What is the True Loss Due to Piracy?: Evidence from Microsoft Office in Hong Kong

Leung, Tin Cheuk

Chinese University of Hong Kong

July 2011

What is the True Loss Due to Piracy? Evidence from Microsoft Office in Hong Kong

Tin Cheuk Leung*

Chinese University of Hong Kong

July 14, 2011

Abstract

Software piracy remains rampant despite the successful measures the Hong Kong government has taken to eradicate street piracy. This is because most people prefer substituting a counterfeit copy of a software CD (street piracy) with an illegal download of the software (Internet piracy). To support this claim, I construct a unique data set from 281 college students in Hong Kong to demonstrate two things. First, I estimate a random-coefficient discrete choice demand system for Microsoft Office from legal and different illegal sources. Estimates obtained from a Bayesian approach, with a mixture of normal priors, indicate a strong substitution pattern between street piracy and Internet piracy. Second, I conduct counterfactuals in which street piracy is absent. Results are twofold. First, most students would switch to Internet piracy. Second, the government, by assuming that each pirated copy represents a lost sale, may over-estimate the gain from eradicating piracy by up to nine times.

*I am indebted to my thesis advisor, Pat Bajari, for his continuous encouragement and support. I thank Tom Holmes, Kyoo il Kim, participants of the HKUST applied micro workshop, the FESAMES 2009, the CES 2010 Annual Meetings, SAET 2010 Conference and 2nd Asia-Pacific Innovation Conference 2011, and two anonymous referees for their valuable comments. I also thank Katherine Leung, Wen-Tai Hsu and Linda Yung for administering surveys; Zoey Li and Aric Chan for their diligent data input; and Ka Fai Li for educating me about software piracy. The usual disclaimer applies. Correspondence: tleung@cuhk.edu.hk

1 Introduction

Innovation induces economic growth. However, the economics profession has not reached a consensus on whether intellectual property rights (IPR) can foster innovation. On the one hand, as Nordhaus (1969) points out, IPR provides commercial incentives to innovate. Chen and Puttitanun (2005) uses a panel data consisting of 64 developing countries and shows that IPR has a positive impact on innovation. Hu and Png (2009) finds that a stronger IPR has a larger positive impact on more patent-intensive industries in over 72 countries between 1981 and 2000. On the other hand, as Boldrin and Levine (2008) argues, IPR can increase the monopoly power of the copyright/patent owner and thus increase the cost of other innovations. Qian (2007) shows that national IPR does not stimulate domestic innovation on pharmaceutical products in the 26 countries in the study.

One determinant of the impact of IPR on innovation is the effectiveness of IPR enforcement. In particular, we would expect that the higher the cost of IPR enforcement, the lesser the net benefit from a stronger IPR. The cost of enforcement changes with IPR infringement technology.[1] In the case of software applications, since the mid 1990s, the emergence of the Internet has been a double-edged sword to software developers. On the one hand, the Internet, by connecting people from different parts of the world, makes people more reliant on computers, and thus increases the demand for software applications. According to Bessen and Hunt (2007), almost 15% of patents issued in the US in 2002 were software patents, compared to 1.1% in 1976. On the other hand, the Internet also increases piracy because it provides people one more channel from which to obtain counterfeit copyrighted goods in addition to street hawkers. To enforce IPR, governments and copyright owners file lawsuits against suppliers and individual pirates, both on the street and on the Internet. However, the enforcement cost is higher on the Internet. While governments have achieved successes in curbing street piracy, they have had setbacks on the Internet. (Section 2)

The emergence of the Internet also lowers the effectiveness of any effort that reduces street piracy, because people can substitute a counterfeit software CD obtained from a street hawker with an illegally downloaded software file from the Internet. Because of the setbacks of the IPR

[1] IPR infringement and piracy are used interchangeably in this paper.

enforcement that governments face on the Internet, the effort that reduces street piracy becomes less effective. While the number of shops engaged in piracy in Hong Kong has dropped 75% in the past six years, the software piracy rate has only dropped 7%. In the extreme case in which street and Internet piracy are perfect substitutes, any effort that reduces street piracy would not be effective in protecting IPR at all.

The potential benefit of eradicating piracy is enormous. Assuming that one pirated copy would, in the absence of piracy, have been a legitimate sale, the Business Software Alliance (BSA) estimates that the revenue loss due to piracy was approximately $53 billion in 2008. Of course, one less pirated copy does not necessarily translate into a legitimate sale. When one source of piracy (street piracy) is not available, people are likely to switch to pirate from another source (Internet piracy). Even when all sources of piracy are not available, people can find substitutes for the software application.

To quantify the true gain from eradicating piracy, it is important to know the substitutability among street piracy, Internet piracy, and substitutes for the software. To the best of my knowledge, this is the first paper to separately estimate the demands for copyrighted products from legal and different illegal sources, which is essential to quantify the substitution pattern. For my empirical analysis, I construct a unique conjoint survey data set (Section 3) on Microsoft Office (henceforth Office), one of the most successful and also heavily pirated software applications. The survey data include 281 college students in Hong Kong. In the survey, students answer two types of questions. First, they provide information on their demographics and consumption of copyrighted goods like Office. Second, in the conjoint survey, they make choices on Office consumption (from legal source, street piracy, or Internet piracy) in ten hypothetical tasks. In each of the ten tasks, prices of Office from different sources are exogenously randomized within a pre-specified range, which provides identification for my empirical model.

My empirical analysis consists of two parts. First, I set up a random-coefficient discrete demand model for Office (Section 4). I follow Rossi, Allenby, and McCulloch (2005) to set up a hierarchical Bayesian discrete demand model for Office from different sources, with a mixture of normal priors. Afterward, I use a hybrid of Gibbs Sampling and Metropolis-Hasting algorithm to implement posterior inference. Second, I use the estimates to conduct counterfactuals to evaluate the effectiveness of a policy that can completely eradicate either street piracy, or Internet piracy, or both (Section 5). Results are twofold. First, absent street piracy, the demand for Office from both the legitimate source and from Internet piracy would increase. Furthermore, because street and Internet piracy are closer substitutes, the demand for Office from Internet piracy would increase more. Second, taking into account the substitution among Office from different sources and the outside options like Google Docs, the estimates indicate the profit gain by eradicating street piracy to be approximately HK$15.2 ($2) per person. Third, BSA, which counts each pirated software (from either street piracy or Internet piracy) as a lost sale, over-estimates the profit loss due to piracy by nine times.

The use of conjoint analysis is not novel. Green and Rao (1971) first introduces conjoint survey analysis as a way to elicit demand estimates. Conjoint survey data are also known as stated-preference data, as opposed to revealed-preference data collected from real-world observations. Several studies argue that conjoint survey data can generate reliable demand estimates.[2] Applications of conjoint survey analysis also abound.[3] There are also several conjoint studies using a hierarchical Bayesian model to estimate random demand coefficients, but most of them do not have a mixture of components of normal priors.[4] Two notable exceptions are Chandukala, Edwards, and Allenby (2010) and Leung (2009). There are two main advantages to using conjoint survey data, instead of real market data, in this research. First, this is possibly the only way to create a panel data set on the consumption of Office from legal and various illegal sources, which is essential to elicit the substitutability between street piracy and Internet piracy and, hence, the real loss due to different types of piracies. Second, conjoint survey can create good instruments for demand estimation because the variation of covariates is purely exogenous as mentioned above.

This paper is related to several papers that estimate the replacement effects of downloaded

[2] Carlsson and Martinsson (2001) and Hensher, Louviere, and Swait (1999) collect both stated-preference data and revealed-preference data of donation choice and freight shipper choice. They show that the hypothesis of parameter equality holds for most parameters across the two data sources.

[3] Leung (2009) uses a similar approach to estimate the complementarity between music and iPods, and to evaluate various copyright policies. Hensher and Louviere (1983) forecasts the choice of attendance at various types of international expositions. Hensher (1994) reviews the development of using conjoint analysis to estimate transportation choice. Many multinational corporations like Marriott, Procter & Gamble, and General Motors also use conjoint survey data to estimate demand for new products (Green, Krieger, and Wind (2004) and Orme (2005)).

[4] See Feit, Beltramo, and Feinberg (2010), Iyengar and Jedidi (2011), and Louviere, Hensher, and Swait (2000).

albums/songs on legitimate CDs in the music industry.[5] Most of them find that piracy has some replacement effects on legitimate CD sales.[6] However, due to data limitation, they do not account for the replacement effect of street piracy and Internet piracy separately.

As this paper has implications on the IPR protection policy, it is also related to several papers that examine the level of protection of IPR across countries. Ginarte and Park (1997) develops a comprehensive index of patent protection and finds that the index is positively related to several economic variables. Marron and Steel (2000) uses the BSA software piracy statistics and finds that IPR protection depends not only on economic concerns but also national culture and institutions. The substitution pattern between street piracy and Internet piracy analyzed in this paper is one determinant not examined in their papers.

The organization of the article is as follows: Section 2 briefly describes the current situation of software piracy. Section 3 discusses the conjoint survey data set. Section 4 sets up the demand for Microsoft Office, and discusses results of the estimation. Section 5 conducts counterfactual experiments using results from Section 4. Section 6 concludes.

2 Software Piracy Across the Globe

Improvements in technology have been a major driving force for growth. Starting from the 1980s, along with the widespread use of computers, software innovation and development has been an important factor for technological improvement.

Software, like music and movies which are also under the protection of copyright laws, has several common properties in production: high fixed cost in terms of R&D expenditure and low marginal cost of production and distribution. Thus, once the expenditure on R&D is spent and the innovation is done, it is easy for people to copy and pirate the finished product. According to BSA, the trade group of the software industry, "35% of the software installed on personal computers worldwide in 2006 was obtained illegally, amounting to nearly $40 billion in global losses due to software piracy."

[5]See Blackburn (2004), Hui and Png (2003), Leung (2009), Oberholzer-Gee and Strumpf (2007), and Rob and Waldfogel (2006)).

[6]One notable exception is Oberholzer-Gee and Strumpf (2007) in which the authors find that piracy has no statistically significant replacement effect.

Table 1: Software Piracy Rates in Selected Regions (%)

	2003	2004	2005	2006	2007
United States	22	21	21	21	20
Canada	35	36	33	34	33
European Union	37	35	36	36	35
Middle East & Africa	56	58	57	60	60
Latin America	63	66	68	66	65
Asia-Pacific	53	53	54	55	59
China	92	90	86	82	82
Hong Kong	52	52	54	53	51

Source: BSA

2.1 Software Piracy and the Emergence of the Internet

Software piracy rates vary across countries. Table 1 shows that software piracy rates are higher in developing regions like Asia.[7] In particular, China's software piracy rate is one of the highest in the world (above 80% between 2003 and 2007). Hong Kong also has a relatively high software piracy rate among developed regions (above 50% between 2003 and 2007).

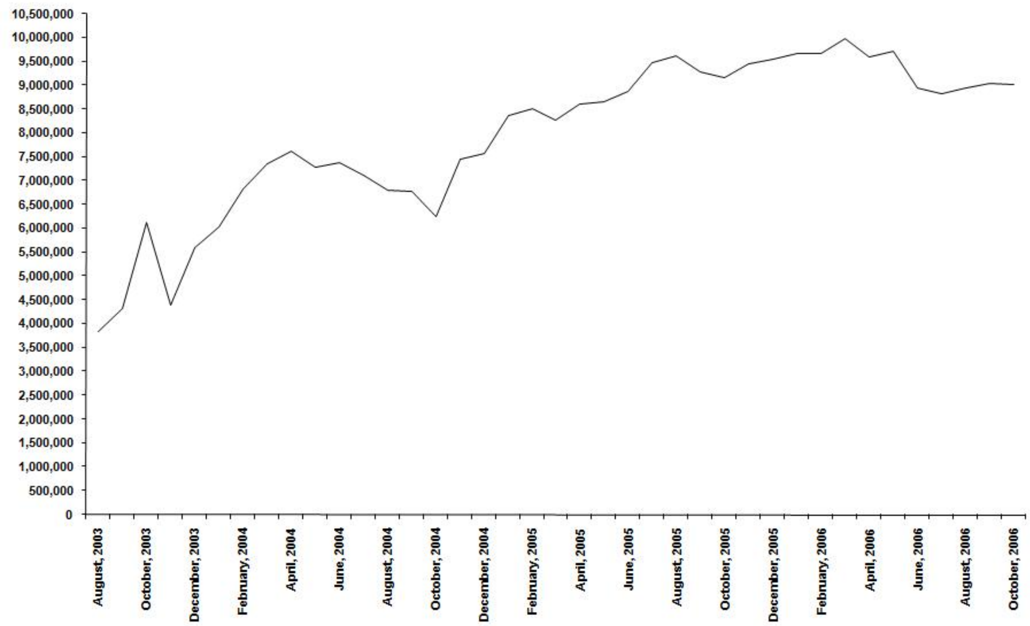
In the 1990s and early 2000s, the sale of counterfeit software CDs was the main form of software piracy. However, the emergence of the Internet has changed the way some people pirate software. According to the World Bank's World Development Indicators, the Internet penetration rate increased worldwide from 7% in 2000 to 23% in 2007. This number has been increasing at a faster rate in many developing countries like China. The Internet penetration rate increased eight times in China, whereas the corresponding numbers increased less than twice in the US between 2000 and 2007.

Easier access and the increase in Internet speed contribute to the spread of Internet piracy. One common technology people use to pirate digital copyrighted goods is Peer-to-Peer (P2P) software. According to Big Champagne, a marketing research firm specializing in Internet products, simultaneous P2P users has been increasing worldwide (Figure 1). At any moment in October 2006, there were about nine million people in the world using P2P software to share files, a significant portion of them being software files.

Under the pressure of international organizations like the World Intellectual Property Organization, the governments of China and Hong Kong have been active in enforcing IPR. According

[7]BSA uses survey data to estimate the number of total software units and pirated software units installed in each country. It then defines piracy rate as the ratio of the pirated software units and the total software units.

Figure 1: Big Champagne: Average Simultaneous P2P Users Worldwide



to Table 2, street piracy in Hong Kong has declined in the past six years: The Hong Kong Customs and Excise Department estimates that the number of shops engaged in piracy decreased from 105 to 25 between 2003 and 2008. The seizure of counterfeit products also declined in the same period.

While enforcement of IPR on the street achieved some success in places like China and Hong Kong, the same is not true on the Internet. IPR enforcement on the Internet is more difficult due to various reasons. First, piracy on the Internet is often trans-national, thus requiring cooperation among different countries for enforcement. For example, even though the founders of Pirate Bay, one of the world's biggest file-sharing sites, were found guilty of breaching copyright law in Sweden, the Website can still operate because its server is located beyond the reach of Swedish or European Union law enforcement.[8] Second, because of the public-relations disaster it created, the Recording Industry Association of America finally stopped its strategy of suing online music pirates.[9] Third, arguing that Internet access is a fundamental human right, French

[8]See The Economist April 17th, 2009 at http://www.economist.com/business/displayStory.cfm?story_id=13518830&source=features_box2

[9]See Wall Street Journal at http://online.wsj.com/article/SB122966038836021137.html.

Table 2: Street Piracy Decline in Hong Kong

	2003	2004	2005	2006	2007	2008
Optical Disc Seizure (in millions)						
Quantity	6.2	7.3	3.8	3.0	4.3	2.2
Value	135.3	157.8	89.2	72.4	99.1	52.3
Shops Engaged in Piracy						
Quantity	105	70	50	45	30	25

Source: Hong Kong Customs and Excise Department

lawmakers rejected an anti-piracy plan which would empower the government to sever pirates' Internet connection.[10]

The situation is similar in China and Hong Kong. Both governments only exert limited effort into enforcement on the Internet. The Chinese government closed down approximately 600 Websites serving as illegal sources of copyrighted goods, out of several hundred thousands, between 2005 and 2007. The Hong Kong Intellectual Property Department set up two seven-man Anti-Internet Piracy Teams to monitor Internet piracy in Hong Kong. Despite the fact that a Hong Kong citizen was convicted for illegal distribution of copyrighted works using P2P software in 2005, the only such case so far in Hong Kong, it is still relatively easy to download an illegal copy of Office in some popular Websites in Hong Kong. In Table 1, the piracy rates in Hong Kong is revealed to be roughly the same, despite the government's effort to eradicate street piracy. This indicates that there is high substitutability between street piracy and Internet piracy.

2.2 Revenue Loss Due to Piracy

The stake of eradicating piracy is high. In 2008, despite China's low income level, the revenue loss due to software piracy in the country was the second highest in the world, behind the United States, at more than $6,000 million (Table 3). For Hong Kong, the same statistics was $225 million, which was among the highest in terms of per capita loss.

These estimates on revenue loss are, however, likely to be higher than they actually are. In particular, in its annual piracy reports, in calculating the "commercial value of unlicensed PC

[10]See New York Times April 13th, 2009 at `http://www.nytimes.com/2009/04/13/technology/internet/13iht-piracy13.html?_r=1&scp=1&sq=french\%20government,\%20peer\%20to\%20peer&st=cse`.

Table 3: Countries with $1,000 Million or More in Piracy Loss ($M)

Country	2008 ($M)
United States	9,143
China	6,677
Russia	4,215
India	2,768
France	2,760
Germany	2,152
Hong Kong	225

Source: BSA

software," the BSA counts each pirated software (both from street piracy and Internet piracy) as a lost sale and calculates losses as follows:

$$\text{\$ Losses} = \text{\# Pirated Software Units} \times \text{Average System Price.} \tag{1}$$

This is not a good estimate of revenue loss because of two reasons. First, a government might over-estimate the benefit of eradicating only street piracy. For example, suppose the total number of pirated software units is 100,000 (40% from street piracy, 60% from Internet piracy), and the price of one legitimate copy is $10. Using the BSA methodology, the government would estimate the gain in eradicating street piracy to be $40,000, while in fact the true gain is lower due to the high substitutability between street piracy and Internet piracy.

Second, even if the government is able to shut off both street piracy and Internet piracy, the potential gain is still lower that what the BSA estimates imply. Because there usually are substitutes. Either a person can choose not to buy and use the software at work or at school, or a person can use free and legitimate substitutes like Google Docs for Office.

3 Data Collection and Description

To estimate the substitutability between street piracy and Internet piracy, I need a panel data set on the consumption choices from legal source, street piracy, and Internet piracy. However, like other illegal activities, data with such features are difficult to come by. This leads me to collect conjoint survey data from college students in Hong Kong.

3.1 Conjoint Survey

I conducted the survey in Fall 2008 and Spring 2009 in seven undergraduate classes in the Open University of Hong Kong (OUHK) and the Chinese University of Hong Kong (CUHK), which allowed for potentially 400 students. Of these, 281 students turned in their surveys. The survey was run at the end of classes. Most students finished the survey within 10 minutes.

I focus on one particular software product in this survey—Office, one of the most popular desktop applications and is also heavily pirated throughout the world.

The whole survey consists of two parts. In the first part, students report their demographic information and consumption behavior of other copyrighted goods like Microsoft Windows and movies.

The second part is the conjoint survey. Green and Rao (1971) first introduces conjoint analysis in marketing. I follow the approach of Louviere and Woodworth (1983) to use choice-based conjoint, which integrates conjoint analysis with discrete choice analysis. Respondents make choices in hypothetical situations with hypothetically set prices and other product attributes for different products. Conjoint survey data are also known as "stated-preference" data, as opposed to "revealed-preference" data, which are collected from real market transactions.

Conjoint analysis is sometimes called "trade-off analysis." It is based on the assumption that purchase decisions are based on multiple criteria and consumers make trade-offs among several product attributes. It is thus important to decide which attributes to include in a conjoint survey.

At the beginning of the conjoint survey, I describe the package of Office, which includes four applications:

- Microsoft Word 2007
- Microsoft Excel 2007
- Microsoft PowerPoint 2007
- Microsoft FrontPage 2007

Figure 2: A Sample Task of the Conjoint Survey

First choice: ☐1 ☐2 ☐3 ☐4 Second choice: ☐1 ☐2 ☐3 ☐4

Option 1:	Option 2:	Option 3:	Option 4
Buy a legal copy of Office CD	**Buy a pirated** copy of Office CD	**Download a pirated** copy of Office on the internet	Do not buy and use Office
$300	**$5**	$0 **30 mins** of search and download time	
Auto update	**No** update	**No** update	

There are ten hypothetical tasks in this conjoint survey.[11] Figure 2 shows a sample of a conjoint task (stimuli). In each task, respondents choose from one of four options to obtain the Office package described above: 1. Buy a legal copy of Office CD; 2. Buy a pirated copy of Office CD; 3. Download a pirated copy of Office on the Internet; and 4. Do not buy Office.[12]

In the survey, each choice is represented by three choice-specific covariates.

- Price (ranges from HKG$5 to HKG$5000)
- Search and download time (ranges from 5 minutes to 5 days)
- Automatic or no update

To ensure students can make choices that reflect their preference, each choice attribute must be presented in a precise and quantifiable way. For instance, the level of price is described by a number, say HK$500, instead of vague term like "expensive." For this reason, I do not include some attributes like the probability of getting caught downloading because it is more difficult to quantify in a way that is understandable to my sample.

The first choice-specific covariate is price (of either the legal or the pirated copy). It is a major criterion when consumers decide the source from which to obtain Office. The market prices of street-pirated Office, legal Office, and student versions of legal Office are approximately HKG$50, HKG$1170, and HKG$500. The second is the time of searching and downloading, which is essentially the price of obtaining the pirated copy on the Internet. It varies from 30 minutes to ten hours depending on the speed of the Internet connection and the availability of BitTorrent (BT) seeds. The last is the regular updates, mostly security updates, provided by Microsoft for most of its products including Office. Some respondents in the test-run of the survey indicated that one of the reasons for paying for a legitimate copy is the availability of updates. People who download Office illegally would have to download additional files to activate the automatic update (which requires additional search and download time). I did a search on some online forums in Hong Kong and found updates with relative ease.

[11] Johnson and Orme (1996) suggests that the reliability of the responses would not diminish with up to 20 tasks. However, due to the time limit for each survey run (ten minutes), I decide to put ten tasks in the survey.

[12] Students choose option 4 if they decide not to use Office in their computers. As we explain to them at the beginning of the survey, this does not prevent them from using Office from other sources (like computers in campus), or using substitutes of Office (like Google Docs).

In the survey, the covariates are drawn randomly and independently across tasks. This exogenous variation of covariates provides clean identification for my demand model. Also, there are I follow the three principles proposed by SawtoothSoftware (2008) to draw the levels of each covariate. The three principles are as follows:

1. **Minimal Overlap**: Each covariate level is shown as few times as possible in a single task.

2. **Level Balance**: Each level of a covariate is shown approximately an equal number of times.

3. **Orthogonality**: Covariate levels are chosen independently of other attribute levels, which ensures that each covariate level's effect on utility may be measured independently of all other effects.

To increase the variation of the covariates, I design five different set of surveys. With ten different tasks in each of the surveys, there are 50 different tasks in total.

3.2 Conjoint Survey Data vs Real Market Data

This section lists the advantages and disadvantages of using conjoint survey compared to using real market transaction data to estimate the demand for Office from different sources.

There are some limitations to using conjoint analysis. First, due to the time limit for each survey run (ten minutes), I do not include some attributes such as software bugs and quality of the software applications. It is reasonable to assume that they do not vary across different sources of Office. Thus, omitting them would not affect the estimates on the substitutability of Office from different sources. Moreover, because I am not interested in investigating which Office product is more likely to be pirated, fixing the Office products (Word, Excel, PowerPoint, and FrontPage) to be the same across options and tasks should not be a problem.

Second, conjoint analysis requires attributes to be quantifiable. However, some attributes, such as the likelihood of getting caught pirating Office, are difficult to quantify. In addition, no one in Hong Kong has ever been caught buying or downloading pirated Office. In the test run of the survey, which included the likelihood of getting caught (in terms of percentage) as one of the product attributes, respondents were confused with the interpretation of the attribute levels. I therefore decide not to include this as one of the product attributes.

People have concerns regarding the validity of conjoint survey data. Some think that real market data are more reliable because they are revealed-preference data. However, as mentioned in the Introduction, various studies in marketing and economics have applied the conjoint survey technique and shown that it can yield reliable demand estimates.

There are several advantages to using conjoint survey data instead of real market data in this research. First, real market data on the consumption of copyrighted goods from legal and various illegal sources are difficult to come by, and conjoint survey is possibly the only way to create a panel data set on this. There are some studies that investigate the impact of counterfeits/illegal downloads on the legitimate market, but they do not estimate the demand for piracy from different sources. Since Grossman and Shapiro (1988a) and Grossman and Shapiro (1988b), there have been numerous empirical studies on the sources and effects of counterfeits. Eisend and Schuchert-Guler (2006) provides a summary of empirical studies that investigate the reasons for counterfeit behavior. Most studies cited in the summary rely on survey or experimental data for their analysis. Qian (2008) analyzes the effect of counterfeits on various market outcomes. The study collects data on counterfeit sale quantities, prices, and costs from the brand-protection office of each authentic company in the shoe industry in China, and then combines the companies' annual financial statements and other relevant company records for the years 1993-2004 gathered from the Chinese Bureau of Statistics Industrial Census and separate surveys. It then uses a natural experiment approach to show that brands with less government protection differentiate their products through innovation, self-enforcement, and subtle high-price signals, among other factors, to reduce counterfeit sales. In the music record industry, there are numerous empirical studies on the effect of music piracy on album sales. Oberholzer-Gee and Strumpf (2007) and Blackburn (2004) gather panel data sets on music piracy by tracking individual illegal downloading behavior on a P2P network. They then combine weekly album sales with their novel data on weekly volumes of downloads to estimate the effect of illegal downloads on album sales. Hui and Png (2003) uses cross-country data on music CD sales and piracy level defined by IFPI from 28 countries, and shows that demand for CDs decreased with piracy between 1994 and 1998. Rob and Waldfogel (2006) conducts surveys in colleges to create a panel data set on legal music consumption and illegal downloading behavior. They use their data set to estimate the same effect. All studies mentioned, however, do not estimate demand for piracy from different sources. This is important for policy analysis because a policy to seize all counterfeit copies of Office may only shift people from street piracy to Internet piracy and may not, therefore, boost legal sales. To the best of my knowledge, this paper is the first paper to construct such a panel data set using conjoint survey.

Second, conjoint survey analysis provides good instruments. There can be two problems using price data from real market data. First, price is endogenously determined. As Berry, Levinsohn, and Pakes (1995) and Nevo (2000) illustrate, prices can be a function of unobserved product characteristics and be correlated with unobserved product heterogeneity. This will lead to a bias of the price estimate. Second, price variation for Office is small. Conjoint survey analysis can avoid these problems due to two reasons. First, prices, as one of the product attributes, are drawn exogenously and independently using the orthogonality principle described in the previous subsection. Second, as the designer of the survey, I can vary the prices of Office within a pre-specified range which can be substantially larger than that in the real market data.

3.3 Data Description

Table 4: Age and Family Income Distribution among Sample

Age	Percentage of the Sample
17-21	76%
22-29	18%
30 or above	6%

Family Income	Percentage of the Sample
HK\$0 - HK\$10,000	25.74%
HK\$10,001 - HK\$20,000	33.46%
HK\$20,001 - HK\$30,000	16.18%
HK\$30,001 - HK\$40,000	10.29%
HK\$40,001 - HK\$50,000	6.25%
HK\$50,001 - HK\$60,000	8.09%

Of the potential 400 students from the five undergraduate classes, 281 turned in their surveys. Table 4 presents the distribution of the students' family income and age. About 60% of the students have family income less than HK\$20,000 (\$2,500) per month, and less than 10% of them have family income more than HK\$60,000 (\$7,500) per month. Most students (76%) are

21 years old or below, which is the normal age of obtaining a bachelor's degree, while 6% of the students are 30 years old or above. The average age is 21. Table 5 summarizes some of the characteristics of the students.

Most students have exposure to Internet technology. On average, they spend about four hours per day on the Internet. Almost 70% of the students have used BT, a P2P software, to share digital files recently. As expected, younger students from CUHK have more exposure to new Internet technology: 70% of CUHK students have used BT recently compared to 60% of OUHK students.

Table 5: Data Description

	Mean (s.d.)	Min	Max
Age	21.37 (4.18)	17	50
Have Used BT Recently	0.68	0	1
Hours Spent on the Internet/Day	4.24 (3.22)	0.5	24
Use Legal Windows	0.72 (0.45)	0	1

N=281

Most students have experience with piracy, be it by purchase of counterfeit copyrighted products or by illegal downloading on the Internet. Table 5 shows that almost 30% of the students are using illegal copies of Microsoft Windows. The proportion is lower than that of other copyrighted goods because most copies of Microsoft Windows are pre-installed on a new computer. Approximately 60% and 70% of respondents are using illegal copies of anti-virus software and Office. Of those, more than half of them obtained the copy from the Internet (Table 6).

To determine how piracy behavior correlates with demographics, I run a logit regression of the sources of copyrighted goods (with 1 being illegal, 0 being legal) on several demographic variables. As Table 7 shows, students with a lower family income are more likely to obtain anti-virus software or Office through an illegal source. In addition, having used BT recently is also associated with a higher likelihood of obtaining Office through an illegal source.

Table 6: Sources of Copyrighted Goods

	Legal	Counterfeit CD	Illegal Download
Anti-Virus Software	30%	5%	65%
Office	38%	24%	38%

N=281

Table 7: Logit Regression of Piracy on Demographics*

	Anti-Virus Software	Office
Age	-0.076	0.506
	(0.214)	(0.450)
Age^2	-0.002	-0.012
	(0.004)	(0.009)
Family Income (HK$10,000)	-0.284***	-0.262**
	(0.092)	(0.092)
Hours Spent on the Internet/Day	-0.003	0.037
	(0.043)	(0.045)
Have Used BT Recently	-0.229	0.563*
	(0.319)	(0.308)
Constant	0.758	-4.675
	(2.900)	(5.444)
N	233	233

* Standard errors are reported in brackets. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

There is a concern that young students compose the majority of my sample, because one expects them to have a different demand for Office from sources compared to their older counterparts. This can bias my results in the counterfactuals. I address this concern by making use of the fact that my sample does cover some older students. Since I can elicit demand estimates based on demographics, I can do counterfactuals using young and old samples separately and test if there is a significant difference. More details can be found in Section 5.

4 A Discrete Choice Demand Model for Microsoft Office

An accurate evaluation of the loss due to piracy requires a thorough understanding of the demand for copyrighted products, both from legal and illegal sources. In this section, I lay out and estimate a model of the demand for Office from different sources using the conjoint survey data. In each task in the conjoint survey, students can choose either to buy a legal copy of an Office CD, to buy an illegal copy of an Office CD, to download an illegal copy of Office on the Internet, or to not buy Office. The standard indirect utility of a choice j for student i in task t is

$$U_{ijt} = \beta_{ij} + \phi_{i,price} P_{jt} + \phi_{i,dt} DT_{jt} + \phi_{i,update} Update_{jt} + \epsilon_{ijt}, \tag{2}$$

where P_j is the price of choice j, DT_j is the search and download time for j, and $Update_j$ is the availability of an update of j. I assume the random utility component, ϵ_{ijt}, is i.i.d. Type I extreme value distributed. We can express demand parameters of student i in Equation (2) as

$$\Theta_i = [\beta_{i1}; \beta_{i2}; \beta_{i3}; \phi_{i,price}; \phi_{i,dt}; \phi_{i,update}],$$

where Θ_i is a 1×6 vector of individual parameter. Then Θ is a $n_i \times 6$ matrix whose ith row is Θ_i, and n_i is the number of students in the sample. Define the covariates of choice j in task t as X_{jt} which is a 1×6 vector. We can then rewrite Equation (2) as

$$U_{ijt} = \Theta_i X_{jt} + \epsilon_{ijt}. \tag{3}$$

Student i's choice probability in task t has the following logit form:

$$Pr_{ijt} = \frac{\exp(U_{ijt})}{\sum_k \exp(U_{ikt}) + 1}. \tag{4}$$

Denote $j^*(t)$ as the choice chosen in task t of the conjoint survey, the likelihood for student i has the following form:

$$Pr_i = \prod_{t=1}^{10} Pr_{ij^*(t)t} \tag{5}$$

As Berry, Levinsohn, and Pakes (1995), Nevo (2000), Petrin (2002), and Rossi, Allenby, and McCulloch (2005) argue, random coefficients models generate better estimates of consumer demands and thus, better own and cross price elasticities compared to homogenous coefficients models. To exploit the panel structure of this conjoint data, I follow Rossi, Allenby, and McCulloch (2005) and use a hierarchical Bayesian model with a mixture of three components of normal priors to flexibly estimate the random coefficients. In addition to the *hyperparameters* that describe the distribution of the heterogeneity, the hierarchical Bayesian approach can also make an inference on the *individual-level* parameters as described below.

Since students provide some of their demographic information in the survey, I include some of these information to control for observed heterogeneity across students. Define Z_i as a $1 \times n_z$ vector of observable characteristics of i which has n_z elements, and Z as an $n_i \times n_z$ matrix. Following Rossi, Allenby, and McCulloch (2005), the demand model, where unobserved heterogeneity is distributed as a K mixture of normal, can be expressed as follows:

$$\begin{aligned} U_{ijt} &= \Theta_i X_{jt} + \epsilon_{ijt} \\ \Theta_i &= Z_i \triangle + u_i \\ u_i &\sim N(\mu_{ind_i}, \Sigma_{ind_i}) \\ ind_i &\sim \text{Multinomial}_K(\gamma) \end{aligned}$$

where γ is a vector giving the mixture probabilities for each of the K components, and $\triangle$ is a $n_z \times 6$ matrix of parameters determining the effects of demographics on each utility coefficients. For ease of illustration, I define $\delta = \text{vec}(\triangle)$. Thus, the individual-level demand parameters for student i, Θ_i, is a function of his demographics (including family income, recent BT experience, and age) and an unobserved factor, u_i. The unobserved factor, u_i, has a flexible distribution of a K-component mixture of normal. The set of hyperparameters that describe the distribution of the heterogeneity include δ (the demographics parameters), γ (the mixture probabilities for each of the K components), and μ_k and Σ_k (the mean and variance-covariance matrix of the kth-component of the distribution of the unobserved heterogeneity, u_i).

The complete specification with priors over the hyperparameters, including the mixture probabilities (α), the demographic coefficients ($\bar{\delta}$ and a_{δ}^{-1}), the means of the unobserved heterogeneity ($\bar{\mu}$ and a_{μ}^{-1}), and the covariance matrices for the unobserved heterogeneity (v and V), can be taken in convenient conditionally conjugate forms:

$$\begin{aligned} \delta &\sim N(\bar{\delta}, a_{\delta}^{-1}) \\ \gamma &\sim Dirichlet(\alpha) \\ \mu_k|\Sigma_k &\sim N(\bar{\mu}, \Sigma_k \times a_{\mu}^{-1}) \\ \Sigma_k &\sim IW(v, V) \\ \{\mu_k, \Sigma_k\} & \quad \text{independent} \end{aligned}$$

where the joint prior on μ_k and Σ_k is independent conditional on γ.

I follow Rossi, Allenby, and McCulloch (2005) by using a hybrid of Gibbs sampling and Metropolis-Hasting method to implement posterior inference for this model. I use a hybrid

Metropolis method that employes customized Metropolis candidate density to draw Θ_i for each student. Conditional on Θ_i, I use an unconstrained Gibbs sampler to draw δ, μ_k, and Σ_k.[13] In particular, I alternately obtain draws between individual-level parameters in (6) and hyperparameters in (7):[14]

$$\Theta_i | ind_i, Z_i \triangle, \mu_{ind_i}, \Sigma_{ind_i} \tag{6}$$

$$\gamma, ind, \triangle, \{\mu_k\}, \{\Sigma_k\} | \{\Theta\} \tag{7}$$

The conditional posterior in (6) is proportional to the product of the likelihood in (5) and the prior of the hyperparameters. I use the Random-Walk Metropolis to obtain draws of Θ_i. The draw of the hyperparameters in (7) can be broken down into a succession of conditional draws:

$$ind | \gamma, Z, \triangle, \{\mu_k, \Sigma_k\}, \{\Theta\} \tag{8}$$

$$\gamma | ind \tag{9}$$

$$\{\mu_k, \Sigma_k\} | ind, \Theta \tag{10}$$

$$\triangle | ind, Z, \{\mu_k, \Sigma_k\}, \Theta \tag{11}$$

where the draw of indicators in (8) is a multinomial draw based on the likelihood ratios with γ_k as the prior probability of membership in each component. The draw of γ given ind in (9) is a Dirichlet draw. The draw of each (μ_k, Σ_k) in (10) can be made using a standard algorithm to draw from a multivariate regression model. The draw of $\triangle$ in (11) requires that we pool data from all K components into one regression model.

There are several advantages with this approach. First, most random coefficient models in economics literature are implemented through an unconditional likelihood approach in which only the hyperparameters are estimated. The hierarchical Bayesian approach, however, can obtain inference on both individual-level parameters and hyperparameters. Second, most econo-

[13] One needs to impose constraints on the Gibbs sampler to fix an identification problem called "label switching" if inference is desired for the mixture component parameters. This is not a problem here because I am interested in estimating individual student parameters and their distribution across students only. An unconstrained Gibbs sampler is enough to ensure identification. See Rossi, Allenby, and McCulloch (2005) for more details.

[14] Interested readers can find the details of the implementation of the MCMC draws in Chapter 5 of Rossi, Allenby, and McCulloch (2005).

metrics models often restrict heterogeneity to subsets of parameters such as intercepts. There is no reason, however, to confine heterogeneity to intercepts because differences in price coefficients are important as well. The hierarchical Bayesian approach can incorporate heterogeneity for all coefficients without additional computation cost because it only requires a draw from a multivariate normal, instead of a univariate normal, distribution in a Gibbs step. Third, it is reasonable to expect a student who prefers Internet piracy (a high intercept coefficient on the choice of Internet-pirated Office) would also prefer street piracy and have a distaste for legal Office. In particular, we expect some correlation among some of the demand coefficients. This approach allows the demand parameters of student i to be correlated without additional computation time. I can back out the correlation among demand coefficients by obtaining draws of the variance-covariance matrix in (10) without restricting the off-diagonal entries to zero. Lastly, it is more flexible than the classical approach because it does not restrict students' unobserved heterogeneity to be normal distributed. Instead, it can be distributed as a mixture of normals, possibly with multi-mode. There are several conjoint studies using the hierarchical Bayesian model to estimate random demand coefficients, but most of them do not have a mixture of components of normal priors.[15]

4.1 Demand Estimates

I now report the empirical estimates of demand from the conjoint data. All basic logit estimates have the expected sign. Instead of reporting the estimates one by one, I offer some interpretations using the estimates. First, other things being constant, a respondent would be willing to pay up to HK$390 to substitute a street-pirated Office (without automatic update) for a legal Office. Second, other things being constant, a respondent would be willing to pay up to HK$122 (equivalent to 1.7 days of searching and downloading time) to substitute an Internet-pirated Office (without automatic update) for a legal Office.

For the random coefficients estimates, the hierarchical Bayesian model specified above includes heterogeneity in all utility coefficients: intercepts, price, download time, and availability of update. I first report the posterior mean of the heterogeneity effects attributed to student

[15] See Feit, Beltramo, and Feinberg (2010), Iyengar and Jedidi (2011), and Louviere, Hensher, and Swait (2000). Two notable exceptions are Chandukala, Edwards, and Allenby (2010) and Leung (2009).

Table 8: Posterior Mean of $\triangle^*$

	Legal Office	Street-Pirated Office	Internet-Pirated Office	Price	DT	Update
Age	-0.011	-0.001	-0.037	0.040	-0.0004	0.028
	[-0.045, 0.026]	[-0.043, 0.036]	[-0.083, 0.001]	[-0.019, 0.061]	[-0.031, 0.029]	[-0.022, 0.060]
BT (0 or 1)	-0.138	0.150	1.017	-0.195	-0.089	-0.145
	[-0.441, 0.227]	[-0.235, 0.528]	[0.690, 1.341]	[-0.367, -0.016]	[-0.298, 0.150]	[-0.508, 0.192]
Income (1-6)	0.105	-0.111	-0.195	0.052	-0.028	0.091
	[-0.028, 0.229]	[-0.222, 0.027]	[-0.328, -0.077]	[-0.011, 0.113]	[-0.096, 0.037]	[-0.008, 0.214]

* The 5th and 95th percentile of the estimates are reported in brackets.

demographics ($\triangle$) in Table 8.[16] I also report the 5th and 95th percentile of the draws. There are three things to note. First, younger students are more price-sensitive. When age increases by 1, the price coefficient would increase by 0.04 on average (which makes the price coefficient smaller in absolute value because the price coefficient is negative). Second, the group of students with recent BT experience (about 2/3 of the sample) exhibit substantial difference to the group without. The former group of students reveal a greater preference for Internet-pirated Office and a distaste for legal Office.[17] They are also more price-sensitive (the price coefficient becomes more negative and decreases by 0.19 on average for this group of students). Students with higher family income are also less likely to engage in Internet piracy.

The above estimation procedure provides a fitted density of utility coefficients across all students. Hence, I report the marginals of this joint distribution to show the need for flexibility in modeling unobserved heterogeneity.

Figure 3 plots the fitted densities of intercepts (μ) from the one- and three-component mixture models for all the six utility coefficients. The vertical line is the basic logit estimate.

The upper panel of Figure 3 provides compelling evidence of the need for a flexible model that can address unobserved heterogeneity. The intercept estimates for all sources of Office exhibit substantial dispersion in the distribution of the unobserved heterogeneity. For legal Office, the intercept estimates even exhibit bimodal distribution, indicating that there are two groups of respondents, one with a strong taste for and another with a strong distaste for legal Office.

Figure 3: Fitted Densities for Random Coefficients

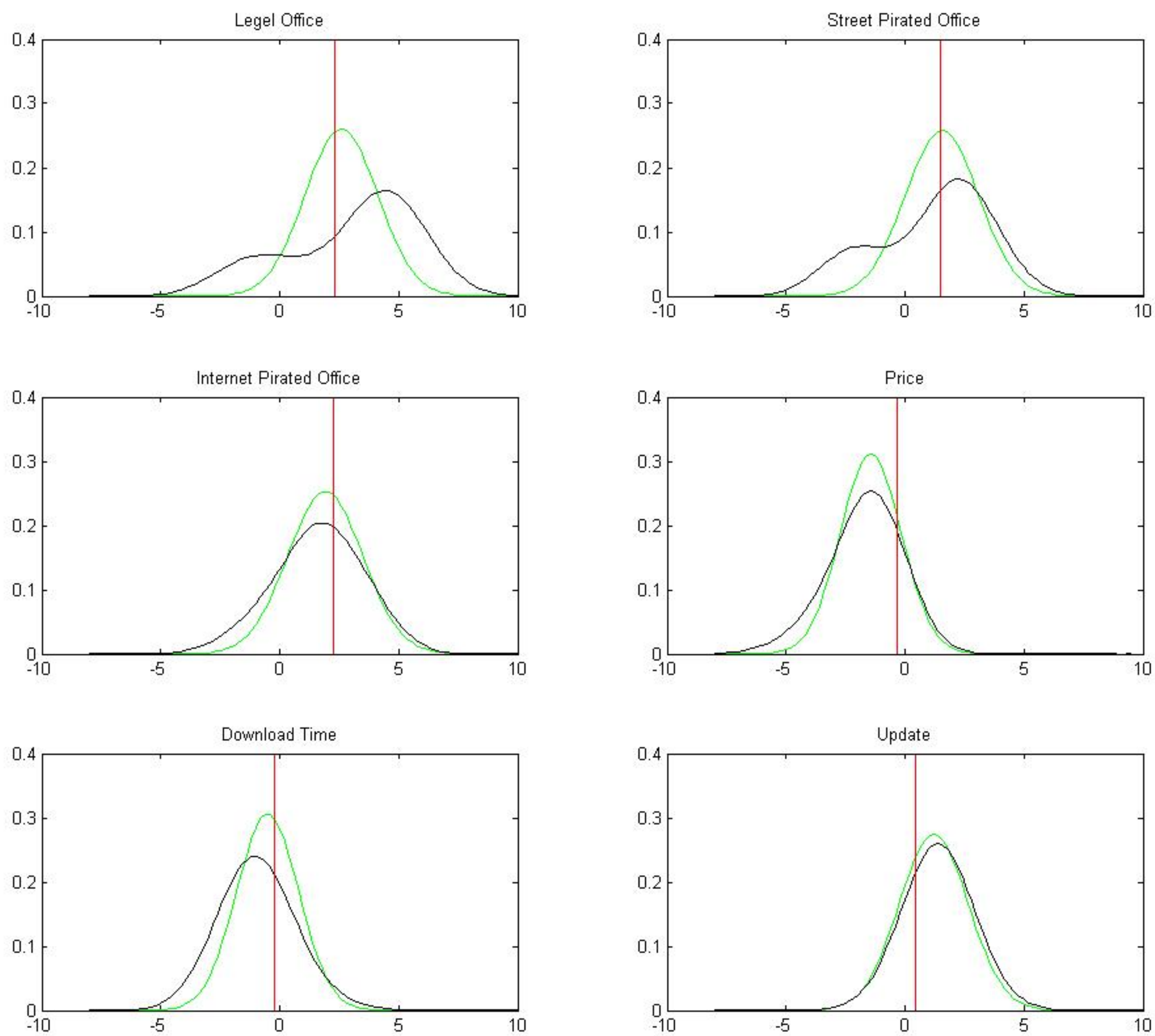


Table 9: Price Elasticities*

Price of Legal Office	Homo. Coef.	1 comp.	3 comp.
Legal Office Share	-1.321	- 1.110	-1.147
	[-1.333, -1.311]	[-1.287, -0.937]	[-1.372, -0.925]
Street Pirated Office Share	0.290	0.286	0.345
	[0.289, 0.292]	[0.224, 0.361]	[0.257, 0.444]
Internet Pirated Office Share	0.290	0.114	0.107
	[0.289, 0.292]	[0.091, 0.139]	[0.080, 0.134]
Price of Street Pirated Office	Homo. Coef.	1 comp.	3 comp.
Legal Office Share	0.035	0.041	0.045
	[0.035, 0.036]	[0.032, 0.051]	[0.034, 0.057]
Street Pirated Office Share	-0.126	-0.360	-0.363
	[-0.127, -0.125]	[-0.396, -0.325]	[-0.409, -0.320]
Internet Pirated Office Share	0.035	0.088	0.080
	[0.035, 0.036]	[0.075, 0.103]	[0.064, 0.097]
Download Time of Internet Pirated Office	Homo. Coef.	1 comp.	3 comp.
Legal Office Share	0.063	0.043	0.042
	[0.063, 0.064]	[0.030, 0.057]	[0.028, 0.059]
Street Pirated Office Share	0.063	0.092	0.109
	[0.063, 0.064]	[0.073, 0.113]	[0.087, 0.133]
Internet Pirated Office Share	-0.053	-0.064	-0.069
	[-0.054, -0.053]	[-0.076, -0.054]	[-0.081, -0.058]

* The 5th and 95th percentile of the estimates are reported in brackets.

4.2 Price Elasticities

Table 9 shows the elasticities implied by the coefficients, which illustrates how prices and download time affect the demand for Office. The three columns are the elasticity estimates under the homogenous coefficients model, one-component model, and three-component model. As we can see, the substitution patterns, demonstrated through the own and cross price elasticities, exhibit substantial difference between the homogenous coefficient and random coefficient models.

First, let us look at the own price elasticity of legal Office. A legal copy of Office is sold at HKG$500 ($60) to students in Hong Kong. Under this price, the own price elasticity for legal Office is approximately slightly above one at -1.147, which implies that the marginal cost for one copy of Office is approximately HKG$64 ($8) using the inverse elasticity rule. The high mark-up is expected in this industry because the fixed cost, in the form of R&D expenditure, is high for software like Office. Note that the own price elasticity is much higher at -1.32 under the homogenous coefficient model, which would over-estimate the marginal cost of legal Office.

Second, the random coefficients model also exhibits a more reasonable substitution pattern.

[16] I report the estimates from the three-component case. The other cases are similar and thus omitted here.
[17] The "distaste for legal Office" is not statistically significant, but it has the expected sign.

As expected, a one percent increase in legal Office would encourage more street and Internet piracy. However, the street piracy would increase more (about 0.4 percent) compared to Internet piracy (about 0.1 percent). Furthermore, when the cost of one type of piracy is higher, people tend to substitute that with another type of piracy rather than switching to purchase a legal copy. This more reasonable substitution pattern cannot be seen using the homogenous coefficients model.

Third, the elasticities of demand with respect to download time are small (less than 0.1%). Since people can do other things (like surfing on YouTube) while downloading Office through BT, the time cost of downloading is low and, thus, the demand is not responsive to download time.

5 Counterfactual

Table 10: Microsoft Office Market in Hong Kong

Price for Legal Office	HKG$500
Price for Street-Pirated Office	HKG$50
Download Time of Internet-Pirated Office	0.5 day

Exchange Rate: HKG$7.8/$

With the demand estimates of Office from different sources, I can proceed to evaluate different copyright policies. In this section, I first examine the effect of the copyright policy of the Hong Kong SAR government, which is to get rid of street piracy. Then I evaluate the true profit loss due to different types of piracies and contrast it with the profit loss calculated by the BSA method.

The counterfactuals are based on the market situation described in Table 10. The official version of Office specified in the survey costs approximately HKG$1170 ($150), but students can purchase a student version at HKG$500 ($64).[18] The prices of pirated Office on the street vary and are HKG$50 ($6.4) on average. The results do not change significantly if I vary the price from HKG$30 to HKG$100. The download time of an illegal copy of Office depends on the Internet connection speed and the popularity of the BT seed that one downloads the Office

[18] I recalculate the optimal price for legal Office under each policy.

file with.

There is a concern that young students compose the majority of the sample. As the estimates in Table 8 implies, they have a different substitution pattern compared to their older counterparts. In particular, older people have a higher tendency to buy Office from a legitimate source, and they are less price-sensitive. Thus, evaluating the effect of this policy based on this sample may underestimate the growth of demand of Office from a legitimate source.

To address this concern, I utilize the true demographic distribution in Hong Kong as basis to simulate utility coefficients of 10,000 individuals using the estimates from Table 8 and Figure 3.[19] There are several things to note in this simulation. First, I restrict the age to be between 15 and 50 (the oldest respondent in my sample is 50 years old). Second, I take the estimates from the Intellectual Property Department in Hong Kong and assume the proportion of people with BT experience to be 20%. Third, due to data limitation, I assume the distribution of the age, family income and BT experience to be independent. Fourth, I assume the simulated sample of students pays a discounted price for Office (HK$500) while the simulated sample of adults pays the standard price (HK$1175).

5.1 Policy: No Street Piracy

As pointed out in Section 2, governments in China and Hong Kong spend a great amount of effort to reduce street piracy. This can become less effective with more widespread Internet piracy because of the substitutability between street piracy and Internet piracy.

In this subsection, I evaluate the effectiveness of the copyright policy that completely gets rid of street piracy. In particular, I remove the option of obtaining counterfeit Office on the street. This copyright policy is effective if most of the demand for street-pirated Office goes to legal Office. I do this counterfactual exercise using the original sample, the simulated sample of students (defined as those simulated individuals with age less than 22), and the simulated sample of adults.

Table 11 shows the result of the counterfactuals. When the government completely gets rid of street piracy, the demand for Internet piracy would increase by 20.8%, about the same as the

[19] I obtain the demographic distribution information in Hong Kong from the Census conducted in 2006, which is publicly available.

Table 11: Changes in Estimated Demand When Street Piracy is Not Available*

	Original Sample	Sim. Sample (Students)	Sim. Sample (Adults)
Legal	↑ 21.7	↑ 20.3	↑ 23.8
	[17.5, 26.3]	[10.8, 37.9]	[4.9, 63.9]
Internet Piracy	↑ 20.8	↑ 20.3	↑ 24.7
	[17.3, 24.0]	[10.7, 38.7]	[8.7, 50.9]

* The estimates are in percentages. The 5th and 95th percentile of the estimates are reported in brackets.

increase in the demand for Office from a legal source (21.7%). The estimates are not drastically different between students and adults. In both simulated samples, the percentage increase in the demand for Internet piracy is about the same as that of the demand for Office from a legal source. Additionally, because the demand for Internet piracy is higher than the demand for legal Office (60% vs. 20% market share), this implies that most students who would choose street piracy would switch to Internet piracy instead of buying the legal version.

5.2 Revenue Loss From Piracy

The US Trade Representatives give prominent attention to piracy loss statistics from BSA. For instance, the 2007 report of the US Trade Representatives state: "According to industry (BSA) estimates, Singapore's piracy rate averaged five percent for music and twelve percent for movies. Business software losses were estimated at nearly $86 million in 2005."[20] These revenue loss estimates by BSA are misleading. While BSA, in its latest annual piracy report, admitted that "... not every unlicensed or stolen software product would be replaced by a paid-for version," it still reports the "commercial value of unlicensed PC software" which counts each pirated software as a lost sale and calculates losses using Equation (1).

Obviously, one less pirated Office does not translate into one more legitimate sale for several reasons. First, if a government only eradicates street piracy, as I show in previous subsections, most of the demand for street piracy would switch to Internet piracy. Second, even when both street piracy and Internet piracy are not available, there can still be various reasons not to buy legal Office. Buyers may already have access to legal Office at schools or at work, or there are

[20] US Trade Representative, 2007 report on Singapore, http://www.ustr.gov/assets/Document_Library/Reports_Publications/2007/2007_NTE_Report/asset_upload_file129_10979.pdf.

Table 12: BSA Over-estimates Piracy Loss*

Loss in Profit (Per Person) due to	This Paper's Estimates			BSA Estimates		
	Original Sample	Sim. Sample Student	Sim. Sample Adult	Original Sample	Sim. Sample Student	Sim. Sample Adult
Street Piracy	15.2	12.6	26.1	85.9	86.1	207.4
	[12.4, 18.2]	[5.5, 21.8]	[1.7, 63.7]	[78.2, 92.7]	[48.1, 136.1]	[80.9, 411.0]
Internet Piracy	15.2	16.6	21.4	236.5	238.1	415.4
	[11.2, 20.0]	[6.0, 27.9]	[1.6, 41.0]	[229.9, 242.8]	[169.5, 274.5]	[216.1, 605.6]
All Piracy	48.6	43.2	64.8	322.4	324.1	622.8
	[40.8, 56.5]	[22.2, 66.4]	[5.0, 108.4]	[315.3, 329.0]	[268.1, 343.7]	[402.0, 778.1]

* The estimates are in HK dollars. The 5th and 95th percentile of the estimates are reported in brackets.

free and legal substitutes like Google Docs provided by Google. These reasons suggest that the BSA's numbers on piracy loss can be substantially inflated.

I can use the estimates I have on the substitution pattern between different types of piracies and legal source of Office to test whether BSA has inflated the piracy loss and by how much. In particular, in this counterfactual, I remove the option of one (or both) type of piracy to calculate the loss in profits due to piracy under two different assumptions. Under the first assumption, consumers freely choose among the remaining options. Under the second assumption, consumers who would choose the piracy now eliminated are forced to buy the legal version of Office (which is the BSA assumption).

Table 12 shows the loss in profit due to different types of piracies under the two different assumptions.[21] There a few things to note. First, taking in account the substitution between street piracy and Internet piracy, the profit loss (per person) due to street piracy is about HK$15.2 (approximately $2). If there is no piracy at all, Microsoft's profit per person would increase by about HK$48.6 (approximately $6).

Second, there is a significant substitution among street piracy, Internet piracy, and the outside option. The profit loss estimates using the BSA assumption, which counts each pirated software as a lost sale, are about seven to nine times higher than my estimates.

Third, the profit loss estimates for the simulated adults group are higher because the price charged to them is higher. The profit loss estimates for both simulated groups are also less precisely estimated because some of the heterogeneity estimates attributed to demographics ($\triangle$) in Table 8 are not very precisely estimated.

[21] I use marginal profit (price - marginal cost), instead of price in Equation (1), to calculate loss in profit.

6 Conclusion

The net welfare effect of enforcing IPR depends on the cost and effectiveness of doing so. This paper argues that a copyright policy to eradicate street piracy can be ineffective due to Internet piracy. Even if people cannot buy counterfeit software copies from street hawkers, they can substitute by downloading software from the Internet. The effectiveness of such a policy is an empirical question. To answer the question, I construct a unique conjoint survey data set from 281 college students in Hong Kong, estimate the demand for Office from legal and different illegal sources, and then use the estimates to conduct counterfactuals. The results from the counterfactuals are twofold. First, most students would switch to Internet piracy if the government can eradicate street piracy. Second, the government might over-estimate the gain from eradicating piracy by up to nine times by assuming that each pirated copy represent a lost sale. With Internet piracy growing and seemingly impossible to stop, the net benefit of maintaining the current IPR system would decrease further.

References

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bessen, J., and R. M. Hunt (2007): "An Empirical Look At Software Patents," *Journal of Economics & Management Strategy*, 16(1), 157–189.

Blackburn, D. (2004): "Does File Sharing Affect Record Sales?," Ph.D. thesis, Harvard University.

Boldrin, M., and D. Levine (2008): *Against Intellectual Monopoly*. Cambridge University Press, New York, NY.

Carlsson, F., and P. Martinsson (2001): "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?," *Journal of Environmental Economics and Management*, 41(2), 179–192.

Chandukala, S. R., Y. D. Edwards, and G. M. Allenby (2010): "Identifying Unmet Demand," http://ssrn.com/abstract=1564290.

Chen, Y., and T. Puttitanun (2005): "Intellectual Property Rights and Innovation in Developing Countries," *Journal of Development Economics*, 78(2), 474–493.

Eisend, M., and P. Schuchert-Guler (2006): "Explaining Counterfeit Purchases: A Review and Preview," *Academy of Marketing Science Review*, 2006(12), 1–22.

Feit, E. M., M. A. Beltramo, and F. M. Feinberg (2010): "Reality Check: Combining Choice Experiments with Market Data to Estimate the Importance of Product Attributes," *Management Science*, 56(5), 785–800.

Ginarte, J. C., and W. G. Park (1997): "Determinants of Patent Rights: A Cross-National Study," *Research Policy*, 26(3), 283–301.

Green, P. E., A. M. Krieger, and Y. J. Wind (2004): "Adventures in Conjoint Analysis: A Practitioner's Guide to Trade-Off Modeling and Applications," http://marketing.wharton.upenn.edu/people/faculty/green/green_monograph.cfm.

Green, P. E., and V. R. Rao (1971): "Conjoint Measurement for Quantifying Judgment Data," *Journal of Marketing Research*, 8(3), 355–363.

Grossman, G. M., and C. Shapiro (1988a): "Counterfeit-Product Trade," *American Economic Review*, 78(1), 59–75.

——— (1988b): "Foreign Counterfeiting of Status Goods," *The Quaterly Journal of Economics*, 103(1), 79–100.

Hensher, D. (1994): "Stated Preference Analysis of Travel Choices: The State of Practice," *Transportation*, 21(2), 107–133.

Hensher, D., and J. Louviere (1983): "Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event," *The Journal of Cultural Research*, 10(3), 348–361.

Hensher, D., J. Louviere, and J. Swait (1999): "Combining Sources of Preference Data," *Journal of Econometrics*, 89(1-2), 197–221.

Hu, A. G., and I. Png (2009): "Patent Rights and Economic Growth," http://www.comp.nus.edu.sg/~ipng/research/patent_grwth.pdf.

Hui, K.-L., and I. Png (2003): "Piracy and the Legitimate Demand for Recorded Music," *Contributions to Economic Analysis and Policy*, 2(1), 1–22.

Iyengar, R., and K. Jedidi (2011): "A Conjoint Model of Quantity Discounts," http://marketing.wharton.upenn.edu/documents/research/Quantity_discounts.pdf.

Johnson, R. M., and B. K. Orme (1996): "How Many Questions Should You Ask in Choice-Based Conjoint Studies?," *Sawtooth Software Research Paper Series*.

Leung, T. C. (2009): "Should the Music Industry Sue Its Own Customers? Impacts of Music Piracy and Policy Suggestions," Mimeo, Chinese University of Hong Kong.

Louviere, J., D. Hensher, and J. Swait (2000): *Stated Choice Methods: Analysis and Application*. Cambridge University Press, New York, NY.

Louviere, J., and G. Woodworth (1983): "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, 20(4), 350–367.

Marron, D. B., and D. G. Steel (2000): "Which Countries Protect Intellectual Property? The Case of Software Piracy," *Economic Inquiry*, 38(2), 159–174.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy*, 9(4), 513–548.

Nordhaus, W. D. (1969): *Invention, growth, and welfare: A theoretical treatment of technological change*. MIT Press, Cambridge, MA.

Oberholzer-Gee, F., and K. Strumpf (2007): "Effect of File Sharing on Record Sales: An Empirical Analysis," *Journal of Political Economy*, 115(1), 1–42.

Orme, B. K. (2005): *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Research Publisher LLC, Madison, WI.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

Qian, Y. (2007): "Do National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analysis of Pharmaceutical Patent Protection, 1978-2002," *Review of Economics and Statistics*, 89(3), 436–453.

——— (2008): "Impacts of Entry by Counterfeiters," *The Quaterly Journal of Economics*, 123(4), 1577–1609.

Rob, R., and J. Waldfogel (2006): "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students," *Journal of Law and Economics*, 49(1), 29–62.

Rossi, P., G. Allenby, and R. McCulloch (2005): *Bayesian Statistics and Marketing*. John Wiley and Sons, Hoboken, NJ.

SawtoothSoftware (2008): "The CBC System for Choice-Based Conjoint Analysis," *Sawtooth Software Technical Paper Series*.