Modeling Software Piracy in Developed and Emerging Economies

Bruce A. Reinig  
Department of Information & Decision Systems  
San Diego State University  
breinig@mail.sdsu.edu

Robert K. Plice  
Department of Information & Decision Systems  
San Diego State University  
rplice@mail.sdsu.edu

Abstract

The software industry loses billions of dollars annually to software piracy and has raised awareness of the high software piracy rates worldwide, particularly in emerging economies. We build a general model of software piracy that includes three economic and social factors suggested by the literature, including per capita GNI, the relative size of a country’s IT market, and government corruption. We then test the model with respect to whether an economy is developed or emerging as designated by OECD membership and find no structural variation. However, a structural break did exist with respect to the relative size of a country’s IT market. The analysis suggests that the classification of an economy as developed or emerging is not necessarily useful for understanding the causal mechanisms that give rise to software piracy. Our findings suggest more insight can be gained by formulating strategies that take into account the relative size of a country’s IT market.

1. Introduction

“While emerging economies account for 45 percent of the global PC hardware market, they account for less than 20 percent of the PC software market. If the emerging economies’ PC software share were the same as it is for PC hardware, the software market would grow by $40 billion a year.” Business Software Alliance, May 2009 [3].

The above quotation is from a joint report of the Business Software Alliance (BSA) and International Data Corporation (IDC), reflecting their belief that software companies suffer losses of their intellectual property rights (IPR) disproportionately in emerging economies. The BSA is the largest IT industry group and is comprised of 29 major software companies, such as Adobe, Microsoft, HP, and SAP. The BSA-IDC annual software piracy report highlights decreasing, but still remarkably high, piracy rates in certain developing economies such as China (80%), Indonesia (85%), and Venezuela (86%) [3]. The BSA suggests that software piracy is inhibiting development in emerging economies and estimates that 600,000 new jobs would be created and $24 billion in tax revenue would be generated if the piracy rate could be lowered by 10 percent over the next four years [4]. The BSA and its members have worked with governments to encourage the enforcement of IPR with notable progress in key economies such as Russia and China. With billions of dollars at stake, few issues are of greater importance to the software industry than the protection of their IPR through a reduction of software piracy.

Whereas the software industry suggests that software piracy inhibits a country’s economic development in terms of per capita gross national income (GNI), a number of researchers argue the direction of this relationship is actually reversed; that is, that increases in per capita GNI reduce software piracy [e.g., 2, 8, 12]. The rationale for this perspective is straightforward: A key motivator of piracy is a desire to save money and the relative cost of software is much higher for people with low disposable income than for people with a high disposable income. We adopt this perspective in the current study, and examine the national software piracy rate as a function of per capita GNI. We also explore the effects of a strong national IT industry and government corruption, as suggested in the literature, on software piracy in both developed and emerging economies.

The objective of this paper is to first build a general model of software piracy by examining key explanatory factors from the literature. We then examine the relationship between these factors and country designations as developed versus emerging economies. Regression analysis is used to validate the model and test for structural breaks that would support the chief implication of the software industry’s analysis, namely, that an economy’s level of economic

---

1 The Center for International Business Education and Research (CIBER) at San Diego State University provided funding to support this research.
development is significant as an explanator of its piracy rate; that is, that it is necessary to model piracy differently in emerging and developed economies. We also examine differences in IT intensive versus IT non-intensive, and high corruption versus low corruption economies.

2. Theory

We begin by taking national software piracy rate as our dependent variable and then include those independent variables that are suggested by the literature. Software piracy is defined by the BSA as the unauthorized copying or distribution of copyrighted software. We are interested not in the absolute amount of software that is pirated in a country, but rather the ratio of pirated software to total software deployed. Thus, our dependent variable is software piracy rate and ranges from zero (all of the piracy in use is authorized) to unity (all of the software in use is unauthorized).

We focus on three factors that have been identified as explanators of software piracy at the country level, namely, per capita GNI, relative size of the IT industry, and the level of government corruption [1, 2, 8, 9]. Because wealthier people have less need to pirate software than poorer people, we expect that software piracy will be a negative function of per capita GNI. Per capita GNI has no theoretical upper bound, so we expect this relationship to be nonlinear. That is, additional increases in GNI cannot reduce piracy below zero, and thus the relationship between the two variables is likely to be asymptotic and approaching zero at high levels of per capita GNI. Such a relationship would best be described by a negative and increasing slope. This assertion is consistent with the findings of Gopal and Sanders [9], who tested for a negative linear relationship between software piracy and per capita GNI. They reported a steeper slope for economies with a per capita GNI of less than $6,000 as compared to economies with a per capita GNI greater than $6,000. We model the relationship through all levels of per capita GNI by including a non-linear component. Thus, our first hypothesis:

H1: National software piracy rate is a negative, curvilinear function of per capita GNI. The impact of additional increases in per capita GNI diminishes as per capita GNI becomes large (i.e., the slope is negative and asymptotically approaching zero).

Some countries have a greater incentive to reduce software piracy than others. Gopal and Sanders [8] argued that a government’s incentive to protect IPR is a function of the size of the domestic software industry, regardless of income levels. It is not surprising, for example, that the country with the largest software industry, the United States, also has the world’s lowest national software piracy rate, at 20 percent [4]. The software industry also notes that an inverse relationship exists between the size of a country’s IT market and its software piracy rate [5]. Thus, our second hypothesis:

H2: National software piracy rate is a negative linear function of the relative size of a country’s IT market.

The rationale for the second hypothesis is that a country with an economy more dependent on IT relative to other industries (e.g., agriculture, manufacturing, tourism) will tend to decrease software piracy because it is in its own interest to do so. To a certain extent, this is likely to be a reciprocal relationship. That is, a country that develops an IT industry will be incentivized to support that industry by protecting IPR and IT companies may be more interested in setting up operations in countries with a reputation for protecting IPR.

Figure 1. Theoretical model of the effects of income levels, IT share of the economy, and government corruption on software piracy.

Government or public sector corruption has been identified as a non-economic factor that is positively related to software piracy [1]. Countries with a high level of corruption are less likely to implement laws designed to protect IPR and are otherwise less likely to actively seek to catch and prosecute copyright offenders [1]. Traphagan and Griffith [14] note that effective enforcement of legal protections must be in place to reduce piracy and that high incomes, in and of
themselves, are not sufficient to predict low piracy rates. Thus, our third hypothesis:

**H3**: National software piracy rate is a positive linear function of the level of corruption.

The three hypotheses are summarized in Figure 1. In the next section we discuss the measures used to test the model and explore differences between different types of economies.

### 2.1 Developed versus Emerging Economies

As noted above, a key objective of this paper is to explore whether it is necessary to model piracy differently in emerging and developed economies. For example, insight into formulating anti-piracy strategies could be gained if the model in Figure 1 must be parameterized differently according to a country’s level of development. On the other hand, if the model is structurally identical for emerging and developed economies, it would suggest that the general level of economic development is less useful than other possible factors for understanding piracy.

It can be difficult to distinguish between emerging and developed economies. Although some countries are clearly developed, such as the United States, Japan, and Germany, and other are clearly emerging, such as India, Brazil, and South Africa, the demarcation is often difficult on what is clearly a multi-dimensional continuum. For this reason, economists and sociologists often use membership in the Organization for Economic Co-operation and Development (OECD) to establish a designation of a developed economy. The OECD is a forum of 30 market democracies that work to help governments “... foster prosperity and fight poverty through economic growth, financial stability, trade and investment, technology, innovation, entrepreneurship and development co-operation [11].” The 30 OECD members produce nearly 60 percent of the world’s goods and services and are committed to a market economy and a pluralist democracy [11]. OECD membership is used as a holistic designation of developed economy status for the purpose of our study.

Although OECD membership is useful for designating a country as developed, the complement of OECD membership (i.e., non-membership) is not particularly useful for establishing a country as an emerging economy. For example, Hong Kong, Singapore, and Israel are not members of OECD but would generally be recognized as a developed economies rather than emerging ones. In recognition of this limitation, we also examine median splits on the independent variables in the model, including per capita GNI, which would place Hong Kong, Singapore, and Israel in the upper half of all countries with respect to income.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OECD (n=28)</th>
<th>Non OECD (n=34)</th>
<th>F-test for equality of $\sigma^2$</th>
<th>t-test for equality of $\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Piracy Rate (percent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M 38.14</td>
<td>S 13.09</td>
<td>M 65.74</td>
<td>S 15.75</td>
</tr>
<tr>
<td>GNI per capita ($1,000s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M 25.92</td>
<td>S 8.43</td>
<td>M 10.93</td>
<td>S 6.94</td>
</tr>
<tr>
<td>IT Sector as a percent of GDP (percent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M 2.35</td>
<td>S .78</td>
<td>M 1.54</td>
<td>S .68</td>
</tr>
<tr>
<td>Corruption Perception Index (CPI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M 6.93</td>
<td>S 2.17</td>
<td>M 4.26</td>
<td>S 1.77</td>
</tr>
</tbody>
</table>

Note: ***p<.001.
3. Analysis and Results

We gathered data on each of the variables represented in Figure 1, then tested for structural variation based on OECD membership. Software piracy rate data was obtained from the Second Annual BSA and IDC Global Software Piracy Study [6]. Piracy rate is the proportion given by dividing the unsold software units that are in use by the total amount of software units in use (both sold and unsold). Per capita GNI data was measured using purchasing power parity and was obtained from the Worldbank database [15]. Relative size of IT market was measured using the IDC Worldwide IT Spending Reports as published in the Expanding the Frontiers of our Digital Future report [5]. This measure consists of IT spending by consumers, business, governments, and educational institutions on IT relative to GDP, but not taking into account exports or imports and excluding telecommunications services [5]. Finally, government corruption was measured using the Transparency International Corruption Perceptions Index 2004 [13].

The corruption perceptions index (CPI) is a composite score developed from multiple surveys (referred to as a “poll of polls”) that ranges from 10 (low corruption) to 0 (high corruption). The CPI reflects the perceptions of business people and country analysts as to the level of corruption that exists among public officials and politicians in each country.

The most recent year in which all of these measures were freely available was 2004. The software piracy rate, CPI, and per capita GNI are published annually and publicly available, but the last year in which the IT market data was freely available was 2004 [appearing in 5]. Thus, we use 2004 data for all measures. The four sources of data intersected on 62 countries, 28 of which are members of OECD (Appendix A).

3.1 Multiple Regression Results for the Overall Model

We use multiple regression to test the hypotheses across all data and then make use of the Chow [7] test to determine whether the regression model performs differently for OECD and non-OECD countries. The software piracy rate, rounded to the nearest percent, is the dependent variable for all models. H1 is tested using a polynomial function that includes a linear component to establish the negative relationship (H1a: $\beta_{GNI} < 0$) and a quadratic component to establish that the relationship is best described by a negative but increasing slope (H1b: $\beta_{GNI^2} > 0$). H2 was tested by including the IT sector as a percent of GDP (H2: $\beta_{IT} < 0$). H3 was tested by including the CPI index as a variable in the model (H3: $\beta_{CPI} < 0$).

The regression results are presented in Table 2. The overall model is statistically significant and accounts for 82.8 percent of the variance in the software piracy rate among the 62 countries (adjusted R^2=.816). H1a, H1b, and H2 were supported at a p<.01 or p<.001 level of significance and H3 was a marginal finding at p=.054. Thus, there was support for the overall model and varying levels of support for the individual hypotheses.

Multicollinearity diagnostics were examined to determine if the regression coefficients were influenced by high correlations among the predictor variables. For example, one would expect corruption and GNI per capita to be correlated because high levels of corruption are typically associated with low levels of per capita GNI. Variance inflation factors (VIFs) exceeded the recommended threshold of 10 [10] for the GNI and GNI squared variables. The VIFs for all variables are below 10 when the GNI quadratic term is eliminated from the model. A subsequent analysis was conducted using the square root of per capita GNI to represent the relationship between software piracy rate as a negative but increasing function and the results were comparable to those appearing in Table 2. Further, no VIF in this subsequent analysis exceeded 5.0. Thus, it was concluded that multicollinearity was a problem created by the inclusion of the quadratic term and not something inherent in the variables themselves.

3.2. Testing for Structural Variation

Although the model gives insight into understanding the factors that influence software piracy, a primary objective of this paper is to determine whether the model performs equivalently in developed versus emerging economies. We begin by conducting a Chow [7] test to determine whether a structural difference exists for the regression model using OECD countries compared to the same model with non-OECD countries. The null hypothesis assumes that the vector of parameter estimates is equal for both groups. In the present case, we have five parameter estimates, including the intercept (i.e., $\beta_0$, $\beta_{GNI}$, $\beta_{GNI^2}$, $\beta_{IT}$, $\beta_{CPI}$). The alternative hypothesis is that one or more of the five parameter estimates is unequal. There was not a statistically significant structural variation based on OECD membership ($F_{(5,52)}=1.63$, p=.168), indicating that the observations for OECD and non-OECD
Table 2. Regression results using all data (n=62)

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>B</th>
<th>s.e.</th>
<th>Model fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>99.73***</td>
<td>4.35</td>
<td>$F_{(4,57)}=68.54^{***}$</td>
</tr>
<tr>
<td>GNI per capita (thousands of dollars)</td>
<td>-2.16***</td>
<td>.54</td>
<td>$R^2=.828$</td>
</tr>
<tr>
<td>GNI per capita squared</td>
<td>.03**</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>IT Sector as a percent of GDP (percent)</td>
<td>-6.47***</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td>Corruption Perception Index (CPI)</td>
<td>-1.71 (p=.054)</td>
<td>1.05</td>
<td></td>
</tr>
</tbody>
</table>

Note: **p<.01 and ***p<.001

countries can be pooled and analyzed as a single model.

We then tested the model for other possible structural variation by conducting a median split on each of the three independent variables. A structural difference based on per capita GNI, for example, would suggest that the influence of the variables on software piracy differs between wealthy (above the median) and poor (below the median) countries. However, there was not a significant structural break on a median split of per capita GNI. We conducted the equivalent test with a median split on CPI and again found no structural break (Table 3). We proceeded to test for a structural break with respect to IT sector as a percent of GDP. The test was less straightforward because there were three observations that shared the median value of 1.70. The three observations were eliminated from the analysis, and the test yielded a significant structural difference at $F_{(5,49)}=2.88$ (p=.023). The results of the four structural variation tests are presented in Table 3.

The presence of a structural difference based on the relative size of IT market indicates that the observations should not be pooled, because one or more of the parameter estimates in the two separate models are unequal. An analysis was conducted to determine where the inequality exists. First, a separate regression model was fitted for each of the two groups (i.e., high IT share countries and low IT share countries). Second, a pooled model was fitted using all of the observations and including five additional independent variables for testing the significance of the slope differentials between the two groups. These variables are set equal to the value of the corresponding observations in the high IT share countries and zero in the low IT share countries. The five variables then have parameter estimates that equal the difference between the two separate regression models, providing a test to determine if the difference is statistically significant. Two variables were found to have significantly different slope estimates between the two models: IT sector as a percent of GDP, which was significant in the high IT share countries and non-significant in the low IT share countries; and corruption, which was non-significant in the high IT share countries and significant in the low IT share countries (Table 4).

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Chow (1960) test</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD Membership</td>
<td>$F_{(5,52)}=1.63$</td>
</tr>
<tr>
<td>Median split on GNI per capita</td>
<td>$F_{(5,52)}=1.14$</td>
</tr>
<tr>
<td>Median split on IT Sector as a percent of GDP</td>
<td>$F_{(5,49)}=2.88^*$</td>
</tr>
<tr>
<td>Median split on Corruption Perception Index</td>
<td>$F_{(5,52)}=1.55$</td>
</tr>
</tbody>
</table>

Note: *p<.01

4. Discussion

Our results provide insight into the factors that influence software piracy rate and a potentially useful perspective for shaping country specific approaches to combating software piracy and protecting IPR. The overall model was supportive of each of the
hypothesized relationships in Figure 1. Software piracy is reduced by increases in per capita GNI and the relative size of a country’s IT industry. Software piracy is also reduced in countries where governments are less corrupt. These results suggest that software piracy is a complex issue and is influenced by a mix of economic and social factors.

The results provide a more nuanced perspective of the oft-cited relationship between per capita GNI and software piracy. Gopal and Sanders [9] provided empirical support of their assertion that income levels reduced software piracy rates and, further, that the impact of an additional thousand dollars of income was more pronounced for low income countries than it was for high income countries. Specifically, Gopal and Sanders [9] conducted separate analyses for countries with a per capita GDP of less than $6,000 and one for countries with a per capita GDP of greater than $6,000. Their results show that the slope for the poorer countries was six times greater than that of the wealthier countries. However, the y-intercepts for these two analyses differ substantially (88.7 for poorer countries and 67.1 for wealthier countries) leading to some odd conclusions if the models were to be used to predict software piracy (such as a country with $5,000 in per capita GDP having an estimated piracy rate of 59.2 percent and a country with $7,000 in per capita GDP having an estimated piracy rate of 60.8 percent).

We contend that there is nothing particularly special about $6000 of income and demonstrate that the relationship can be modeled with a single y-intercept if the non-linear nature of the relationship is taken into account. The marginal benefit of additional per capita GNI becomes small as per capita GNI increases. For example, an increase in per capita GNI from $4,000 to $5,000 is associated with a reduction of 1.88 percent in software piracy whereas an increase in per capita GNI from $24,000 to $25,000 is associated with a reduction of .61 percent in software piracy. Our model further validates and then builds on the work of Gopal and Sanders [9] in terms of the relationship between income and software piracy.

There are two unanticipated findings with respect to structural variation in the model. First, there is no structural difference when comparing OECD members to non-OECD members. OECD membership clearly differentiated between the mean value of each of the four variables considered (Table 1) but the estimated parameter estimates in the hypothesized relationships do not differ from one group to the next. This was also true when separating the countries by income level. This may suggest that the distinction of developed versus emerging economy is not particularly useful for understanding the complex relationships that give rise to software piracy. Much is made of this differentiation by the software industry, but in terms of software piracy it may be more useful to see where a country resides on a variety of economic and social dimensions rather than its classification as developed or emerging. These designations can be ambiguous and fail to consider the underlying causal mechanisms that give rise to software piracy. It is true that developed nations tend to enjoy higher incomes and less government corruption than emerging nations, for example, but this is not true in all cases and there exists substantial variance on these measures within developed and emerging economies.

The second unanticipated finding is the structural difference that exists for a median split on the relative IT share of the economy. That this occurred in the absence of a structural difference based on per capita GNI provides additional support for the assertion by Gopal and Sanders [8] that the relative size of a domestic software industry influences software piracy independent of income. We note, however, that we examined the relative size of the IT industry [6] whereas Gopal and Sanders [8] focused exclusively on the software industry. It is also notable that the break revealed significant differences on two parameters (Table 3). Within high IT countries, IT share is a significant predictor of software piracy rates: A one-percent increase in the relative size of the IT market would imply that over 10 percent of all software would convert from unauthorized to authorized. By contrast, relative IT share is not statistically significant in low IT countries. Corruption, on the other hand, has a nearly opposite effect. Corruption is non-significant in high IT countries while a one-point increase in CPI would result in a reduction in piracy of over four percent of total software in low IT countries.

These results suggest that different approaches to reducing piracy may be called for in high IT economies versus low IT economies. In both groups, piracy rates decline when there are increases in per capita GNI, but the software industry might do well to focus on countering government corruption in low IT countries and seeking to expand the IT market in high IT economies. At the very least, IT share appears to be a more meaningful distinction to draw between economies than does the concept of general economic development when the goal is to understand how software piracy is influenced by economic and social factors.

4.1 Limitations

There are a number of limitations in this study that should be considered when interpreting and generalizing the results. First, there is measurement error associated with each of the variables. For
example, per capita GNI fails to take into account income inequality. The mean income for a given country may not be a typical income in that country. Government corruption is difficult to measure and, as such, we need to rely on a perceptual measure rather than an observed measure. The CPI is a sound measure of perceived corruption but it is nonetheless subject to measurement error inherent in any link between perceptions and reality. Second, the data analyzed is cross-sectional and limited to one year. Additional research can examine these relationships over multiple years to see if they continue to hold. Finally, although we gathered data from as many countries as possible, our model may not accurately reflect the relationships that exist between these variables in countries not included in the model.

5. Conclusion

We report on a study that examined the influence of three factors on the software piracy as suggested by the literature. We developed a regression model that supported a hypothesized nonlinear relationship between per capita GNI and linear relationships with relative IT market and corruption with software piracy. The model performed equally well in developed versus emerging economies as designated by OECD membership. The results suggest that the demarcation of developed versus emerging was not particularly meaningful in terms how the independent variables in this study influence software piracy. Future research may examine additional independent variables to see if structural breaks exist among them. We hope that our model will contribute insight into the development of successful strategies for reducing software piracy.

| Table 4. Regression models for median split grouping on IT as a percent of GDP |
|---------------------------------|------------------|------------------|------------------|
|                                 | High IT Countries (n=30) | Low IT Countries (n=29) |
| Model Fit:                      | \( F_{(4,25)}=18.90^{***}, R^2=.867 \) | \( F_{(4,24)}=16.93^{***}, R^2=.859 \) |
| Parameter Estimates:            | \( \beta \) | s.e. | \( \beta \) | s.e. | Differential slope estimates |
| Intercept                       | 92.62^{***} | 9.89 | 89.48^{***} | 7.76 | 3.14 |
| GNI per capita (thousands of dollars) | -2.00^* | .94 | -1.71^* | .88 | -.29 |
| IT Sector as a percent of GDP (percent) | .02 | .02 | .02 | .02 | .01 |
| Corruption Perception Index (CPI) | -10.12^{**} | 3.15 | 8.64 | 6.27 | -18.76^{***} |
|                                 | .65 | 1.33 | -4.13^{**} | 1.43 | -4.78^* |

Note: Sig. levels for F tests and two-tailed t tests are indicated with ^*p<.05, ^{**}p<.01 and ^{***}p<.001
References


Appendix

Countries or regions included in the study.

OECD Members:
Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, South Korea, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States.

Non OECD Members:
Argentina, Brazil, Bulgaria, Chile, China, Colombia, Costa Rica, Croatia, Dominican Republic, Egypt, Estonia, Hong Kong, India, Indonesia, Israel, Kuwait, Latvia, Lithuania, Malaysia, Panama, Peru, Philippines, Romania, Russia, Saudi Arabia, Singapore, Slovenia, South Africa, Thailand, UAE, Ukraine, Uruguay, Venezuela, Vietnam

Note: Two OECD members (Iceland and Luxembourg) were not included in the study.