A survey on software piracy empirical literature: Stylized facts and theory

Nicolas Dias Gomes [a,*], Pedro André Cerqueira [b], Luís Alçada Almeida [a]

[a] Faculty of Economics, University of Coimbra and INESC-Coimbra, Portugal
[b] Faculty of Economics, University of Coimbra and GEMF, Portugal

ARTICLE INFO

ABSTRACT

As software is central in today's world, the problem of software piracy is of increasing importance. It reduces the revenues of firms that develop new software and, therefore, it may hinder innovation and growth. To understand better the causes of this phenomenon this paper presents a survey of the empirical literature regarding software piracy and discusses if the findings are coherent with the theoretical literature. Overall we are able to identify eight stylized facts in five dimensions: the Economic, Cultural, Educational, Technological and Legal. Moreover we argue that most of these findings are coherent with that which the theoretical models imply.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Technology has evolved over the years and is present in almost everything we use. The areas where this influence is most evident are those that involve the use of computers and the Internet. They increase the productivity of firms and make life easier for households allowing, for instance, home banking or online shopping. Other examples can be added; perhaps one of the events that improved most significantly the productivity of enterprises was the replacement of the typewriter by the computer. However, computers (and related devices such as tablets and smartphones) cannot be used without the proper software and only with it can we exploit their full potential. An operating system starts and controls these machines, but tools such as the productivity packs, to produce professional documents, are also required.

Copyright laws protect software and hardware and it is in the case of software that the protection of intellectual propriety is more challenging due to its nature: (i) it can be reproduced at virtually no cost and with the same quality as the original, (ii) it is easily modified by hackers that break its protective barriers and (iii) it is easily distributed.

Infringement of the intellectual propriety of software products, usual known as software piracy, occurs when there is an unauthorized use or sale of commercially available software (Moores and Dhillon, 2000) that is protected under national and international copyright laws. This piracy comes in many forms.[1]

---

[1] Softlifting: purchasing a single licensed copy of software and loading the same copy onto several computers, contrary to the license terms; Internet: making unauthorized copies of copyrighted software available to others electronically; Software counterfeiting: the illegal duplication and distribution of copyrighted software in a form designed to make it appear to be legitimate; OEM unbundling: selling stand-alone software that was intended to be bundled with specific accompanying hardware; Hard disk loading: installing unauthorized copies of software onto the hard disks of personal computers, often as an incentive for the end user to buy the hardware from that particular hardware dealer, and Renting: unauthorized rental of software for temporary use as if it were a video.

* Corresponding author at: Faculdade de Economia, Universidade de Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal.
E-mail address: nicolasdiasgomes@gmail.com (N.D. Gomes).

Software piracy affects first of all firms' profits, as the sales of legal software units tend to decrease. These losses are reported annually by the *Business Software Alliance* (BSA) and, although the complete methodology applied is not publicly available, the estimates reported are based on confidential information provided by BSA members such as Adobe®, AVG®, Intel®, Microsoft® or Symantec®, for instance, covering both the hardware and software industry.[2] However software piracy does not affect only the firms, the medium and long run impact in the economy is also important. If the copyright laws are not enforced it reduces the incentive to innovate and, therefore, to produce newer and better software tools which would improve productivity. Therefore, this impact may hinder the medium and long term growth of economies as shown by Andrés and Goel (2012).

To have a better understanding of how to fight this phenomenon we must separate two types of piracy: the commercial type in which we buy a DVD on the black market – in this case the reseller has profits and compete with the "legal" firms (the competition is asymmetrical[3]); and the end-user piracy, when consumers use, at home or work, software that was not "legally" sold, which is more difficult to detect as it does not involve physical means. There are some actions that firms and governments can implement to protect from these two forms of piracy. Either in the courts, enforcing anti-piracy laws, or by introducing mechanisms that can detect pirated products making them unusable to the user.

However, without knowing why people engage in this illegal activity, the effectiveness of the actions taken by companies or governments is reduced. With this in mind, and in order to get a better understanding of what causes software piracy, this paper offers a survey of the empirical literature. It classifies the factors in five dimensions (Economic, Cultural, Educational, Technological and Legal), analyses how each of these dimensions affect software piracy and systemizes the findings into eight stylized facts. Furthermore, this paper also analyses how these stylized facts relate to the theoretical literature,[4] this is, if they are coherent or in the case of different theoretical results, which one seems to prevail. By doing this, we hope to offer a systematic view which can help companies and governments to design policies and actions to better deal with the issue.

The paper is organized as follows: Section 2 reviews the empirical literature on software piracy, describing the methods used and their limitations, the stylized facts found and how they can be anchored in the existent theoretical models and finally, Section 3 concludes and gives advice on how to overcome the limitations found.

2. Empirical literature

The empirical literature on piracy software has used an array of methodologies: surveys using respondents from universities and from the labor market; or cross-country analysis using panel or cross-section data relying on macroeconomic sources.

Empirical literature that uses surveys can obtain richer results, being able to model each parameter (age, sex, income), but it has some drawbacks. First, results rely on the willingness of the respondents to answer truthfully and even if the inquiry is anonymous, due to the nature of the crime, the respondents may underreport their activity. Also, surveys are used in a particular group of the population in a particular city (for instance Gopal and Sanders (1998), Butt (2006), Higgin (2006) or Gan and Koh (2006) survey college students and Lau (2004) surveys business users) which leads to the well-known population bias problem, which can influence the main findings and make the extension of the results problematic. Finally, most of the questionnaires rely on a Likert scale.[5] When respondents answer questions it is possible that they go to the extremes or the middle, which can bias the conclusions.

To overcome these problems, authors such as Gopal and Sanders (1998) or Holm (2003) used a cross-sectional model that explains the phenomenon at a country level, complementing the results from the surveys.

Other studies make cross-country analysis, using either panel or cross-section data and related methodologies. Before we discuss the advantages and problems of each methodology we should refer that most of these studies, regarding the software piracy level, use data from the *Business Software Alliance*. To our knowledge, these estimates are the only ones that provide a historical view of software piracy from 1994 to 2014. In spite of this, these estimates suffer from some drawbacks, which begin in the computation itself of their rates.

The *Business Software Alliance* relies on external consultants to do surveys, such as the *International Data Corporation* (BSA, 2014) and *Ipsos Public Affairs* (BSA, 2012). As part of the estimations rely on surveys, at least part of the data will suffer from the same problems previously described. Another set of important variables that serve to compute piracy rates derives from the proprietary information of its members. This complements the surveys. Again, in this case, external individual/corporate consultants cannot verify information concerning their validity. Additionally, these input variables are only available for a small group of countries, and vary over time. For the remaining countries of the world, estimates are projected based on macroeconomic variables, such as income.

Another drawback of these estimates is derived from their lack of clarity and consistency over the years. It is

---

[2] We should note that some findings suggest that some level of piracy can be beneficial for the software developer. See, for instance, Lahiri and Dey (2013) or Lu and Poddar (2012).

[3] Some authors that model this phenomenon are Peitz and Waelbroeck (2004), Peitz and Waelbroeck (2006a), Duchêne and Waelbroeck (2005) and Zhang (2002).

[4] For surveys on the theoretical literature see Peitz and Waelbroeck (2006b) or Belleflamme and Peitz (2010).

[5] A Likert scale is a psychometric scale commonly involved in research studies employing questionnaires. It is the most widely used approach to scaling responses in survey research, such that the term is often used interchangeably with rating scale. Usually it is divided into 5 ordinal values: 1. Strongly disagree, 2. Disagree; 3. Neither agree nor disagree; 4. Agree and 5. Strongly agree. See Wuensch (2005).