Textbook Piracy: Exploring the Decision to Pirate or Not

Francisca Rebelo

September 15, 2014

Supervisor: Catarina Reis

Dissertation submitted in partial fulfillment of requirements for the degree of MSc in Economics, at the Universidade Católica Portuguesa.

Abstract

While much has been written about the consequences of the digitization of entertainment media (music, movies, etc.) and the best strategies to cope with copyright infringement, there is still little research on the effect of the digitization of books. This raises the question of how the publishing industry can learn from the experience of the entertainment industry. As tablets and e-readers become more prevalent, increasing the attractiveness of digital copies of textbooks and other materials, the issue of copying will become increasingly salient for academic texts. Demand for academic books differs from that of experience goods such as music, movies, or the wider range of books since the primary consumers, students, see textbooks not as cultural items but as a learning tool.

This thesis investigates the key factors in students' decision to acquire textbooks using panel data from a survey of students at Católica-Lisbon School of Business Economics. We model the decision to buy a textbook first using multinomial logit models in which the students' choice depends not only on the price and quality of the book but also their valuation of a textbook as a study tool. We then attempt to measure each student's tendency to pirate and incorporate that estimate into the multinomial model. The decision to pirate or not a given book seems to be based primarily on the perceived usefulness of the book, more strongly than price.

Contents

Acknowledgements

First and foremost, I would like to thank my advisor, Catarina Reis, for guiding me through the research process these past months. Additionally, I am indebted to all of the professors in the MSc in Economics program, especially those who attended a preliminary presentation in May, for their comments and advice throughout the year. In particular, I would be remiss not to mention Teresa Lloyd Braga for her constant support and guidance, as well as Leonor Modesto for her availability and advice.

Finally, writing this dissertation and my entire MSc experience would not have been the same without the support of my friends and classmates. And, of course, a big thank you goes to all my family (on both sides of the Atlantic), who patiently endured the ups and downs of the past few months, not to mention more than one overly detailed discussion of econometric models.

1 Introduction

We live in an increasingly digitized and networked world, in which new technology continually transforms the way we consume and share media. Most new creative content (text, photography, music, video) is available in digital form. Older content is increasingly made accessible in digital form. At the same time, new ways to share files have created new opportunities to copy and illegally distribute goods. This digital revolution has challenged the traditional means of distribution for information goods and reopened the question of the optimal level of copyright in society (Varian 2005).

Since music files tend to be smaller in size and the CD format is easily digitized, the music industry was the first to feel the effect of digital piracy. Napster and its successors are often blamed for rapidly declining CD sales, although empirical evidence shows these claims were largely exaggerated. The music industry was the first to attempt to control the proliferation of file sharing through litigation and technological solutions such as digital rights management (DRM). With the launch of iTunes, the music industry was also the first to offer a legal digital channel, directly competing with pirate channels (Rob & Waldfogel 2004, Varian 2005, Wu & Chen 2008, Scorcu & Vici 2013). As we discuss in the literature review, there has been in the last decade a growing body of theoretical and empirical research on the effects of piracy. This research seeks to find statistical evidence of sales displacement and the degree to which pirated goods caused the aforementioned loss of revenue.

The music industry was not the only one to experience piracy. It was joined quickly by other media industries, such as the software and movie industry. While there is a growing literature on the consequences of the digitization of entertainment media (music, movies, etc.), there is very little initial research on the effect of digitization of books. In part this scarcity of research reflects the fact that books have more recently been made available in digital form. This delay gives the publishing industry an opportunity to avoid errors made by the entertainment industry and learn from its successes. As tablets and e-readers become more prevalent, increasing the attractiveness of digital books, the issue of copyright protection will become increasingly salient. This thesis focuses on an important subset of the publishing industry: the textbook industry, which for decades has curated students' learning experiences at all grade levels.

Textbooks have a particularity in that their main buyers, students, may not recognize their cultural value, seeing them instead as a tool to facilitate passing a class or attaining a higher grade. Therefore, demand for academic books differs from that of experience goods such as music, movies, or non-academic books. The student determines the value of the book as a study tool, using as a basis of their choice the recommendation of the professor and their own study habits. In parallel, the student can choose from a variety of channels and formats in which to acquire the book, deciding whether to get the books in digital or physical form, and whether to obtain them through a legal or illegal channel (downloading or photocopying). This choice is conditioned by access to technology; ethics or attitudes toward pirating; and price, or willingness to pay for the book.

We investigate the key factors in students' decision to acquire or pirate textbooks using panel data from a survey of students at Católica-Lisbon School of Business Economics. We find that the most significant variables impacting the probability of pirating are gender (though not in the way we expected) and the

year students began the course, as well as price, GPA, and the relative importance of textbooks to the student.

Textbook piracy appears to be a small issue in the greater realm of economic concerns. But, we know that ideas and innovation are the engines of growth, and so incentivizing the development of new ideas is key to making countries more prosperous. Textbooks and other technical manuals play an important role in training society's future researchers and entrepreneurs. If digital piracy has a significant impact on the demand for scholarly works, then it may affect the incentive to write (on the part of the authors) and distribute (on the part of the publisher) these works. The short-term gains to students from obtaining a textbook for free can be offset by the long run welfare loss due the smaller supply. By understanding attitudes towards copyright infringement and the decision-process of consumers, we can design better policies that balance protecting the rights of the author with the incentives for consumers.

The following chapter reviews the literature on digital piracy. It begins with an overview of the transition to digital technology. It then summarizes the key insights from the experience of the music industry, and reviews the potential welfare implications of piracy. Chapter 3 describes the methodology we use and the survey data we collected. Chapter 4 discusses the results of the multinomial logit model we use to study the factors affecting the probability of three alternative decisions regarding each textbook: not acquiring the textbook, purchasing it, or pirating. Chapter 5 considers the sample probability of pirating, using it to measure a "propensity to pirate" for individuals and incorporating that estimate into the multinomial model. Finally we present some conclusions and suggestions for further research in Chapter 6.

2 Literature Review

Sharing digital files is a costless, non-rivalrous activity since the individual sharing files can keep a copy of the file they distribute. In addition, the Internet makes physical distance is irrelevant to the process of distribution, so it is perhaps unsurprising then that file sharing quickly became one of the most common online activities. In 1999, the year Napster was founded, there were few participants of file sharing networks. By 2006 there were 10 million simultaneous users on the major peer-to-peer (P2P) networks, with one in five Internet users using P2P services (33 million people). That year, US households swapped more than 300 million files monthly, a figure which continued to grow due to the low cost and network externalities. More recently, this trend has reversed slightly. The use of P2P networks fell to 11% of Internet users (21 million people) in 2012[1] due to a combination of increased legal pressure and the availability of legal downloading and streaming services (Oberholzer-Gee & Strumpf 2007; NPD 2012; Sandvine 2013).

As discussed above, since the music industry was the first to be affected by the onset of digital piracy, much of the literature in this area attempts to unravel the effect of piracy on music sales, as well as evaluating the strategies implemented to prevent or mitigate the impact from piracy. While music files

[1] In Europe, P2P networks still account for a large portion of traffic, largely because of the lack of availability of legal alternatives in many European countries.

and other entertainment goods are not directly comparable to textbooks, we can derive some lessons from the strands of research that have evolved.

We begin with a characterization of digital goods and piracy, including a discussion of the factors leading to or inhibiting piracy. We briefly review of the theoretical and empirical literature on the impact of piracy on sales, including an examination of some positive effects of piracy that counterbalance sales displacement. We then summarize the research on strategies firms can use to prevent piracy or mitigate its effects.

From there we discuss a broader, largely theoretical strand of literature on the welfare implications of piracy, both in the short and long run. This discussion leads us to a debate about the optimal level of copyright in society, in the context of the question of whether in the presence of file sharing there should be a change in our understanding of the relative costs and benefits of granting copyright.

2.1 Characterization of Digital Goods

Digital Goods

Music, movies, software, and books are all examples of information goods. They are characterized by high fixed costs of production and, even before the digital era, low marginal costs of distribution. Digital versions of these goods also share some important characteristics. First, they can be compressed without losing much information or quality, which means that copies can have a level of quality similar to the original. As a result, the marginal cost of reproducing digital products is effectively zero.

Additionally, the Internet and other recent technologies have driven down the cost of distributing digital goods, which, augmented by social network effects, facilitates product distribution. This reduction in the marginal cost of reproducing and distributing media goods, independent of piracy, should lead to an increase in both producer and surplus, as prices fall and demand increases. This effect is amplified by the fact that digitization also reduces the fixed costs of producing new goods, increasing the number of products available in the market. Finally, most digital goods are fairly complex, and therefore frequently have sampling effects, as consumers need substantial information to know how to value the good (Peitz & Waelbroeck 2006; Waldfogel 2012; Wu & Chen 2008).

Of course, it is the very same technologies that make it inexpensive to produce digital goods also make it easier to illegally copy and distribute the same goods (Wu & Chen 2008; Peitz & Waelbroeck 2006). File sharing is a low-cost, non-rivalrous activity with significant network externalities. While some of these files are shared between family and friends, the spread of digital copies occurs largely on peer-to-peer networks, which connect strangers across the globe. File sharing is a truly global phenomenon: while 90% of users are in developed countries, file sharing occurs in 150 countries and the correlation between user share and Internet use is fairly loose. Contrasting with this geographic dispersion is the fact that, in the case of music downloading, there is only a limited number of songs frequently copied, largely those on the top of the Billboard charts (Bakos, Brynjolfsson, & Lichtman 1999; Oberholzer-Gee & Strumpf 2007).

The Cost of Piracy

These characteristics largely explain the explosion of digital piracy, particularly since 1999 when Napster developed of peer-to-peer (P2P) file sharing technology. But while media obtained through P2P networks is free of financial costs, it is not a costless technology nor are its copies necessarily perfect substitutes for originals. If it were so, legal digital distribution channels such as iTunes or the Kindle store wouldn't have a customer base. A part of the cost is determined by whether the copy is a good substitute for the original file: there is some disutility from using a copy that has lower quality than the original (Danaher et al. 2010; Peitz & Waelbroeck 2006).

Danaher et al suggest that this cost of pirating is a fixed cost which is technology-specific rather than media-specific. This implies that if a consumer pays this fixed cost to download for a particular television show, the effect spills over to other television shows and possibly other media. We can understand this cost in a variety of ways: it could be the cost to learn to use the technology, the cost of overcoming moral qualms about piracy, the probability of being caught and punished, or even a convenience cost (it may take longer to find files). These costs could be seen as either fixed or variable. For example, one's moral qualms may depend on how much one downloads. Danaher et al use the removal of NBC content from iTunes in 2009 as an event study, concluding that the cost of piratingis a fixed cost–either learning or moral–since the increase in piracy after the removal of NBC content exceeded the legal sales before the removal. They further argue that, as technology becomes more sophisticated, the learning cost will be driven to zero, indicating that the bulk of this cost will be based on the perceived morality of piracy.

Who Pirates?

An important the question is: who engages in digital piracy? This question is particularly relevant for the textbook industry since students are among the main culprits of pirating generally, as they are typically poor in money but rich in free time. Pirates are more likely to pirate goods with higher prices, when there is more bandwidth available (since this lowers the time cost of piracy), and, in the case of music, more popular songs (Sinha & Mandel 2008; Battacharjee et al 2003)

What makes a consumer choose to pirate a good instead of purchasing it? Theories about the formation of ethics suggest that individuals' attitude towards piracy – the belief or lack thereof that downloading illegally is immoral – is influenced by perceived risks and benefits, habit, and social norms. This relates to the interpretation that the fixed cost to downloading can be interpreted as a moral cost. It also explains evidence from surveys which have found most Internet users don't believe that downloading is stealing, though this attitude is slowly changing. However, those who have never illegally downloaded music are more likely to perceive downloading as unethical. We can describe piracy then as a habit: once someone begins downloaded illegally (paying that fixed moral cost), they are more likely to continue engaging in piracy. This is analogous to the persistence in the decision to "go digital", Once a consumer converts to using digital media, short of lack of availability, he or she is likely to continue consuming that type of media (NPD 2012; Scorcu & Vici 2013; McCorkle et al. 2012; Lenhert & Fox 2000; Taylor 2004, Yoon 2002)

Finally, one of the hypotheses of this thesis is that there is likely more piracy of textbooks than entertainment goods, because textbooks are viewed differently by their main consumers, students. This idea is based on an observation by Scorcu and Vici, who note that there is a distinction between voluntary and compulsory reading. Wherein the former a consumer implicitly has a positive willingness to pay for the book, compulsory reading (such as textbooks) is characterized by a lack of recognition of the cultural value of the good. In particular, textbooks function as inputs into a human capital production function. They are useful to help students learn the material and pass a class. Therefore, the rights of the author (or publisher) are invisible in the decision making process of the student. Combined with the consideration that most textbooks are used by students for a short period of time, we would expect textbooks to be more commonly pirated than other media (Scorcu & Vici 2013).

2.2 Does Piracy Have a Real Impact on Sales?

Since music can be encoded into smaller files than films or books, music files were the first to be digitized, and therefore the first at risk of digital piracy. The music industry, followed quickly by the Motion Picture Association of America (MPAA) and other entertainment industry organizations, has long blamed piracy for a steep loss of sales in the first decade of the new millennium. However, both the theoretical and empirical literature studying the phenomenon reaches mixed conclusions. Most empirical studies find evidence of some sales displacement, though far less than trumpeted by the industry (Danaher et al. 2010; Oberholzer-Gee & Strumpf 2007; Rob & Waldfogel 2004). Theoretical results are also ambiguou, pointing to positive externalities from piracy: in particular information effects from sampling and the broadening of demand from deepened network effects and bringing low-value customers into the market.

Theoretical Literature

By making available a free version of a good, particularly one that is almost identical to the original good, pirate channels inevitably cause some consumers to opt for piracy, thus lowering sales. Nonetheless, the theoretical literature identifies a number of effects that could counteract revenue loss due to sales displacement. There may be some limits to the substitutability of downloads for legal sales: legal purchases may have superior quality or contain additional features such as liner notes or instructions (Oberholzer-Gee & Strumpf 2007).

In order for there to be sales displacement, there must exist some consumers who would've paid for the good and are now downloading it instead. The theoretical literature points out that a large portion of the consumers who practice downloading are, in fact, "low-value" consumers who may not have bought the good if digital copies were not available. Oberholzer-Gee and Strumpf note that, as file-sharers are generally time-rich but cash-poor, the fact that downloaders buy fewer records may be the result of a selection effect. Rob and Waldfogel describe this segmentation as a crude form of third degree price discrimination, in which consumers distribute themselves according to Internet access and willingness to pirate. Revenue only declines if those consumers who segment themselves into the "downloader" category would've otherwise bought the good. In this context, access to downloads actually can cause

the deadweight loss to shrink, improving overall welfare (Rob & Waldfogel 2004; Oberholzer-Gee & Strumpf 2007).

More indirectly, downloading changes the willingness-to-pay for music, though the direction of this effect is ambiguous. On one hand, the availability of a free version of the good lowers consumers' willingness to pay. However, a number of authors suggest that one strategy in the face of copying is to actually increase prices, incorporating the right to copy into the value of the good. The seller can increase their profit if the sales displacement is less than the willingness to pay for the right to copy (Besen 1986; Varian 2005; Takeyama 1994). This was the strategy academic journals followed in the 1970s after photocopiers became widespread: they raised subscription prices for institutions (such as universities and libraries) as individual consumers began to acquire mainly copies (Liebowitz 1985).

Another set of papers discusses the demand-augmenting effects of network externalities from P2P networks. Oberholzer-Gee and Strumpf note that in file server chat rooms it is common to discuss music, which may expose consumers to music they wouldn't be exposed to otherwise (Oberholzer-Gee & Strumpf 2007). Takeyama notes that, when there are network externalities, firms want to expand output either because of higher marginal revenue or to create installed bases. In this context, copying can be an efficient means of doing so, enabling price discrimination and increasing the number of users of the original good, as well as of complementary goods and services (Takeyama 1994; Peitz & Waelbroeck 2006; Xie & Sirbu 2005). There are also sampling effects which are related to network externalities. Consumers can use file sharing to sample music, and, if they like what they hear, they may subsequently buy the music or complementary goods, such as concerts. This effect is particularly relevant for digital goods which, as mentioned above, are often complex goods about which consumers require substantial information (Chellappa & Shivendu 2005; Vernik, Purohit, & Desai 2011; Shapiro & Varian 1999, Duchene & Waelbreock 2002). Copies can serve an informational role in this context, partially replacing marketing and promotion (Peitz & Waelbroeck 2003).[2]

Empirical Literature

Given the ambiguous effects identified in the theoretical literature, the issue of what effect piracy has actually had on the industries affected becomes an empirical one. On the one hand, there is a negative effect from sales displacement as consumers choose to pirate rather than buy. On the other hand, piracy creates or amplifies sampling and network effects which augment demand. Consequently, there is a long stream of studies using various methodologies to determine which effect dominates.

While these studies reach different on the effect of piracy, most find that there is some displacement, but generally less than expected or predicted. Liebowitz concludes that downloading is responsible for reductions in album sales. Zentner, using international time series aggregate data, finds that places with more Internet connections had bigger sales reductions and that self-reported downloaders were less likely to have purchased music recently. Rob and Waldfogel use a sample of US college students

[2]Here the differentiation between textbooks and other digital goods is essential, as many of these effects may not apply in our case. The main consumers of textbooks are the students themselves, and they typically restrict their search to the textbooks recommended by their professors. Therefore sampling effects and network externalities likely have little impact on the demand for textbooks.

and find that each downloaded song reduces purchases by 20%, with the effect being less accentuated for hit albums as opposed to the whole sample of music acquired by those surveyed. They also note that the students surveyed value the downloaded music significantly less than the music they purchase, supporting the idea that some of that music would not have otherwise been purchased. On the other hand, Oberholzer-Gee and Strumpf look at weekly downloads and sales of hit albums over several years and find that there is no statistical relationship between downloads and sales. Additionally, they reject what they call the "dropout" hypothesis (that is, that file sharers and buyers are two separate groups, and when a consumer begins to download they stop buying CDs), citing industry studies that indicate the downloaders continue to purchase legal CDs. They conclude that P2P networks are not the primary cause of the CD sales slump, holding responsible other recent adjustments in the music industry[3] (Rob & Waldfogel 2004; Oberholzer-Gee & Strumpf 2007; Liebowitz 2004; McCorkle et al. 2012; Zentner 2003).

2.3 Countering Piracy

Since the extent and rapid evolution of person-to-person networks took recording companies by surprise, the strategies the music industry implemented were highly reactionary. Some of the early tools used to attempt to control music piracy include litigation against downloaders, damaging file-sharing networks, and–the most popular–protecting media content through Digital Rights Management (DRM) systems. DRM systems are technologies that control the use of a file, limiting how many times a file can be copied, where it can be played, etc.

The literature on DRM has largely concluded that it is an ineffective and even counter-productive tool against piracy. The technological constraints on files lower the value of the good, which can actually lead consumers to copy more. While initially many music companies insisted on using DRM technology to protect digital files, eventually the trend began to reverse. Since DRM imposes restrictions on legal users, eliminating it can have a positive impact both on consumer surplus and profit since it increases the utility and willingness to pay of consumers who value more flexible files (Vernik, Purohit, & Desai 2011; Sinha, Machado, & Sellman 2010).

Recently, the model the industry has increasingly adopted is to compete directly with pirates by offering legal digital services (either paid downloading or streaming, subsidized by advertising revenues). The success of iTunes and other legal sites indicates that at least some consumers are willing to pay a positive amount to download music legally (Sinha & Mandel 2008; Wade 2004). A parallel effort is the use of advertising campaigns to change attitudes about piracy, increasing the moral cost to piracy.

2.4 Welfare Effects

Beyond the consideration of the effect that piracy has on sales, it is important to look at the overall effect on welfare in society. We now discuss some literature on the welfare effects of piracy in the short and long run.

[3]These changes include the end of a transition to CDs and a change in the way most albums are distributed from specialized stores to mega-stores such as Wal-Mart and Target.

Short Run Effects

As we've discussed above, there is certainly evidence of some sales displacement due to piracy, and consequently lower profit and producer surplus in the short run. However the literature suggests that aggregate welfare probably increases with the existence of piracy. On one hand, copying media goods is not a welfare loss, but a transfer from the producer to the consumer who acquires the value of the good without paying for it. Additionally, we saw that in the case of the music industry, much of the music consumed through piracy is "low value". That is, absent the ability to download, consumers would not have acquired the files, as their willingness to pay is below the current price (or even close to zero). In this case, access to pirated music allows the transformation of deadweight loss into consumer surplus, albeit without offsetting the previously mentioned losses in producer surplus (Rob & Waldfogel 2006, Oberholzer-Gee & Strumpf 2007, Waldfogel 2012).

One relevant strategy in the short run is to apply indirect appropriation, as in the previously mentioned case of the photocopier and academic journals (Liebowitz 1985). The issue with digital goods is that indirect appropriation requires some estimate of how many copies will be made of the original or who mainly copies in order to assess how to incorporate that "right to copy" in the price of the original or how to effectively discriminate prices. With digital goods, this is difficult (if not impossible) to determine, as a file may be copied thousands of times or merely dozens (Peitz & Waelbroeck 2006, Besen & Kirby 1989).

Long Run Effects

The analysis above assumes that the supply of information goods remains constant, which is a reasonable assumption in the short run. However, downloading does not only change the demand curve (with the availability of an alternative source), but can also change the supply of available goods. Which goods are offered in the marketplace are the product of demand (what consumers want to buy) and the ability of firms to appropriate this valuation as revenue (Rob & Waldfogel 2006). Piracy reduces the ability of firms to directly appropriate consumer valuations into revenue, possibly reducing supply.

That said, there is little evidence to date of a fall in the supply of media. On the contrary, the technological innovations of past decades reduced the cost of producing and distributing goods for copyright holders. The fixed cost of producing creative content has dramatically declined, faciliating an explosion of content.

Finally, digitilization has challenged the current intellectual property rights framework. The classical argument for intellectual property rights is based on a concern for incentivizing creators and distributors to create and market new goods and products. Granting a temporary monopoly to the authors or inventors of new products through copyright or patent laws, allowing them to price the good above marginal cost for a fixed time period, offsets the elevated fixed costs and risk of investing in these enterprises. In effect, copyright law is a trade off between higher incentives to produce with (temporarily) higher consumer prices (Landes & Posner 1989). In the current context, the cost of copying information is dramatically lowered, reducing the protection from copyright (Oberholzer-Gee & Strumpf 2007), but simultaneously reopening the question of what is the ideal level of copyright.

3 Methodology & Data

3.1 Data Collection

There is little direct data on piracy habits outside surveys conducted to study downloading behavior. We collected data from undergraduate students at Católica-Lisbon School of Economics in March and April of 2014. The surveys were administrated in two courses, a 2nd and 3rd year class (Econometrics and Finance II respectively) at the beginning of each class period. Finance II was chosen for being a third year class required for both Economics and Business Administration (BA) majors. Econometrics was chosen to supplement the data acquired previously because it is a second year course well frequented by both Economics and BA majors .

The survey included four sets of questions designed to measure the level of textbook piracy among the sampled students, along with a series of explanatory and control variables.[4] The first set of questions covered general demographics (gender, year, major and GPA) followed by a series of questions seeking to identify study habits. The students were asked to rank the relative importance of textbooks, lecture notes, and other study materials, as well as identifying the main drivers of the decision to obtain or not a textbook.

We then asked the students about their access to technology: whether they have access to a tablet, knowledge of P2P networks, and their reading preferences. We also inquired whether students downloaded a variety of media (music, movies, books, etc.) and, if so, their most common sources for that media.

Finally, we selected seven classes, all part of the core curriculum of the two majors and which have a single recommended textbook. For these classes we asked the students to identify whether they acquired the textbook and, if so, from where did they aquire the textbook.

Data on textbook price and quality was collected online. Prices (in Euros) were collected from the school bookstore website. This implicitly assumes that the textbooks sampled have roughly the same price differences in other stores, but this assumption is questionable. Textbook prices vary widely depending on the source, on the edition (international or not) and type of book (hardcover or paperbck). There is also a robust secondhand market for textbooks. Online versions or supplements of the textbooks often have different prices. These aspects make the measure of the actual cost of textbooks very noisy. By collecting data from the university bookstore, we hope to minimize that noise since the bookstore only makes available the most recent international edition and does not sell secondhand books.

The quality of the books is based on the customer ratings on Amazon.co.uk, a five-star rating scale. Quality is defined as the weighted average of the rating for the current edition and the previous edition.[5] We also collected additional information, such as the number of pages and editors, from Amazon.

[4] An copy of the survey distributed can be found in Appendix C.

[5] Since the textbook for *Introdução à Economia* only has a Portuguese edition and it not available on Amazon.co.uk, we do not have a measure of quality for that particular textbook.

3.2 Descriptive Statistics[6]

Individual Characteristics

Looking at a general characterization of the sample, we see that around half the students are male, 65% are Business Administration majors (as opposed to Economics majors), and the average GPA is 13.25 out of 20 (with the median being 14 out of 20). There is little difference in the GPA between male and female students (13.39 and 13.29 respectively), although there is some difference by declared major, with Economics majors on average having a higher GPA by around one grade point than Business Administration students (13.90 and 13.04, medians 14 and 12 respectively). The two majors differ in their distribution of gender, with 60% of Economics major being male as opposed to 41.11% of Business Administration students. Correspondingly, of the female students sampled almost three-fourth chose Business Administration (73.3%) while their male colleagues show a weaker preference for that major (57.14%). We don't expect choice of major to be particular significant, while GPA likely varies negatively with piracy, as students with higher GPA may be more "responsible" or law-abiding than their colleagues. Similarly, we expect students with a higher GPA to be more likely to acquire the textbook overall.

The availability of other materials, teacher recommendation, and the overall relevance of the book were most frequently identified as the most important factors in choosing to acquire a textbook. Additionally, textbooks were most often ranked last in the a list of relevant study materials. This implies that one of the more important factors in the decision to acquire a given textbooks is its perceived usefulness to the course. Interestingly, cost is not cited as one of the main considerations: only 21.5% of students stated it was the main impediment to getting a textbook. Cost was ranked in third place by most students, behind access to other materials (such as lecture notes and course exercises) and teacher recommendation. Price likely affects principally the decision to purchase (instead of pirating) a textbook, with less of an impact on the decision to acquire the textbook.[7]

	Have access to tablet	Prefer to read on paper	Familiar with P2P networks
2010	20.0%	100%	100%
2011	57.14%	84.13%	52.38%
2012	78.26%	82.61%	67.39%

Table 1: Technological Penetration

Regarding the importance of technological access, we see some clear distinctions between different groups. Looking at the access to tablets, reading preferences, and familiarity with P2P networks, there is a trend towards greater technological penetration, which can be seen when we disaggregate the data by

[6] In Appendix A, we have several tables that summarize the following information.

[7] It would be interesting to compare the results in this paper from surveying business and economics students to other disciplines. Many of the students rely far more on lecture notes and other resources for their studies as compared to students from other areas, such as the physical sciences. Scorcu and Vici, who survey students at the University of Bologna across majors, have an extended discussion of how these habits differ.

the year students began their course. We see below that students who began their college studies later have more access to tablets and slightly weaker preferences for reading on paper (as opposed to being indifferent or prefering reading on screens). Although from one year to the next access to technology is likely to be similar, we find some impact of time in this sample.

Overall, a majority of the students have access to a tablet or eReader (63.79%). However only 60% of female students had access to a tablet, compared to 67.89% of male students. The difference is even starker between the two majors: 57.89% of BA majors have access to a tablet, far less than the 75% of their economics colleagues. This effect may be driven in part by the gender differences, although there are likely other factors at play as well. Most students prefer to read on paper (87.07%) as opposed to on a screen (2.59%), with slight differences. 90% of female students prefer reading on paper compared with 80.3% of males. Again, probably due to the way gender is distributed across the two majors, there is a discrepancy there as well with 89.5% of business students prefering reading on paper as opposed to 82.5% of economics students. Even though many students have access to tablets, the penetration of the device has not yet substantially changed reading preferences.

Finally, sixty percent of all students sampled state that they are familiar with P2P networks, which, as we discussed above, have been one of the main medium for obtaining pirated digital information goods. While there is not substancial difference between majors, there is a substancial gender gap: 75% of male students are familiar with P2P networks while only 46.67% of female students are. This follows the literature on piracy which shows that more consumers who engage in piracy are male.

Figure 1 below shows the most frequent sources of downloaded media of the students sampled: online stores, internet search, P2P networks, or personal networks (such as family and friends). These sources vary substantially between different types of media. For example, around 40% of students reported that they acquire books and textbooks from friends and classmates, while only around 10% acquire movies and TV series from friends and classmates. This pattern is likely due to convenience – how individuals consume media – as well as availability. Additionally, individuals have different opinions about how much, what, and when in which it is acceptable to pirate, which naturally is reflected in the variation of what and how they actually pirate.

Internet searches are the preferred method of finding media. As expected, male students use pirate technology more than female students. Roughly 35% of males cite P2P networks as their main source for a given media while only 21.84% of female students do so. Female students are correspondingly more likely to rely on personal networks (24.68% for females vs. 20.08% for males) or online stores (8.8% vs. 3.03%).

Textbook Characteristics

Most of the textbooks recommended for the seven courses sampled are international editions of textbooks by American publishers, with the exception being the textbook for Introduction to Economics, which is a Portuguese edition.

The textbooks ranged in price from 36 euros to 72.40 euros, with an average of 59.30 euros. Removing the textbook for Introduction to Economics, which is a slight outlier, the average price of the textbooks

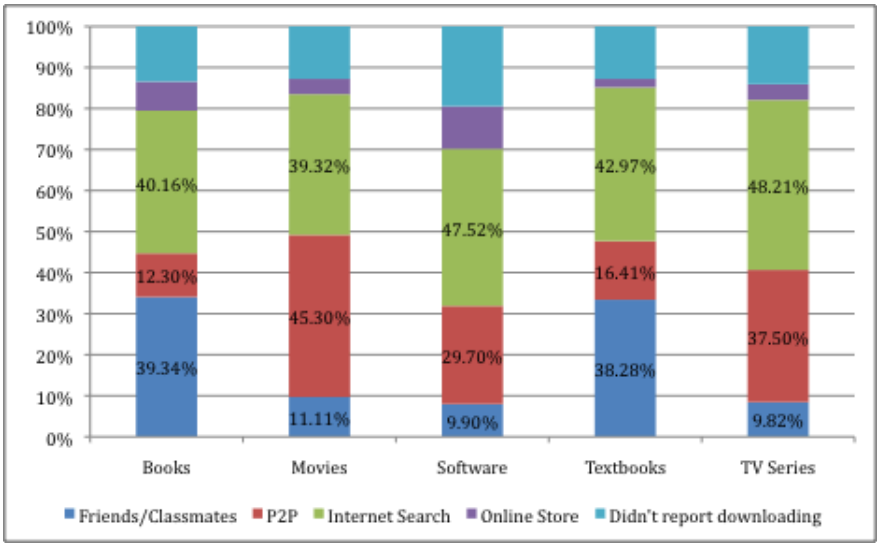

Figure 1: Sources of Downloaded Media

	Price (euros)	Quality (stars)	Pages
Introdução à Economia	36.00	-	672
Introduction to Marketing	53.58	4.40	720
Microeconomics I	55.57	4.50	672
Econometrics	61.11	4.68	900
Finance II	67.95	4.33	946
Macroeconomics I	68.64	4.13	720
Management Control Systems	72.40	3.42	896

Table 2: Book Characteristics

increases to 63.19 euros and all of the textbooks in the sample are within one standard deviation of the average with the exception of the Management Control Systems textbook, which is slightly more expensive but within two standard deviations.

In terms of quality, the books were rated on average with 4.24 stars out of five, with a range of 3.42 stars (Management Control Systems) to 4.68 stars (Econometrics). Again all of the textbooks are within one standard deviation of the average rating, with Management Control Systems again being an outlier, rated lower but within two standard deviation of the mean.

3.3 Observed Choices over Textbooks

On average students obtained 3.21 books, and bought 2 of those books. In percentage terms students acquired 70.24% of the books recommended for the classes indicted, and purchased 61.55% of those books.

We find, contrary to the literature, that in our sample female students pirate textbooks more than male students on average. While they obtain the textbooks recommended for class more frequently than their male colleagues (76.09% vs 63.73%), they buy fewer of those books (53.59% vs 70.41%).

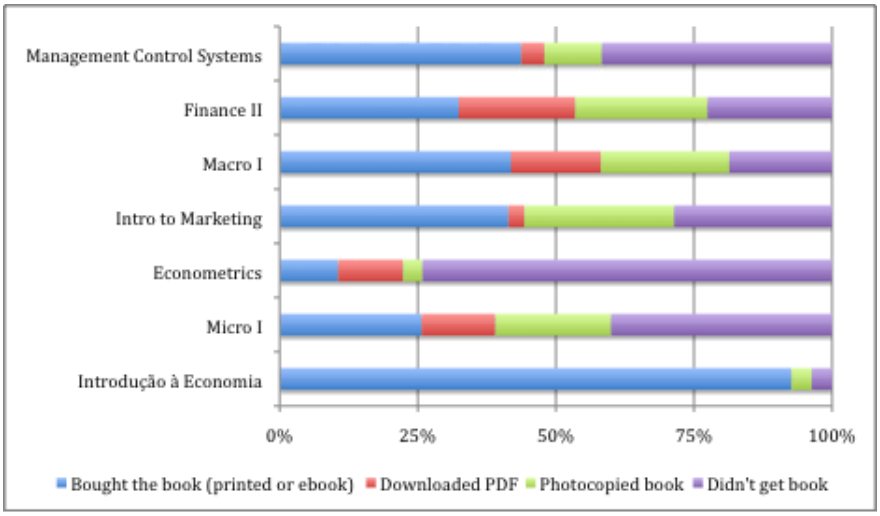

Figure 2: Source of Textbooks as % of students enrolled

Roughly 58% of students in each class who obtain the book, purchase it[8]. The remaining students were more likely to photocopy the textbook than to download it as a pdf. One exception is Introdução à Economia, which is the only class in this sample with a textbook written in Portuguese. The high proportion of students buying this textbook may also be related to its value for the class: the textbook is written by the professor who teaches the class, which is one of the first courses students attend at Católica-Lisbon. Other outliers include Econometrics, which has three times as many downloaders than photocopiers, and Management Control Systems, with the second highest proportion of purchased books.

	% Pirated	% Purchased	% Didn't Obtain
Introdução à Economia	3.67%	92.66%	3.67%
Microeconomics I	34.49%	27.71%	40.00%
Econometrics	15.29%	10.59%	74.12%
Marketing	30.00%	41.42%	27.14%
Macroeconomics I	39.55%	41.86%	18.60%
Finance II	45.07%	32.39%	21.13%
Management Control Systems	14.58%	43.75%	41.67%

Table 3: Book Sources, Enrolled Students

[8] 41.2% of those enrolled buy the book on avearge, and 32.74% of enrolled students didn't acquire the book at all.

In Table 2 we further summarize the source of each textbook by the percentage pirated, purchased, or not obtained, the three decisions that will be analyzed in the following chapters. We see again that Introduction to Economics and Econometrics are outliers, with the former being purchased by 92.66% of enrolled students (more than double the next highest observation) and Econometrics students abstaining from acquiring the book 74.12% of the time. These discrepencies are likely to be related to largely unobservable characteristics such as the usefulness of the book for the course.

4 Factors Impacting Decision to Acquire, Purchase, or Pirate Textbooks

When starting any course, a student faces two interrelated choices. They must decide whether to acquire the course materials at all. If they decide to acquire them, they must decide whether to purchase the materials (in the present case, textbooks, though a similar decision is faced when discussing software or other study tools) or obtain them by other means.[9] These intertwined decisions are naturally studied using multinomial models, where the dependent variable is a discrete set of qualitative choices.

4.1 Student Preferences: The Additive Random Utility Model[10]

As we describe above, each student faces a discrete choice for each course they take. Formally, for each course t, student i faces three alternatives (indexed by j): not acquiring the book at all, purchasing it, or pirating it. The student's utility depends on their choice over these alternatives, and this choice in turn depends on the characteristics of the student (study habits, etc.), characteristics of the book (price, quality, etc.), and attributes particular to each choice. The characteristics of the student are constant over t and j, while characteristics of the choice may vary over t as well as j. Analogously, book characteristics are constant over i and j. While many authors impose the assumption that pirated copies are of lesser quality, and therefore provide less utility than the legal version, we will assume that the utility is constant over j for simplification.

The utility student i receives from choice j for textbook t is therefore

$$U_{itj} = V_{itj} + \epsilon_{itj} \tag{1}$$

$$V_{itj} = f(\mathbf{z}_{itj}, \text{unknown paramters})$$

[9] These means are not necessarily illegal: students may borrow materials from others, for example. However the proportion of students doing so is small enough to ignore in our analysis. Similarly, we also define a "purchased" book as newly purchased textbooks or those bought from the second-hand market, since the number of students that purchased used books is statistically irrelevant in this sample. Additionally, due to the constraints of our dataset, we will treat photocopied and downloaded textbooks as equivalent.

[10] The additive random utility model (ARUM) is widely used in the multinomial model literature due to the advantage of having a direct connection to utility thoery, facilitating the interpretation of the estimates. We follow Greene 2007 and Cameron and Trivedi 2010 in the following analysis.

where $\mathbf{z}_{kij}$ includes the attributes of the student, the textbook and the choice made as described above. If the student chooses option j, then we assume U_{itj} is the maximum of the three utilities. The probability that individual i makes choice j for course t is given by

$$\Pr(U_{itj} > U_{itm} \mid \mathbf{z_{itj}}) \, \forall m \neq j \tag{2}$$

Combining Equation 1 with Equation 2, we obtain

$$\Pr(\epsilon_{itm} - \epsilon_{itj} \leq V_{itj} - V_{itm} \mid \mathbf{z_{itj}}) \, \forall m \neq j \tag{3}$$

Let Y_{it} denote the choice made by student i for book t. We assume that the errors are iid and follow some distribtuion $F_j(.)$. We can rewrite Equation 3 as the probability that $Y_{it} = j$ given the distribution of the errors. That is,

$$\Pr(Y_{it} = j \mid \mathbf{z_{itj}}) = F_j(\mathbf{z}_{\text{itj}}) \tag{4}$$

This model has a log likelihood function of

$$\ln L(\theta) = \sum_i \sum_j y_{ij} \ln F_j(z_{\text{itj}}, \theta) \tag{5}$$

where θ is a vector of coefficients. We can parition $\mathbf{z}_{\text{kij}}$ and its coefficient, θ into individual characteristics and choice attributes where

$$\mathbf{z}_{\text{itj}} = [\ \mathbf{x}_{\text{itj}} \quad w_{it}\ ]$$

$$\theta = [\ \beta' \quad \alpha_j'\ ]'$$

We allow the effect of the individual characteristics w_{it} to vary across alternatives, indexing α_j by the choice of alternative (j): the effect of gender on the probability of pirating a book is likely different than its effect on the probability of purchasing or not acquiring.

4.2 Multinomial Logit Model

The multinomial logit model applies when the data contains only individual-specific characteristics, as is our case. Using this model imposes the assumption that the decision to choose one alternative over another is based not on characteristics of the alternatives themselves but diferences in consumers' preferences, based on individual characteristics.[11] Since we have a panel structure, where we observe the decision of the same students across various books, included in these "individual-specific" characteristics are the

[11] This is not a totally natural assumption in our case. Thre exists a model known as the alternative specifical conditional logit model that can be used in the case of having data on both individual and choice specific attributes. While it would be convenient to use such a model, since it is likely that the choice here affects utility, we do not have sufficient information about the implicit cost of pirated books to do so.

variables pertaining to the textbooks. Therefore, folowing the structure discussed in the previous section, we will have $\mathbf{z_{itj}} = w_{it}$, as there are no choice-specific variables.

We assume that the errors are iid and follow a logistic distribution. Additionally, we assume independence of irrelevant alternatives (IIA). This condition requires that removing one of the alternatives (say the ability to pirate) should not change the relation between the remaining choices (here purchasing or not obtaining the book). In other words, it imposes the property that preferences over the choices follow the transivity property and are not inconsistent.

Following Equation 4, the probability of a student i choosing action j for the text of a particular textbook t can be written as

$$\Pr(Y_{it} = j \mid w_{it}) = \frac{\exp(w'_{it}\alpha_j)}{\sum_{j=1}^{3} \exp(w'_{it}\alpha_j)} \tag{6}$$

To ensure model identification we set $\alpha_j = 0$ for the choice of purchasing the book ($j = 3$).[12] Coefficients must then be interpreted with respect to the base category, in this case "purchased book". We therefore rewrite Equation 6 as

$$\Pr(Y_{it} = j \mid w_{it}) = \frac{\exp(w'_{it}\alpha_j)}{1 + \exp(w'_{it}\alpha_1) + \exp(w'_{it}\alpha_2)} \tag{7}$$

For a more direct interpretation of the effects of our explanatory variables on the probabilities, we can compute the marginal effects of the regressors on the probability of each choice. Note that the marginal effects may not have the same sign as the coefficients and will differ across alternatives.[13] In general, we compute

$$\frac{\partial P_{itj}}{\partial w_{it}} = P_{itj}\left(\alpha_j - \bar{\alpha}\right) \tag{8}$$

where $\bar{\alpha}$ is the coefficients averaged over the alternatives. This forumulation ensures that the marginal effects sum to zero across alternatives.

We use the following explanatory variables: GPA and the relative importance of textbooks (ranked out of three against lecture notes and course exercises where "1" indicated the most importance study tool). We will also use three dummy variables that measure technological access: access to a tablet, preferences for reading on paper, or on a screen, and familiarity with P2P networks. Finally we have three demographic variables: the year each student began their university studies, gender, and major.[14] In addition to the the characteristics of the students, we include two characteristics of the textbooks: price and quality.

[12] Only two of the probabilities can be freely specified since $\sum_{j=1}^{3} \mathrm{F}_j(\mathbf{z}'_{\mathrm{itj}}\theta) = 1$ for any distribution. Any of the three alternatives would work to interpret the results.

[13] Since the marginal effect depends on the probability which in turn depends on w_{it}, the impact of a change in one of the regressors varies along the distribution. In the tables below we present the marginal effect at the mean.

[14] For these two dummy variables, Female=1 and Business Administration=1

We estimated four models, varying w_{it} in each regression, with the dependent variable being the discrete decision made by each student for each course in which they enrolled.

Only Individual Characteristics: Models 1 & 2

In Model 1 we consider only the individual characteristics. These models implicitly assume that the choice in question depends solely on individual preferences, an assumption we relax below. Therefore we define w_{it}^1 as

$$w_{it}^1 = [Year_i, Gender_i, Major_i, GPA_i, TextbookRank_i, Tablet_i, Reading_i, P2P_i]$$

In Model 2 we re-estimate without the explanatory variable $TextbookRank_i$, in particular to see the effect on the way $Gender_i$ and GPA_i impact the probabilities.

$$w_{it}^2 = [Year_i, Gender_i, Major_i, GPA_i, Tablet_i, Reading_i, P2P_i]$$

Incorporating Course and Textbook Characteristics: Models 3 & 4

In the third and fourth models we include variables related to the textbooks for each course.

First, in Model 3, we use data collected on the price of each book. The results in the following section are fairly robust to various functional forms, with the results in both levels and logs generally equivalent: we append these in Appendix B. We also ran several regressions using quality, however these regressions had the unexpected result of quality varying in the same direction as price. As we mentioned previously, measures of textbook prices are very noisy, and it could be that quality is here a proxy for the secondhand market price. In Appendix D we show the results of Model 3 regressions with Quality.

$$w_{it}^3 = [Year_i, Gender_i, Major_i, GPA_i, TextbookRank_i, Tablet_i, Reading_i, P2P_i; \ln Price_t]$$

In the fourth model instead of price and quality, we used dummy variables for the courses as well as the variables contained in w_{it}^1.

4.3 Estimation Results

Below we present the marginal effects of the explanatory variables on the probability of each choice for the four multinomial logit regressions we estimated.[15] These estimates allow us to guage the impact of each explanatory variable on the probability of choosing each option. Looking at each alternative in turn, we see slight variations in the effect of the explanatory variables.

[15] In Appendix B we present additional results for each model, including global measures of fit, the estimated coefficients, and results of a Wald test to verify statistical significance across alternatives.

Pr(*Pirated*)	Model 1	Model 2	Model 3	Model 4
Year	0.057	0.068*	0.082*	0.092*
Gender	0.175***	0.171***	0.180***	0.194***
Major	0.006	0.006	0.025	0.059
GPA	-0.020**	-0.015	-0.024**	-0.025**
Ranked importance of textbooks	-0.072**	-	-0.085**	-0.092***
Access to tablet	-0.023	-0.012	-0.019	-0.023
Reading preferences (paper/screen)	0.049	0.024	0.060	0.064
Familiarity with P2P networks	0.076*	0.045	0.079*	0.085*
ln *Price*	-	-	0.547***	-
Introduction to Economics	-	-	-	-0.463***
Macroeconomics I	-	-	-	-0.022
Marketing	-	-	-	-0.238***
Econometrics	-	-	-	-0.364***
Microeconomics I	-	-	-	-0.162**
Management Control Systems	-	-	-	-0.372***

Table 4: Marginal Effects on Pr(*Pirated*)

In Table 4 we see the marginal effects of the regressors on the probability of pirating the book.[16] Most models indicate that students beginning their studies later will be more inclined to pirate, which may be related to the penetration of technology we observed in the previous chapter. Female students are more likely to pirate, with the increase in probability ranging from 0.17 to 0.27. This finding is consistent with the observation in the previous chapter that female students pirated more than male students by 16.9 percentage points.

Students with a higher GPA were less likely to pirate by -0.021 on average across the models. Additionally, students who ranked textbooks as more important were less likely to pirate; as these students likely value textbooks more highly than their colleagues, they probably have a higher willingness to pay.

Familiarity with P2P networks is here statistically significant and positively correlated with the probability of pirating, increasing the probability 0.076 to 0.11, depending on the model. This variable is not significant in explaning the other alternatives: knowledge of the technology used for pirating only impacts the probability of choosing to pirate, implying that knowing how to pirate doesn't ex ante make someone more or less likely to purchase or obtain the book.

In Model 3 price is strongly statistically significant, with an 1% increase in price per page leading to a 0.547% increase in the probability of pirating. In Model 4, the course dummies variables are all

[16] * p<0.1 ** p<0.05 *** p<0.01 for all tables. We present the results according to the standard error used in each particular table, here estimated by Stata using the delta method.

significant except for Macro I. All the textbooks were pirated less than Finance II, the dropped dummy, with the magnitudes of the marginal effects capturing the order of the observed proportion pirated for each course. These dummies may be capturing differences between the course textbooks not entirely captured by price and quality, such as the strength of the teacher's recommendation and the extent to which the book is used in class or for assignments.

Pr(*Purchased*)	Model 1	Model 2	Model 3	Model 4
Year	-0.013	-0.017	-0.046	-0.022
Gender	-0.033	-0.024	-0.049	-0.062
Major	0.006	0.010	-0.022	-0.054
GPA	0.021	0.018	0.032*	0.029
Ranked importance of textbooks	0.143***	-	0.179***	0.216***
Access to tablet	0.015	0.013	0.009	0.032
Reading preferences (paper/screen)	-0.013	0.011	-0.040	-0.005
Familiarity with P2P networks	-0.009	0.030	-0.016	-0.016
$\ln Price$	-	-	-1.146***	-
Introduction to Economics	-	-	-	0.622***
Macroeconomics I	-	-	-	0.053
Marketing	-	-	-	0.150*
Econometrics	-	-	-	-0.224***
Microeconomics I	-	-	-	-0.064
Management Control Systems	-	-	-	0.177**

Table 5: Marginal Effects on Pr(*Purchased*)

Looking next at the impact on the probability of purchasing, we see that the importance of textbooks is the most consistently statistical significant determinant of the probability of purchasing. This implies that the usefulness of textbooks as a study tool, relative to other (free) alternatives, has the most bearing on whether a student is willing to pay for a textbook. The signs of the marginal effects are reversed relative to the previous table. This reversal is likely to reflect the fact that purchasing and pirating are opposite choice, given the decision to acquire the book.

Unsurprisingly we see a larger impact of price in Model 3 than before on the probability of pirating, since here the price is the actual cost of the choice analyzed. As expected, more expensive books are purchased far less, with a unit increase in price per page decreasing the probability of purchasing a book by 1.15%.

Again, the marginal effects of the course dummies, less significant than for the previous alternative, follow the observed ranking of purchased books. Introduction to Economics has a substantially higher probability of purchase than the remaining books (0.622 more likely than Finance II), and the probability

of purchasing an Econometrics textbook falls 0.224 percentage points compared to Finance II. Unlike the previous alternative, the probability of purchasing seems largely unaffected by characteristics of the individual (such as gender, etc.), with aspects of the textbooks themselves and study habits having a major impact on whether students purchase the textbooks.

Pr(*Didn't Obtain*)	Model 1	Model 2	Model 3	Model 4
Year	-0.044**	-0.051**	-0.035	-0.070**
Gender	-0.142***	-0.147***	-0.131**	-0.132*
Major	-0.012	-0.016	-0.003	-0.005
GPA	-0.002	-0.003	-0.009	-0.005
Ranked importance of textbooks	-0.071*	-	-0.093**	-0.124**
Access to tablet	0.008	-0.001	0.00	-0.008
Reading preferences (paper/screen)	-0.036	-0.035	-0.020	-0.059
Familiarity with P2P networks	-0.067	-0.074	-0.063	-0.070
ln *Price*	-	-	0.599***	-
Introduction to Economics	-	-	-	-0.158***
Macroeconomics I	-	-	-	-0.031
Marketing	-	-	-	0.088
Econometrics	-	-	-	0.588***
Microeconomics I	-	-	-	0.226***
Management Control Systems	-	-	-	0.195**

Table 6: Marginal Effects on Pr(*Didn't Obtain*)

Finally, looking at the marginal effects on the probability of obtaining the book, we see that the effects are similar to those above. Students who began college later are more likely to obtain their textbooks (a difference of 0.044 to 0.086 across the models) and, as we saw in the descriptive statistics in the previous chapter, female students are subtantially more likely to acquire the textbook of the course than their male colleagues. Additionally, students with higher GPA and those who value textbooks more highly are less likely to abstain from acquiring the textbook. A higher price increases the probability of not obtaining the book by 0.599%.

Overall we see that many of the factors that predict whether a student will pirate are innate to the student, such as their gender, the year they begin college. Students with higher GPA are more likely not only to acquire the book, but to purchase rather than pirate it, as are students who ranked textbooks more highly compared to lecture notes and course exercises. This result suggests that students who are more "responsible" or who value textbooks as a study tool not only are willing to acquire the textbook, but do so legally. Technological access seems to have little effect in the decision, with the exception of

familiarity with P2P networks which increase the probability of pirating, but not the decision to acquire the textbook.

Finally we observe that the course dummies in Model 4 were largely significant and fall broadly in the order of the observed probabilities, which we will discuss further in the next chapter. There may be something unobservable through characteristics such as the price that impact students' decisions, such as the way professors use textbooks in class, that varies with the courses themselves rather than the textbooks.

In this model we treat the decision to purchase, pirate, or not obtain a textbook as equivalent choices. However, there are a number of ways of modeling this choice. In the next chapter we will consider the case where consumers have a propensity to pirate, which determines in part their ultimate decision of whether and how to acquire textbooks.

5 The Propensity to Pirate

In the previous chapter we investigate the effect of a variety of individual and choice-related factors on a multinomial decision of how (and whether) to acquire a textbook. We now approach this question from a slightly different perspective. In the above analysis we abstract from any possible costs that pirating may hold for the consumer. However, while downloading a book may have a marginal cost of zero (and photocopying close to zero), it is reasonable to assume that there is some fixed cost to pirating. As discussed in the literature review, this cost could be associated in part to learning to use the technology, but also can be interpreted as a moral cost that the consumer pays when they decide to acquire pirated material for the first time.

Therefore, it is interesting to re-examine the choice students make over how (or whether) they acquire textbooks incorporating this "fixed cost" into the analysis. In order to approximately estimate this effect, we used a two-stage regression, where we first attempt to measure the propensity to pirate of each student, using the individual factors contained in w_{it}^1. We then use the fitted values of this regression in a multinomial model with all three choices and the characteristics of the textbooks.

5.1 Measuring the Propensity to Pirate: Minimum χ^2 Method

To estimate this "fixed cost", we will use a linear probability function estimated using weighted least squares, or a minimum chi-squared method. We assume that the theoretical probability of pirating for each student is given by $p_i = \beta' \mathbf{w_i}$, where w_i comprises the characteristics of the students used in the previous chapter. However, we don't directly observe this p_i. Instead, because we have panel data, we observe the number of times they pirate over the courses they were enrolled in, $\hat{p_i}$.

We can write a linear probability function as

$$\hat{p_i} = \gamma_i + \beta' w_i + u_i \tag{9}$$

where the error is heteroskedastic, such that $u_i \sim \left[0, \frac{1-p_i}{n_i p_i}\right]$. In the table below we present the results for the estimation of Equation 9, using $w_i = \sqrt{\frac{n_i}{\hat{p_i}(1-\hat{p_i})}}$ as weights (Maddala 1983).

	Coefficient	t-statistic	p-value
Year	0.044***	4.27	0.000
Gender	0.171***	9.03	0.000
Major	0.003	0.13	0.893
GPA	-0.023***	-5.06	0.000
Ranked importance of textbooks	-0.072***	-5.10	0.000
Access to tablet	-0.027	-1.14	0.158
Reading preferences (paper/screen)	0.040	1.50	0.133
Familiarity with P2P networks	0.075***	3.85	0.000
Constant	-88.016	-4.26	0.000

Table 7: First Stage Regression

We note first that the coefficients are generally consistent with those estimated in model 1 of the multinomial model. Reading preferences and access to a tablet, as well as choice of major are not statistically significant. The remaining explanatory variables are all strongly significant. Students with a higher GPA tend to pirate less and those familiar with P2P networks were more likely to pirate. Students in later years also were slightly more likely to pirate than those who began their studies earlier. This may be because as the technology is further difused, students use it more (following the theory cited above that pirating has some fixed cost associated with it). Finally, as before, we see that female students are substantially more likely to pirate than male students.

The fitted values associated in this model predict that on average students will pirate 24.95% of their textbooks, with a range of -0.026[17] to 63.35%. As the regression groups individuals, each student is associated with a fitted value, their individual propensity to pirate, mapped in the histogram below. With the exception of the outlier who pirates 63.35% of their textbooks, most students are clustered between 0 and 50%. While this baseline propensity to pirate varies from individual to individual, we observe that no individual chooses to always pirate. This is an odd result given the idea of a fixed cost: in principle students should, once they pirate one textbook, choose to pirate all the books they obtain. Instead, they vary their decision, implying that the characteristics of the textbooks themselves have important bearing on the decision.

[17]Since this model was estimated using a linear probability function, the fitted values are not restricted to being strictly positive.

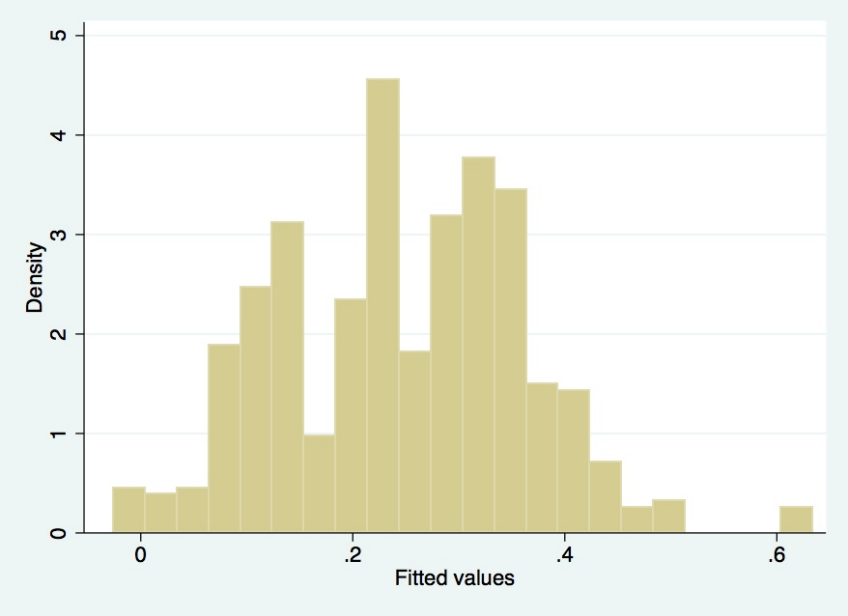

Figure 3: Histogram of Fitted Values

5.2 Applying Estimates to the Multivariate Choice Model

Using these fitted values of the general least squares regression, we used a multinomial model to estimate the probability of choosing which media to obtain (or not) a textbook. Here, instead of using individual characteristics directly, we will use the fitted values estimated above, as well as the characteristics of the textbooks. In essence, we will re-estimate models 3 and 4 from the previous chapter. Overall, the propensity to pirate is highly significant across all alternatives, as are the remaining variables concerned with textbook characteristics (with the exception of the dummy variable for Macroeconomics I in model 4, as before).

Pr(*Pirated*)	Model 3	Model 4
Fitted valued, $\hat{p}_i$	1.132***	1.215***
$\ln Price$	0.521***	-
Introduction to Economics	-	-0.443***
Macroeconomics I	-	-0.040
Marketing	-	-0.203***
Econometrics	-	-0.336***
Microeconomics I	-	-0.142*
Management Control Systems	-	-0.354***

Table 8: Marginal Effects on Pr(*Pirated*)

As to be expected, students with a higher propensity to pirate are more likely to choose to pirate, and which substantially explains the variation in the probability of pirating. The price varies positively with the probability of piracy, and while positive and statistically significant, has a smaller magnitude, increasing the probability of pirating by 1.3% for every 1% price increase. In model 4 we see that all of the course dummies have negative coefficients, which implies the probability of pirating the corresponding textbook relative to Finance II is lower; they again follow the ranking observed.

Pr($Purchased$)	Model 3	Model 4
Fitted valued, $\hat{p_i}$	-0.839***	-0.862**
$\ln Price$	-1.115***	-
Introduction to Economics	-	0.624***
Macroeconomics I	-	0.079
Marketing	-	0.135*
Econometrics	-	-0.200***
Microeconomics I	-	-0.049
Management Control Systems	-	0.162*

Table 9: Marginal Effect on Pr($Purchased$)

Observing the effect on the probability of purchasing, as expected students with a higher propensity to pirate are less likely to purchase textbooks. Following the interpretation that there is a fixed cost of pirating, these students with a lower propensity to pirate (who we could say haven't paid that fixed cost) are much more likely to purchase the book in question. A higher price decreases the probability of purchasing the textbook, controlled for individual effects, and the course dummies again capture the rankings that the book were purchased in the sample.

Pr($Didn't Obtain$)	Model 3	Model 4
Fitted valued, $\hat{p_i}$	-0.293	-0.352
Price per Page (log)	0.594***	-
Introduction to Economics	-	-0.181***
Macroeconomics I	-	-0.039
Marketing	-	0.070
Econometrics	-	0.535***
Microeconomics I	-	0.191***
Management Control Systems	-	0.192**

Table 10: Marginal Effect on Pr($Didn't Obtain$)

Here, the coefficient for the propensity to pirate is not statistically significant, while price and course dummies are highly significant as before. This strengthens the idea that students decide first whether or not to obtain the textbook for a particular course, and then decide whether to purchase or pirate the textbook in question, depending on their particular preferences.

In this second approach, we considered the possibility that pirating has an implicit fixed cost, which we sought to measure by defining a propensity to pirate for each student and using that variable to explain the end choice. What we find is that while individual traits are significant in tending students towards or against pirating, the characteristics of the book continue to be important.

6 Conclusion

This paper studies the determinants of piracy of academic textbooks used at the college level. We study the demand side of this market, analyzing which students are more likely to pirate textbooks and why. We find that, while most students have pirated at least one book, no student pirates all or even close to all the textbooks they use.

Our two main results are as follows. First, the probability of pirating is negatively related to the price of the book and positively related to the diffusion of technology that facilitates pirating. Second, the most important variable that affects the decision to pirate is the usefulness of the book and its relevance to the course. Books that are very useful and are integrated into the way the professor teaches the course, are less likely to be pirated. Books that are perceived as less useful are either pirated or not obtained at all. It is reasonable to conjecture that, when the option of pirating is not available, students simply do not obtain books that they perceive as less useful.

An example of our second result is the difference in the extent to which two books included in our sample are pirated: the Introduction to Economics and the Econometrics textbook. The former was written by the professor who teaches the course as a guide to the class that follows closely the material discussed in the lectures. In contrast, the latter is described by the instructor as being a useful reference but not a close guide to the material taught in the course. The Introduction to Economics textbook is very likely to be obtained by the students, with most students purchasing the book instead of pirating it. In contrast, fewer students obtain the Econometrics textbook and, those who do, are more likely to pirate it.

There is a parallel between our results and those obtained by empirical research on pirating in the music industry. Consumers are more likely to purchase music that they value highly and more likely to pirate music that they expect to have lower value. These results suggest that, in the absence of the pirating option, consumers might simply choose not to obtain music that they value less.

This research can help suppliers and distributers of textbooks understand how to incentivize students to acquire legally the textbooks they use in their classes. Textbook publishers cannot control the diffusion of the pirating technology. But they can control the usefulness of their books and influence the extent to which they are integrated into academic curricula. They can also control the sales price to make the option to pirate less attractive.

Understanding what motivates students to purchase or pirate textbooks is important for two reasons. First, it allows us to devise better policies to incentivize the production and distribution of the textbooks that will educate future generations of scientists and entrepreneurs. Second, the lessons from the textbook market contain valuable information about general attitudes towards intellectual property rights. Understanding these attitudes is important, since economic growth depends on innovative ideas and the creation of these ideas arguably depends on the right set of property rights.

7 References

Bakos, Yannis, Erik Brynjolfsson, and Douglas Lichtman. 1999. 'Shared Information Goods', Journal of Law and Economics, 42: 117-56.

Besen, Stanley M. 1986. 'Private copying, reproduction costs, and the supply of intellectual property', Information Economics and Policy, 2: 5-22.

Cameron, A. Colin, and Pravin K. Trivedi. Microeconometrics Using Stata, Revised Edition. Stata Press, 2010. Print.

Chellappa, Ramnath K., and Shivendu Shivendu. 2005. 'Managing Piracy: Pricing and Sampling Strategies for Digital Experience Goods in Vertically Segmented Markets', Information Systems Research, 16: 400-17.

Danaher, Brett, Samita Dhanasobhon, Michael D. Smith, and Rahul Telang. 2010. 'Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy', Marketing Science, 29: 1138-51.

Greene, William H. "Econometric Analysis." 6 ed: Prentice Hall, 2007. Print.

Liebowitz, Stan J. 1985. 'Copying and indirect appropriability: Photocopying of journals', The Journal of Political Economy: 945-57.

———. 2004. 'Will MP3 downloads annihilate the record industry? The evidence so far', Advances in the Study of Entrepreneurship, Innovation & Economic Growth, 15: 229-60.

Maddala, G.S. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, 1983. Print.

McCorkle, Denny, James Reardon, Douglas Dalenberg, Andrew Pryor, and John Wicks. 2012. 'Purchase or Pirate: A Model of Consumer Intellectual Property Theft', Journal of Marketing Theory & Practice, 20: 73-86.

NPD. 2012. "Music File Sharing Declined Significantly in 2012." In Increased use of free music streaming services takes a bite out of illegal peer-to-peer music file sharing activity. Port Washington, New York: The NPD Group.

Oberholzer-Gee, F., and K. Strumpf. 2007. 'The Effect of File Sharing on Record Sales: An Empirical Analysis', Journal of Political Economy, 115: 1-42. Peitz, Martin, and Patrick Waelbroeck. 2003.

"Making use of file sharing in music distribution." mimeo, University of Mannheim and ECARES, Free University of Brussels.

———. 2006. 'Piracy of digital products: A critical review of the theoretical literature', Information Economics and Policy, 18: 449-76.

Rob, Rafael, and Joel 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 49.1 (April) (2006): 29-62.

Sandvine (2013). "Global Internet Phenomena Report." (2): 35.

Scorcu, Antonello E., and Laura Vici. 2013. "Economic and cultural factors and illegal copying in the university textbook market." In ACEI working paper series. Bologna, Italy: Department of Economics, University of Bologna.

Sinha, Rajiv K, and Naomi Mandel. 2008. 'Preventing digital music piracy: the carrot or the stick?', Journal of Marketing, 72: 1-15.

Sinha, Rajiv K., Fernando S. Machado, and Collin Sellman. 2010. 'Don't Think Twice, It's All Right: Music Piracy and Pricing in a DRM-Free Environment', Journal of Marketing, 74: 40-54.

Takeyama, Lisa N. 1994. 'The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities', The Journal of Industrial Economics, 42: 155-66.

Varian, H. 2005. 'Copying and Copyright', The Journal of Economic Perspectives, 19: 121-38. Vernik,

Dinah A., Devavrat Purohit, and Preyas S. Desai. 2011. 'Music Downloads and the Flip Side of Digital Rights Management', Marketing Science, 30: 1011-27.

Waldfogel, Joel. 2012. 'Copyright Research in the Digital Age: Moving from Piracy to the Supply of New Products', The American Economic Review, 102: 337-42.

Wu, Shin-yi, and Pei-yu Chen. 2008. 'Versioning and Piracy Control for Digital Information Goods', Operations Research, 56: 157-72.

Appendix

A Descriptive Statistics

% that obtained from...	Books	Movies	Software	Textbooks	TV Series
Friends/Classmates	39.34	11.11	9.90	38.28	9.28
P2P Networks	12.30	45.30	29.70	16.41	37.50
Internet Search	40.16	39.32	47.52	42.97	48.21
Online Store	8.20	4.27	12.87	2.34	4.46
Didn't Report	15.52	14.66	24.14	14.66	16.38

Table 11: Sources of Downloaded Media

	% of sample	average gpa	average price
Purchased Textbook	42.94%	13.45	51 euros
Pirated Textbook	24.48%	12.94	61 euros
Didn't Obtain Textbook	32.58%	13.26	61 euros

Table 12: Overall Textbook Sources in Sample

		Average	Median	Mode
Gender		48.24% male		
Major		65.52% Business		
Year Entered Program		54.31% entered in 2011		
GPA		13.25	14	12
Importance of study materials (rank out of 3)	Course Exercises	1.32	1	1
	Lecture Notes	2.04	2	2
	Textbooks	2.61	3	3
Self-reported frequency obtain textbook		53.48%	0.5	0.25
Factors in obtaining textbook (rank out of 4)	Other Materials Available	2.11	2	1
	Peer Opinion	2.73	3	4
	Price	2.70	3	3
	Teacher Recommendation	2.24	2	1
Main reason for not getting a textbook	Isn't Useful	46.55%		
	Too Costly	21.55%		
	Used Other Materials	22.41%		
	Already Had Access	13.79%		
Do you have access to a tablet or e-Reader?		63.79% had access		
Do you prefer to read on a screen or on paper?		87.07% prefer paper		
		10.34% indifferent		
		2.59% prefer on a screen		
Are you familiar with P2P networks?		60.24% said yes		

Table 13: Descriptive Statistics

		Male	Female	Business Administration	Economics
GPA		13.39	13.29	13.04	13.90
% of Economics		60%	40%		
% of Business Administration		42.11%	57.89%		
% of Males				57.14%	42.86%
% of Females				73.33%	25.67%
Access to Tablet		67.89%	66%	57.89%	75%
Prefer to read on paper		80.3%	90%	89%	82.5%
Familiar with P2P networks		75%	46.67%	60.5%	60%
Mainly obtain downloaded goods from...	Friends	20.08%	24.68%	25.19%	16.76%
	P2P Networks	34.85%	21.84%	25.44%	32.96%
	Internet Search	43.05%	44.62%	43.64%	43.02%
	Online Store	3.03%	8.86%	5.75%	7.26%
Number of books obtained, overall		2.94	3.46	3.34	2.97
Number of books purchased, overall		2.19	1.90	2.08	1.95
% of books obtained out of enrolled courses		63.73%	76.09%	71.13%	68.59%
% of books purchased out of books obtained		70.41%	53.59%	59.98%	64.49%

Table 14: Descriptive Statistics, disaggregated (averages unless otherwise indicated)

B Additional Regression Results

Multinomial Logit Models

We present here additional information from the outputs of the multinomial models estimated in chapter 4, to supplement the marginal effects reported in the text of the paper.

The first table shows measures of fit and global statistics for all four models.

	Model 1	Model 2	Model 3	Model 4
Number of Observations	513	525	513	513
Wald Statistic (χ^2_{obs})	49.27***	35.82***	114.72***	247.27***
Pseudo R^2	0.049	0.035	0.127	0.267
AIC	1083.527	1118.534	1000.98	866.73
BIC	1159.852	1186.748	1085.786	993.938

(a) Measures of Fit

		Model 1	Model 2	Model 3	Model 4	Actual
$\Pr(\hat{PiratedBook})$	Observed	24.95%	24.58%	24.95%	24.95%	24.48%
	Difference in pp	+0.0047	-0.001	+0.0047	+0.0047	-
$\Pr(\hat{PiratedBook})$	Observed	31.77%	32.38%	31.77%	31.78%	32.53%
	Difference in pp	-0.008	-0.002	-0.0076	-0.008	-
$\Pr(\hat{PiratedBook})$	Observed	43.27%	43.20%	43.27%	43.27%	42.94%
	Difference in pp	+0.0034	+0.0026	+0.0033	+0.0034	-

(b) Fitted Values, in percent

Table 15: Global Statistics

The following tables shows the estimated coefficients across all four models. In all regressions the base alternative was "purchased book", and the estimated coefficients below must be read with respect to the base outcome. Therefore a positive coefficient means that we are more likely to choose alternative j than purchase the book. For example, since GPA has a negative coefficient, a higher GPA decreases the probability of pirating with respect to purchasing the book.

	Model 1	Model 2	Model 3	Model 4
Year	0.274	0.339	0.379	0.465
Gender	0.823***	0.807***	0.920***	1.010***
Major	0.011	0.004	0.226	0.379
GPA	-0.132**	-0.109*	-0.170**	-0.174**
Ranked importance of textbooks	-0.626***	-	-0.776***	-0.866***
Access to tablet	-0.134	-0.080	-0.103	-0.173
Reading preferences (paper/screen)	0.240	0.081	0.326	0.300
Familiarity with P2P networks	0.346	0.130	0.378	0.425
Price per Page	-	-	6.534***	-
Introduction to Economics	-	-	-	-4.043***
Macroeconomics I	-	-	-	-0.202
Marketing	-	-	-	-1.053***
Econometrics	-	-	-	-0.145
Microeconomics I	-	-	-	-0.177
Management Control Systems	-	-	-	-1.867***
Constant	-553.3059	-681.668	-747.761	-936.902

Table 16: Estimated Coefficients: Pr($Pirated$) relative to Pr($Purchased$)

	Model 1	Model 2	Model 3	Model 4
Year	-0.109	-0.117	-0.058	-0.190
Gender	-0.371	-0.397	-0.313	-0.314
Major	-0.050	-0.071	0.012	0.095
GPA	-0.053	-0.049	-0.081	-0.077
Ranked importance of textbooks	-0.542**	-	-0.645***	-0.866***
Access to tablet	-0.008	-0.031	0.005	-0.095
Reading preferences (paper/screen)	-0.085	-0.131	-0.024	-0.186
Familiarity with P2P networks	-.191	-0.295	-0.176	-0.199
Price	-	-	4.197***	-
Introduction to Economics	-	-	-	-2.817***
Macroeconomics I	-	-	-	-0.327
Marketing	-	-	-	-0.010
Econometrics	-	-	-	2.619***
Microeconomics I	-	-	-	1.001***
Management Control Systems	-	-	-	0.258
Constant	218.3265	236.505	126.716	381.580

Table 17: Estimated Coefficients: Effect on Pr($Didn't Obtain$) relative to Pr($Purchased$)

Since the statistical significance of the coefficients varies with the choice of base outcome, we performed a joint Wald test across the three alternatives to verify the statistical significance. Below we present the observed χ^2 statistics of this Wald test where $H_0 : \alpha_j = 0 \, \forall \, j$, and note the resulting statistical significance of the coefficients. In general, the results here do not vary largely from those of based on the standard errors of the coefficients and marginal effects (a result which rules out issues of multicollinearity).

	Model 1	Model 2	Model 3	Model 4
Year	5.33*	4.82*	4.63*	5.23*
Gender	17.87***	18.12***	17.58***	17.52***
Major	0.05	0.08	0.58	1.20
GPA	4.77*	3.09	5.80*	5.32*
Ranked importance of textbooks	11.08***	-	10.78***	12.39***
Access to tablet	0.30	0.09	0.15	0.25
Reading preferences (paper/screen)	0.63	0.23	0.75	1.02
Familiarity with P2P networks	4.34	2.97	4.22	4.31
Price per Page (log)	-	-	62.31***	-
Introduction to Economics	-	-	-	57.18***
Macroeconomics I	-	-	-	0.44
Marketing	-	-	-	8.46**
Econometrics	-	-	-	44.19***
Microeconomics I	-	-	-	15.48***
Management Control Systems	-	-	-	15.84***

Table 18: Observed χ^2 statistics for Wald test

Model 3: Alternative Functional Forms for Price and Quality

Besides the elasticity of price per page shown in the main text, we also estimated the effect of price along and price per page both at levels (linearly) and with logs. As a large portion of the pirating we observe is through photocopying, we consider also the price per page (in both logs and level). We present now the results of these additional regressions, beginning with the measures of fit for the global models.

	Linear Price	Linear $\frac{Price}{Page}$	$\ln Price$	log $\frac{Price}{Page}$
Number of Observations	513	513	513	513
Wald Statistic (χ^2_{obs})	144.55***	104.08***	159.5***	114.72***
Pseudo R^2	0.137	0.117	0.1516	0.127
AIC	989.999	1012.654	974.132	1000.98
BIC	1074.805	1097.46	1058.937	1085.786

Table 19: Measures of Fit Global Statistics

Below we present the marginal effects from the three versions of Model 3. The estimations are statistically equivalent, as in most cases the estimated marginal effect is within the standard error of each model. Therefore these results are robust to the three functional forms estimated.

	Linear Price	Linear $\frac{Price}{Page}$	$\ln Price$	$\ln \frac{Price}{Page}$
Year	0.084*	0.067*	0.082*	0.068*
Gender	0.181***	0.180***	0.180***	0.178***
Major	0.027	0.044	0.025	0.037
GPA	-0.023**	-0.022**	-0.024**	-0.023**
Ranked importance of textbooks	-0.084*	-0.083**	-0.085**	-0.083***
Access to tablet	-0.018	-0.018	-0.019	-0.018
Reading preferences (paper/screen)	0.061	0.057	0.060	0.057
Familiarity with P2P networks	0.079*	0.077*	0.079*	0.077*
$Price$	0.010***	-	-	-
$\frac{Price}{Page}$	-	10.727***	-	-
$\ln Price$	-	-	0.547***	-
$\ln \frac{Price}{Page}$	-	-	-	0.793***

Table 20: Marginal Effects on Pr($Pirated$)

	Linear Price	Linear $\frac{Price}{Page}$	ln $Price$	ln $\frac{Price}{Page}$
Year	-0.047	-0.025	-0.046	-0.027
Gender	-0.047	-0.040	-0.049	-0.040
Major	-0.025	-0.031	-0.022	-0.023
GPA	0.031*	0.027*	0.032*	0.028*
Ranked importance of textbooks	0.174***	0.168***	0.179***	0.171***
Access to tablet	0.008	0.009	0.009	0.009
Reading preferences (paper/screen)	-0.039	-0.026	-0.040	-0.027
Familiarity with P2P networks	-0.015	-0.009	-0.016	-0.009
$Price$	-0.020	-	-	-
ln $\frac{Price}{Page}$	-	-16.411***	-	-
ln $Price$	-	-	-1.146***	-
ln $\frac{Price}{Page}$	-	-	-	-1.258***

Table 21: Marginal Effects on Pr($Purchased$)

	Linear Price	Linear $\frac{Price}{Page}$	ln $Price$	ln $\frac{Price}{Page}$
Year	-0.037*	-0.042**	-0.035	-0.041*
Gender	-0.134**	-0.140**	-0.131**	-0.138**
Major	-0.002	-0.014	-0.003	-0.014
GPA	-0.008	-0.005	-0.009	-0.006
Ranked importance of textbooks	-0.090*	-0.085*	-0.093**	-0.088*
Access to tablet	0.010	0.009	0.00	0.009
Reading preferences (paper/screen)	-0.021	-0.031	-0.020	-0.030
Familiarity with P2P networks	-0.064	-0.068	-0.063	-0.067
$Price$	0.010***	-	-	-
$\frac{Price}{Page}$	-	5.639***	-	-
ln $Price$	-	-	0.599***	-
ln $\frac{Price}{Page}$	-	-	-	0.465***

Table 22: Marginal Effects on Pr($Didn't\,Obtain$)

Additionally we present the observed χ^2 statistics testing individual significance across the alternatives, where again the three forms of the model agree substantially.

	Linear Price	Linear $\frac{Price}{Page}$	$\ln Price$	$\ln \frac{Price}{Page}$
Year	4.00	4.78*	3.92	4.63*
Gender	17.61***	17.60***	17.65***	17.58***
Major	0.32	0.81	0.26	0.58
GPA	5.98**	5.67**	6.03**	5.80*
Ranked importance of textbooks	10.31***	10.76***	10.52***	10.78***
Access to tablet	0.16	0.15	0.17	0.15
Reading preferences (paper/screen)	0.88	0.74	0.86	0.75
Familiarity with P2P networks	4.21	4.23	4.22	4.22
$Price$	79.85***	-	-	-
$\frac{Price}{Page}$	-	52.56***	-	-
$\ln Price$	-	-	93.80***	-
$\ln \frac{Price}{Page}$	-	-	-	62.31***

Table 23: Observed χ^2 statistics for Wald test

Minimized Chi-Squared Model

We present the measures of fit for the first and second stage regressions in this model. Since we used a multinomial model for the second stage regression, as before we show the estimated coefficients. Finally, we present the observed χ^2 statistics testing individual significance across the alternatives.

	$\ln Price$	Model 4
Number of Observations	513	513
Wald Statistic (χ^2_{obs})	129.73***	191.84***
Pseudo R^2	0.135	0.243
AIC	964.074	865.275
BIC	989.515	933.120
Fitted valued, $\hat{p}_i$	33.68***	34.17***
$\ln Price$	50.97***	-
Introduction to Economics	-	61.83***
Macroeconomics I	-	0.89
Marketing	-	6.55**
Econometrics	-	40.75***
Microeconomics I	-	11.49***
Management Control Systems	-	15.55***

Table 24: Global Statistics and Observed χ^2 statistics for Wald test

	$\ln Price$	Model 4
Fitted valued, $\hat{p_i}$	6.936***	7.432***
$\ln Price$	83.779***	-
Introduction to Economics	-	-3.984***
Macroeconomics I	-	-0.314
Marketing	-	-0.928**
Econometrics	-	-0.231
Microeconomics I	-	-0.187
Management Control Systems	-	-1.825***
Constant	-21.787	-1.446

(a) Pr($Pirated$) relative to Pr($Purchased$)

	$\ln Price$	Model 4
Fitted valued, $\hat{p_i}$	1.010	0.650
$\ln Price$	52.119***	-
Introduction to Economics	-	-2.851***
Macroeconomics I	-	-0.420
Marketing	-	-0.085
Econometrics	-	2.262***
Microeconomics I	-	0.799**
Management Control Systems	-	0.209
Constant	-17.783	-0.514

(b) Pr($Didn't Obtain$) relative to Pr($Purchased$)

Table 25: Estimated Coefficients

C Survey Administered

Survey on Study Methods

This is a survey for a thesis investigating the way students study and which materials they choose to use. Please answer each question as accurately as possible. Thank you for participatin!

1. What is your gender? ____ M ____ F

2. What are you studying? ____ Business Administration ____ Economics

3. What year did you enter your program? _____

4. What is your grade average so far?

____10-11 ____12-13 ____14-15 ____16-17 ____18+

5. Please rank the importance of the following study materials, assigning 1 to the most important, 2 to the second most important, and 3 to the least important.

____ Course exercises
____ Lecture notes
____ Textbooks

6. If there is a recommended textbook for a class, how frequently to you acquire it?

____0-25% ____26-50% ____51-75% ____76%-100%

7. Please rank the following factors according to their importance in your decision to acquire the textbook for a particular class:

____ Availability of other study materials ____ Price
____ Opinion of peers ____ Teacher recommendation

8. If you choose not to purchase a textbook for a class, what is the main driver of that decision?

____ Isn't useful ____ Too costly
____ Used other materials ____ Already had access

9. Do you have access to a tablet or e-Reader? ____Y ____N

10. Do you prefer to read on paper or on a screen? ____ Paper ____ Screen ____ Indifferent

11. Are you familiar with P2P networks, such as BitTorrent? ____Y ____N

12. If you have ever downloaded the following media, please indicate your source.

	Friends/Classmates	P2P Network	Internet Search	Online store, such as iTunes
Books				
Movies				
Software				
Textbooks				
TV Series				

13. Below there is a list of courses offered at Católica-Lisbon. For each course, to the best of your knowledge, please indicate if you obtained the textbook recommended for that class, and if so, from where. As a reminder, below this table there is a list of the books recommended for each of these courses.

Courses	Have not taken the class	Did not get book	Bought the printed copy	Photocopied book*	Bought eBook version	Downloaded PDF
Introdução à Economia I/II						
Microeconomics I						
Econometrics						
Introduction to Marketing						
Macroeconomics I						
Finance II						
Management Control Systems						

**If you acquired photocopies of the book, please indicate the percentage of the book photocopied.*

Introdução à Economia I/II César das Neves: Introdução à Economia

Microeconomics I Baye and Price: Managerial Economics and Business Strategy

Econometrics Wooldridge: Introductory Econometrics

Introduction to Marketing Kotler, Armstrong, Harris, and Piercy: Principles of Marketing

Macroeconomics I Williamson: Macroeconomics

Finance II Brealey, Myers, and Allen: Principles of Corporate Finance

Management Control Systems Horngren, Datar, Rajan: Cost Accounting - A Managerial Emphasis

D Regressions Including Quality

As we discussed in the text, our measure for quality is the average number of stars over the last two editions of each textbook on Amazon.co.uk. Due to issues with finding a robust measure for price, quality behaves oddly in these regressions, often in the same direction as price. This leads us to conclude that the quality of the textbooks as measured is a proxy for the price on the second-hand market. Below we present the marginal effects of these regressions in the second stage regression. Although quality itself is a highly significant variable, it seems to distort the results for price and the regressions globally have a lower pseudo-R^2 than those presented in the main text.

	Linear	$\frac{Price}{Quality}$	Logarithmic
Number of Observations	513	407	407
Wald Statistic (χ^2_{obs})	53.20***	43.66***	51.61***
Pseudo R^2	0.0747	0.0596	0.072
AIC	838.0726	847.474	840.113
BIC	870.1431	871.527	872.183

Table 27: Global Statistics

	Linear	$\frac{Price}{Quality}$	Logarithmic
Fitted valued, $\hat{p_i}$	1.405***	1.364***	1.404***
$Price$	0.005	-	-
$Quality$	0.087	-	-
$\frac{Price}{Quality}$	-	-0.004	-
$\ln Price$	-	-	0.333
$\ln Quality$	-	-	0.427

(a) On Pr($Pirating$)

	Linear	$\frac{Price}{Quality}$	Logarithmic
Fitted valued, $\hat{p_i}$	-0.850***	-0.784***	-0.553**
$Price$	-0.010**	-	-
$Quality$	-0.373***	-	-
$\frac{Price}{Quality}$	-	0.021***	-
$\ln Price$	-	-	0.224
$\ln Quality$	-	-	0.960**

(b) On Pr($Purchased$)

	Linear	$\frac{Price}{Quality}$	Logarithmic
Fitted valued, $\hat{p_i}$	-0.554**	-0.580**	-0.851***
$Price$	0.005	-	-
$Quality$	0.286***	-	-
$\frac{Price}{Quality}$	-	-0.017**	-
$\ln Price$	-	-	-0.557**
$\ln Quality$	-	-	-1.387***

(c) On Pr($Didn't Obtain$)

Table 26: Marginal Effects in Models with Quality