On the Reliability of Software Piracy Statistics

I.P.L. Png*

Revised, March 2008

Abstract

Despite tremendous debate and policy interest in software piracy, the accuracy of piracy statistics compiled by the Business Software Alliance (BSA) has been accepted at face value.

Based on a review of the BSA methodology and empirical analysis, I conclude the following. First, prior cross-country studies of software piracy in the years 2002 and earlier were mis-specified: they more likely explained the demand for legitimate software relative to computers in use rather than piracy rates.

Second, BSA statistics were biased on a cross-country basis either in the years 2002 and earlier, or the years 2003 and after, or both.

Third, from 2003 onward, following a change in the BSA consultant and methodology, piracy rates across countries were inflated by an average of almost 4% points.

* Department of Information Systems, National University of Singapore, 3 Science Drive 2, Singapore 117543, Tel: +65 6516-6807; http://www.comp.nus.edu.sg/~ipng/. I thank Hui Kai-lung, Barrie Nault, V. Sambamurthy, and participants at seminars at the Research Institute for Economy, Trade, and Industry, Japan, for helpful discussions and advice, and the Ministry of Health, Welfare, and Labor of Japan for financial support, administered through the Institute of Economic Research, Hitotsubashi University. This version expands the data analysis from 49 countries in the previous version to 86 countries.

1. Introduction

Major copyright-protected industries -- business software, movie, music, electronic games, books, broadcasting -- publish statistics of piracy against their respective industries. These statistics play an important role in public policy. They are cited by industry to argue for more government enforcement and increased penalties against piracy.

Piracy statistics play a central role in U.S. international trade policy. Special 301 of the U.S. Trade Act of 1974 requires the U.S. Trade Representative to 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. 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. Piracy statistics reported by the International Intellectual Property Alliance (IIPA) are the key evidence in Special 301 reviews. See, for instance, IIPA (2007).

Piracy statistics are systematically cited in free trade negotiations between the United States and other countries. For instance, Table 7-1 of the U.S. International Trade Commission report on the economic effects of the U.S.-Chile free trade agreement quotes IIPA statistics of piracy in Chile for 2002: movies -- 40%, recorded music -- 35%, business software -- 51%, and electronic games -- 78%.[1]

[1] U.S. International Trade Commission, U.S.-Chile Free Trade Agreement: Potential Economywide and Selected Sectoral Effects, Investigation No. TA-2104-5, USITC Publication 3605, June 2003.

The U.S. Trade Representative regularly cites industry piracy statistics in its annual reports on foreign countries. For instance, the 2007 report on Singapore states: "According to industry estimates, Singapore's piracy rate averaged 5 percent for music and 12 percent for movies. Business software losses were estimated at nearly $86 million in 2005".[2]

The public-policy implications of piracy have motivated numerous academic studies into the causes of piracy. These have focused on business software piracy and investigated how it differs across countries (e.g., Gopal and Sanders 1998; Marron and Steel 2000; Hufsted 2000; van Depken and Simmons 2004; Kranenburg and Hogenbirk 2005; Rodriguez 2005; Fischer and Rodriguez 2005; Rodriguez 2006; Chellappa et al. 2006; Robertson et al. 2007) and within the United States (Bezmen and Depken 2006).[3]

The widespread use of piracy statistics in public policy and academic research prompts the question: How accurate and reliable are these statistics, and in particular, are they biased in any way?

In this paper, I review the methodology, coverage, and implementation of piracy statistics published by the Business Software Alliance (BSA). Further, I investigate and find evidence of systematic bias in time and across countries in the published piracy statistics. I also show that most prior studies of software piracy were mis-specified: the

[2] U.S. Trade Representative, 2007 report on Singapore, http://www.ustr.gov/assets/Document_Library/Reports_Publications/2007/2007_NTE_Report/asset_upload_file129_10979.pdf [Accessed, November 10, 2007].

[3] Other academic studies have used piracy to explain various business and marketing strategies and public policy issues – including sales of legitimate software (Givon et al. 1995) and recorded music (Hui and Png 2003), enforcement by software publishers (Gu and Mahajan 2005), and the incentive to produce movies (Png and Wang 2007).

BSA methodology of estimating software piracy rates was such that these prior studies actually estimated the legitimate demand for software relative to computers in use.

2. Methodology

Copyright law governs the expression contained in, *inter alia*, books, music, movies, software, electronic games, databases, and designs. Industry associations for producers of business software, music, electronic games, and movies publish international statistics of piracy. The statistics differ in the scope of coverage, frequency of publication, methodology, and sources of primary information.

In Table 1, I report various aspects of the piracy statistics published by the four industry associations.[4] The statistics published by the Business Software Alliance (BSA) provide the widest geographical coverage (97 countries), 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, viz., International Data Corporation (IDC) (BSA 2006).

--- 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

[4] This Table relies, in part, on Hui and Png (2005).

focused on three groups of business PC software – general productivity applications, professional applications, and utilities.

The IPRC estimated piracy using an indirect methodology. For each country, the quantity of pirated software was estimated as being 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, page 11).

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" countries.

In 2003, BSA engaged a new consultant, IDC, and the methodology underlying the business software piracy statistics was refined. IDC expanded the scope of measurement to cover all PC software, including operating systems, systems software such as databases and security packages, and general and specific applications software.

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. In turn, the quantity installed was estimated as the number of computers in use multiplied by norms for the software load.

By contrast with the IPRC which applied U.S. norms for the software loads, the IDC calculated the software loads per new computer and per existing computer 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 (BSA 2006, page 17).[5]

IDC explained that the "results of these surveys were used to populate IDC's input models for the other countries. For 2004 and 2005, IDC updated these models based on additional local country research, interviews with users and the channels, and spot inventories" (BSA 2006, page 17).

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 2006, page 17).

With regard to methodology, IDC has been more transparent (5 pages) than the previous consultant, IPRC (just 2 pages).

From the viewpoint of policy and research, the key question is whether the change in consultant from IPRC to IDC, and the revision of the estimation methodology had any

[5] 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.

impact on piracy statistics. Table 2 reports two regressions of the rate of business software piracy between 2000-05 on year indicators.[6] I focused on the period 2000-2005 as it included equal numbers of years before and after the change in consultant and methodology.

--- Table 2: BSA methodology revision ---

Table 2, column (i), reports a regression of piracy rates on a constant and year indicators. The coefficient of the indicator for year 2002 was negative and significant. Table 2, column (ii), reports a regression with just one indicator – for years 2003 and after. The coefficient of the indicator was 3.955(±1.688), suggesting that the change in consultant and methodology was associated with a systematic increase in piracy rates by almost 4% points.

Compared to the mean piracy rate of 58.66% during the period, the impact of the change in consultant and methodology was quite substantial. The implication is that any government pronouncement or action, and study of piracy should take account of changes in the BSA consultant and methodology.

3. Software piracy studies

The choice between buying and copying a copyrighted item depends on a trade-off between the price of the item against the consequences of possible enforcement and ethical costs, which trade-off might depend on user's economic characteristics and values

[6] Regressions reported in Tables 2 and 6 used ordinary least squares.

(Chen and Png 2003), infrastructure, and technology to protect intellectual property rights (Sundararajan 2004). Hence, given the copyright owner's pricing, enforcement efforts, and protection technology, the end-user's choice depends on five groups of factors – user's economic characteristics, personal values, law and enforcement, and infrastructure.

Scholars in fields as diverse as information systems, economics, and ethics have studied the determinants of software piracy using the BSA statistics. Table 3 summarizes these studies.

--- Table 3: Studies of software piracy ---

Gopal and Sanders (1998) hypothesized that piracy would be lower in wealthier countries and those with a larger domestic software industry. They regressed piracy rates in 13 countries between 1995-96 (means 58% and 53%) on national income, as measured by GDP per capita (means $21,643 and $22,081), and size of domestic software industry relative to GDP (means 14.88×10^{-6} and 12.22×10^{-6}). They found support for both hypotheses.

Husted (2000) hypothesized that piracy would depend on both economic factors – income levels and distribution – as well as personal values. The sociologist, Geert Hofstede (2001) famously developed four indexes of national culture from surveys of over 72,000 IBM employees over the period 1967–1973 in 40 countries (Hofstede 1983, 2001). The indexes measured individualism (and its complement, collectivism), power distance, masculinity, and uncertainty avoidance. The indexes were later extended to other countries.

Husted (2000) regressed piracy rates in 39 countries for 1996 (mean 0.61) on GNP per capita (mean $12,169), income inequality (mean Gini coefficient 31.11%), individualism (mean 46.15), power distance (mean 58.36), masculinity (mean 49.54), and uncertainty avoidance (mean 61.10). Among the explanatory variables, only income, income inequality, and individualism were significant.

Marron and Steel (2000) regressed piracy in 77 countries between 1994-97 (mean 71%) on GDP per capita (mean $10,000), individualism (mean 4.4),[7] the International Country Risk Guide composite index of property and contract institutions (mean 6.7), expenditures on R&D relative to GDP (mean 1.1), and average years of schooling in population aged 25 and older (mean 6.7). They found that piracy was negatively associated with income, individualism, institutions, and education.

Shin et al. (2004) updated the Gopal and Sanders (1998) and regressed piracy rates in 49 countries for 1999 (mean 53%) on GDP per capita (mean $14,800) and collectivism (mean 56). Their results were consistent with previous studies. In addition, they found that income and collectivism had stronger effects among low-income countries (those with GDP per capita less than $6,000 a year).

Depken and Simmons (2004) regressed piracy in 65 countries for 1994 (mean 75.19%), on GDP per capita (mean $8,806), individualism (mean 42.35), power distance (mean 61.01), interaction between individualism and power distance, literacy (mean 2232.7), the inflation rate (59.84%), and other variables.[8] Consistent with previous

[7] Since the lowest value of Hofstede's individualism index was 6 (for Guatemala), this mean was probably a typographical error.

[8] The high mean inflation rate was probably due to an outlier with an inflation rate of 2,700%.

studies, piracy was decreasing in income and individualism. The new result was that piracy decreased with inflation. However, Depken and Simmons (2004) did not check whether the latter result depended on including an outlier country with a 2,700% inflation rate.

van Kranenburg and Hogenbirk (2005) investigated the determinants of piracy of four copyrightable products – business software, recorded music, movies, and entertainment software. They regressed piracy rates in 35 countries between 1999-2001 on GDP, Euromoney's country risk rating (which ranges from 1, representing very high risk, to 100, representing no risk), and a customized index of copyright law. They did not report the means of variables. Piracy increased with country risk and decreased with the index of copyright law. GDP was not significant, but it should be noted that GDP was specified as an aggregate rather than on a per-capita basis.

Rodriguez (2006) used a fixed-effects estimator to regress piracy rates in 23 European countries for 1994, 1997, and 2000 (mean 56%) on GDP per capita (mean $18,526), a customized index of copyright protection for software (mean 1.523), and other explanatory variables. Piracy decreased with income and the copyright index.

Fischer and Rodriguez (2005) investigated the impact of income inequality on piracy. They regressed the logarithm of piracy rates in 71 countries between 1994-2002 (mean 4.11) on the logarithm of GNP per capita (mean 7.83), and the square of the logarithm of GNP per capita, income inequality (mean Gini coefficient 39.43), individualism (mean 41.52), and two measures of institutions – the World Bank Rule of

Law index (mean 0.51) and the Ginarte and Park (1997) index of intellectual property rights (mean 3.00).

The "Rule of Law" 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" (Kauffman et al. 2007, page 4). The index has been compiled biennially from 1996-2000, and annually from 2002 onward.

Fischer and Rodriguez's (2005) study was impressive for its detail and care, including various robustness checks and using 2-stage least squares to account for the possible endogeneity of income inequality. They found that piracy increased with log income and decreased with the square of log income, income inequality, individualism, and institutions.

Bagchi et al. (2006) regressed piracy rates in 37 countries for 1996, 2001, and 2003 on GDP per capita, individualism, the Transparency International index of corruption, measures of IT infrastructure and internet penetration, and other explanatory variables. Piracy decreased with individualism and corruption, but GDP per capita was only partly significant. Bagchi et al. (2006) did not report any statistics – means, coefficients or standard errors.

Chellappa et al. (2006) modeled the user's decision to copy software as a two-stage process which depends on the price of the item, law and enforcement, and ethical costs. They used factor analysis to construct an index of legal costs from the World

Bank's Rule of Law Index and Control of Corruption Index (Kauffman et al. 2007), and an index of moral costs from Hofstede's indexes of collectivism, power distance, and the interaction between collectivism and power distance.

Similar to the Rule of Law index, the Control of Corruption index is a perceptual measure of "the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests" (Kauffman et al. 2007, page 4).

Chellappa et al. (2006) regressed the logarithm of the ratio of the piracy rate to the legitimate demand for 53 countries between 1994-2004 (mean -0.43) on the index of legal costs (mean 2.00), the index of moral costs (mean 3.00), and other explanatory variables. Piracy decreased with both legal and moral costs.

Robertson et al. (2007) focused on economic determinants of software piracy in Latin America. They regressed piracy in 20 countries between 2000-04 (mean 69.1%) on gross national income per capita (mean $2,712.5), foreign direct investment, collectivism (mean 18.47), internet users per capita (81.80%), and other explanatory variables. Surprisingly, they found that income was only partly significant – perhaps because it was highly correlated with internet usage. In addition, collectivism was not significant.

Generally, all previous studies of the determinants of business software piracy accepted BSA statistics at face value. They focused on developing hypotheses regarding economic, cultural, and technical determinants of software piracy and testing the hypotheses using the BSA statistics.

Referring to Table 3, in the prior studies, three factors – national income per capita, individualism, and institutions/law/enforcement – stand out. The empirical effect of these three factors appeared to be robust to alternative specifications, geographical coverage, and time periods.

4. Specification

Formally, the methodology applied by BSA consultants, IPRC, in years up to and including 2002 was to estimate quantity pirated in country *i* for year *t* as

$$P_{it} = \lambda_{bn} B_{nit} + \lambda_{be} B_{eit} + \lambda_{hn} H_{nit} + \lambda_{he} H_{eit} - S_{it}\,, \qquad (1)$$

where λ_{bn}, λ_{be}, λ_{hn}, and λ_{he} were the norms for software load on new and existing computers for business and home users respectively, B_{nit}, B_{eit}, H_{nit}, H_{eit}, were the numbers of new and existing computers in use among business and home users in country *i* respectively, and S_{it} was the quantity of software legitimately acquired in country *i*. IPRC then calculated the piracy *rate* in country *i* for year *t* as the ratio of the pirated quantity to total consumption,

$$r_{it} = \frac{P_{it}}{\lambda_{bn} B_{nit} + \lambda_{be} B_{eit} + \lambda_{hn} H_{nit} + \lambda_{he} H_{eit}} = \frac{\lambda_{bn} B_{nit} + \lambda_{be} B_{eit} + \lambda_{hn} H_{nit} + \lambda_{he} H_{eit} - S_{it}}{\lambda_{bn} B_{nit} + \lambda_{be} B_{eit} + \lambda_{hn} H_{nit} + \lambda_{he} H_{eit}}$$
$$= 1 - \frac{S_{it}}{\lambda_{bn} B_{nit} + \lambda_{be} B_{eit} + \lambda_{hn} H_{nit} + \lambda_{he} H_{eit}}. \qquad (2)$$

Generally, the prior studies of the determinants of business software piracy estimated models of the form,

$$r_{it} = f(\text{VARIABLES}_{it}), \qquad (3)$$

where f is a function. However, substituting from (2), the preceding equation becomes

$$1 - \frac{S_{it}}{\lambda_{bn}B_{nit} + \lambda_{be}B_{eit} + \lambda_{hn}H_{nit} + \lambda_{he}H_{eit}} = f(\text{VARIABLES}_{it}),$$

which simplifies to

$$\frac{S_{it}}{\lambda_{bn}B_{nit} + \lambda_{be}B_{eit} + \lambda_{hn}H_{nit} + \lambda_{he}H_{eit}} = g(\text{VARIABLES}_{it}), \qquad (4)$$

where $g(.) \equiv f(.) - 1$.

Note that, in the methodology as implemented by IPRC, the norms for software load, λ_{bn}, λ_{be}, λ_{hn}, and λ_{he} were constants that did not vary across countries. By comparing (3) with (4), the obvious and very serious implication is that the prior studies had not explained business software piracy. Rather they had explained the sales of legitimate software relative to the stocks of new and existing computers, or equivalently, the demand for legitimate software relative to the stocks of new and existing computers.

This is clearest in the simple case where $\lambda_{bn} = \lambda_{be} = \lambda_{hn} = \lambda_{he} = \lambda$, and so, (4) simplifies to

$$\frac{S_{it}}{B_{nit} + B_{eit} + H_{nit} + H_{eit}} = \lambda \cdot g(\text{VARIABLES}_{it}). \qquad (5)$$

That the IPRC applied different norms to the various segments is secondary. Because the same norms were applied to each country, the variable truly being explained was not the piracy rate but rather the demand for legitimate software relative to computer stocks.

Intriguingly, Marron and Steel (2000) did anticipate this issue: “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?" However, Marron and Steel dismissed the concern "[a]fter speaking with the consultant who prepared the trade groups' figures" (footnote 8).

Husted (2000) also considered the limitations of BSA statistics, but concluded that they were reliable for *cross-country* study as there was "no indication that BSA has an interest in demonstrating lower rates of piracy in certain areas rather than in others" (pp. 207-208).

However, with effect from 2003, BSA engaged new consultants, IDC, and revised the methodology. Formally, the IDC methodology was to estimate quantity pirated in country *i* for year *t* as

$$P_{it} = \lambda_{bni} B_{nit} + \lambda_{bei} B_{eit} + \lambda_{hni} H_{nit} + \lambda_{hei} H_{eit} - S_{it}, \qquad (6)$$

where λ_{bni}, λ_{bei}, λ_{hni}, and λ_{hei} were the norms for software load on new and existing computers for business and home users in country *i* respectively, and the remaining variables are as defined above. IDC then calculated the piracy *rate* in country *i* for year *t* as the ratio of the pirated quantity to total consumption,

$$r_{it} = 1 - \frac{S_{it}}{\lambda_{bni} B_{nit} + \lambda_{bei} B_{eit} + \lambda_{hni} H_{nit} + \lambda_{hei} H_{eit}}. \qquad (7)$$

Hence, with the IDC methodology, the equation corresponding to (4) would be

$$\frac{S_{it}}{\lambda_{bni} B_{nit} + \lambda_{bei} B_{eit} + \lambda_{hni} H_{nit} + \lambda_{hei} H_{eit}} = g(\text{VARIABLES}_{it}). \qquad (8)$$

By (7), since the norms for software load vary by country and year, the software piracy rates produced by IDC were not merely the complement of the demand for legitimate software relative to computer stocks. To the extent that the norms for software load were accurately calibrated for the respective countries, the software piracy rates should be more accurate than those produced in earlier years by the IPRC.

To check my criticism of the studies using pre-2003 software piracy rates, I estimated specifications (4) and (8) using the three most robust "determinants" of business software piracy as identified in the prior studies, *viz*., income per capita, individualism, and institutions/law/enforcement. Table 4 reports summary statistics of the data, while Table 5 reports pairwise correlations.

-- Table 4: Descriptive statistics --

-- Table 5: Correlations --

Before discussing the results, it should be emphasized that my regressions differ from the previous studies in terms of the dependent variable. While the previous studies specified the dependent variable to be the software piracy rate, my regressions specified the dependent variable to be the demand for legitimate software relative to computers in use.

Figures 1(a) and 1(b) depict scatter plots of the demand for legitimate software against income, as measured by GDP per capita, for 2000-02 and 2003-05 respectively. Apparently, the demand for legitimate software was increasing with income, which is consistent with software being a normal good. Moreover, the relationship seemed to

differ: in the later period, it seemed to be steeper, the intercept was smaller, and the fit better than in the earlier period.

-- Figure 1(a): Income and legitimate demand, 2000-02 --

-- Figure 1(b): Income and legitimate demand, 2003-05 --

Table 6, columns (a) and (b), report the corresponding regressions. The regressions confirmed the impressions from the scatter plots. In the later period, the demand was more sensitive to income (coefficient of 1.799 (±0.0528) vis-à-vis 1.293 (±0.0780)), the intercept was smaller (15.99 (±0.831) vis-à-vis 25.72 (±1.291)), and the fit better (R^2 of 0.786 vis-à-vis 0.605) than in the earlier period.

--- Table 6: Demand for legitimate software ---

In Table 6, column (c), I checked whether the difference in the relationship between the legitimate demand and income between the two time periods was statistically significant by adding an interaction between GDP per capita and an indicator for years 2003-05. The constant and coefficient of GDP per capita were identical to those in column (a). Importantly, the coefficient of the interaction was 0.506 (±0.0950) and the coefficient of the year 2003-05 indicator was -9.737 (± 1.535), both of which were significant. The evidence suggests that the change in consultant and revision in methodology from 2003 onward were associated with the legitimate demand becoming apparently more sensitive to income.

By (2) and (7), piracy rates are just the complement of the legitimate demand. Hence, the results reported in Table 6, column (c), suggests that the change in consultant

and revision in methodology from 2003 onward were associated with piracy rates becoming *apparently* more sensitive to income.

In prior studies, piracy rates were negatively related with Hofstede's index of individualism. Hofstede (1983, 2001) produced only one set of indexes, which I deemed to apply to the year 2000. Figure 2 depicts a scatter plot of the demand for legitimate software against individualism. Apparently, the legitimate demand increased with individualism.

-- Figure 2: Individualism and legitimate demand, 2000 --

Table 6, column (d), reports the corresponding regression. As Hofstede's index did not vary over time, the regression was a pure cross-section analysis. The coefficient of individualism was 0.584 (±0.0583), which was significant, confirming that the legitimate demand increased with individualism.

Referring to Table 5, GDP per capita and individualism were highly correlated. Accordingly, if I had attributed the Hofstede individualism index to all years, 2000-05, and regressed a specification similar to Table 6, column (c), including a year 2003-05 indicator and interaction with individualism, the results would be similar. The indicator would have been negative and the coefficient of individualism would have been larger in the later period.

In prior studies, piracy rates were negatively related to various measures of institutions, law, and enforcement. Copyright law is that most pertinent to the demand for legitimate software. Park (2003) pioneered a comprehensive cross-country index of

copyright law. This index was based on a detailed analysis of national copyright law on four dimensions – coverage, duration, restrictions, and treaties. Reynolds (2003) revised and updated the index, and extended its coverage to range from 37 countries in 1965 to 146 countries in 2002. The index is not available for 2003 and after.

Figure 3 depicts a scatter plot of the demand for legitimate software against Reynolds' (2003) copyright index for the years 2000-02. Apparently, the legitimate demand increased with the strength of copyright law.

-- Figure 3: Copyright and legitimate demand, 2000-02 --

Table 6, column (d), reports the corresponding regression. The coefficient of copyright index was 50.44 (±6.005), which was significant, confirming that the legitimate demand increased with the strength of copyright law.

Reynolds' (2003) copyright index characterized the state of law. However, the extent of piracy depends on both law and its enforcement. With regard to respect for and enforcement of law, the obvious measure is the World Bank's Rule of Law index, which was used by Fischer and Rodriguez (2005) and Chellapa et al. (2006).

Figures 4(a) and 4(b) depict scatter plots of the demand for legitimate software against the "Rule of Law" index for 2000-02 and 2003-05 respectively. Apparently, the legitimate demand increased with the rule of law. Moreover, the relationship seemed to differ: in the later period, it seemed to be steeper, the intercept was smaller, and the fit better than in the earlier period.

-- Figure 4(a): Rule of law and legitimate demand, 2000-02 --

-- Figure 4(b): Rule of law and legitimate demand, 2003-05 --

In Table 6, column (h), I checked whether the difference in the relationship between the legitimate demand and the rule of law between the two time periods was statistically significant by adding an interaction between rule of law and an indicator for years 2003-05. The constant and coefficient of rule of law were identical to those in column (f). Importantly, the coefficient of the interaction was 3.909 (±0.988) and the coefficient of the year 2003-05 indicator was -3.527 (±1.225), both of which were significant.

The evidence suggests that the change in consultant and revision in methodology from 2003 onward were associated with the legitimate demand and piracy rates becoming apparently more sensitive to the rule of law.

Based on my review of the BSA methodology and empirical analysis, it seems apparent that prior cross-country studies of software piracy in the years 2002 and earlier were more likely explaining the demand for legitimate software relative to computers in use rather than piracy rates *per se*.

Moreover, the extent of the "explanation" – whether in terms of income or institutions/law/enforcement or other factors – shifted between 2002-03 when the BSA changed its consultant and methodology. The conclusion must be that the "explanation" in either the studies of 2002 and earlier, or those of 2003 and after were biased, or indeed, all were biased.

5. Conclusions

U.S. government pronouncements and actions as well as prior 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:

- Prior cross-country studies of software piracy in the years 2002 and earlier were more likely explaining the demand for legitimate software relative to computers in use rather than piracy rates;

- BSA statistics were biased on a cross-country basis either in the years 2002 and earlier, or the years 2003 and after, or both;

- From 2003 onward, following a change in the BSA consultant and methodology, piracy rates across countries were inflated by an average of almost 4% points.

In prior research, two authors did briefly discuss the reliability of BSA statistics. Husted (2000), pp. 207-08, felt that the BSA had no reason to bias statistics across countries. By contrast, I found evidence of systematic bias. This might not have been intentional, but rather the outcome of BSA assumptions regarding software load.

Marron and Steel (2000), footnote 8, asked rhetorically then dismissed the possibility that any correlation between piracy rates and income had been induced by BSA methodology. My empirical analysis seems to suggest that this might very well have been true.

Marron and Steel's (2000) rhetorical question points to an important direction for future research. Based on the IPRC methodology (BSA 2003), the U.S. software loads were applied to all other countries. The U.S. norms are likely to have been higher than actual software loads in other countries. Referring to (4), the IPRC methodology would have resulted in the ratio of legitimate demand to computers in use in other countries being under-estimated. Hence, as I show in Figure 5, the true relation between the legitimate demand and income would have been gentler, as illustrated by the broken line.

-- Figure 5: Estimation bias – Income --

For years 2003 and after, following the change in consultant to IDC and revision of the methodology, different software loads were applied to each country. This should have involved smaller software loads for countries poorer than the U.S. However, the resulting statistics yielded a relation with a smaller intercept and steeper slope (Table 6, column (c)), as illustrated in Figure 5 by the dotted line. This result seems inconsistent with the expected result based on the methodology as stated by IPRC.

The important 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 key implication of my analysis is that BSA statistics should be used with great caution.

References

Bagchi, Kallol, Peeter Kirs, Robert Cerveny, "Global software piracy: can economic factors alone explain the trend? *Communications of the ACM*, Vol. 49 No. 6, June 2006, 70-76.

Bezmen, Trisha L., and Craig A. Depken II, "Influences on software piracy: Evidence from the various United States", *Economics Letters*, Vol. 90, No. 3, March 2006, 356-361.

Business Software Alliance, *Third Annual BSA and IDC Global Software Piracy Study*, May 2006, http://www.bsa.org/globalstudy/upload/2005%20Piracy%20Study%20-%20Official%20Version.pdf [Accessed, July 27, 2006].

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

Depken II, Craig A., and Lee C. Simmons, "Social construct and the propensity for software piracy", *Applied Economics Letters*, Vol. 11 No. 2, February 2004, 97–100.

Chellappa, Ramnath K., Yuanyuan Chen, and Sriram Venkataraman, "An Empirical Examination of Global Software Piracy: Pricing and Policy Implications", Goizueta Business School, Emory University, Atlanta GA 30322, 2006.

Chen, Yeh-ning, and I.P.L. Png, "Information Goods Pricing and Copyright Enforcement: Welfare Analysis", *Information Systems Research*, Vol. 14 No. 1, March 2003, 107-123.

Fischer, Justina A.V., and Antonio Rodríguez Andrés, "Is Software Piracy a Middle Class Crime? Investigating the Inequality-Piracy Channel", Discussion paper no. 2005-18, Department of Economics, University of St. Gallen, August 2005.

Ginarte, Juan C., and Walter G. Park, "Determinants of patent rights: a cross-national study", *Research Policy*,Vol. 26 No. 3, October 1997, 283–301.

Givon, Moshe, Vijay Mahajan, and Eitan Muller, "Software Piracy: Estimation of Lost Sales and the Impact of Software Diffusion", *Journal of Marketing*, Vol. 59, No. 1, January 1995, 29-37.

Gopal, Ram D., and G. Lawrence Sanders, "Global software piracy: you can't get blood out of a turnip", *Communications of the ACM*, Volume 43, Issue 9, September 2000, 82 – 89.

Gopal, Ram D. and G. Lawrence Sanders, "International Software Piracy: Analysis of Key Issues and Impacts", *Information Systems Research*, Vol. 9 No. 4, December 1998, 380-397.

Gu, Bin, and Vijay Mahajan, "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, Geert, "Dimensions of national cultures in fifty countries and three regions", in J. B. Deregowski, S. Dziurawiec and R. C. Annis, eds., *Expications in Cross-Cultural Psychology*, Lisse, Netherlands: Swets and Zeitlinger, 1983, 333–355.

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

Hofstede, Geert, Cultural Dimensions, http://www.geert-hofstede.com/hofstede_dimensions.php, [Accessed, November 8, 2007].

Hui, Kai-Lung, and Ivan Png, "Piracy and the Legitimate Demand for Recorded Music", *Contributions to Economic Analysis & Policy*, Vol. 2 No. 1, 2003, Article 11.

Hui, Kai-Lung, and Ivan Png, "Measuring Piracy of IP-Protected Goods: A Methodology", World Intellectual Property Organization: Working Group of Experts on Developing a Methodology

for Measuring Copyright Piracy, November 1-4, 2005, http://www.econaa.com/resources/methodology_piracy_rate_sector.pdf [Accessed, July 27, 2006].

Husted, Bryan W., "The Impact of National Culture on Software Piracy", *Journal of Business Ethics*, Vol. 26 No. 3, August 2000, 197-211.

International Federation of the Phonographic Industry (IFPI), *The Recording Industry: Piracy Report, 2006*, London.

International Intellectual Property Alliance (IIPA), Press Release: The Copyright Industries in the International Intellectual Property Alliance (IIPA) Submit to USTR their 2007 Report on Piracy in 60 Countries/Territories, Washington, DC, February 17, 2007.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, "Governance Matters III: Governance Indicators for 1996-2002", World Bank, June 30, 2003.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, "Governance Matters IV: Aggregate and Individual Governance Indicators 1996–2006", Policy Research Working Paper 4280, World Bank, July 2007.

Marron, Donald B., and David G. Steel, "Which Countries Protect Intellectual Property? The Case of Software Piracy", *Economic Inquiry*, Vol. 38 No. 2, April 2000, 159-174.

Park, Walter G., "Do Intellectual Property Rights Stimulate R&D and Productivity Growth? Evidence from Cross-National and Manufacturing Industry Data", in Jon Putnam, ed., *Intellectual Property Rights and Innovation in the Knowledge-Based Economy*, Calgary: University of Calgary Press, 2003.

Png, I.P.L., and Qiuhong Wang, "Copyright Duration and the Supply of Creative Work: Evidence from the Movies", Dept of Information Systems, National University of Singapore, revised, November 2007.

Reynolds, Taylor F., "Quantifying the Evolution of Copyright and Trademark Law", PhD dissertation, American University, Washington, D.C. 20012, 2003.

Robertson, Christopher J., K. Matthew Gilley, Victoria Crittenden, and William F. Crittenden, "An analysis of the predictors of software piracy within Latin America", *Journal of Business Research*, forthcoming 2007.

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

Shin, Seung Kyoon, Ram D. Gopal, G. Lawrence Sanders, and Andrew B. Whinston, "Global software piracy revisited", *Communications of the ACM*, Vol. 47 No. 1, January 2004, 103-107.

Sundararajan, Arun, "Managing Digital Piracy: Pricing and Protection", *Information Systems Research*, Vol. 15, No. 3, September 2004, 287–308.

van Kranenburg, Hans, and Annelies Hogenbirk, "Multimedia, Entertainment, and Business Software Copyright Piracy: A Cross-National Study", *Journal of Media Economics*, Vol. 18, No. 2, April 2005, 109-129.

Figure 1(a): Income and legitimate demand, 2000-02

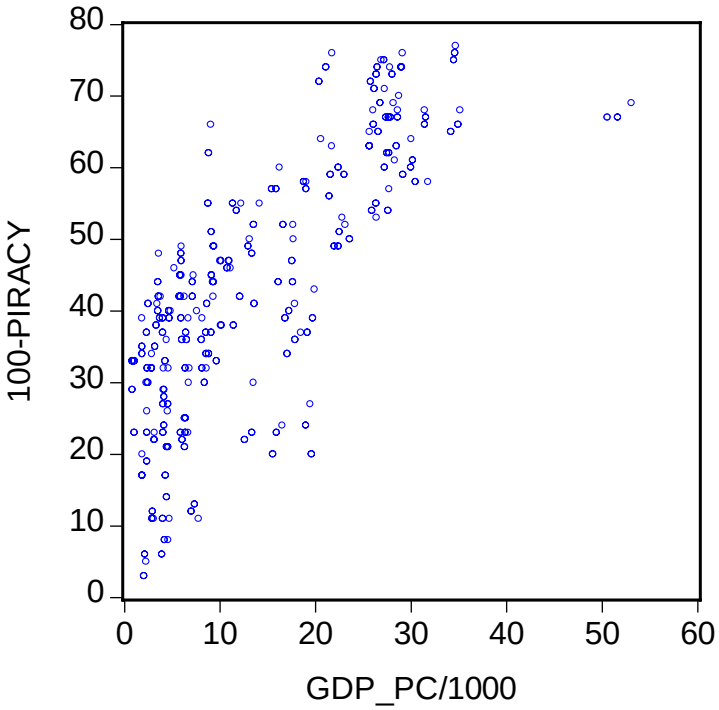

Figure 1(b): Income and legitimate demand, 2003-05

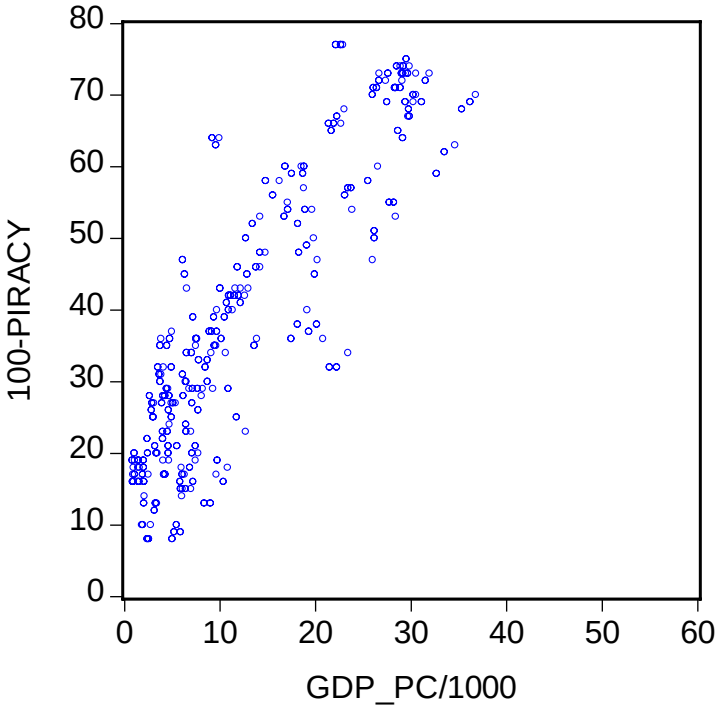

Figure 2: Individualism and legitimate demand, 2000

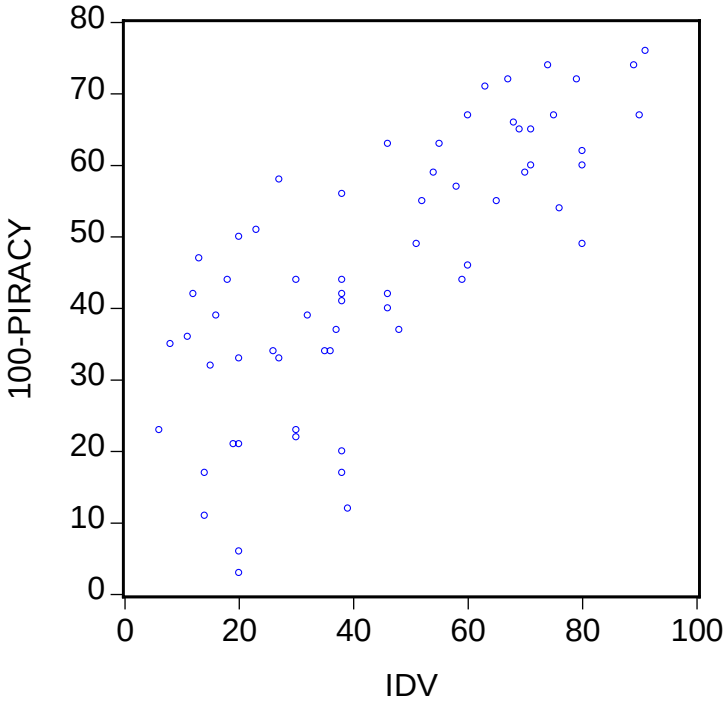

Figure 3: Copyright index and legitimate demand, 2000-02

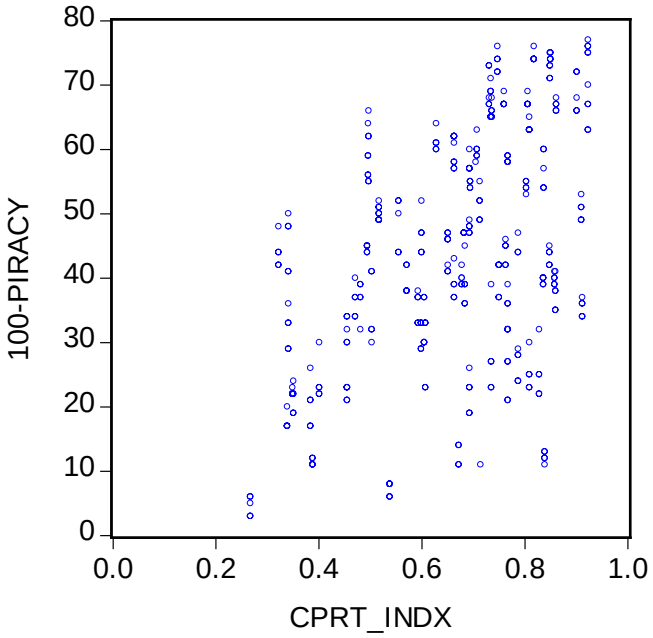

Figure 4(a): Rule of law and legitimate demand, 2000-02

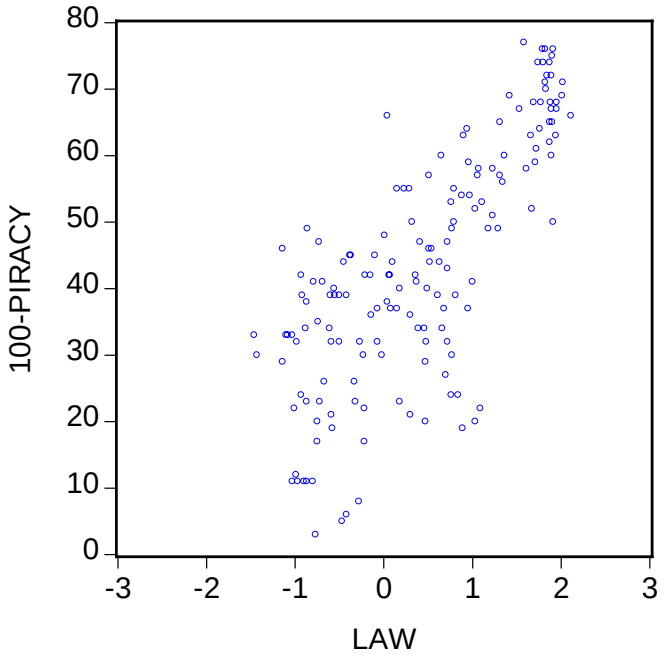

Figure 4(b): Rule of law and legitimate demand, 2003-05

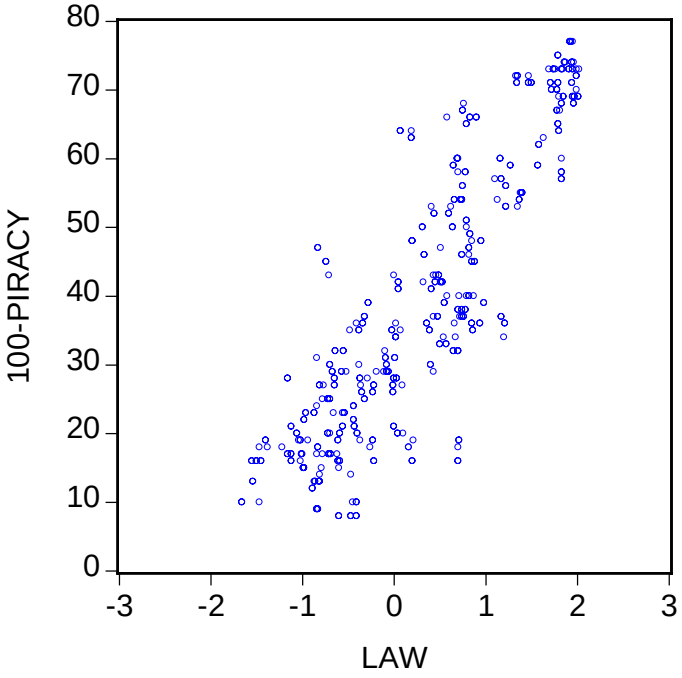

Figure 5: Estimation bias – Income

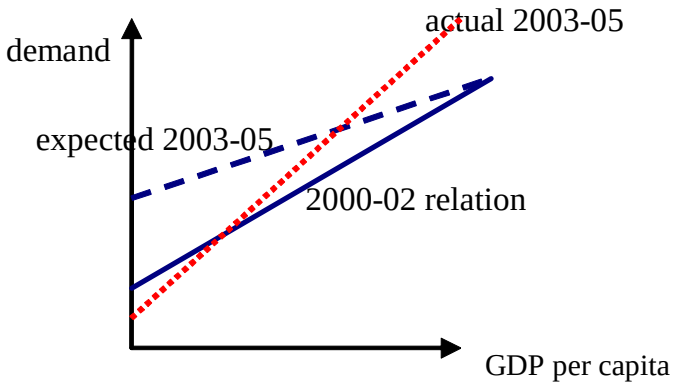

Table 1: Industry piracy statistics

	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. The Entertainment Software Alliance published statistics only for countries subject to potential investigation under Special 301.
2. The Motion Picture Association study estimated piracy in an additional 42 countries by projection from survey results in 22 countries.

Table 2: BSA methodology revision

	(i)	(ii)
Constant	60.37***	56.59***
	(1.989)	(1.206)
Year 2000	-2.411	
	(2.908)	
Year 2001	-3.458	
	(2.880)	
Year 2002	-5.446**	
	(2.880)	
Year 2003	-0.010	
	(2.837)	
Year 2004	0.505	
	(2.813)	
Year 2003-05		3.955**
		(1.668)
No. of obs.	516	516
Adjusted R^2	0.0033	0.0089
F-statistic	1.344	5.626

Table 3: Studies of software piracy

Study	Coun-tries	Period	Explanatory variables				
			Income per capita ($'000)	Individualism	Power distance	Institutions /law/enforc ement	Other
Gopal and Sanders (1998)	13	1995-96	-0.016*** (0.0028)				Size of domestic software industry
Husted (2000)	39	1996	-0.014*** (0.0026)	-.0033*** (0.0009)	n.s.		Income inequality
Marron and Steel (2000)	77	1994-97	-1.04*** (0.28)	-3.55*** (0.79)		negative	education
Shin et al. (2004)	49	1999	-0.76(5)	-0.295(5)			
Depken and Simmons (2004)	65	1994	-0.001*** (0.0002)	-0.188** (0.08)	p.s.		Inflation, literacy
van Kranenburg and Hogenbirk (2005)	35	1999-2001				negative	Region
Fischer and Rodriguez (2005)	71	1994-2002	-0.029*** (0.0093) (6)	-0.005*** (0.0011)		negative	Income inequality
Rodriguez (2006)	23	1994, 1997, 2000	-1.156*** (0.180)			negative	
Bagchi, et al. (2006)	37	1996, 2001, 2003	p.s.	negative(7)		negative(7)	

Chellappa et al. (2006)	53	1994-2004		negative[8]	negative[8]	negative[8]	
Robertson et al. (2007)	20	2001-04	p.s.	n.s.			Foreign direct investment, internet usage

Notes:

1. n.s. = not statistically significant, p.s. = partially significant;
2. Standard errors in parentheses; ***significant at 1%, **significant at 5%, significant at 10%;
3. Cells are blank if the variable was not included in the respective study;
4. If the study included multiple estimates, I report the median estimate;
5. Standard errors were not reported;
6. Fischer and Rodriguez (2005) specified piracy rate in logarithm and included both the logarithm of GNP per capita and its square as explanatory variables. The coefficients reported here are for the square of the logarithm of GNP per capita;
7. Bagchi et al. (2006) specified income as an aggregate rather than on per capita basis. Coefficients and standard errors not reported;
8. Chellapa et al. (2006) regressed the logarithm of the ratio of the piracy rate to the legitimate demand on two indexes: an index of legal costs from the World Bank's Rule of Law Index and Control of Corruption Index (Kauffman et al. 2007), and an index of moral costs from Hofstede's indexes of collectivism, power distance, and the interaction between collectivism and power distance. The index of collectivism is 100 less the index of individualism.

Table 4: Descriptive statistics

	Units	Source	Observ-ations	Mean	Std dev	Min.	Max.
Business software piracy	%	BSA	516	58.66	19.00	23	97
GDP per capita	'000 USD at PPP	World Bank	982	9.016	9.734	0.069	57.02
Individualism[(1)]		Hofstede (2007)	77	42.32	22.88	6	91
Copyright index[(2)]	[0,1]	Reynolds (2003)	423	0.5650	0.1973	0.0780	0.9230
Rule of Law	[-3, 3]	Kauffman et al. (2007)	944	-0.0887	0.989	-2.370	2.110

Notes:

1) Deemed to be for year 2000;
2) Copyright index covered only years up to 2002;

Table 5: Correlations

	Piracy	GDP per capita	Individualism	Copyright	Rule of law
Piracy	1.00	-0.83	-0.75	-0.48	-0.83
GDP per capita	-0.83	1.00	0.73	0.57	0.86
Individualism	-0.75	0.73	1.00	0.37	0.72
Copyright	-0.48	0.57	0.37	1.00	0.50
Rule of law	-0.83	0.86	0.72	0.50	1.00

Table 6: Demand for legitimate software

	(a) 2000-02	(b) 2003-05	(c) 2000-05	(d) 2004	(e) 2000-02	(f) 2000-05	(g) 2000-05	(h) 2000-05
Constant	25.72*** (1.291)	15.99*** (0.831)	25.72*** (1.291)	19.25*** (3.557)	10.90*** (3.868)	37.68*** (1.069)	34.15*** (0.600)	37.68*** (1.068)
GDP per capita	1.293*** (0.0780)	1.799*** (0.0528)	1.293*** (0.0789)					
Individualism (Hofstede 2007)				0.584*** (0.0583)				
Copyright index (Reynolds 2003)					50.44*** (6.005)			
Rule of Law (Kauffman et al. 2007)						13.85*** (0.871)	17.75*** (0.469)	13.85*** (0.869)
GDP per capita x Year 2003-05			0.506*** (0.0950)					
Rule of law x Year 2003-05								3.909*** (0.988)
Year 2003-05			-9.737*** (1.535)					-3.527*** (1.225)
No. of observations	240	262	502	64	228	163	270	433
Adjusted R^2	0.605	0.786	0.713	0.563	0.227	0.575	0.766	0.705
F-statistic	367.6	959.2	415.7	82.14	67.74	220.6	881.3	345.7

Notes:

1) Standard errors in parentheses; ***significant at 1%, **significant at 5%, significant at 10%;
2) Standard errors calculated using White's adjustment for heteroscedasticity.