Measuring Piracy of IP-Protected Goods: A Methodology

Kai-lung Hui and I.P.L. Png

Economic Analysis Associates & National University of Singapore

November 2005

1. Introduction

Intellectual property (IP) includes patents, copyrights, trademarks, and trade secrets. Any good protected by some form of intellectual property is an "IP-protected" good. Examples of IP and the corresponding goods that might possibly be IP-protected include

- Patent – pharmaceuticals and electronics,
- Copyright – software, movies, music, and video-games, and
- Trademark – clothing, handbags, watches, and cosmetics.

Generally, IP law aims to strike a balance between two considerations (Plant 1934; Nordhaus 1969):

- Stronger protection increases the return to innovators, and over the long run, encourages more innovation;
- Weaker protection increases the use of existing innovations, and hence, in the short run, raises the benefit to end-users and also encourages incremental innovations that build upon existing innovations. [1]

The return to innovators depends on the demand for the item. At the simplest level, the innovator's return depends on the expected sales and the expected price of the item.

Even absent the institution of intellectual property, the sale of products such as pharmaceuticals and software would involve substantial margins of selling price over

[1] An alternative is to replace intellectual property rights with a system of rewards for inventors and creators (Shavell and van Ypersele 2001).

© 2005, Economic Analysis Associates

cost of production. These margins are necessary to compensate the inventors/creators for their investment in developing and marketing the product. Absent such margins, such products could only be provided through charity.

With the institution of intellectual property, the sale of products such as pharmaceuticals and software would involve even higher price-cost margins. These margins naturally attract "piracy", which is the production of copies of IP-protected goods without the permission of the IP owner. Not having invested in development and marketing, the pirate producers can offer the item at a lower price than the IP owner and still break even.

In this paper, we consider piracy and counterfeiting to be equivalent. However, these are different from parallel imports or grey-market products, which are goods produced by the IP owner but diverted to other markets.

Piracy erodes the legitimate producer's profit in two ways – reducing the legitimate sales and reducing the price (Hui and Png 2003). To the extent that investment in new innovations is sensitive to expected future profit, piracy erodes the incentive for investment in new innovations.

Consequently, a key issue in national and international IP policy is to measure the extent of piracy. However, there is no agreed methodology for the measurement of piracy. Indeed, this is a controversial issue, involving heated debate between IP owners and consumers of IP-protected products and policy-makers.

Accordingly, our objective in this paper is to develop a general methodology for the measurement of piracy of IP-protected goods. This methodology should satisfy three properties: It should be

- Robust to differences in national and local conditions,
- Scaleable to accommodate markets of different size, and
- Flexible to encompass multiple categories of IP-protected goods.

2. Overview of Methodologies

© 2005, Economic Analysis Associates

Generally, there are two ways to measure the rate of piracy of an IP-protected good. One *directly* measures the quantity of pirated units consumed, while the other measures the pirated quantity *indirectly* by inferring it as the difference between total consumption and the legitimate sales. Below, we describe and compare the two methodologies.

2.1 Direct Methodology

The direct methodology is the more obvious approach. It measures the quantity of pirated goods consumed, $Pirated\_Quantity\_Consumed$, as well as the legitimate consumption, $Legitimate\_Quantity\_Consumed$. It then calculates total consumption as the sum of the consumption of pirated and legitimate goods,

$$\begin{aligned} &Total\_Quantity\_Consumed \\ &= Pirated\_Quantity\_Consumed + Legitimate\_Quantity\_Consumed. \end{aligned}$$

Finally, the direct approach obtains the piracy rate, as the ratio of consumption of pirated units to the total consumption,

$$Piracy\_Rate = \frac{Pirated\_Quantity\_Consumed}{Total\_Quantity\_Consumed}.$$

The critical challenge in the direct methodology is to measure the $Pirated\_Quantity\_Consumed$. This requires information from either the manufacturer or distributor of the pirated items, or end-users, or both. Typically, manufacturing, distribution, and consumption of pirated goods are all illegal. Generally, it is difficult to acquire accurate and representative information about illegal activities. The same applies to information about sale, distribution, and consumption of pirated goods.

As between manufacturing, distribution, and consumption of pirated goods, the typical view would be that manufacturing is the worst crime, distribution a moderate crime, and consumption the least culpable crime. Certainly, authorities worldwide focus enforcement and courts impose highest penalties on manufactures. Correspondingly, authorities and courts pay relatively least attention to consumers.

As a result, it would be most difficult to acquire accurate and representative information from manufacturers of pirated goods. Moreover, although they are not targeted for enforcement, consumers might still not report their consumption of pirated goods truthfully. Why should they?

© 2005, Economic Analysis Associates

The following example illustrates some of the challenges that arise even when respondents report truthfully. A leading international consumer goods manufacturer engaged a consumer research organization to survey particular supermarkets in China to estimate the volume of counterfeits of their products. However, some of the counterfeits were so like the legitimate item in appearance that the surveyors had difficulty distinguishing them. As a result, the manufacturer believes that the survey substantially under-estimated the actual volume of counterfeits.

2.2 Indirect Methodology

Given the difficulties with the direct methodology, IP owners have developed an alternative which measures piracy through an *indirect* methodology. This approach does not directly measure the quantity of pirated units consumed. Instead, it infers the pirated quantity as the difference between the total consumption and the legitimate consumption.

Specifically, the indirect methodology measures the total consumption and the legitimate consumption, and then calculates the quantity of pirated units consumed as

$$\begin{aligned}&Pirated\_Quantity\_Consumed\\&= Total\_Quantity\_Consumed - Legitimate\_Quantity\_Consumed.\end{aligned}$$

Then, the indirect methodology calculates the piracy rate as ratio of consumption of pirated units to the total consumption,

$$Piracy\_Rate = \frac{Pirated\_Quantity\_Consumed}{Total\_Quantity\_Consumed}.$$

The key rationale for the indirect methodology is that it is more cost-effective to acquire accurate and representative information about *total consumption* than *consumption of pirated goods*. Obviously, this is an empirical issue that varies with the facts and circumstances of each market.

2.3 Indirect Methodology Based on Complementary Platform

In the case of particular IP-protected goods, the indirect methodology has been further developed to exploit a strong relation between the consumption of the IP-protected good and a complementary "platform". For instance, the use of personal-computer (PC) software is complementary with the PC, the use of video-games is complementary with

© 2005, Economic Analysis Associates

the corresponding console, and the use of proprietary batteries is complementary with the corresponding electronic device (e.g., mobile phone, camera, and camcorder).

The indirect methodology based on a complementary platform proceeds as follows. First, it measures the total quantity of the complementary platform in service, $Stock\_Platform$, and the annual consumption of the IP-protected good (including both legitimate and pirated goods) *per platform*, $Consumption\_Rate$. Note that the quantity of the platform in service is a stock, while the annual consumption is a flow.

It then calculates the total consumption by multiplying the stock of the platform by the annual consumption per platform,

$$Total\_Quantity\_Consumed = Consumption\_Rate \times Stock\_Platform.$$

With this information, it then proceeds as the indirect methodology described above in Section 2.2.

The key rationale for the indirect methodology based on a complementary platform is that

- It is more cost-effective to acquire accurate and representative information about the complementary platform than about the IP-protected good itself, and
- There is a stable relation between the consumption of the IP-protected good and the complementary platform.

As in the choice between the direct and indirect methodologies, the choice between the simple indirect methodology and the indirect methodology based on a complementary platform is an empirical issue that varies with the facts and circumstances of each market.

2.4 Industry Practice

In the following Table, we list various industries and which of the various methodologies their respective trade associations or particular IP-owners have used to measured piracy. The Table is based on published reports, and presentations at the meeting of the WIPO Working Group of Experts on Developing a Methodology for Measuring Copyright Piracy, November 1-4, 2005.

© 2005, Economic Analysis Associates

Table – Industry practices

Method	**Direct**	**Indirect**	**Indirect based on complementary platform**
Business software			Business Software Alliance (2005)
Recorded music[2]	International Federation of the Phonographic Industry	International Federation of the Phonographic Industry	
Pre-recorded video-tapes, VCDs, and DVDs[3]	Motion Picture Association	Motion Picture Association	Motion Picture Association
School textbooks[4]		Textbook publishers, Hong Kong	
Video-games[5]		Entertainment Software Association	Entertainment Software Association

3. Application to Business Software

Below, we illustrate how to apply the indirect methodology based on a complementary platform to measure piracy of business PC software. In this case, the platform is the PC. Our approach is similar to that used by the International Data Corporation for the

[2] The International Federation of the Phonographic Industry relies on member national associations for estimates of piracy at the national level. The national associations use a variety of methods, including both direct and indirect methodologies (Jopling 2005).

[3] For developed countries, the Motion Picture Association estimates piracy by sampling stores that rent or sell the items, while for high-piracy markets, the Association estimates total consumption from the sales of blank media or from the quantity of players in use (Olsen 2005).

[4] Hong Kong textbook publishers estimate total consumption for books adopted by schools as the student enrolment less estimates for used books and sharing (Bammel 2005).

[5] The Entertainment Software Association estimates total consumption from the quantity of game consoles (Morell 2005; Olsen 2005).

© 2005, Economic Analysis Associates

Business Software Alliance. Among several differences in our approach, the key difference is the method that we propose for estimating the consumption rate (Step 1 below).

Step 1: Consumption Rate

To collect the information about the consumption rate relative to the platform, administer a survey to two groups of respondents: households (H), and businesses (B).

In the household survey, ask each respondent *i* to report the following information:

- Total number of computers at the beginning of the year, *BASE(i)*, and
- Number of new computers purchased in the year, *COMP(i)*, of which the
  - Number that were *replacement* computers, *COMPR(i)*.

Using the above information, calculate the

- Number of new computers purchased that were not replacements, *COMPN(i)* = *COMP(i)* – *COMPR(i)*.

In the same survey, ask respondents to report the quantities of the 10 representative software titles listed below that they have installed including upgrades (or received bundled with a PC):[6]

(1) MS-Windows,
(2) MacOS,
(3) Linux,
(4) Internet Explorer,
(5) Mozilla Firefox,
(6) MS Office,
(7) Star Office,
(8) Acrobat,
(9) Photoshop, and
(10) AutoCAD.

For each respondent *i*, denote the total quantity of software installed as *DEMD*(1,*i*), …, *DEMD*(10,*i*), where 1 to 10 denote the various software titles. Then, for each representative software title *j*, calculate the average quantities of each title acquired

[6] We have used these 10 titles for illustration purposes only. In practice, market research should be applied to select titles that are representative of the demand for business PC software.

© 2005, Economic Analysis Associates

per existing, new replacement, and new non-replacement computer using a multiple regression with the specification,

- Dependent variable = *DEMD(j,i*),
- Independent variables = *BASE(i), COMPR(i), COMPN(i)*.

Then, for *household* respondents, the average units of software title *j* installed

- In existing PCs is the coefficient, *DEMDS_H(j)*, of *BASE(i)* in the multiple regression,
- In *replacement* PCs is the coefficient, *DEMDR_H(j)*, of *COMPR(i)* in the multiple regression,
- In *non-replacement* PCs is the coefficient, *DEMDN_H(j)*, of *COMPN(i)* in the multiple regression.

Finally, calculate the average units of all 10 representative software titles installed

- In existing PCs as the sum, $DEMDS\_H = \sum_j DEMDS\_H(j)$,
- In *replacement* PCs as the sum, $DEMDR\_H = \sum_j DEMDR\_H(j)$, and
- In *non-replacement* PCs as the sum, $DEMDN\_H = \sum_j DEMDN\_H(j)$,

where the sums are computed over the 10 software titles, indexed by *j*.

Likewise, for *business* respondents, collect the relevant data and then use a multiple regression to estimate the average units of all representative software titles installed

- In existing PCs, *DEMDS_B*,
- In *replacement* PCs, *DEMDR_B*, and
- In *non-replacement* PCs, *DEMDN_B*.

Step 2: Platform Stock

To scale the business PC software consumption rate to the national level, we need estimates of the total number of new PCs sold to households (*PLAT_H*) and the installed base of PCs among households (*BASE_H*) in the country for the year. Similarly, we need corresponding estimates, *PLAT_B* and *BASE_B*, for businesses in the country for the year. Such data can be purchased from a market research organization such as Gartner or the International Data Corporation (IDC), or possibly obtained from the national statistical agency.

© 2005, Economic Analysis Associates

From these data, estimate the total number of new *non-replacement* computers sold to households in the country over the preceding year as:[7]

$$PLATN\_H = BASE\_H(T) - BASE\_H(T-1)\,,$$

where T represents the beginning of the current year, and $T-1$ represents the beginning of the preceding year.

Using the data on the total number of new PCs sold and the above calculation of the number of new non-replacement computers, the total number of new *replacement* computers can be calculated as

$$\begin{aligned} PLATR\_H &= PLAT\_H - PLATN\_H \\ &= PLAT\_H - \left[BASE\_H(T) - BASE\_H(T-1)\right]. \end{aligned}$$

Similarly, for *businesses*, the total number of new *non-replacement* computers sold in the preceding year is:[8]

$$PLATN\_B = BASE\_B(T) - BASE\_B(T-1)\,,$$

and, hence, the total number of new *replacement* computers sold in the preceding year is

$$PLATR\_B = PLAT\_B - \left[BASE\_B(T) - BASE\_B(T-1)\right].$$

<u>Step 3: Total Consumption</u>

For *households*, the total quantity of software acquired for the various categories of PCs can be computed as follows,

- For *existing* computers
  $$DEMDS\_H\_AGG = DEMDS\_H \times BASE\_H(T-1)\,,$$
- For *replacement* computers
  $$DEMDR\_H\_AGG = DEMDR\_H \times PLATR\_H\,,$$
- For *non-replacement* computers
  $$DEMDN\_H\_AGG = DEMDN\_H \times PLATN\_H\,.$$

Finally, the total quantity of software acquired by *all households* is

$$DEMD\_H\_AGG = DEMDS\_H\_AGG + DEMDR\_H\_AGG + DEMDN\_H\_AGG\,.$$

Similarly, the total quantity of software acquired by businesses

[7] We implicitly ignore the dissolution of households through death, emigration, divorce, and other causes.
[8] We implicitly ignore the dissolution of businesses.

© 2005, Economic Analysis Associates

- For *existing* computers

  $$DEMDS\_B\_AGG = DEMDS\_B \times BASE\_B(T-1),$$

- For *replacement* computers

  $$DEMDR\_B\_AGG = DEMDR\_B \times PLATR\_B,$$

- For *non-replacement* computers

  $$DEMDN\_B\_AGG = DEMDN\_B \times PLATN\_B.$$

Finally, the total quantity of software acquired by *all businesses* is

$$DEMD\_B\_AGG = DEMDS\_B\_AGG + DEMDR\_B\_AGG + DEMDN\_B\_AGG.$$

Total software acquired for the country is then the sum of household and business acquisitions,

$$Total\_Quantity\_Acquired = DEMD\_H\_AGG + DEMD\_B\_AGG.$$

Step 4: Legitimate Quantity

To collect the information about the legitimate quantity, the most direct and probably the most reliable way is to ask the respective software publishers.

For each representative software title *j*, the respective publisher should report

- The quantity sold (including upgrades) through *direct* marketing in the country, *SALES_D(j)*. The quantity should include all direct sales through the publisher's salesforce, the Internet, and other direct channels, and also site licensing.
- The average price of direct sales (including upgrades), *PRICE_D(j)*. The price should be that paid by the end-user.
- The quantity sold (including upgrades) through *indirect* marketing in the country, *SALES_I(j)*. The quantity should include all sales through indirect distribution channels, including hardware vendors bundling the software with hardware, value-added resellers, and conventional retail stores.
- The average price of indirect sales (including upgrades), *PRICE_I(j)*. The price should be the (wholesale) price paid by the distribution channel.

Then, for each representative software title *j*, the legitimate quantity sold is

$$SALES(j) = SALES\_D(j) + SALES\_I(j) .$$

Hence, the total legitimate quantity sold of all the representative titles is

$$Legitimate\_Quantity\_Consumed = \sum_j SALES(j) ,$$

where the sum is computed over the 10 software titles, indexed by *j*.

Having calculated the total acquisitions in Step 3 and now the legitimate acquisitions, the formulas in Section 2.2 can be used to calculate the piracy rate.

Notes

The methodology presented above is subject to one important limitation. In countries with a relatively high price for software, substantial quantities of the legitimate product may be imported through parallel channels. Similarly, in countries with a relatively low price for software, substantial quantities of the legitimate product may be exported through parallel channels. Such parallel trade would bias the estimates of the national piracy rate.

The reason to collect information about prices is two-fold. One reason is to support analysis of the impact of piracy on the pricing of the IP-protected good in direct and indirect sales (Hui and Png 2003). For each representative software title *j*, the average price is the weighted average of the prices in direct and indirect sales,

$$PRICE(j) = \frac{SALES\_D(j)}{SALES(j)} \times PRICE\_D(j) + \frac{SALES\_I(j)}{SALES(j)} \times PRICE\_I(j) .$$

The average price over all the representative titles is

$$PRICE = \frac{1}{\sum_j SALES(j)} \sum_j SALES(j) \times PRICE(j) ,$$

where the sums are computed over the 10 software titles, indexed by *j*.

The other reason to compute the average prices is to support any analysis of the impact of parallel trade on the piracy rate. Depending on whether the average price is higher or lower than in other countries, the piracy rate may be over- or under-estimated respectively.

© 2005, Economic Analysis Associates

Over time, the identity of the representative software titles should be researched. In particular, if the correlation among the piracy rates of the various titles is sufficiently high, the study could focus on a smaller set of titles, which would still be sufficiently representative.

References

Bammel, Jens, Presentation – Working Group of Experts on Developing a Methodology for Measuring Copyright Piracy, World Intellectual Property Organization, Geneva, November 1-4, 2005.

Business Software Alliance, *Second Annual Global Software Piracy Study*, Washington, DC, May 2005

Hui, Kai-lung, and I.P.L. Png, "Piracy and the Legitimate Demand for Recorded Music", *Contributions to Economic Analysis & Policy*, Vol. 2 No. 1 (2003), Article 11.

Jopling, Keith, Presentation – Working Group of Experts on Developing a Methodology for Measuring Copyright Piracy, World Intellectual Property Organization, Geneva, November 1-4, 2005.

Morell, Jonathan, Presentation – Working Group of Experts on Developing a Methodology for Measuring Copyright Piracy, World Intellectual Property Organization, Geneva, November 1-4, 2005.

Nordhaus, William, *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change*, M.I.T. Press, 1969.

Olsen, Karsten, Presentation – Counterfeiting and Piracy: WIPO/OECD Expert Meeting on Measurement and Statistical Issues, Geneva, October 17-18, 2005.

Plant, Arnold, "The Economic Aspects of Copyright", *Economica*, Vol. 1 No. 2, May 1934, 167-195.

Shavell, Steven, and Tanguy van Ypersele, "Rewards versus Intellectual Property Rights", *Journal of Law and Economics*, Vol. 44 No. 2, October 2001, 525-546.

© 2005, Economic Analysis Associates