

# WHAT IS THE TRUE LOSS DUE TO PIRACY? EVIDENCE FROM MICROSOFT OFFICE IN HONG KONG

Tin Cheuk Leung\*

*Abstract*—Using a unique conjoint data set drawn from 281 college students in Hong Kong, I estimate a random-coefficient discrete choice demand system for Microsoft Office from legal and various illegal sources. Counterfactual results show two things. First, most students would switch to Internet piracy even if the government eradicated street piracy. This explains why software piracy in Hong Kong remains well above 40% despite the government's successful measures to bring down street piracy. Second, the true gain from shutting off all sources of piracy is HK\$48.6 (US\$6) per person, only 15% of the Business Software Alliance's estimated cost of piracy.

## I. Introduction

AS a member of the World Trade Organization since 1995, Hong Kong has been actively enforcing the copyright law required by the Agreement on Trade-Related Aspects of Intellectual Property Rights. The Prevention of Copyright Piracy Ordinance, enacted in 1998, requires all optical disc factories in Hong Kong to obtain manufacturer licenses. These factories are also subject to frequent inspection. An amendment of the same ordinance in 2002 further requires manufactures of stampers (“master discs”) in Hong Kong to obtain licenses from the Hong Kong Customs and Excise Department. The ordinance, together with efforts to crack down on syndicates of retail shops selling pirated software applications, has to a large extent eradicated street piracy in Hong Kong. As table 1 shows, the number of retail shops engaged in piracy dropped dramatically from about one thousand in 1998 to fewer than ten in 2010.

The almost total disappearance of street piracy, however, has not been accompanied by a corresponding slump in overall software piracy. According to the Business Software Alliance (BSA), the software piracy rate in Hong Kong declined from 59% in 1998 to only 45% in 2010.<sup>1</sup> One of the main reasons is that people increasingly substitute a pirated copy of software from a retail shop with a pirated copy downloaded from the Internet.

The lower cost of accessing the Internet, peer-to-peer (P2P) file-sharing software such as BitTorrent (BT), and close-to-zero risk of persecution are among the factors contributing

to the emergence of Internet piracy.<sup>2</sup> Whatever the reason, the widespread of Internet piracy had a stifling effect on the effectiveness of Hong Kong's copyright enforcement policy.

This is hardly an issue relevant only to Hong Kong. In many developing countries, such as mainland China, street piracy is a main concern of copyright holders. At the same time, the explosion of Internet use means that Internet piracy is also growing. For Hong Kong as well as these developing countries, the degree of substitutability between street and Internet piracy is an important determinant of the effectiveness of any copyright enforcement policy. For instance, in the extreme case in which street and Internet piracy are perfect substitutes, the level of software piracy would stay the same even if a government completely eradicated street piracy. How much the substitutability between different sources of piracy has stifled the effectiveness of such policy is an empirical question that this paper addresses.

An estimate of the substitutability of, and more general estimates of demand for, copyrighted products from legal and illegal sources will help us evaluate the effectiveness of different copyright enforcement policies and to estimate the “correct” profit gain from eradicating piracy. In its public campaign against piracy, the BSA often quotes estimates of profit loss due to piracy under the assumption that one fewer pirated copy will translate into a legitimate sale. An accurate estimate of substitutability among different sources of piracy and the outside options will correct the BSA's estimates.

This paper estimates the demand for Microsoft Office (hereafter Office) from legal sources and different sources of piracy, quantifies the substitution pattern among the different sources, and evaluates the effectiveness of different copyright enforcement policies in a world with more than one source of piracy. To the best of my knowledge, this is the first study to conduct such an analysis for any copyrighted product. For the empirical analysis, I designed a conjoint survey to collect a unique data set (section III) on Office, one of the most successful and also heavily pirated software applications. The survey data are from 281 college students in Hong Kong. In the survey, students provided information on their demographics and consumption of copyrighted products such as Office and then indicated the way in which they would acquire Office (from a legal source, through street piracy, or through Internet piracy) in ten hypothetical tasks. In each of the ten tasks, the prices of Office from different sources were exogenously randomized within a prespecified 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

Received for publication September 9, 2010. Revision accepted for publication March 29, 2012.

\* Leung: Chinese University of Hong Kong.

I am indebted to my thesis advisor, Pat Bajari, for his continuous encouragement and support. I thank Tom Holmes, Kyoo il Kim, participants in the HKUST applied micro workshop, and at FESAMES 2009, the CES 2010 Annual Meeting, the SAET 2010 Conference, and the second 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. Any errors are my own.

<sup>1</sup> The BSA uses survey data to estimate the number of total software units and pirated software units installed in each country. It then defines the piracy rate as the ratio of pirated software units to total software units.

<sup>2</sup> Only one person has ever been convicted of the illegal distribution of copyrighted works using BT in Hong Kong. The case occurred in 2005.

TABLE 1.—HONG KONG STREET PIRACY IN DECLINE

	1998	2003	2004	2005	2006	2007	2008	2009	2010
Optical disc seizure (in millions)									
Quantity	NA	6.2	7.3	3.8	3.0	4.3	2.2	2.0	0.89
Value	NA	135.3	157.8	89.2	72.4	99.1	52.3	71.0	26.0
Number of shops engaged in piracy	1,000	105	70	50	45	30	25	15	<10

Hong Kong Customs and Excise Department.

(section IV). 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. Implementation of posterior inference takes the form of a hybrid of Gibbs sampling and the Metropolis-Hasting algorithm. I then use the estimates to conduct counterfactuals to evaluate the effectiveness of various copyright enforcement policies (section V).

The main result is that the two sources of piracy are close substitutes for each other. When the government completely eradicates street piracy, the demand for Office from both the legitimate source and from Internet piracy will increase. Furthermore, because street and Internet piracy are close substitutes, the demand for Office from Internet piracy would increase more. Microsoft's expected profit will increase by only HK\$15.2 (\$2) per person in the sample.

In addition, even in a world without piracy, those who would usually acquire Office illegally would decide not to buy. By assuming one fewer pirated copy will translate into a legitimate sale, the BSA overestimates the loss due to piracy. The results in the counterfactuals show that the true gain from shutting off all sources of piracy is only 10% to 15% of what the BSA would claim it to be.

The remainder of this paper is organized as follows. Section II briefly reviews the related literature. Section III discusses the conjoint survey data set. Section IV sets up the demand model for Office and discusses the results of the estimation. Section V describes counterfactual experiments using the results from section IV. Section VI concludes the paper.

## II. Literature Review

This study is related to several others that have investigated the impact of counterfeits and illegal downloads on legitimate markets. Since the work of Grossman and Shapiro (1988a, 1988b), there have been numerous empirical studies on the sources and effects of counterfeits. Eisend and Schuchert-Guler (2006) provide 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 using data on counterfeit sale quantities, prices, and costs collected from the brand protection offices of authentic companies in the Chinese shoe industry and then combining the companies' annual financial statements and other relevant company records for 1993 to 2004 gathered from the Chinese Bureau of Statistics

Industrial Census and separate surveys. A natural experimental approach is then used 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.

Numerous empirical studies have also been conducted 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 downloads on a P2P network. They then combine weekly album sales with their novel data on weekly download volumes to estimate the effect of illegal downloads on album sales. Hui and Png (2003) use cross-country data on music CD sales and piracy levels defined by the International Federation of the Phonographic Industry from 28 countries and show that the demand for CDs decreased with piracy between 1994 and 1998. Rob and Waldfogel (2006) conduct surveys in colleges to create a panel data set on legal music consumption and illegal downloading. Due to data limitations, these papers do not separately account for the replacement effect of street piracy and Internet piracy.

This paper is also related to several studies that examine the level of intellectual property rights (IPR) protection across countries. Ginarte and Park (1997) develop a comprehensive index of patent protection and find that it is positively related to several economic variables. Marron and Steel (2000) use the BSA software piracy statistics and find that IPR protection depends not only on economic concerns but also on national culture and institutions. The substitution pattern between street piracy and Internet piracy analyzed in this paper is a determinant not examined in these studies.

The lack of real market data led to the use of conjoint survey analysis in this paper. Since the work of Green and Rao (1971), several studies have provided compelling evidence that conjoint survey data can generate reliable demand estimates.<sup>3</sup> Applications of conjoint survey analysis abound.<sup>4</sup> There are also several conjoint studies using a hierarchical Bayesian model to estimate random-demand coefficients,

<sup>3</sup>Carlsson and Martinsson (2001) and Hensher, Louviere, and Swait (1999) collect both stated-preference 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.

<sup>4</sup>Leung (2009) uses a similar approach to estimate the complementarity between music and iPods and to evaluate various copyright policies. Hensher and Louviere (1983) forecast 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 such as Marriott, Procter & Gamble, and General Motors also use conjoint survey data to estimate demand for new products (Green, Krieger, & Wind (2004) and Orme (2005)).

FIGURE 1.—A SAMPLE TASK OF THE CONJOINT SURVEY

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

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

but most of them do not have a mixture of components of normal priors.<sup>5</sup> Two notable exceptions are the studies of Chandukala, Edwards, and Allenby (2011) and Leung (2009).

### III. Data Collection and Description

One of the main contributions of this paper is that it provides estimates of the degree of substitutability between street piracy and Internet piracy. This requires a panel data set on the consumption choices from a legal source, street piracy, and Internet piracy. However, similar to other illegal activities, data with such features are difficult to come by for software piracy. This led me to collect conjoint survey data from college students in Hong Kong.

#### A. Conjoint Survey

I conducted the survey in fall 2008 and spring 2009 in seven undergraduate classes at the Open University of Hong Kong (OUHK) and the Chinese University of Hong Kong (CUHK). The survey was administered at the end of classes, and most students finished it within ten minutes. Of 400 students in the classes, 281 turned in their surveys.

I focused on one particular software product in the survey, Office, which is one of the most popular, and also heavily pirated, desktop applications.

The survey consisted 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 was the conjoint survey. Green and Rao (1971) were the first to use conjoint analysis in marketing. I follow the approach of Louviere and

Woodworth (1983) in using choice-based conjoint analysis, 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 made according to 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 described the Office package, which included four applications:

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

There were ten hypothetical tasks in the conjoint survey.<sup>6</sup> Figure 1 shows a sample of a conjoint task (stimuli). In each task, the respondents chose from one of four options to obtain the Office package: (a) buy a legal copy of the Office DVD, (b) buy a pirated copy of the Office DVD, (c) download a pirated copy of Office from the Internet, and (d) do not buy Office.<sup>7</sup>

<sup>6</sup> Johnson and Orme (1996) suggest that the reliability of the responses would not diminish with up to twenty tasks. However, due to the time limit for each survey run (ten minutes), I decided to put ten tasks in the survey.

<sup>7</sup> Students chose option 4 if they did not want to use Office on their personal computers. As explained to them at the beginning of the survey, this would not prevent them from using Office in other places (like computers in campus) or using substitutes (such as OpenOffice).

<sup>5</sup> See Feit, Beltramo, and Feinberg (2010), Iyengar and Jedidi (2012), and Louviere, Hensher, and Swait (2000).

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

- Price (ranging from HK\$5 to HK\$5,000)
- Search and download time (ranges from five minutes to five days)
- Automatic or no update

To ensure that the students could make choices that reflected their preferences, each choice attribute had to be presented in a precise and quantifiable way. For instance, the price level was described by a number, say HK\$500, instead of vague term like *expensive*. For this reason, I did not include attributes such as the probability of being caught downloading, which is difficult to quantify.

The first choice-specific covariate was price (of either the legal or the pirated copy). Price is a major criterion when consumers decide the source from which to obtain Office. The market prices of street-pirated Office, legally obtained Office, and student versions of legally obtained Office are approximately HK\$50, HK\$1,170, and HK\$500, respectively. The second covariate was the time spent searching and downloading, which was the price of obtaining the pirated copy on the Internet. It varied from thirty minutes to ten hours depending on the speed of the Internet connection and the availability of BitTorrent (BT) seeds. The last covariate was 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 was the availability of updates. People who download Office illegally would have to download additional files to activate the automatic updater, which requires additional search and download time. I searched online forums in Hong Kong and found updates with relative ease.

In the survey, the covariates were drawn randomly and independently across tasks. This exogenous variation of covariates provided clean identification for my demand model. I also followed the three principles proposed by Sawtooth Software (2008) to draw the levels of each covariate.

1. Minimal overlap: Each covariate level should be shown as few times as possible in a single task.
2. Level balance: Each covariate level should be shown approximately an equal number of times.
3. Orthogonality: Covariate levels should be chosen independent of other attribute levels, which will ensure that the effect of each covariate level on utility can be measured independent of all other effects.

To increase the variation of the covariates, I designed five different sets of surveys. With ten different tasks in each of the surveys, there were fifty different tasks in total.

### B. Conjoint Survey Data versus Real Market Data

There were some limitations to using conjoint analysis. First, due to the time limit for each survey run (ten minutes), I could not include 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 across tasks is not a problem.

Second, conjoint analysis requires attributes to be quantifiable. However, some attributes, such as the likelihood of being caught pirating Office, are difficult to quantify. In addition, no one in Hong Kong has ever been punished for 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 thus decided 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, 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 or illegal downloads on the legitimate market, but they do not estimate the demand for piracy from different sources. This is important for policy analysis because a policy to seize all street pirated copies of Office may only shift people from street to Internet piracy and may not boost legal sales. To the best of my knowledge, this paper is the first 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 can be correlated with unobserved product heterogeneity. This leads to bias in the price estimate. Second, price variation is small for Office. Conjoint survey analysis can avoid these problems because prices, as one of the product attributes, were drawn exogenously and independently using the orthogonality principle described in the previous subsection. Moreover, as the designer of the survey, I could vary the prices of Office within a prespecified range that was substantially larger than what it would have been in real market data.

TABLE 2.—AGE AND FAMILY INCOME DISTRIBUTION AMONG SAMPLE

	Percentage of the Sample
Age	
17–21	76%
22–29	18%
30 or above	6%
Family income	
HK\$0–10,000	25.74%
HK\$10,001–20,000	33.46%
HK\$20,001–30,000	16.18%
HK\$30,001–40,000	10.29%
HK\$40,001–50,000	6.25%
HK\$50,001–60,000	8.09%

TABLE 3.—DATA DESCRIPTION

	Mean (SD)	Minimum	Maximum
Age	21.37 (4.18)	17	50
Have used BT recently	0.68	0	1
Hours spent on the Internet per day	4.24 (3.22)	0.5	24
Use legally obtained Windows	0.72 (0.45)	0	1

*N* = 281.

TABLE 4.—SOURCES OF COPYRIGHTED GOODS

	Legal	Counterfeit CD	Illegal Download
Antivirus software	30%	5%	65%
Office	38%	24%	38%

*N* = 281.

### C. Data Description

Table 2 presents the distribution of the student respondents' family income and age. About 60% of the students had family incomes of less than HK\$20,000 (\$2,500) per month, and less than 10% of them had family incomes of more than HK\$60,000 (\$7,500) per month. Most of the students (76%) were aged 21 years or less, the normal age of obtaining a bachelor's degree; 6% of the students were aged 30 years or more. The average age was 21. Table 3 summarizes certain student characteristics.

Most of the students had exposure to the Internet and, on average, spent about four hours per day on it. Almost 70% of the students had used BT to share digital files recently. Among them, 70% of the CUHK students, who were on average younger, had used BT recently compared to 60% of OUHK students.

Most of the students had experience with either one or both types of piracy. Table 3 shows that almost 30% of the students were using pirated copies of Microsoft Windows. The proportion is lower than that of other copyrighted goods because most copies of Microsoft Windows are preinstalled on new computers. Approximately 60% and 70% of the students were using pirated copies of antivirus software and Office. Of those, more than half of them had obtained copies from the Internet (table 4).

To illustrate how piracy correlates with demographics, I ran a logit regression of the sources of copyrighted goods (with 1 being illegal, 0 being legal) on several demographic variables.

TABLE 5.—LOGIT REGRESSION OF PIRACY ON DEMOGRAPHICS

	Antivirus Software	Office
Age	−0.076 (0.214)	0.506 (0.450)
Age <sup>2</sup>	−0.002 (0.004)	−0.012 (0.009)
Family income (HK\$10,000)	−0.284*** (0.092)	−0.262** (0.092)
Hours spent on the Internet/day	−0.003 (0.043)	0.037 (0.045)
Have used BT recently	−0.229 (0.319)	0.563* (0.308)
Constant	0.758 (2.900)	−4.675 (5.444)
<i>N</i>	233	233

Standard errors are reported in parentheses. Significant at \*\*\*1%, \*\*5%, and \*10%.

As table 5 shows, students with a lower family income are more likely to obtain antivirus software or Office through an illegal source. Having used BT recently is also associated with a higher likelihood of obtaining Office through an illegal source.

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

### IV. A Discrete Choice Demand Model for Microsoft Office

An accurate evaluation of different copyright enforcement policies necessitates a thorough understanding of the demand for copyrighted products from both 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 could choose to buy a legal copy of an Office DVD, buy a pirated copy of an Office DVD, download a pirated copy of Office from the Internet, or not buy Office. The standard indirect utility of a choice *j* for student *i* in task *t* is

$$U_{ijt} = \beta_{ij} + \phi_{i,price} P_{jt} + \phi_{i,dt} DT_{jt} + \phi_{i,update} Update_{jt} + \epsilon_{ijt}, \quad (1)$$

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 that the random utility component,  $\epsilon_{ijt}$ , is i.i.d. type 1 extreme value distributed. We can express the demand parameters of student *i* in equation (1) as

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

where  $\Theta_i$  is a  $1 \times 6$  vector of individual parameters.  $\Theta$  is a  $n_i \times 6$  matrix whose  $i$ th row is  $\Theta_i$ , 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  $1 \times 6$  vector. Equation (1) can then be rewritten as

$$U_{ijt} = \Theta_i X_{jt} + \epsilon_{ijt}. \quad (2)$$

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

$$Pr_{ijt} = \frac{\exp(U_{ijt})}{\sum_k \exp(U_{ikt}) + 1}. \quad (3)$$

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

$$Pr_i = \prod_{t=1}^{10} Pr_{ij^*(t)t}. \quad (4)$$

As Berry et al. (1995), Nevo (2000), Petrin (2002), and Rossi et al. (2005) argue, random coefficient models generate better estimates of consumer demands and, thus, better own- and cross-price elasticities compared to homogeneous coefficients models. To exploit the panel structure of this conjoint data, I follow Rossi et al. (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 make an inference on the individual-level parameters as described below.

As the students provided some of their demographic information in the survey, I include aspects of that 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 et al. (2005) the demand model, where unobserved heterogeneity is distributed as a  $K$  mixture of normal, can be expressed as

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

where  $\gamma$  is a vector giving the mixture probabilities for each of the  $K$  components and  $\Delta$  is a  $n_z \times 6$  matrix of parameters determining the effects of demographics on each utility coefficient. For ease of illustration, I define  $\delta = \text{vec}(\Delta)$ . Thus, the individual-level demand parameters for student  $i$ ,  $\Theta_i$ , is a function of his or her 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 describes the distribution of the heterogeneity includes  $\delta$  (the demographics parameters),  $\gamma$

(the mixture probabilities for each of the  $K$  components), and  $\mu_k$  and  $\Sigma_k$  (the mean and variance-covariance matrix of the  $k$ th component of the distribution of the unobserved heterogeneity,  $u_i$ ).

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

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

where the joint prior on  $\mu_k$  and  $\Sigma_k$  is independent conditional on  $\gamma$ .

I follow Rossi et al. (2005) by using a hybrid of Gibbs sampling and the Metropolis-Hasting method to implement posterior inference for this model. I use a hybrid Metropolis method that employs customized Metropolis candidate density to draw  $\Theta_i$  for each student. Conditional on  $\Theta_i$ , I use an unconstrained Gibbs sampler to draw  $\delta$ ,  $\mu_k$ , and  $\Sigma_k$ .<sup>8</sup> In particular, I alternately obtain draws between individual-level parameters in equation (5) and hyperparameters in equation (6):<sup>9</sup>

$$\Theta_i | ind_i, Z_i \Delta, \mu_{ind_i}, \Sigma_{ind_i}, \quad (5)$$

$$\gamma, ind, \Delta, \{\mu_k\}, \{\Sigma_k\} | \{\Theta\}. \quad (6)$$

The conditional posterior in equation (5) is proportional to the product of the likelihood in equation (4) and the prior of the hyperparameters. I use the random-walk Metropolis to obtain draws of  $\Theta_i$ . The draw of the hyperparameters in equation (6) can be broken down into a succession of conditional draws,

$$ind | \gamma, Z, \Delta, \{\mu_k, \Sigma_k\}, \{\Theta\}, \quad (7)$$

$$\gamma | ind, \quad (8)$$

$$\{\mu_k, \Sigma_k\} | ind, \Theta, \quad (9)$$

$$\Delta | ind, Z, \{\mu_k, \Sigma_k\}, \Theta, \quad (10)$$

where the draw of indicators in equation (7) is a multinomial draw based on the likelihood ratios with  $\gamma_k$  as the prior probability of membership in each component. The draw of  $\gamma$  given  $ind$  in equation (8) is a Dirichlet draw. The draw of

<sup>8</sup>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 et al. (2005) for more details.

<sup>9</sup>Interested readers can find the details of the implementation of the MCMC draws in chapter 5 of Rossi et al. (2005).

each  $(\mu_k, \Sigma_k)$  in equation (9) can be made using a standard algorithm to draw from a multivariate regression model. The draw of  $\Delta$  in equation (10) requires that we pool data from all  $K$  components into one regression model.

There are several advantages to this approach. First, most random coefficient models in the 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 econometrics 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 also important. The hierarchical Bayesian approach can incorporate heterogeneity for all coefficients without additional computation cost because it requires only a draw from a multivariate normal, instead of 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 a legal copy of Office. In particular, I expect 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 equation (9) without restricting the off-diagonal entries to 0. Finally, it is more flexible than the classical approach because it does not restrict students' unobserved heterogeneity to being normal distributed. Instead, it can be distributed as a mixture of normals, possibly with multiple modes.

#### A. Demand Estimates

I now report the empirical estimates of demand from the conjoint data. Table 6 reports the log-marginal density for different model specifications. The posterior probability, a measure of the goodness of fit, is monotone in the log-marginal density. It should be noted that the log-marginal density already includes an automatic penalty for adding additional parameters (Rossi et al., 2005). There is significant improvement in fit from the simple logit model to random coefficient models. The goodness of fit is, however, about the same between the one- and three-component models. Indeed, the price elasticity estimates from the one- and three-component models are not statistically different from each other (see section IVB).

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 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 copy of Office (without automatic update) for a legal copy of Office. Second, other things being constant, a respondent would be willing to pay up to HK\$122

TABLE 6.—GOODNESS OF FIT

	Log Marginal Density
Homogeneous coefficient	-2,490.7
1 component	-1,000.8
3 components	-1,001.8

(equivalent to 1.7 days of search and download time) to substitute an Internet pirated copy of Office (without automatic update) for a legal copy of 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 demographics ( $\Delta$ ) in table 7.<sup>10</sup> I also report the 5th and 95th percentiles of the draws. There are three things to note. First, younger students are more price sensitive. When age increases by 1, the price coefficient increases 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 two-thirds of the sample) exhibits substantial difference from the group without recent BT experience. The former group has a greater preference for Internet-pirated Office and a distaste for legally obtained Office.<sup>11</sup> It is also more price sensitive (the price coefficient becomes more negative and decreases by 0.19 on average for this group). Students with higher family incomes are also less likely to engage in Internet piracy.

The estimation procedure above 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 2 plots the fitted densities of intercepts ( $\mu$ ) from the one- and three-component mixture models for all six utility coefficients. The vertical line is the basic logit estimate.

The upper panel of figure 2 provides compelling evidence of the need for a model that can address unobserved heterogeneity. The intercept estimates for all sources of Office exhibit substantial dispersion in the distribution of the unobserved heterogeneity. The random-coefficient model's ability to capture the level of dispersion of the unobserved heterogeneity is the main reason for its significant improvement in fit compared to the basic logit model.

#### B. Price Elasticities

Table 8 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 homogeneous coefficients model, one-component model, and three-component model. As can be

<sup>10</sup> I report the estimates from the three-component case. The other cases are similar and thus omitted here.

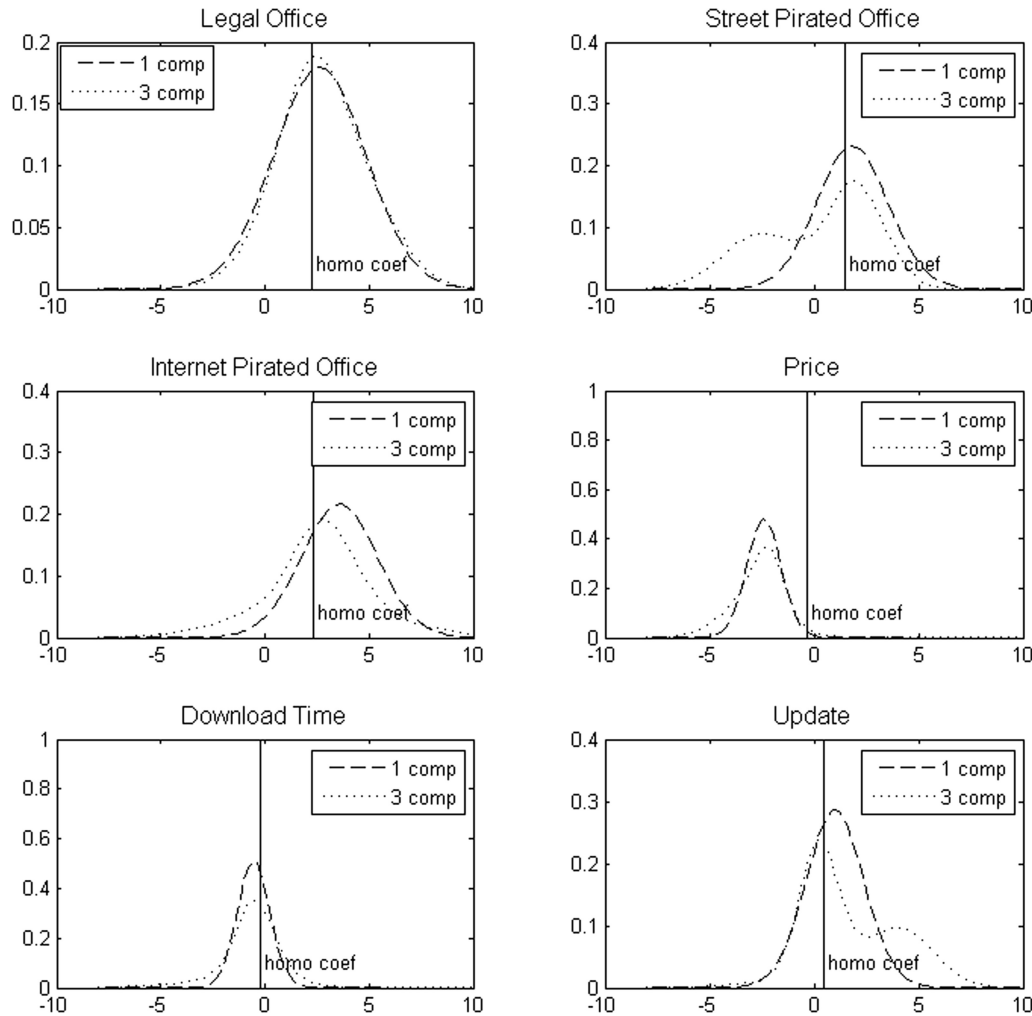
<sup>11</sup> The "distaste for legal Office" is not statistically significant, but it has the expected sign.

TABLE 7.—POSTERIOR MEAN OF  $\Delta^*$

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

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

FIGURE 2.—FITTED DENSITIES FOR RANDOM COEFFICIENTS



observed, the substitution patterns, demonstrated through own and cross-price elasticities, exhibit substantial difference between the homogeneous coefficient and random-coefficient models.

First, let us look at the price elasticity of legally obtained Office. A legal copy of Office is sold at HK\$500 (\$60) to students in Hong Kong. Under this price, its price elasticity is slightly above 1 at  $-1.147$ , which implies that the marginal cost for one copy of Office is approximately HK\$64 (\$8) using the inverse elasticity rule. The high markup 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 price elasticity is much higher at  $-1.32$  under the homogeneous coefficient model, which would overestimate the marginal cost of legally obtained Office.

Second, the random coefficient model also exhibits a more reasonable substitution pattern. As expected, a 1% increase in the price of legally obtained Office would encourage more street and Internet piracy. However, the street piracy would increase more (about 0.35%) than Internet piracy (about 0.1%). Furthermore, when the cost of one type of piracy is higher, people tend to substitute that with another type

TABLE 8.—PRICE ELASTICITIES

	Homogeneous Coefficient	One Component	Three Components
Price of legally obtained Office			
Legally obtained Office share	-1.321 [-1.333, -1.311]	-1.110 [-1.287, -0.937]	-1.147 [-1.372, -0.925]
Street pirated Office share	0.290 [0.289, 0.292]	0.286 [0.224, 0.361]	0.345 [0.257, 0.444]
Internet pirated Office share	0.290 [0.289, 0.292]	0.114 [0.091, 0.139]	0.107 [0.080, 0.134]
Price of street-pirated Office			
Legally obtained Office share	0.035 [0.035, 0.036]	0.041 [0.032, 0.051]	0.045 [0.034, 0.057]
Street-pirated Office share	-0.126 [-0.127, -0.125]	-0.360 [-0.396, -0.325]	-0.363 [-0.409, -0.320]
Internet-pirated Office share	0.035 [0.035, 0.036]	0.088 [0.075, 0.103]	0.080 [0.064, 0.097]
Download time of Internet-pirated Office			
Legally obtained Office share	0.063 [0.063, 0.064]	0.043 [0.030, 0.057]	0.042 [0.028, 0.059]
Street-pirated Office share	0.063 [0.063, 0.064]	0.092 [0.073, 0.113]	0.109 [0.087, 0.133]
Internet-pirated Office share	-0.053 [-0.054, -0.053]	-0.064 [-0.076, -0.054]	-0.069 [-0.081, -0.058]

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

TABLE 9.—MICROSOFT OFFICE MARKET IN HONG KONG

Price for legally obtained Office	HKG\$500
Price for street-pirated Office	HKG\$50
Download time of Internet-pirated Office	0.5 day

Exchange rate: HK\$7.8/US\$1.

of piracy rather than switch to purchase a legal copy. This more reasonable substitution pattern cannot be seen using the homogeneous coefficients model.

Third, the elasticities of demand with respect to download time are small (less than 0.1%). Because 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.

## V. Counterfactual

With the demand estimates of Office from different sources, I can 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 eliminate street piracy. Then I evaluate the true profit loss due to piracy from all sources and contrast it with the profit loss calculated by the BSA method.

The counterfactuals are based on the market situation described in table 9. The official version of Office specified in the survey costs approximately HK\$1,170 (\$150), but students can purchase a student version for HK\$500 (\$64). The prices of pirated Office on the street vary and are HK\$50 (\$6.4) on average. The results do not change significantly if I vary the price from HK\$30 to HK\$100. The download time of an illegal copy of Office depends on the Internet connection speed and the popularity of the BT seed that the Office file is downloaded with.

There could be a concern that young students compose the majority of the sample. As the estimates in table 7 imply, 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 the street piracy elimination policy based on this sample could underestimate the growth of demand for Office from legitimate sources.

To address this concern, I use the true demographic distribution in Hong Kong as a basis to simulate the utility coefficients of 10,000 individuals using the estimates from table 7 and figure 2.<sup>12</sup> There are several things to note in this simulation. First, I restrict the age to between 15 and 50, the age of the oldest respondent in my sample. 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 limitations, 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\$1,175).

### A. The No-Street-Piracy Policy

Hong Kong and many developing countries have put a great deal of effort into reducing street piracy. However, such efforts can become less effective with more widespread Internet piracy because of the substitutability between street and Internet piracy.

In this section, I evaluate the effectiveness of a copyright policy that completely eliminates street piracy. In particular,

<sup>12</sup>The demographic distribution information is from the Hong Kong Census conducted in 2006, which is publicly available.

TABLE 10.—EFFECTIVENESS OF A POLICY THAT ERADICATES STREET PIRACY

	Original Sample	Simulated Sample of Students	Simulated Sample of Adults
Increases in estimated demand (%)			
Legal	21.7 [17.5, 26.3]	20.3 [10.8, 37.9]	23.8 [4.9, 63.9]
Internet piracy	20.8 [17.3, 24.0]	20.3 [10.7, 38.7]	24.7 [8.7, 50.9]
Increases in estimated profit for Microsoft (HK\$ per person)			
This paper's estimate	15.2 [12.4, 18.2]	12.6 [5.5, 21.8]	26.1 [1.7, 63.7]
BSA's estimate	85.9 [78.2, 92.7]	86.1 [48.1, 136.1]	207.4 [80.9, 411.0]

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

I remove the option of obtaining counterfeit copies of Office on the street. This copyright policy is effective if most of the demand for street-pirated Office goes to legally obtained Office. I conduct this counterfactual exercise using the original sample, the simulated sample of students (defined as those simulated individuals younger than 22), and the simulated sample of adults.

The upper half of table 10 shows the result of the counterfactuals. When the government completely eliminates street piracy, the demand for Internet piracy increases by 20.8%, about the same as the 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 legally obtained Office (60% versus 20% market share), most students who would choose street piracy would switch to Internet piracy instead of buying the legal version of Office.

The lower half of table 10 shows the increase in Microsoft's expected profit under such a policy. Taking the substitution effect into account, the expected profit would increase by about HK\$15 per person in the sample if there was no street piracy. This estimate is much lower than the estimate calculated using the BSA assumption that each piece of pirated software is a lost sale, which is about HK\$86. The substitution of street piracy with Internet piracy and the outside option significantly undermines the effectiveness of the policy that eradicates street piracy alone.

### B. Other Copyright Enforcement Policies

To further illustrate the effectiveness of various copyright enforcement policies in a world with more than one channel by which to pirate software, I simulate the demand and the expected profit using the original sample under various piracy prices. In the simulation, the price of a pirated copy of Office on the street ranges from HK\$0 to HK\$500 (the highest in the conjoint survey for a pirated copy), and the download time ranges from zero to five days (the highest in the conjoint survey). As table 9 states, the current market price of a

pirated copy of Office is HK\$50 and the downloading time is 0.5 day.

Figure 3 shows the expected profit under various piracy prices. It provides a compelling case that a successful copyright enforcement policy necessitates action against all sources of piracy. If the government tackles only one type of piracy, the price of that type of piracy will go up, while the price of another type will stay the same. Figure 3 shows that the expected profit will increase only marginally. For instance, when the price of street-pirated Office increases from HK\$50 to HK\$500, with the download time staying at 0.5 day, the expected profit will increase only from HK\$70 to HK\$80. However, the same increase in expected profit can be reached when the price of street-pirated Office increases to HK\$200 and download time increases to one day.

### C. Revenue Loss from All Piracy

Suppose that the Hong Kong government could enforce copyright to perfection and completely eradicate piracy of all sorts. What would be the gains for Microsoft? The BSA answers this question by assuming that each piece of pirated software is a lost sale. However, absent any form of piracy, a person can choose not to buy and use the software at work or at school or can use free and legitimate substitutes such as OpenOffice. Ignoring the substitution of outside options could significantly overestimate the true gain from eradicating piracy.

I use the estimates of the substitution pattern between different sources of piracy and legal sources of Office to test whether the BSA has inflated the piracy loss and by how much. In particular, in this counterfactual, I remove the option of both sources of piracy to calculate the loss in profits due to piracy under two different assumptions. Under the first assumption, consumers can choose the outside option. Under the second assumption, consumers who would choose either type of piracy are forced to buy the legal version of Office (the BSA assumption).

Table 11 shows that ignoring the substitution effect can substantially inflate the estimates of gains from eradicating piracy. Because a significant portion of people would decide

FIGURE 3.—SIMULATED PROFIT UNDER DIFFERENT PIRACY PRICES

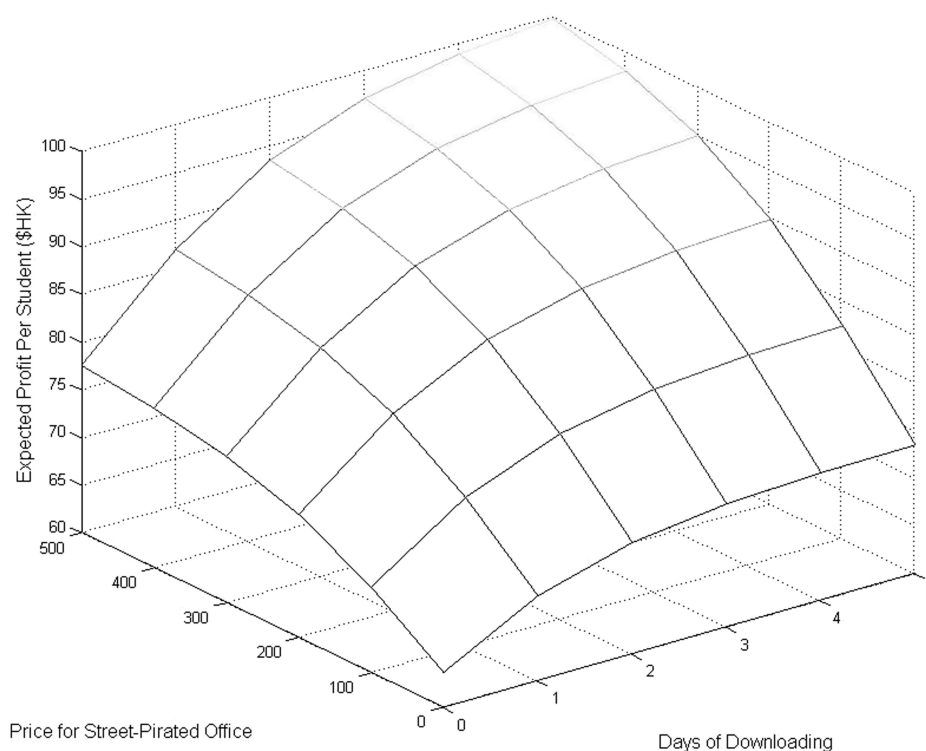


TABLE 11.—THE BSA OVERESTIMATES GAINS FROM ERADICATING PIRACY

Gains in Profit (per Person) Absent Piracy	Original Sample	Simulated Sample of Students	Simulated Sample of Adults
This paper's estimates	48.6 [40.8, 56.5]	43.2 [22.2, 66.4]	64.8 [5.0, 108.4]
BSA estimates	322.4 [315.3, 329.0]	324.1 [268.1, 343.7]	622.8 [402.0, 778.1]

Estimates are in Hong Kong dollars. The 5th and 95th percentiles of the estimates are reported in brackets.

not to buy if piracy was not possible, the true gain in profit would be 10% to 15% of what the BSA claims it to be.

## VI. Conclusion

Hong Kong's main copyright enforcement policy is to crack down on retail shops engaging in street piracy. Although these shops have almost disappeared, the software piracy rate is still well above 40%. This begs the question of how effective the copyright enforcement policy is in an environment where there is another form of piracy—Internet piracy. To answer the question, I construct a unique set of conjoint survey data from 281 college students in Hong Kong, estimate the demand for Office from legal sources and different sources of piracy, and then use the estimates to conduct counterfactuals.

There are two main results. First, most students would switch to Internet piracy even if the government eradicated street piracy. The expected profit for Microsoft would

increase by only HK\$15.20 (\$200) per person in the sample. This has profound implications for many developing countries where both street and Internet piracy exist. As the counterfactual results suggest, tackling both types of piracy would be more effective than focusing all resources on tackling street piracy alone.

Second, governments and international organizations such as the World Trade Organization, which often rely on the BSA's estimates on gain from eradicating piracy to guide their trade policies, need to treat those estimates with caution. The BSA's estimates ignore the substitution between piracy and outside options and are thus hugely inflated. The counterfactual exercise shows that the true gain from eradicating piracy is only 10% to 15% of the BSA's estimates.

## REFERENCES

- Berry, S., J. Levinsohn, and A. Pakes, "Automobile Prices in Market Equilibrium," *Econometrica* 63 (1995), 841–890.
- Blackburn, D., "Does File Sharing Affect Record Sales?" Ph.D. dissertation Harvard University, 2004.
- Carlsson, F., and P. Martinsson, "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?" *Journal of Environmental Economics and Management* 41 (2001), 179–192.
- Chandukala, S. R., Y. D. Edwards, and G. M. Allenby, "Identifying Unmet Demand," *Marketing Science* 30 (2011), 61–73.
- Eisend, M., and P. Schuchert-Guler, "Explaining Counterfeit Purchases: A Review and Preview" *Academy of Marketing Science Review* 2006:12 (2006), 1–22.
- Feit, E. M., M. A. Beltramo, and F. M. Feinberg, "Reality Check: Combining Choice Experiments with Market Data to Estimate the Importance of Product Attributes," *Management Science* 56 (2010), 785–800.
- Ginarte, J. C., and W. G. Park, "Determinants of Patent Rights: A Cross-National Study," *Research Policy* 26 (1997), 283–301.

- Green, P. E., A. M. Krieger, and Y. J. Wind, "Adventures in Conjoint Analysis: A Practitioner's Guide to Trade-Off Modeling and Applications" (2004), [http://marketing.wharton.upenn.edu/people/faculty/green/green\\_monograph.cfm](http://marketing.wharton.upenn.edu/people/faculty/green/green_monograph.cfm).
- Green, P. E., and V. R. Rao, "Conjoint Measurement for Quantifying Judgment Data," *Journal of Marketing Research* 8 (1971), 355–363.
- Grossman, G. M., and C. Shapiro, "Counterfeit-Product Trade," *American Economic Review* 78 (1988a), 59–75.
- , "Foreign Counterfeiting of Status Goods," *Quarterly Journal of Economics*, 103:1 (1988b), 79–100.
- Hensher, D., "Stated Preference Analysis of Travel Choices: The State of Practice," *Transportation*, 21 (1994), 107–133.
- Hensher, D., and J. Louviere, "Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event," *Journal of Cultural Research* 10 (1983), 348–361.
- Hensher, D., J. Louviere, and J. Swait "Combining Sources of Preference Data," *Journal of Econometrics* 89 (1999), 197–221.
- Hui, K.-L., and I. Png, "Piracy and the Legitimate Demand for Recorded Music," *Contributions to Economic Analysis and Policy* 2 (2003), 1–22.
- Iyengar, R., and K. Jedidi, "A Conjoint Model of Quantity Discounts," *Marketing Science* 31 (2012), 334–350.
- Johnson, R. M., and B. K. Orme, "How Many Questions Should You Ask in Choice-Based Conjoint Studies?" Sawtooth Software research paper (1996).
- Leung, T. C., "Should the Music Industry Sue Its Own Customers? Impacts of Music Piracy and Policy Suggestions," unpublished manuscript (2009).
- Louviere, J., D. Hensher, and J. Swait, *Stated Choice Methods: Analysis and Application* (Cambridge: Cambridge University Press, 2000).
- Louviere, J., and G. Woodworth, "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research* 20 (1983), 350–367.
- Marron, D. B., and D. G. Steel, "Which Countries Protect Intellectual Property? The Case of Software Piracy," *Economic Inquiry* 38 (2000), 159–174.
- Nevo, A., "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics and Management Strategy* 9 (2000), 513–548.
- Oberholzer-Gee, F., and K. Strumpf, "Effect of File Sharing on Record Sales: An Empirical Analysis," *Journal of Political Economy* 115 (2007), 1–42.
- Orme, B. K., *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research* (Madison, WI: Research Publisher, 2005).
- Petrin, A., "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy* 110 (2002), 705–729.
- Qian, Y., "Impacts of Entry by Counterfeiters," *Quarterly Journal of Economics* 123 (2008), 1577–1609.
- Rob, R., and J. Waldfogel, "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 (2006), 29–62.
- Rossi, P., G. Allenby, and R. McCulloch, *Bayesian Statistics and Marketing*. (Hoboken, NJ: Wiley, 2005).
- Sawtooth Software, "The CBC System for Choice-Based Conjoint Analysis," Sawtooth Software technical paper (2008).