On the Reliability of Software Piracy Statistics

I.P.L. Png*

Electronic Commerce Research and Applications,
Vol 9 No 5 (September-October 2010)

Abstract

The Business Software Alliance (BSA) changed its consultant and methodology for measurement of software piracy in 2002-03. Based on a review of the methodology and empirical analysis, I conclude that the change had systematic effects on published piracy rates. The trend rate of decrease of piracy rates fell from 2.0% points per year to 1.1% points per year. The impact of the change was larger among developing than advanced countries. Software usage in countries where usage was not directly measured was projected on the basis of national income. Any government policy or academic study using the BSA software piracy statistics should take account of the BSA's change in methodology.

Key words: Piracy; Intellectual property; Software

* I thank Mark Rogers, Kai-lung Hui, Qiu-hong Wang, the co-chairs and reviewers of HICSS-43 minitrack on Competitive Strategy, Economics and Information Systems, and the ECRA editor and reviewers, and participants at seminars at the Research Institute for Economy, Trade, and Industry, Japan, and WISE 2008 for helpful discussions and advice. Financial support from the NUS Academic Research Fund (R-313-000-089-112) is gratefully acknowledged. I disclose that I am as a member of Microsoft Corporation's Trustworthy Computing Academic Advisory Board, for which I receive a consultant fee.

1. Introduction

The Omnibus Trade and Competitiveness Act of 1988, Section 1303, requires the U.S. Trade Representative to produce the Special 301 Report annually on countries that do not provide adequate and effective protection of intellectual property (IP) rights, or deny fair and equitable market access to U.S. exporters of IP-protected items.[1] Countries whose laws, policies, or practices are deemed to adversely affect U.S. producers or products may be subject to investigation, trade sanctions, or other penalties. Special 301 reports give prominent attention to piracy statistics published by the Business Software Alliance (BSA) and other intellectual property groups. The BSA (2009) reported that, in 2008, software vendors lost $53.0 billion to piracy.

The public-policy implications of piracy motivated the World Intellectual Property Organization and the Organization for Economic Cooperation and Development to commission studies to measure piracy (Olsen 2005). As well, numerous academic studies have investigated the causes of business software piracy (Gopal and Sanders 1998 and 2000, Marron and Steel 2000, Husted 2000, Depken and Simmons 2004, van Kranenburg and Hogenbirk 2005, Fischer and Rodriguez 2005, Rodriguez 2006, Chellappa et al. 2006, Robertson et al. 2008, Bezmen and Depken 2006). Major factors include national income per capita, national culture, and law and institutions.[2]

Here, I review the methodology, coverage, and implementation of the BSA software piracy statistics from 1997 to 2007. Around 2002-03, the BSA changed the

---

1 Public Law No. 100-418, 102 Stat. 1107 (1988).

2 Other academic studies have used piracy to explain various business and marketing strategies and public policy issues – including sales of recorded music (Hui and Png 2003), enforcement by software publishers (Gu and Mahajan 2005), and as a measure of intellectual property rights (Zhao 2006).

consultant responsible for compiling software piracy statistics and the new consultant revised the statistical methodology. Based on various econometric tests, I conclude that the change in methodology had systematic effects on published piracy rates. First, the trend rate of decrease of piracy rates fell from 2.0% points per year to 1.1% points per year. The impact of the change was substantially more pronounced among developing than advanced countries. Second, I infer that, in estimating software piracy in countries where software usage was not directly measured, the software usage was projected on the basis of national income.

The central implication of my analysis is that BSA statistics should be used with caution in any government policy or academic study. Comparisons of software piracy over time periods that span 2002-03 should take account of the change in methodology. International comparisons of software piracy should take account of projections based on national income.

2. Methodology

Copyright law governs the expression contained in books, music, movies, software, electronic games, databases, and designs. Inherently, these items are costly to create but relatively cheap to reproduce, whether legitimately or otherwise. Illicit copying reduces the economic return to the creator and thereby reduces the incentive to create new works (Chen and Png 2003).

Generally, the objective of piracy statistics is to measure the extent of consumption of illicit copies (whether self-made or purchased). Industry associations use piracy statistics to calculate the losses that copyright owners suffer from piracy. Table 1 reports various aspects of the piracy statistics published by international associations of producers of business software, music, electronic games, and movies. The statistics differ in the scope of coverage, frequency of publication, methodology, and sources of primary information.

The BSA statistics provide the widest geographical coverage, are published annually, and, appear to be the most transparent in terms of methodology, data sources, and implementation. The methodology is published in some detail, uses both internal and external data, and, since 2003, has been implemented by a well-reputed consultant, International Data Corporation (IDC) (BSA 2004).

--- Table 1. Industry piracy statistics ---

For 2002 and earlier years, the business software piracy statistics were produced by the International Planning and Research Corporation (IPRC) (BSA 2003). IPRC focused on three groups of business PC software – general productivity applications, professional applications, and utilities.

The IPRC estimated piracy using the following indirect methodology. For each country, the quantity of pirated software was estimated as the difference between the quantity installed and the quantity legitimately acquired. In turn, the quantity installed was estimated as the number of computers in use multiplied by corresponding norms for the "software load" in four customer segments -- new and existing residential computers, and new and existing business computers.

Software load is the quantity of software installed per computer. The norms for software load for the four segments were based on U.S. market research (BSA 2003, p. 11). However, the IPRC did not explain whether, and if so, how it adjusted the U.S. norms to compute the software load in other countries.

The IPRC directly estimated the numbers of computers in use “for the major countries … from proprietary and confidential data supplied by BSA member companies”, while “[t]he “rest of region” data was used to develop piracy estimates outside of the major markets” (BSA 2003, pp. 11-12). The IPRC did not specify the “major” markets or method by which it developed the “rest of region” data.

In 2003, the BSA engaged a new consultant, IDC, to compile the statistics of business software piracy. IDC is an international provider of market intelligence and advisory services for information technology industries. The IDC applied the same basic methodology, estimating the quantity of pirated software indirectly as the difference between the quantity installed and the quantity legitimately acquired, with the quantity installed estimated as the number of computers in use multiplied by norms for the software load.

The IDC refined the methodology in three ways (BSA 2004, pp. 9-12). First, it expanded the scope of measurement to include operating systems and consumer applications. Second, the IDC calculated the software loads from surveys of consumers and business users in 15 countries -- Bolivia, Brazil, Chile, China, Colombia, Costa Rica, Dominican Republic, Guatemala, Kuwait, Malaysia, Mexico, Romania, Spain, Taiwan and the United States.[3] IDC did not explain whether it adjusted the norms from the 15 countries to compute the software load in the other countries, and, if so, how it made the adjustments.[4] Third, the IDC distinguished the software load according to new, existing, versus retired computers, shareware and open source, and Windows and non-Windows operating systems.

As for the numbers of computers, IDC collected information on PC shipments for "more than 75 countries", while for the "additional 25-plus countries and markets, the data were either collected in-country or modeled regionally based on IDC's rest-of-region estimates" (BSA 2004, p. 10). The IDC did not elaborate on how it modeled the number of computers in the other 25-plus countries.[5]

Overall, comparing the IPRC and IDC approaches, it seems that the IDC methodology and data were more comprehensive. The IDC covered more categories of software, accounted for open source and shareware, and measured the software load in 15 countries. Further, the IDC was more transparent about its methodology (4.5 pages) as compared with the IPRC (just 2 pages).

From the viewpoint of policy and research, the key questions are: (i) whether the change in consultant from IPRC to IDC, and the consequent revision of the methodology had any systematic impact on piracy statistics; and (ii) whether the methods used by BSA

[3] IDC did not explain the selection of the 15 countries for calculating the software loads. In terms of geography, Europe and Asia seem under-represented, while Central and South America seem over-represented. In terms of economic development, advanced countries seem under-represented and less developed countries seem over-represented.

[4] Subsequently, the IDC expanded the survey. By 2008, it encompassed 24 countries (BSA 2009, p. 17).

[5] By contrast, the IPRC was silent on whether they actually measured the numbers of each segment of computers in every country, or somehow projected these numbers.

to project the software load and number of computers from countries in which they compiled these data to other countries had any systematic impact on piracy statistics.

To address these questions, I compiled national piracy rates for 81 to 103 countries over the eleven-year period 1997-2007 from BSA publications (the panel began with 81 countries in 1997 and ended with 103 countries in 2007). The period of study included seven years before and five years after the change in methodology. Accordingly, it would provide good coverage of any impact of the change. For brevity, I refer to the years, 1997-2002, before the change as "pre-change" and the years, 2003-07, after the change as "post-change".

3. Piracy Trend

For a high-level preview, Figure 1 illustrates the evolution of the average piracy rate. Evidently, the average piracy rate declined by about 2% points each year until 2002. Between 2002 and 2003, the average piracy rate jumped by over 2% points, and thereafter, continued a downward trend, but at a much lower rate of decline than until 2002.

--- Figure 1. Average piracy rate ---

The break in the trend decrease of piracy coincided exactly with the change in methodology. Table 2 reports various econometric tests to confirm the impact on national piracy rates. Models (1) and (2) were simple ordinary least-squares (OLS) regressions of the piracy rate on year indicators and pre- and post-change year trends respectively.[6] The results from both models suggest that piracy rates followed a significant downward trend.

Referring to model (1), the year indicators show a clear decreasing trend in the average piracy rate from 1999-2002. The results from model (2) are even sharper: prior to the change in consultant and methodology, the average piracy rate fell by 2.0% points annually, while, after the change, the average rate fell at the slower rate of 0.7% points annually. The difference between the pre-change and post-change trends was statistically significant ($F(1, 989) = 15.30$, $p < 0.0001$).

--- Table 2. Piracy trend ---

The OLS regression pooled all countries, regardless of economic, institutional, or cultural differences, into a single estimate. Obviously, it would be more appropriate to account for any systematic national differences. Models (3) and (4) replicated the analysis, using country fixed effects. The fixed effects are intended to account for any systematic national differences which did not vary over time. The results were even stronger than with the OLS estimates.

Referring to model (3), the year indicators show a clear decreasing trend from 1998-2002, and then an upward shift by about 1.7% points between 2002-03. Referring to model (4), the trend rate of decrease of piracy rates was significantly higher before the change in methodology (–2.0% points per year) than after the change (–1.1% points per

[6] The data exhibited serial correlation within panels ($F(1, 102) = 142.78$, $p < 0.0001$, using the test proposed by Wooldridge (2002, p. 282), as implemented by Drukker (2003)). Accordingly, all standard errors for fixed-effects estimates in Tables 2, 6, and 7 were calculated by country-level cluster.

year). The difference between the pre-change and post-change trends was statistically significant (F (1, 102) = 60.18, $p < 0.0001$).

One possible reason why the downward trend of piracy rates decelerated around 2002-03 was the expansion of BSA coverage to include countries with higher piracy rates. Specifically, in 1997, the BSA piracy statistics covered 81 countries, while in 2007, they covered 103 countries. The coverage was expanded to include countries with relatively high piracy, such as Albania (78%), Kazakhstan (79%), and Zambia (82%), where the numbers in parentheses are the respective piracy rates in 2007.

To avoid any bias due to the expanded coverage, models (5) and (6) limited the fixed effects estimates to those countries covered throughout the period, 1997-2007. The results from the balanced sample were similar. Referring to model (6), the pre-change trend rate of decrease of piracy rates (–2.1% points per year) was significantly higher than the post-trend change (–1.1% points per year) (F(1, 80) = 58.55, $p < 0.0001$).

To give context to this change in trend, Figure 1 shows, in square dots, the projection of the average piracy rate from 2003 onwards, based on the year trend, as reported in Table 1, model (6). In 2007, the average piracy rate was 38%. If the average piracy rate had continued its pre-2002 downward trend, then, by 2007, it would have been (2.064 – 1.148) x 5 = 4.58% points lower, that is, about 33.4% rather than 38%, which is a substantial disparity.

To further explore the impact of the BSA's change of its consultant and the related methodology, Figure 2 shows the trend in the average piracy rate, for countries with above and below the median national income per capita as of 1997. This was motivated by the various studies showing that piracy was correlated with income (Husted 2000, Marron and Steel 2000, Gopal and Sanders 1998, Depken and Simmons 2004, Fischer and Rodriguez 2005, Rodriguez 2006). Figure 2 suggests that the change in the trend decline in piracy rates was steeper among lower-income countries.

--- Figure 2. Piracy rates: Advanced versus developing countries ---

To investigate the difference in the impact by national income in a rigorous way, I conducted quantile regressions for the piracy rate. Table 3 reports the results for the 25th percentile, median, and 75th percentile regressions. In all three, the post-change trend was smaller than the pre-change trend. The difference was largest and statistically significant in the 75th percentile regression for the segment with highest piracy rates ($F(1, 888) = 8.58$, $p < 0.005$). The difference was smaller and also statistically significant in the median regression ($F(1, 888) = 6.73$, $p < 0.01$). These tests confirm the intuition from Figure 2 that the impact of the BSA's change in methodology was largest among the lower-income countries, which are those with the higher piracy rates.

--- Table 3. Piracy trend: Quantile regressions ---

It is also interesting to note from Table 3 that the standard errors of the post-change trends were smaller than the standard errors of the pre-change trends. Apparently, after the change in methodology, the trends in piracy rates across countries tended to be less dispersed.

Could the change in the trend decrease of piracy rates, particularly among countries with lower incomes, be due to changes in some factor other than the BSA's change in methodology? In the sections below, I present various statistical tests to rule out this possibility. Specifically, I included other covariates which have been shown to influence software piracy – income per capita, national culture, and institutions. I found the same conclusion: the trend rate of decrease of piracy rates was significantly higher before the change in methodology than after the change.

One possible explanation for the change in trend that I could not rule out is that software publishers worldwide substantially stepped up enforcement against piracy around 2002-03 and persisted with the revised enforcement policy thereafter. However, a review of the BSA publications and a meeting with a BSA official did not provide any evidence of a change in enforcement policies around that time.

4. Projection of Software Usage

Formally, the methodology applied by BSA's consultants, IPRC and IDC, was to estimate the quantity pirated, P_{it}, in country i during year t as the difference between usage of software, U_{it}, and the quantity of software legitimately acquired, S_{it},

$$P_{it} = U_{it} - S_{it}\,. \qquad (1)$$

In essence, the usage was computed as

$$U_{it} = \lambda_i N_{it}\,, \qquad (2)$$

where λ_{it} was the norm for software load and N_{it} was the number of computers in country i in year t.[7]

[7] Equation (2) is a simplification of the IPRC/IDC methodology, which distinguishes four segments of computers – new and existing business computers, and new and existing home computers. The following analysis can be generalized to take account of the four segments.

The *piracy rate* in country *i* for year *t* was then calculated as the ratio of the pirated quantity to usage,

$$r_{it} = \frac{P_{it}}{U_{it}} = \frac{U_{it} - S_{it}}{U_{it}} = 1 - \frac{S_{it}}{U_{it}}. \tag{3}$$

As noted above, neither IPRC nor IDC revealed whether, and if so, how they adjusted the norms for software load from the sampled countries to compute the norms in other countries. Further, IDC did not reveal how it projected the number of computers in the sampled countries to other countries. IPRC did not even disclose whether it used any projection. The interesting and important question is how they generated these estimates.

Multiple academic studies have pointed to income as being the single most important influence on the rate of software piracy (Husted 2000, Marron and Steel 2000, Gopal and Sanders 1998, Depken and Simmons 2004, Fischer and Rodriguez 2005, Rodriguez 2006). Referring to equation (3), these observations are consistent with income being an important determinant of both legitimate sales and usage.

However, Marron and Steel (2000, p. 162) did query the observed correlation between software piracy and income: “the trade groups’ estimation procedure involves significant assumptions. This raises the question of whether our empirical analysis might uncover artifacts of the estimation procedure rather than true relationships among the variables. For example, did the analysts assume that high-income countries have lower piracy rates?”

Income would be a very intuitive basis on which to project the number of computers and software load. To investigate whether the number of computers and software load were projected on the basis of income, I used instrumental variable methods. Instrumental variable methods are used to estimate relationships where one or more explanatory variables are endogenous to the dependent variable. In the present case, as I explain below, if the number of computers or software load were projected on the basis of income, then, in a regression of the rate of legitimate consumption on income, income would be endogenous.

First, I used equation (3) to obtain the rate of *legitimate consumption* is

$$c_{it} = 1 - r_{it} = \frac{S_{it}}{U_{it}}. \tag{4}$$

In logarithms, this becomes

$$\ln c_{it} = \ln S_{it} - \ln U_{it}. \tag{5}$$

Referring to (2), suppose that software usage was projected as an increasing function of income, and specifically, in the reduced form,

$$U_{it} = Y_{it}^{\alpha} \tag{6}$$

with $\alpha > 0$. Then, substituting from (6) in (5), the logarithm of the legitimate consumption rate, $\tilde{c}_{it}$, as estimated by BSA would be

$$\ln \tilde{c}_{it} = \ln S_{it} - \alpha \ln Y_{it}. \tag{7}$$

Consider a regression of $\ln \tilde{c}_{it}$ on $\ln Y_{\text{it}}$,

$$\ln \tilde{c}_{it} = \ln S_{it} - \alpha \ln Y_{it} = \beta_1 \ln Y_{it} + \beta_2 X_{it} + \varepsilon_{it}, \tag{8}$$

where β_1 and β_2 are coefficients, and X_{it} represents other covariates which might affect legitimate consumption. By (7), the error term, ε_{it}, in (8), would be correlated with the explanatory variable, $\ln Y_{\text{it}}$. Equivalently, income would not be exogenous. The essential reason is the projection of software usage on the basis of income.

A Hausman test (Wooldridge 2006, pp. 532-533) of the endogeneity of income in (8) would proceed as follows. Perform first stage regressions for every income variable – regressing every income variable on instruments for the income variables, Z_{it}, and exogenous variables, X_{it}, that affect piracy.

$$\ln Y_{it} = \gamma_1 X_{it} + \gamma_2 Z_{it} + \eta_{it}, \tag{9}$$

where γ_1 and γ_2 are coefficients, and η_{it} is a random error. Extract the residuals, u_{it}, from the first-stage regression,

$$u_{it} = \ln Y_{it} - \gamma_1 X_{it} - \gamma_2 Z_{it}. \tag{10}$$

In the second stage, regress the rate of legitimate consumption on the income variables, exogenous variables that affect piracy, and the residuals from the first-stage regressions,

$$\ln \tilde{c}_{it} = \theta_1 \ln Y_{it} + \theta_2 X_{it} + \theta_3 u_{it} + \upsilon_{it}, \tag{11}$$

where θ_1, θ_2, and θ_3 are coefficients. If the first-stage residuals are significant in the second-stage regression, then the corresponding first-stage regressor is not exogenous.

The implication is that, indeed, legitimate consumption and hence piracy rates were based on income.[8] [9]

-- Table 4. Mathematical notation --

-- Table 5. Descriptive statistics --

-- Table 6. Correlations --

Tables 4 and 5 present descriptive statistics of the data and pairwise correlations respectively. The key issue in any instrumental variables analysis is the choice of instrument. Generally, the instrument must be correlated with the endogenous explanatory variable but not correlated with the dependent variable. I identified per capita residential electricity consumption as an instrument. Residential electricity consumption would depend on income, but, intuitively, there would not seem to be any direct relation between electricity consumption and legitimate consumption of software, except to the trivial extent that software is used on computers and computers consume electricity.

Referring to Table 6, residential electricity consumption was relatively more correlated with income than with piracy.[10] Since legitimate consumption is the

---

[8] Information systems researchers have applied Granger causality in situations of an endogenous explanatory variable (e.g,. Dutta (2001) and Barry et al. (2006)). However, econometricians emphasize that Granger causality should *not* be used as a test of endogeneity, for example, "has nothing to say about contemporaneous causality … does not allow us to determine whether z_t is an exogenous or endogenous variable in an equation relating y_t to z_t" (Wooldridge 2006, p. 660), and see also Enders (2004, pp. 283-284). Nevertheless, for completeness, I conducted the Granger causality tests: for 20 of 77 countries, the null hypothesis of no Granger causality from lagged income to the legitimate consumption rate could not be rejected.

[9] Table 4 summarizes the mathematical notation used in this paper.

complement of piracy, this would also imply that residential electricity consumption was relatively more correlated with income than with legitimate consumption.

Table 7 reports tests of endogeneity, using post-change residential electricity consumption per capita as the instrument for post-change income per capita, in two sub-samples. One sub-sample comprised the 15 base countries in which the IDC surveyed consumers and business users to compute norms for software loads and measure computer ownership. The other sub-sample comprised all other countries, for which the IDC somehow projected the norms for software loads and modeled computer ownership. (In Tables 6 and 7, except as otherwise noted, all variables except year trends were specified in logarithms (Wooldridge 2006, pp. 197-200), and all regressions included country fixed effects, and to be conservative, were limited to the balanced panel.)

--- Table 7. Endogeneity of income ---

Referring to Table 7, models (2) and (3), in the first-stage regression of post-change income, the coefficient of post-change electricity consumption was significant, while in the second-stage regression, the coefficient of the residuals from the first-stage regression was significant. This suggests that, over the years 2003-07, following the change in methodology, income was endogenous.[11] This finding was consistent with my

---

[10] Referring to Table 6, the correlation between electricity consumption and piracy was -0.589. This was probably a spurious correlation due to the separate correlations between income and electricity consumption (0.765) and between income and piracy (-0.815). Indeed, in a regression of the logarithm of the legitimate consumption rate on the logarithms of income and electricity, the coefficient of the logarithm of income was significant but the coefficient of the logarithm of electricity was not significant.

[11] In the similar Hausman test (unreported) of both income and post-change income, only the residuals of post-change income were significant in the second stage. Accordingly, I inferred that only post-change income was endogenous and that income *per se* was not endogenous.

conjecture that the norm for software load or number of computers was projected on the basis of income per capita.

For completeness, Table 7, model (4), reports the 2SLS (2-stage least squares) estimates. The coefficient of post-change income was positive and significant (0.098 (±0.029)) while the post-change trend was -0.086 (±0.031). By comparison with the OLS estimate in model (1), the major differences were that the coefficient of post-change income and the post-change trend were about double in magnitude. Apparently, not only did the change in methodology affect the trend in piracy rates, but it also caused legitimate software consumption to become apparently more sensitive to income per capita. There seems to be no obvious reason for this except the assumptions underlying the BSA consultant's projection of software usage.[12]

The significance of post-change income also helps to confirm that the endogeneity of income was due to the consultant's projection of software usage. An alternative explanation – that both income and usage were related to some omitted factor (e.g., computer literacy or education) – is implausible as there is no obvious reason for the relation between usage and the hypothetical omitted factor to have changed around 2002-03.

By investigating the relation between legitimate consumption and income in the base countries, I can further characterize the impact of the projection of software usage. Recall that IDC calculated the software loads from surveys of users in 15 base countries,

[12] I explored two other instruments – exports per capita and infant mortality, but found them to be weak instruments.

while using projections for the other countries. Accordingly, among the base countries, income should not be endogenous.

Table 4, models (5) to (8), reports estimates for the base countries. Referring to model (7), the first-stage residuals were not significant in the second-stage regression. Consistent with my argument above, post-change income was not endogenous among the base countries. This provides indirect corroborating evidence for my conjecture that the number of computers or software load were projected on the basis of income.

Since income is exogenous to legitimate consumption among the base countries, the most appropriate estimate for the base countries would be the OLS regression in model (5). Comparing this estimate for the base countries with the 2SLS regression for the projection countries, reported in model (3). I make two observations.

- The post-change trend, which was marginally significant at -0.040 (± 0.020), among the base countries was more than 2 standard errors smaller than the post-change trend, -0.086 (± 0.031), among the projection countries.

- The coefficient of post-change income, which was marginally significant at 0.056 (± 0.026) among the base countries, was about 1.5 standard errors smaller than the coefficient, 0.098 (± 0.029), among the projection countries.

The disparity between the estimates for the base vis-à-vis projection countries suggests two systematic biases in the IDC’s measurement of piracy from 2003 onward.

Apparently, the projection of software usage on the basis of income caused (i) a stronger downward trend in legitimate consumption, and (ii) legitimate consumption to be more sensitive to income.[13]

By equation (4), the piracy rate is just the complement of the legitimate consumption rate. Accordingly, I infer that IDC's projection of software usage resulted in the trend decline in the piracy rate being lower and the sensitivity of piracy to changes in income being higher among the projection countries relative to the base countries.

Finally, I note that I did not find any evidence that income was endogenous to published rates of software piracy between 1997-2002 among the projection countries or the base countries. This is consistent with the disclosure by the consultant, IPRC, that it directly applied U.S. norms for software loads to other countries, without any adjustment (BSA 2003, p. 11).

5. Robustness

In previous studies of the determinants of business software piracy (e.g., Gopal and Sanders 1998, Husted 2000, Marron and Steel 2000, Rodriguez 2006, and Fischer and Rodriguez 2005), three factors stood out as being robust to alternative specifications, geographical coverage, and time periods.[14] They were national income per capita, individualism (a dimension of national culture), and law and institutions.

[13] While the IPRC did disclose that it estimated the number of computers "outside of the major markets" (BSA 2003, pp. 11-12), it did not reveal the respective markets, hence I could not conduct a similar test of endogeneity for the years before 2002 as with the years after 2003.

[14] Of the multiple studies, only Fischer and Rodriguez (2005) included country fixed effects. All others applied cross-section analysis without considering variation over time.

Could the variation in one or more of these factors possibly explain the changes in the pattern of piracy rates and their relation with income around 2002-03 or the endogeneity of income? In the following discussion, I consider the impact of including these other factors relative to the 2SLS estimate in Table 7, model (4), which is treated as a benchmark. The robustness checks, presented in Table 8, used post-change residential electricity consumption per capita as the instrument for post-change income per capita.[15]

The results reported in Table 7 suggest that changes in income could not account for the changes in the pattern of piracy rates around 2002-03. Indeed, the empirical results suggest the opposite – that the relationship between piracy rates and income itself changed around 2002-03.

The sociologist, Geert Hofstede (1983, 2001) famously developed four indexes of national culture from surveys of IBM employees over the period 1967–1973 in 40 countries. The indexes were later extended to other countries. One index measured individualism and its complement, collectivism. Hofstede's (1983, 2001) indexes are static. Accordingly, the individualism index could not be used to account for changes in piracy rates or their relation with income over time. However, could differences in individualism account for the endogeneity of income?

Table 8, model (1), reports a 2SLS regression of legitimate consumption including individualism as an additional covariate. To complete each country's panel for individualism, the same value was replicated for every year, hence, fixed effects could

[15] I also conducted the corresponding robustness checks using the entire sample, comprising both projection and base countries. The results (unreported) were very close to those reported in Table 8.

not be used. In the estimates, legitimate consumption was significant and increasing in individualism. The coefficient of post-change income was positive, about one-third of the magnitude as in the benchmark, but was not precisely estimated. This imprecision might be due to the high correlation between income and individualism. Moreover, with individualism being constant over time, I could not use fixed effects, hence, the estimates were driven by across-country variation as well as within-country variation.[16] For these reasons, the estimate including individualism should be treated with caution.

--- Table 8. Robustness (projected countries) ---

With regard to law and institutions, the obvious measure is the World Bank's *rule of law index*, which was also used by Fischer and Rodriguez (2005) and Chellappa et al. (2006). This index is compiled from multiple primary sources, and is a perceptual measure of "the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence" (Kaufmann et al. 2007, p. 4). The index was compiled biennially from 1996-2000, and annually from 2002 onward.

Table 8, model (2), reports a 2SLS regression of legitimate consumption including the rule of law index. The results were qualitatively similar to the benchmark estimate: the coefficient of post-change income was positive and significant, and the legitimate consumption rate was subject to an increasing trend before the change in

[16] Fixed-effects estimates are essentially estimates relative to the mean value of the dependent variable for each country, and hence, driven by within-country variation.

methodology but a decreasing trend afterward. The rule of law itself was not significant.[17]

In the previous section, I mooted the possibility that post-change income appeared to be endogenous in regressions of legitimate consumption because usage was determined by an omitted factor, such as computer literacy or education, which was correlated with income. Table 8, models (3)-(4), reports 2SLS regressions of legitimate consumption including the percentage of personal computer users among the population and the literacy rate respectively. The results were very close to benchmark: the coefficient of post-change income was positive and significant, and legitimate consumption was subject to an increasing trend before the change in methodology but a decreasing trend afterward. Neither computer usage nor the literacy rate was significant.[18]

6. Conclusions

U.S. government pronouncements and actions as well as many academic studies have taken BSA software piracy statistics at face value. Based on a review of the BSA methodology and empirical analysis, I conclude that a change in the BSA consultant and methodology around 2002-03 had systematic effects on published piracy rates.

[17] One possible reason why the coefficient of income was lower in this estimate than in the benchmark is that improvements in the rule of law might raise both income and legitimate consumption of software. Hence, in the benchmark estimate, with the rule of law excluded, the income variable would absorb part of the effect of the rule of law.

[18] In another unreported test, I obtained similar results using the percentage of Internet users among the population.

- The trend rate of decrease of piracy rates fell from 2.0% points per year to 1.1% points per year, so, raising piracy rates from the levels that would have been implied had they followed the trend before the change.
- The magnitude of the impact on piracy rates was larger for developing as compared with advanced countries.
- In countries for which software usage was not directly measured, it was projected on the basis of national per capita income.

These results were robust to the sample of countries, choice of instruments, and inclusion of alternative explanatory variables.

The key direction for future research is to gain access to the BSA methodologies and data so as to better understand the biases in their statistics, and so that future policy and research can be appropriately calibrated. Meanwhile, the central implication of my analysis is that government policy and academic study should use BSA statistics with caution. Comparisons of software piracy over time periods that span 2002-03 should take account of the change in methodology. International comparisons of software piracy should take account of projections based on national income.

References

Bagchi, K., Kirs, P., and Cerveny, R. Global software piracy: can economic factors alone explain the trend? *Communications of the ACM*, 49, 6, 2006, 70-76.

Barry, E. J., Kemerer, C.F., and Slaughter, S.A. Environmental volatility, development decision, and software volatility: A longitudinal analysis. *Management Science*, 52, 3, 2006, 448-464.

Bezmen, T. L., and Depken, C.A. Influences on software piracy: Evidence from the various United States, *Economics Letters*, 90, 3, March 2006, 356-361.

Business Software Alliance. *First Annual BSA and IDC Global Software Piracy Study*, BSA Worldwide Headquarters, Washington, DC, 2004.

Business Software Alliance, *Third Annual BSA and IDC Global Software Piracy Study*, BSA Worldwide Headquarters, Washington, DC, 2006.

Business Software Alliance, *Sixth Annual BSA and IDC Global Software Piracy Study*, BSA Worldwide Headquarters, Washington, DC, May 2009.

Business Software Alliance, *Eighth Annual BSA Global Software Piracy Study: Trends in Software Piracy, 1994-2002*. BSA Worldwide Headquarters, Washington, DC, June 2003.

Chellappa, R.K., Chen, Y., and Venkataraman, S. An empirical examination of global software piracy: Pricing and policy implications. Working paper, Goizueta Business School, Emory University, Atlanta, GA, 2006.

Chen. Y.N., and Png, I.P.L. Information Goods Pricing and Copyright Enforcement: Welfare Analysis, *Information Systems Research*, 14, 1, 2003, 107-123.

Depken, C.A., and Simmons, L.C., Social construct and the propensity for software piracy, *Applied Economics Letters*, 11, 2, February 2004, 97–100.

Drukker, D.M. Testing for serial correlation in linear panel-data models, *Stata Journal*, 3, 2, 2003, 168–177.

Dutta, A. Telecommunications and economic activity: An analysis of Granger causality. *Journal of Management Information Systems*, 17, 4, 2001, 71-95.

Enders, W. *Applied Econometric Time Series*, 2nd Ed. John Wiley and Sons, New York, NY, 2004.

Fischer, J.A.V., and Rodríguez Andrés, A. Is software piracy a middle class crime? Investigating the inequality-piracy channel. Discussion paper 2005-18, Department of Economics, University of St. Gallen, Switzerland, August 2005.

Gopal, R.D. and Sanders, G.L. Global software piracy: you can't get blood out of a turnip. *Communications of the ACM*, 43, 9, 2000, 82 – 89.

Gopal, R.D. and Sanders, G.L. International software piracy: Analysis of key issues and impacts. *Information Systems Research*, 9, 4, 1998, 380-397.

Gu, B., and Mahajan, V. How much anti-piracy effort is too much? – A study of the global software industry, Red McCombs School of Business, University of Texas at Austin, 2005.

Hofstede, G. Dimensions of national cultures in fifty countries and three regions. In J. B. Deregowski, S. Dziurawiec and R.C. Annis (eds.), *Explications in Cross-Cultural Psychology*, Swets and Zeitlinger, Lisse, Netherlands, 1983, 333–355.

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

Hui, K.L., and Png, I.P.L. Piracy and the legitimate demand for recorded music. *Contributions to Economic Analysis and Policy*, 2, 1, 2003, Article 11 (www.bepress.com/bejeap/contributions/vol2/iss1/art11/).

Hui, K.L., and Png, I.P.L. Measuring piracy of IP-protected goods: A methodology. World Intellectual Property Organization, Working Group of Experts on Developing a Methodology for Measuring Copyright Piracy, Economic Analysis Associates LLP, Singapore, November 1-4, 2005 (www.econaa.com/resources/methodology_piracy_rate_sector.pdf).

Husted, B.W. The impact of national culture on software piracy. *Journal of Business Ethics*, 26, 3, 2000, 197-211.

Kaufmann, D., Kraay, A., and Mastruzzi, M. Governance matters IV: Aggregate and individual governance indicators, 1996–2006. Policy research working paper 4280, World Bank, New York, NY, July 2007.

Marron, D.B., and Steel, D.G. Which countries protect intellectual property? The case of software piracy. *Economic Inquiry*, 38, 2, 2000, 159-174.

Olsen, K. Counterfeiting and piracy: Measurement issues. Background report for the WIPO/OECD Expert Meeting on Measurement and Statistical Issues, Geneva, Switzerland, October, 17-18, 2005.

Robertson, C.J., Gilley, K.M., Crittenden, V., and Crittenden, W.F. An analysis of the predictors of software piracy within Latin America. *Journal of Business Research*, 61, 6, 2008, 651-656.

Rodríguez Andrés, A. The relationship between copyright software protection and piracy: Evidence from Europe. *European Journal of Law and Economics*, 21, 1, 2006, 29-52.

Shin, S.K., Gopal, R.D., Sanders, G.L., and Whinston, A.B. Global software piracy revisited. *Communications of the ACM*, 47, 1, 2004, 103-107.

van Kranenburg, H., and Hogenbirk, A. Multimedia, entertainment, and business software copyright piracy: A cross-national study. *Journal of Media Economics*, 18, 2, 2005, 109-129.

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2002.

Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, 3rd ed. Thomson Southwestern, Mason, OH, 2006.

Zhao, M. Conducting R&D in countries with weak intellectual property rights protection. *Management Science*, 52, 8, 2006, 1185–1199.

Author's contact information

Professor Ivan Png
NUS Business School
National University of Singapore
Mochtar Riady Building, BIZ1-6-38
15 Kent Ridge Drive
Singapore 119245
Singapore

Tel: +65 6516 6807
Fax: +65 6779-5059
iplpng@gmail.com and ipng@nus.edu.sg

About the author

Dr Ivan Png is Lim Kim San Professor in the NUS Business School and Professor of Economics and Information Systems at the National University of Singapore. His research focuses on the economics of intellectual property, privacy, and pricing. He is the author/co-author of Managerial Economics, which has been published in multiple editions. Dr Png is an independent director of Hyflux Water Trust Management Pte Ltd and Healthway Medical Corporation Ltd. He was a nominated MP (10th Parliament of Singapore), 2005-06.

Table 1. Industry piracy statistics

	Business Software Alliance	International Federation of the Phonographic Industry	Entertainment Software Alliance	Motion Picture Association
Coverage (countries and territories)	97	73	Varies[1]	22[2]
Precision of statistics	2 digits, eg, "37%"	Bands (< 10%, 10-24%, 25-50%, >50%)	2 digits, eg, "37%"	2 digits, eg, "37%"
Frequency	annual	annual	annual	2004 only
Methodology	Disclosed in detail	Disclosed broadly	Disclosed broadly	Disclosed broadly
Raw data	Not disclosed	Not disclosed	Not disclosed	Not disclosed
Data sources	Internal (software sales) and external (computer sales)	Internal (member associations)	Internal and external (consumer survey)	Internal and external (consumer survey)
Compilation and analysis	Third party (International Data Corp)	Association itself	Association itself	Third party (LEK)

Notes:

1. This Table relies, in part, on Hui and Png (2005).
2. The Entertainment Software Alliance published statistics only for countries subject to potential investigation under Special 301.
3. The Motion Picture Association study estimated piracy in an additional 42 countries by projection from survey results in 22 countries.

Table 2. Piracy trend

VARIABLES	(1) Years - all OLS	(2) Trend - all OLS	(3) Years - all	(4) Trend - all	(5) Years - balanced	(6) Trend - balanced
yr98	-1.756		-1.756***		-1.765***	
	(3.045)		(0.296)		(0.300)	
yr99	-5.098*		-5.098***		-5.160***	
	(3.045)		(0.414)		(0.415)	
yr00	-7.463**		-7.553***		-7.617***	
	(3.035)		(0.594)		(0.599)	
yr01	-8.533***		-8.595***		-8.667***	
	(3.009)		(0.713)		(0.719)	
yr02	-10.579***		-10.641***		-10.716***	
	(3.009)		(0.812)		(0.824)	
yr03	-5.781*		-8.324***		-8.247***	
	(2.968)		(1.035)		(1.032)	
yr04	-4.730		-8.097***		-8.222***	
	(2.925)		(1.141)		(1.152)	
yr05	-4.223		-8.552***		-8.605***	
	(2.898)		(1.177)		(1.193)	
yr06	-5.065*		-9.394***		-9.395***	
	(2.898)		(1.207)		(1.231)	
yr07	-6.758**		-10.568***		-10.519***	
	(2.898)		(1.231)		(1.270)	
Pre-change trend		-2.032***		-2.034***		-2.064***
		(0.484)		(0.155)		(0.160)
Post-change trend		-0.733***		-1.138***		-1.148***
		(0.224)		(0.133)		(0.137)
Constant	65.451***	66.886***	67.302***	68.703***	65.815***	67.279***
	(2.153)	(1.881)	(0.719)	(0.811)	(0.689)	(0.800)
Observations	992	992	992	992	891	891
R-squared	0.019	0.017	0.302	0.285	0.300	0.285
No. of countries			103	103	81	81

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: Average piracy rate; Robust cluster standard errors in parentheses; Regressions (iii)-(vi) included fixed effects.

Table 3: Piracy trend: Quantile regressions

	(1) 25%	(2) Median	(3) 75%
Pre-change trend	-1.667*	-2.333***	-3.000***
	(0.915)	(0.726)	(0.775)
Post-change trend	-1.296***	-1.030***	-1.444***
	(0.429)	(0.340)	(0.371)
Constant	51.667***	69.333***	86.000***
	(3.527)	(2.795)	(3.018)
Observations	891	891	891

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: Average piracy rate; Balanced panel;
Standard errors in parentheses; All regressions included fixed effects

Table 4. Mathematical notation

Notation	Definition
P_{it}	quantity pirated
U_{it}	software usage
S_{it}	quantity of software legitimately acquired
λ_{it}	the norm for software load
N_{it}	number of computers
r_{it}	piracy rate
c_{it}	rate of legitimate consumption
α	coefficient
$\tilde{c}_{it}$	BSA estimate of the rate of legitimate consumption
β_1, β_2	coefficients
X_{it}	other covariates which might affect legitimate consumption
ε_{it}	random error
Z_{it}	instruments for income
γ_1, γ_2	coefficients
η_{it}	random error
u_{it}	residuals from first-stage regression
θ_1, θ_2, and θ_3	coefficients

Note: i represents county and t represents year

Table 5. Descriptive statistics

	Units	Source	Observations	Mean	Std dev	Min.	Max.
Business software piracy	%	BSA	992	59.98	19.59	20	98
GDP per capita (income)	'000 USD (2000 prices)	World Bank	998	9.331	11.029	0.300	54.178
Residential electricity consumption per capita	kWh per capita	GMID[(a)]	980	1.280	1.466	0.021	7.955
Individualism		Hofstede (1983, 2001)	726	44.333	23.622	6.000	91.000
Rule of law		Kauffman et al. (2007)	818	0.285	0.991	-1.677	2.051
PC users per capita		GMID[(a)]	945	0.163	0.180	0.002	0.767
Literacy rate	%	GMID[(a)]	1042	88.550	14.376	34.600	99.900

[(a)] Euromonitor, Global Market Information Database

Table 6. Correlations

	Piracy	GDP per capita	Residential electricity consumption per capita	Individu alism	Rule of law	PC users per capita	Literacy rate
Piracy	1.000						
GDP per capita	-0.815	1.000					
Residential electricity consumption per capita	-0.589	0.765	1.000				
Individualism	-0.767	0.720	0.568	1.000			
Rule of law	-0.842	0.850	0.671	0.777	1.000		
PC users per capita	-0.802	0.879	0.653	0.706	0.827	1.000	
Literacy rate	-0.440	0.464	0.309	0.386	0.489	0.515	1.000

Table 7. Endogeneity of income

	Projected countries				15 base countries			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	1st stg	2nd stg	2SLS	OLS	1st stg	2nd stg	2SLS
Log GDP per capita	1.309***	-3.237***	1.439***	1.439***	1.357***	-2.847**	1.414**	1.414***
	(0.360)	(0.977)	(0.354)	(0.322)	(0.420)	(1.024)	(0.481)	(0.451)
Log post-change	0.058***		0.098***	0.098***	0.056*		0.074	0.074
GDP per capita	(0.015)		(0.028)	(0.029)	(0.026)		(0.080)	(0.076)
Pre-change trend	0.057***	-0.049**	0.059***	0.059***	0.062**	-0.054***	0.063**	0.063***
	(0.011)	(0.021)	(0.012)	(0.011)	(0.024)	(0.016)	(0.024)	(0.023)
Post-change trend	-0.045**	1.041***	-0.086***	-0.086***	-0.040*	1.029***	-0.058	-0.058
	(0.020)	(0.025)	(0.031)	(0.031)	(0.020)	(0.026)	(0.076)	(0.072)
Log post-change		1.030***				0.771***		
electricity		(0.058)				(0.129)		
1st stage residuals			-0.066**				-0.024	
			(0.027)				(0.080)	
Constant	-7.852**	27.992***	-8.978***		-8.017**	23.732**	-8.493**	
	(3.063)	(8.298)	(3.010)		(3.408)	(8.408)	(3.924)	
Observations	626	626	626	626	139	139	139	139
Number of countries	63	63	63	63	14	14	14	14
R-squared	0.350	0.972	0.366		0.453	0.972	0.455	

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: Log legitimate consumption rate; Instrument: Log residential electricity consumption; Robust cluster standard errors in parentheses; All regressions included fixed effects

Table 8. Robustness (projected countries)

VARIABLES	(1) Individualism	(2) Rule of law	(3) PC Users	(4) Literacy
Log GDP per capita	0.268***	0.395*	1.431***	1.425***
	(0.067)	(0.219)	(0.364)	(0.365)
Log post-change GDP per capita	0.019	0.039***	0.068***	0.075***
	(0.047)	(0.012)	(0.014)	(0.019)
Pre-change trend	0.064***	0.045***	0.042***	0.051***
	(0.010)	(0.011)	(0.015)	(0.013)
Post-change trend	0.010	-0.017	-0.070***	-0.071**
	(0.049)	(0.011)	(0.022)	(0.028)
Log Individualism	0.227***			
	(0.086)			
Log Rule of Law		0.040		
		(0.025)		
Log PC users per capita			0.085	
			(0.106)	
Log literacy rate				1.851
				(2.031)
Constant	0.253			
	(0.562)			
Observations	509	253	581	558
Number of countries		41	60	59

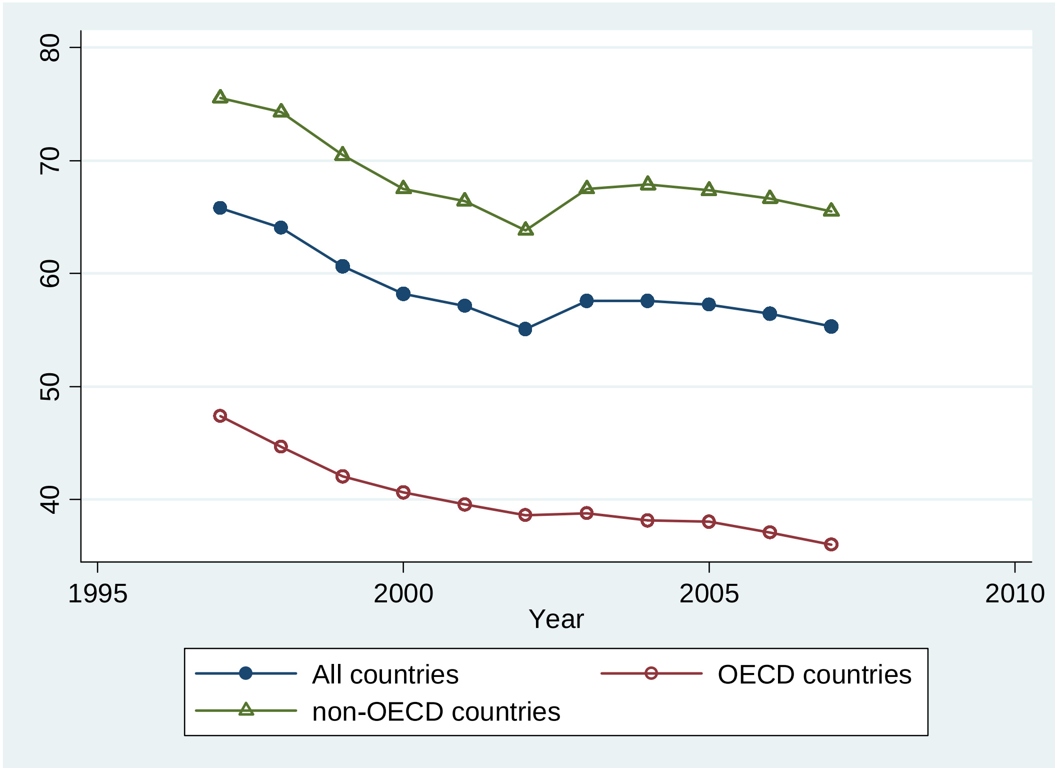
*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: Log legitimate consumption rate; Instrument: Log residential electricity consumption; Robust standard errors in parentheses; Models (2), (3) and (4) included fixed effects

Figure 1. Average piracy rate

Note: Figure 1 was drawn for the balanced panel of 81 countries, for which the BSA published piracy statistics throughout the period of study.

Figure 2. Piracy rates: Advanced versus developing countries



Note: Figure 2 was drawn for the balanced panel of 81 countries, for which the BSA published piracy statistics throughout the period of study.