Expert Systems with Applications 38 (2011) 8782–8803

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A neuro-computational intelligence analysis of the global consumer software piracy rates

Mohamed M. Mostafa

Gulf University for Science and Technology, College of Business, Kuwait

ARTICLE INFO

Keywords:
Global software piracy
Ethical behavior
Neural networks
Bayesian regression
Evolutionary computation models

ABSTRACT

Software piracy represents a major damage to the moral fabric associated with the respect of intellectual property. The rate of software piracy appears to be increasing globally, suggesting that additional research that uses new approaches is necessary to evaluate the problem. The study remedies previous econometric and methodological shortcomings by applying Bayesian, robust and evolutionary computation robust regression algorithms to formally test empirical literature on software piracy. To gain further insights into software piracy at the global level, the study also uses five neuro-computational intelligence methodologies: multi-layer perceptron neural network (MLP), probabilistic neural network (PNN), radial basis function neural network (RBF), generalized regression neural network (GRNN) and Kohonen's self-organizing maps (SOM) to classify, predict and cluster software piracy rates among 102 nations. At the empirical level, this research shows that software piracy is significantly affected by the wealth of nation as measured by gross domestic product (GDP), the nation's expenditure on research and development and the nation's judicial efficiency. At the methodological level, this research shows that neuro-computational models outperform traditional statistical techniques such as regression analysis, discriminant analysis and cluster analysis in predicting, classifying and clustering software piracy rates due to their robustness and flexibility of modeling algorithms.

1. Introduction

Software piracy has been defined as the unauthorized copying, distribution, and downloading of computer programs and applications (Globerman, 1988). Even though piracy and counterfeiting occur across almost all types of products, software piracy is a unique form of piracy because in most cases the pirate is both producer and consumer of the software (Andres, 2006a). Unlike brand piracy or patent infringements, "producers usually do not sell the bogus product – they simply share it" (Nill & Shultz, 2009, p. 290). Software piracy has recently evolved into a truly global problem. For example, a survey by the Business Software Alliance (BSA) estimated that illegal software duplication cost the worldwide software industry approximately $50 billion in losses (BSA, 2009). In a study across 16 countries over the period 1998–2002, Peitz and Waelbroeck (2004) concluded that digital piracy would be responsible for a 20% decrease in sales.

Out of the 102 countries BSA studied, countries such as Vietnam and Indonesia maintain some of the highest software piracy rates (around 90%). Such countries are sometimes called "one disc" countries because a few discs are legitimately purchased and then many copies are illegally made from the original (Correa, 1995). Advanced countries such as the US and the UK have one of the lowest rates (around 20%), while transition economies such as Romania and Bulgaria are somewhere between the two extremes. This might suggest that the relationship between software piracy rates and the wealth of a nation may well be nonlinear – a reminiscent of what is known in the literature as the Environmental Kuznets Curve (EKC), which describes the commonly observed inverted-U-shape relationship between environmental degradation and per-capita income. Fig. 1 illustrates this relationship. However, it should be noted that the size of the pirated software in developing economies is small compared to advanced economies. For example, Vietnam and Indonesia account for losses of $96 million and 350$ million, respectively, whereas losses in Western Europe and the US are estimated at $10.6 billion and $8.1 billion, respectively (Correa, 1995).

Some authors argue that software industry may benefit from piracy through network effects, which enhance exposure and facilitate adoption (e.g. Jain, 2008; Jaisingh, 2009). However, the net effect is believed to be negative (Limayem, Khalifa, & Chin, 2004). Software piracy is also regarded as a serious social and political problem. For example studies have shown that pirated and counterfeited products lower consumers' confidence in legitimate brands and negatively affect companies' reputation (Wilk and Zaichkowsky, 1999), impact upon consumers' perception of

E-mail address: moustafa.m@gust.edu.kw

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.01.090

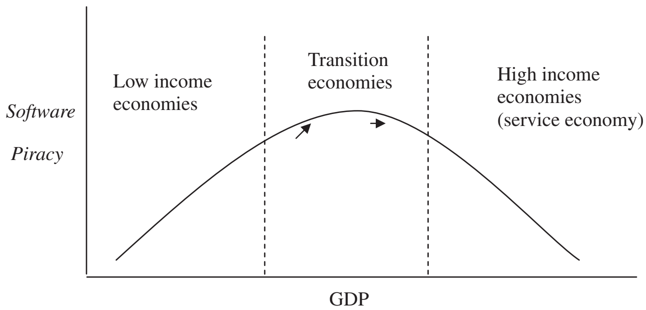

Fig. 1. Software piracy EKC-type curve.

genuine articles (Chakraborty, Allred, & Bristol, 1996), and pose threat to consumer health and safety (Cordell, Wongtada, & Kieschnick, 1996). The advent of the Internet has exacerbated the problem because piracy over the Internet is extremely difficult to detect (Hinduja, 2008). Wall (2005) noted four characteristics of the Internet that have facilitated an increase in digital piracy in recent years: It allows anonymous communication, it is transnational, it has created a shift in thinking from the ownership of physical property to the ownership of ideas, and it is relatively easy to use.

Although software piracy has been studied at various levels, including global (e.g. Husted, 2000), national (e.g. Bezmen & Depken, 2006) and individual levels (e.g. Shore et al., 2001), no previous studies have attempted to use neuro-computational and evolutionary computation (EC) techniques to predict, classify and cluster software piracy across nations. In this research we aim to fill this research gap by predicting, classifying and clustering software piracy rates across 102 nations through the use of intelligent modeling techniques. More specifically, the purpose of this research is twofold:

- To determine the major factors that affect the software piracy at the global level; and
- To benchmark the performance of neuro-computational and evolutionary computation (EC) models against traditional statistical techniques.

Thus, this study makes at least three important contributions to the broader literature on software piracy. First, most studies involve comparisons of software piracy rates across a relatively small number of countries. Our study includes 102 nations, which makes it the most comprehensive study so far. By doing so the study adds depth to the knowledge base on software piracy. Second, by employing neuro-computational and EC methods such as neural networks and evolutionary robust regression, the study adds breadth to the debate over the causes of software piracy at the global level. Finally, by focusing solely on software piracy, rather than on other forms of digital piracy, this study enriches the knowledge base of this under-represented issue.

This paper is organized as follows. The next section summarizes software piracy literature and develops research hypotheses. The methodology used to conduct the analysis follows. The subsequent section presents empirical results of the analysis. Next, the paper sets out some implications of the analysis. This section also deals with the research limitations and explores avenues for future research.

2. Literature review and hypotheses development

Drawing on research from North America, Europe and Australasia, there is a wealth of evidence that suggest that a wide variety of factors influence software piracy. These can be characterized as affluence, level of human capital, level of corruption, level of inventive activity, domestic income inequality, and judicial efficiency.

Previous research found GDP to be a robust predictor of software piracy. The presumption is that the higher the income level, the lower the rate of software piracy. For example, using data for 59 nations from 2000 to 2005, Yang, Sonmez, Bosworth, and Fryxell (2009) found that economic well-being as measured by gross national income (GNI) explains between 59% and 74% of the variation in software piracy rates during the study period. In a study of 39 nations, Husted (2000) found that the higher the level of economic development, the lower the rate of software piracy. Applying a macro level panel data technique, Andres (2006a) studied factors affecting software piracy rates across 23 European countries. Results indicate that income and copyright software protection are the most important factors in determining software piracy rates in Europe. Depken and Simmons (2004) studied software piracy rates across 75 nations. The authors concluded that GDP per capita was negatively and significantly correlated with software piracy rates at the 5% significance level. Based on a study of 50 nations, Ronkainen and Guerrero-Cusumano (2001) found that purchasing power parity adjusted GNI accounted for 73% of the variation in software piracy. Burke (1996) examined software piracy across signatories to Berne, Rome and Phonographic copyright conventions. The authors concluded that GDP indirectly influences software piracy rates because it allows police and judiciary to actively implement intellectual property protection conventions. Bezmen and Depken (2006) also found that a 1% increase in per capita income in the US lead to 0.25% decrease in software piracy rate. Other empirical research that corroborates the negative impact of income on software piracy includes Yang and Sonmez (2007), Robertson, Gilly, Crittenden, and Crittenden (2008), Bagchi, Kirs, and Cerveny (2006), Banerjee, Khalid, and Sturm (2005), Gopal and Sanders (2000) and Marron and Steel (2000).

Country level results were also confirmed at the individual consumer level. For example, in a multination study including 627 consumers from Honk Kong, New Zealand, Pakistan and the US, Shore et al. (2001) found that affordability did influence consumers' decision to buy pirated products. In a study on 200 US consumers, Bloch, Bush, and Campbell (1993) concluded that pirated product buyers tend to be less "well-off" financially. Finally, Sims, Cheng, and Teegen (1996) found a negative relationship between household income and softlifting. Cheng, Sims, and Teegen (1997) found that household income is significantly related to "can't afford software" as a reason to pirate software among university students. The theoretical underpinning of the relationship between income and software piracy is best captured by the 'compensatory consumption' theory (Hill, 2002). This theory posits that financially-deficient consumers have the same materialistic goals as well-off consumers. Thus, low income consumers use a compensatory consumption strategy by purchasing pirated low priced products to achieve their materialistic goals. The previous discussion suggests the following hypothesis.

H1: The higher the GDP per capita, the lower the rate of software piracy.

There has been little research on the association between education level and software piracy rates. In general, this research provides evidence that education has a negative effect on piracy rates. In a study linking the average years of schooling and piracy rates, Marron and Steel (2000) found that education may lead to a tendency toward lower piracy rates. Yang (2007) studied piracy rates across 91 countries and found partial support for the relationship between expenditure on education and software piracy rates. However, the author concluded that "one stark contrast between