, generate network effects, and foster adoption by word-of-mouth. Thus, it makes sense to imagine a possible balance between negative and positive effects. The three situations are represented by figures 1, 2 and 3.

Table 2 presents the results regressing a random-effects linear model, with Generalized Least Squares, for a panel data for 66 countries between 1995 and 2000. The resulting model, in equation [7] from regression 9, helps to analyze the effect on the software market, given a change in illegal copying. The linear model has quadratic form, which reaches a minimum when the rate of illegal copying equals 81.29%. Thus, it generates overall positive effects on the software market when is above 81.29%, and negative effects when it is below. The policy implications of this result add to those of Section IV. Illegal copying shows both negative and positive effects and, above this point, the overall effect is positive.

The network effects of the illegal user base are large and, given the high sophistication of the software sector (a network industry), it is difficult to believe that no company is taking advantage of this.

From a public policy perspective, this is relevant for countries with low-developed software markets that are looking for nationwide-agreements with major software companies. These agreements could look for fostering diffusion of information and communication technologies and better develop the local software market, along with a medium-term plan for controlling the development of the illegal user base. It is unlikely that countries would continue to act as if they were guilty for having illegal copying, and not understanding that software vendors benefit from these effects. From a double bottom line perspective, making evident the double effect of illegal copying would help addressing the problem in a more open, and transparent manner.

Final Conclusions and Future Directions for Research

The combined result of the last three sections of this paper has major implications. First, empirical evidence suggests that illegal copying is far more complicated that having "bad guys" who don't want to pay for software. Illegal copying is partially generated by this reason, as well as by social characteristics, poorly served markets and software products that don't fit local needs.

Analytical and empirical evidence suggests that allowing illegal copying in early stages of development of a software market helps to achieve a faster development of the market, to the benefit of the copyright owner, and may be a dominant strategy for a closed source software company in presence of open source competition. From this perspective, a game theoretical analysis of this relationship is being developed to present more conclusive results.

Resulting from this, an additional implication is that proprietary software companies enforce their copyright in a way that appears to be contingent to the possibilities of using illegal copying as source of market development and competing with open source, for later trying to capture the market by legal means. This "strategy of market development" is likely to be more profitable than a proper "sponsorship" strategy, in the sense of Katz and Shapiro (1986).

From a public policy perspective, a major question remains: How to consider the legal revenues generated by illegal copying in policy discussions about compensation to copyright owners. From a business policy standpoint, the question is how to design strategies that would better take advantage and control of the, now illegal, user base. Developing countries engagement is critical, and they need help from and local presence of IT companies.

There are different issues that have not been addressed but would generate interesting discussions and questions, such as deeply analyze the evidence about selective enforcement of copyrights across different regional markets. Additionally, the lack of data about open source software presents a major problem for empirical analysis. In the same way, the lack of the data used in Section II constraints the quality of cross-country time-series analysis, which would be done by considering endogenous relationships between variables and illegal copying by two-stages least squares.

Exhibit 1: Key of Variables

ln(ill_copying) = Logarithm of the rate of illegal copying

ln(GDPpercapita) = logarithm of the GDP per capita in PPP

Legal = Composed variable for legal framework

leg_brit =dummy variable for British legal framework

leg_fren =dummy variable for French legal framework

leg_soci =dummy variable for Socialist legal framework

leg_germ =dummy variable for German legal framework

leg_scan =dummy variable for Scandinavian legal framework

soft_fitting = Software fitting local needs

IT_services = Availability of IT services

Soft_Comp = Competition in the local software market

Cons = Constant

References

1. Al-Jabri, I., and Abdul-Gader, A. (1997) "Software Copyright Infrigements: an Exploratory Study of the Effects of Individual and Peer Beliefs" International Journal of Management Science, Vol. 25, No. 3, 335-344.

2. Bakos, Y., Brynjolfsson, E. and Lichtman, D. (1999) "Sharing Information Goods" Journal of Law and Economics, Volume 42, Number 1, Part 1, April 1999.

3. Business Software Alliance (2001) "Sixth Annual BSA Global Software Piracy Study", < http://www.bsa.org/resources/2001-05-21.55.pdf > last viewed on December 12, 2001

4. Chen, Y. and Png, I. (2001) "Software Pricing and Copyright Enforcement", National Taiwan University (http://www.comp.nus.edu.sg/~ipng/research/ last visited on November 12, 2001).

5. DeLong, J. B and Froomkin, A. M. (2000) "Speculative Microeconomics for Tomorrow's Economy" in Kahin, B. and Varian, H. *Internet Publishing and Beyond: the economics of digital information and intellectual property*. MIT Press.

6. Easterly, W. and Sewadeh, M. (2001) "Global Development Network Growth Database", The World Bank, Version updated on June 2001. < http://www.worldbank.org/research/growth/pdfiles/GDN/Social%20Indicators%20and%20Fixed% 20Factors.xls> Last visited on December 2001.

7. Givon, M., Mahajan, V., Muller, E. (1997) "Assessing the Relationship between the User-Based Market Share and Unit Sales-Based Market Share for Pirated Software Brands in Competitive Markets" Technological Forecasting and Social Change, 55, 131-144.

8. Gayer, A. and Shy, O. (2001a) "Freeware, Downloading and Piracy in Markets for Digital Media", Dept. of Economics, University of Haifa.

9. Gayer, A. and Shy, O. (2001b) "Copyright Protection and Hardware Taxation", Dept. of Economics, University of Haifa.

10. Greene, W. (2000) "Econometrics Analysis", 4th Edition, Prentice Hall.

11. Katz, M. and Shapiro, C. (1986) "Technology Adoption in the Presence of Network Externalities" The Journal of Political Economy, Volume 94, Issue 4, 822-841

12. Katz, M. and Shapiro, C. (1994) "Systems Competition and Network Effects", The Journal of Economic Perspectives, Volume 8, Issue 2, 93-115.

13. Kirkman, G., Osorio, C. and Sachs, J. (2002) "The Network Readiness Index: Measuring the Preparedness of Nations for the Networked World" in Kirkman, G., Cornelius, P., Sachs, J., and Schwab, K. eds *Global Information Technology Report: Readiness for the Networked World*, Harvard University and World Economic Forum, Oxford University Press. (forthcoming 2002)

14. Kirkman, G., Cornelius, P., Sachs, J., and Schwab, K. "Global Information Technology Report: Readiness for the Networked World", Harvard University and World Economic Forum, Oxford University Press. (forthcoming 2002)

15. Kumar, S. and Swaminathan, J. (2001) "Diffusion of Innovations under Supply Constraints" revised and resubmitted to Operations Research, August 2001. <http://www.stanford.edu/~skumar/ORResubmit.pdf> last viewed on December 12, 2001.

16. Mas-Collell, A., Whinston, M. and Green, J. (1995) "Microeconomic Theory", Oxford Press.

17. Novos, I., Waldman, M. (1984) "The Effects of Increased Copyright Protection: an Analytic Approach" The Journal of Political Economy, Volume 92, Issue 2, 236-246.

18. Rogers, E.M. (1995) "Diffusion of Innovations", Free Press, New York.

19. Varian, H. (1993) "Economic Incentives in Software Design", University of California at Berkeley < http://www.sims.berkeley.edu/~hal/Papers/Software.pdf > last viewed on December 12, 2001.

20. Varian, H. (1994) "Buying, Sharing and Renting Information Goods", University of California at Berkeley, Version August 5, 2000. Last viewed on December 12, 2001 < http://www.sims.berkeley.edu/~hal/Papers/sharing.pdf>

21. Varian, H. (1998) "Markets for Information Goods", University of California at Berkeley, <http://www.sims.berkeley.edu/~hal/Papers/japan/japan.html>

22. Shapiro, C. and Varian, H. (1999) "Information Rules", Harvard Business School Press

23. Varian, H. (2001) "Economics of Information Technology", University of California at Berkeley. <http://www.sims.berkeley.edu/~hal/Papers/mattioli.pdf> last viewed on December 12, 2001.

[i] So far, this is true, but it is likely to change with the popularization of Application Service Provision, and software activation codes that would require online interaction.

[ii] Varian (1998), Bakos, Brynjolfsson, and Lichtman (1999)

[iii] Business Software Alliance (2001) is the source of data used for estimating illegal copying of software. By the assumption of aggregation and existence of a representative software piece, BSA data may be interpreted as percentage of software pieces that potentially would have been sold but are lost to illegal copying and, therefore, percentage of user base that is illegal.

[iv] These questions were included as part of the annual Executive Opinion Survey submitted by the World Economic Forum and Harvard University. See Cornelius and McArthur (2001) for details in the methodology. From the Executive Opinion Survey, "circling 1.....means you agree wholeheartedly with the answer on the left-hand side. Circling 7.....means you agree wholeheartedly with the answer on the right-hand side. Circling 2.....means you largely agree with the left-hand side. Circling 3.....means you agree somewhat with left-hand side. Circling 4.....means your opinion is indifferent between the two answers. Circling 5.....means you agree somewhat with the right-hand side. Circling 6.....means you largely agree with the right-hand side".

[v] Katz and Shapiro (1986)

[vi] The Bass Model has been successfully used to explain the diffusion of a large number of innovations. Givon, Mahajan and Muller (1997) successfully use it in modeling the diffusion of spreadsheets (Excel and Lotus) and word processors (Word and Word Perfect).

[vii] BSA (2001)

[viii] For a model on the effects of supply constraints in diffusion, please see Kumar and Swaminathan (1999).

[ix] Givon, Mahajan and Muller (1997).

[x] See http://www.bsa.org/. The seventh version of BSA's Annual Global Study was released during the final stages of completing this draft, and data for 2001 was not included. Email the author for an updated version at carlos_osorio@post.harvard.edu .

[xi] An updated version of this paper containing data from 2001 will be available shortly. Please email the author if you are interested.

[xii] See Gayer and Shy (2001b).