

Pitfalls in Measuring the Impact of File -sharing on the Sound Recording Market

Stan J. Liebowitz*

***Abstract:** This paper surveys the extant literature on the impact of file-sharing on the recording industry. It begins by examining the theory behind the impact of file-sharing. One novelty from this analysis is the finding that the effect of 'sampling' of copyrighted materials can be expected to have a negative impact on copyright owners, quite the opposite of the impact normally assumed. Overall, although one can create hypothetical situations where network effects might allow file-sharing to be beneficial to copyright owners, the expectation that file-sharing would harm copyright owners appears a far more realistic outcome. The analysis then turns to the empirical work that has been performed up to the time this paper was written. The various approaches are critiqued and compared. I posit some explanations for the variations in results and conclude that the evidence strongly suggests that file-sharing harms the sound recording industry. (JEL K0, L0, L5, L8)*

The sales of music CDs have fallen dramatically in the last few years. A leading candidate among possible causes of this decline is the recent rise in file-sharing. The recording industry has tried to stem this decline by suing or threatening to sue individuals heavily engaged in file sharing. These law suits have attracted a good deal of publicity and discussion. A rather influential school of thought has formed in opposition to the corporate players in these industries, and these lawsuits have provided fodder for their critiques of traditional copyright.¹

The motion picture industry has expressed concerns that its sales are likely to suffer a similar fate to that of the sound recording industry if nothing is done to stem the unrestricted use of file-sharing software. Owners of advertising-based

* School of Management, University of Texas at Dallas, Richardson, TX 75083, e-mail: liebowit@utdallas.edu.

¹ These copyright critics, who are sometimes associated with the concept of the 'creative commons,' argue that copyright laws are being used by the sound recording, movie, and software industries so as to thwart competitive forces that would open up the market to new competition. This is the thesis of Laurence Lessig's recent book *Free Culture* which views the current controversies as extensions of long-running debates regarding the power of cartels to monopolize access to creative works. In this view of the world, file-sharing is a wealth enhancing innovation, likely to democratize the entertainment industry by allowing artists to broadcast and distribute their works without intermediaries such as record companies. In this view, file-sharing systems should be promoted and if necessary, copyright law should be altered to allow file-sharing to proceed apace.

mediums, such as television, are also concerned about their future viability if the advertising is stripped out of the programming and made available through file sharing networks, as is sometimes contemplated.

The very high profile of the music and movie industries has given the file-sharing issue a great deal of visibility. Although the revenues from these industries are only a small portion of GDP, these industries provide products that consume a large portion of our leisure time and tend to be recognized throughout the world.

At the center of the file-sharing debate is the empirical issue of whether or not file-sharing decreases sales.

Empirical work should always be informed by theory. In this paper I first explore some of the theoretical issues involved with file-sharing. I examine several theoretical factors that have been proposed to explain the behavior of file-sharers – substitution, sampling, network effects and indirect appropriability. One of the novelties of this analysis is the conclusion that sampling, the behavior most relied upon by those suggesting that file-sharing might have a benign impact on the industry, would actually have a negative impact. I conclude that of these factors, only network effects might have a positive impact on sales but suggest that this is unlikely to be an important factor. Thus economic theory would lead to an expectation that file-sharing should lead to a decline in sales relative to what they would have been in the absence of file-sharing.

I then turn to different methodologies that have been chosen in empirical examinations of this issue and I review these studies in some detail. I conclude that these studies tend to confirm what the theory predicts. Although some studies reach different conclusions, those studies appear to suffer from weak methodological choices or are based on questionable empirical specifications.

1 Background

Sound recordings, movies, and television occupy the large bulk of our time spent on leisure activities, with the average American watching four and a half hours of television and listening to more than three hours of music each day.² The advertising industry is primarily based around these industries, as is the consumer electronics industry, which would include all forms of stereo equipment, televisions, DVD players, VCRs, and so forth. Thus even beyond its

² See Table No.909 “Media Usage and Consumer Spending: 1993 to 2003” in the 2000 US Statistical Abstract.

share of GDP, its share of the collective consciousness and behavior of consumers is very large.³

The products of these entertainment industries have proven amenable to digitization, allowing them to be transferred over peer-to-peer file-sharing networks. Napster was the first well-known peer-to-peer file-sharing system, but others have followed in the wake of the preliminary injunction that effectively shut Napster down. Current replacements have surpassed Napster in popularity. Some Napster replacements are amazingly popular. As an example, Kazaa has been downloaded 350,000,000 times, Morpheus has been downloaded 125,000,000 times and iMesh has been downloaded 75,000,000 times from a single website.⁴ It has been estimated that file-sharing represents over one third of all material transferred over the Internet, and that music files are downloaded to the tune of .8 billion files per month although recent work (Liebowitz 2005) questions the reliability of some of these figures .

In what follows I will focus on the sound recording industry because that is that market that has attracted the most attention. Since most computer users have enough bandwidth to download MP3 files and also have in place the requisite CD burners with which to allow the listening of this music in locations not tied to a computer, this is the arena where most file sharing is taking place. When bandwidth allows the downloading of high quality video files to take place in a matter of minutes, instead of the current length of hours, and when DVD burners become as ubiquitous as are CD writers, then movies, and to a lesser extent television, may become equally subject to file-sharing.

2 Economic theory of file-sharing's impact

One topic that has received too little attention in the recent literature is the theory underlying predicted impacts of file-sharing on the marketplace.⁵

One set of theories largely mirror the claims made in the Napster case by the opposing sides but there are additional theories as well. On the one hand, a downloaded file can substitute for the purchase of an original CD or single song. Substitution of a free alternative is easily understood to have a negative impact on sales. On the other hand, the claim has been made that users might

³ It is also possible that its value to consumers might be much larger than that indicated by market revenues. Although Paul Romer (2002) suggests that the main interest in these topics is that they create "teachable moments" I would suggest that he places too much confidence in market revenues being related to total surplus.

⁴ On the download.com website, as of mid May 2004.

⁵ For discussions of the economics of copying and copyright see Peitz and Waelbroek (2004a), Watt (2004) or Varian (forthcoming).

merely use downloaded songs to become more familiar with potential music. Although this was originally referred to as the *exposure* effect (see Liebowitz 1981; 1985), it is currently called the *sampling* effect. Under this scenario users sample from available music and then purchase those songs and albums that are found to be most suitable to matching the tastes of the users. This sampling hypothesis is usually associated with a claim that sales will increase if consumers are allowed to become more familiar with the product before they purchase it, although, there has not been much analysis of this claim.

There is also a claim of potential network effects. As more downloaders listen to music, this theory goes, consumers derive greater value from their legitimate purchases. It is suggested that this might lead to an increase in the sales of CDs.

Finally, there is a possibility that sellers of original files can capture the value from later copiers indirectly in the price of originals, a concept known as indirect appropriability.

I examine each of these claims in turn.

2.1 Sampling

Although it is often asserted that sampling would obviously be expected to have a positive impact on sales, the impacts of sampling are far more subtle. Indeed, a more complete analysis tends to lead one to expect that sampling would lead to a decrease in sales in this market.

The sampling story basically argues that file sharing allows consumers to experience music in a more complete manor prior to purchase than they would have been able to do were they to use the more traditional methods of learning about music: hearing it on the radio or at a friend's house. With file-sharing, listeners can become as familiar with the music as they wish, listening to it over and over again until they are sure they like it.

At that point, according to the sampling theory, the listeners go out and purchase the music. A natural question is to ask why they would make a purchase when they already have the item for free. There are several possible answers. First, they might be uncomfortable listening to music which they have not purchased. This discomfort might arise from a sense of honesty or a sense of wishing to support their favorite musicians. Alternatively, listeners might get to know three or four songs on an album which then allows them to feel comfortable buying the entire album and avoid the efforts involved with downloading another 6 songs.

Assuming that sampling occurs in the manner described above, what would be the likely economic impacts of sampling?

Assume that those engaged in sampling have no intention of listening to MP3 files after the sample period. Instead, they either purchase the music or throw it away. This is a pure analysis of sampling independent of any pirating motive.

I am not aware that any proponents of the sampling story have attempted to fill in the details of the economic impacts of sampling.⁶ Presumably, after sampling, consumers have more information about which CDs to purchase, allowing them to purchase CDs that provide greater utility than they would have purchased without sampling. Although it is natural to think that consumers would be led to purchase more CDs if CDs can provide greater utility than they did without sampling this is not necessarily the case.

To see this it helps to think of a CD as a candy bar, following a line of reasoning developed by Jack Hirshleifer (1971).⁷ Each individual consumer has particular tastes in music and some CDs are better than others as far as satisfying these tastes. Consumers, after all, do not derive utility from the CDs *per se* but derive enjoyment from listening to the music contained within the CD. The underlying demand can be thought of as the demand for music-listening services, which is met to differing degrees of success by various CDs. Those CDs that better satisfy the consumer can be thought of as providing more music-listening services within the fifty or so minutes of music contained within the CD. Since those CDs contain more of what the consumer wants, they can be analogized to providing consumers larger candy bars containing more of the candy which the consumer ultimately desires.

An equivalent analysis is to think of sampling as reducing the risk associated with the purchase of a CD. If consumers are not sure whether they will like the music on a CD they will consider it a somewhat risky purchase. Sampling has

⁶ The closest might be a paper by Gopal, Bhattacharjee, and Sanders (forthcoming). They attempt to analyze the theoretical impacts of file-sharing using a fairly typical model. Sampling plays an important role in their model but they do not analyze the impacts of sampling by itself. If the full cost of sampling were zero, consumers in their model would sample all music, in order to find the most highly valued music. Whether they would then purchase the preferred music depends on other costs, such as costs of punishment if caught pirating, the sound quality differentials between sampled and purchased music, and the revealed value of music. If sound quality was identical between original and copies and if there were no punishment if copying, consumers would completely pirate and purchase no music.

⁷ By working with the underlying characteristic of the good we can avoid the heterogeneity problem brought about by the fact that CDs are not perfectly homogenous. This is done by assuming that CDs are differentiated by containing different quantities of the underlying music-listening service characteristic. This keeps the analysis simple. It also allows the analysis to treat the indirect valuation of copiers as altering the quantity of music services created by the original as an analog to increasing durability of a product. I provide this type of analysis in Liebowitz (1982a).

the possibility of reducing or removing this risk. What this means is that sampling allows consumers to achieve greater expected utility with a CD purchase than they would have achieved without sampling. At a given price of CDs, the cost of the expected satisfaction gained by consumers falls. Again, it is equivalent to putting more expected yummy music services in each CD.

It is natural to think that if candy bars remain constant in price while increasing in size, that the quantity of candy bars sold will increase. After all, each large candy bar provides more utility than a small candy bar. This is apparently the thinking of those claiming that sampling MP3s increases the sales of CDs.

The story, however, is not so simple.

First note that the price of candy is effectively lowered when the bar becomes bigger holding the price of the candy bar constant. If the demand *for candy* is elastic, then revenue in the market will increase when candy goes down in price, as it does when constant-price bars become larger. If revenues increase in this way, and the price of bar is unchanged, then more candy-bars are purchased. The inverse of this story will hold when the demand for candy is inelastic. Making candy bars larger rotates the demand curve so that the price is higher at small quantities but lower at large quantities—in other words, the demand curve becomes steeper. Satiation occurs at a smaller number of bars since each bar is bigger.

A simple analogy can be made to the introduction of cable television. Cable allowed television viewers who were previously accustomed to having a choice from only a handful of broadcast signals to instead have a choice from dozens of channels. This should have increased the probability that viewers would find, in any half hour period, a program more to their liking than they were likely to have found with only the limited original choices. The analogy to the sampling hypothesis is very strong since allowing viewers to find a better program in a thirty minute television time slot is similar to allowing listeners the ability to find better selections in their choices of CD, as sampling is claimed to do.

Previous research examining cable's impact of increased viewer choice on total time viewing television tended to find that that providing more choice to consumers did *not* increase the time they spent viewing television.⁸ Thus the claim

⁸ See Liebowitz (1982) who compares the link between viewing hours and cable penetration across different Canadian metropolitan areas and finds an insignificant but sometimes negative relationship. Also see Weimann (1996) who examines viewers in Israel after the introduction of a multi-channel cable system where previously there had been but a single public channel (a more extreme increase in choice than would normally be found). After a year, there was virtually no difference in changes in viewing between a group with cable and a control group that did not receive cable (the cable group increased its viewing by 16 minutes over the control

that providing consumers the ability to fine-tune their product selections need not increase their consumption is seen to have real-world explanatory power.

Understood in this way, it becomes clear that better information about the product might lead to either an increase or a decrease in the number of units sold. Equivalently, it might lead to either an increase or a decrease in revenues. There is, however, some additional information in this market that helps to resolve this imprecision.

CDs are thought to have low variable costs of production and high fixed costs. It is common in theoretical models of markets like software or music to assume a zero marginal cost of production. Although this is merely a theoretical convenience, since the variable costs are clearly not zero, variable costs do appear to be quite low in the case of sound recordings.

What are some of these variable costs? The cost of a blank CD is only a few cents and the pressing of CDs appears to be less than a dollar. Although the artists normally receive a royalty that is expressed as a function of sales, those payments are usually paid up-front as a non-refundable advance against future royalties, so for most units sold, marginal royalties paid by the producers are effectively zero. Promotional costs for CDs are also usually also taken out of up-front advances, removing another potential variable cost from the variable cost column. There is a variable payment made to the composers of songs that are included in the CD, however, with a statutory maximum payment of approximately seventy cents per CD.⁹

It seems reasonable to conclude, therefore, that variable costs are quite low relative to the wholesale price of CDs, which is in the vicinity of twelve dollars.¹⁰

This datum of variable costs being a relatively unimportant component of costs provides some important additional information about elasticity of demand facing each CD. Profit maximization, when marginal costs are zero, is equivalent to revenue maximization. The elasticity of demand is one at the profit maximizing price.

group). There are several papers looking at the impact of cable on different European countries and reaching largely the same conclusions that can be found in Becker and Schönbach (1989).

⁹ This is a compulsory license which amounts to ninety cents for each CD. In the common case where the performer is the composer, however, it is typical for there to exist a "controlled composition" clause which pays less than (75% appears to be a typical rate) the statutory payment.

¹⁰ If the composer is not the performer of the song, and if the song has not been recorded, this payment will be negotiated. After a song has been recorded by one performer, however, anyone can record the song if they purchase a 'compulsory license' which is the statutory payment described in the text. According to statistics on the RIAA web page, the average price in 2003 was \$11.91.

The elasticity of demand for CDs maps directly into the elasticity of demand for music-listening services. If the price elasticity of demand for CDs equals one, so too must the price elasticity of demand for music-listening services. After all, if the marginal revenue of another unit of music service is negative, so too must be the marginal revenue of the CD containing that unit.

Because there is competition between record titles, we should expect that the elasticity of demand for music-listening services (whether defined by musical genre or the entire industry) will be less than the elasticity of demand individual firms or individual record titles, for the same reasons that price elasticities of demand for industries are lower elasticities than their constituent firms. This implies that the elasticity for the industry will be less than one.

It matters little, however, whether the price elasticity is less than one or equal to one. In either case revenues (and profits) fall when the price is lowered.¹¹ The effect of sampling (more music-listening services at a constant CD price) is to lower the price of music-listening services. The net effect should be to lower the revenues generated by music-listening services. With a price per CD that is independent of the sampling effect, this implies that the quantity of CDs will fall due to sampling. This analysis assumes, as does the candy bar problem, that the price of CDs is exogenous to the change in music-listening services created by CDs.

This is quite the opposite of the claim that is most frequently made for sampling.

2.2 Substitution effect

The substitution effect is quite simple to analyze. The copy is treated as a substitute for the original. If the copy is identical or close in quality to the original, and if the cost of making the copy is low, the copy for a price of zero dominates the original at its positive price.

This is most appropriately modeled as a decrease in demand. The substitution effect can only work to reduce the price and quantity sold in the market.

2.3 Network effects

Some products have network effects. These occur for a product when consumers' values of the product change depending on the number of other users there

¹¹ Boldrin and Levine (2003) assume that elasticity is greater than one in order to generate their results, discussed below.

are of the product. Telephones and fax machine are two examples of products where the value of those products depends on the number of individuals using those products.

It has sometimes been claimed that network effects might be important to understand the impact of copying. Conner and Rumelt (1991), Takeyama (1994) and Shy and Thisse (1999) each examine models where the existence of unauthorized users creates additional value to the purchasers of legitimate copies and thus might increase the profits of the seller. These models are usually put forward in the context of software, although they might also be applied to file-sharing.¹²

There are several issues to be addressed in the context of possible network effects for sound recordings. The first is whether there are network effects at work. Contingent on network effects existing for sound recordings, a different issue is whether those network effects work merely to shift demand among different sound recordings or whether they work to alter the size of the market. A third issue is the nature of the impact of network effects – are they local or global – does the totality of usage matter, or just the number of individuals using the product who personally know one another?

The network effects story applied to file-sharing is relatively straightforward. File-sharing is likely to increase the number of music listeners since file-sharing provides access even to those who are unwilling to pay for it.¹³ If there are more listeners to the music, even if the additional users were merely among the file-sharers, the value of music for non-file-sharing individuals might increase and the non-file-sharers would then purchase more music.

Before we accept this possibility of network effects, however, we need to understand possible mechanisms that might lead to network effects. The question is: do consumers derive value from other individuals listening to the same music they listen to? Unlike telephones, where network effects are obvious, or software, where the ability to transfer files might be important, the linkage of values between different music listeners is far less clear.

¹² It is unclear how strong network effects are for different categories of software. For some categories, such as personal finance software, network effects might be expected to be weak or nonexistent. In other cases, such as spreadsheets network effects are thought likely to be large. Although there have been attempts to measure the strength of network effects for spreadsheets, these attempts were marred by using Lotus 1-2-3 file compatibility as measure of network effects when such compatibility was also important for upgrading spreadsheet users wanting to remain compatible with their old work independent of any network effects.

¹³ It is possible that listening to file-shared music might merely replace listening to the radio or listening to purchased music. File-sharing, however, does allow more precise choice of music than could be achieved by radio listening.

Local or global network effects?

As normally modeled, network effects depend on the number of other users of the product. This would be a global network effect where only the total number of users counts, not the identity or locations of specific users. If you are in the habit of selling products over the telephone, network effects from telephones might well be global – you don't care who or where your customers are, as long as there are a lot of them.¹⁴

On the other hand, most potential network effects are likely to be local – you care mainly about calling certain individuals, you plan to transfer files with a few particular colleagues, you discuss movies and music with your friends. In these cases the total number of users of a product is largely irrelevant to you.

The first question is whether the activity of music-listening is subject to either local or global network effects to any important extent. This is a difficult question to answer. When a musical composition/performance becomes popular it might be due to some innate quality within the music, combined of course with some form of information, be it radio play, newspaper and magazine reviews, or word of mouth. On the other hand, it might be due some self-reinforcing network effect.

If network effects are global in nature then it isn't clear how file-sharing could impact sales since file-sharing doesn't usually show up in global measures of music popularity, which are normally based upon purchased CD sales or airplay.¹⁵ If consumers care about the total universe of users they can only estimate the number of users from these publicly available statistics.

Local network effects, on the other hand, suggest that CD purchasers interact with friends who engage in file-sharing and then are influenced by their friends file-sharing musical tastes.

There is also the question of whether network effects, if they exist for musical works, have the impact mainly of shifting output from one musical composition to another or whether they have the impact of changing the overall size of the market. Normally, network effects are thought to increase the value of products as networks get larger. Products have some autarky value independent of network effects which can then be enhanced as the size of the network

¹⁴ It is likely that for indirect network effects, where the price of some secondary item (such as repair service) is impacted by the number of users, that the identity of the users doesn't matter much and that the network effects are global. In the case of file-sharing, however, network effects, to the extent that there are network effects, would seem to be of the direct variety.

¹⁵ Websites such as BigChampagne.com purport to measure popularity of downloads (although it is apparently uploads that they measure) and it is possible that such measures might become part of the public information about popularity.

increases, as would be the case for computers and the Internet. In this instance network effects can only increase the value of products and the size of markets. It is possible, however, that network effects can lower utility below the autarky value – e.g., an unpopular restaurant that loses patrons who do not want to be seen as ‘uncool’. In this latter case network effects need not increase the overall value or size of markets. If network effects work mainly to allocate utility from one CD to another depending on shifts in popularity, but do not increase the overall utility of CDs in general, then they can have a neutral or negative impact on overall sales.

The nitty gritty of network effects in music

What might be the nature of network effects in music? Although it is almost certainly a common event that one person hears music at a friend’s house and decides to buy it, that is not sufficient for a network effect to exist. Just hearing new music is merely a form of sampling. For it to be a network effect, the value of the music once it has been heard must be higher because friends like it than would otherwise be the case. Good candidates for network effects might be dancing music since such music is involved with group activities. Even so, music listened to as a part of group activities implies that one of the group members already has that music. Whether the other members are likely to purchase the music themselves to enhance group consumption would seem to be questionable. In the case of file-sharing, a network effect would exist when a person who doesn’t file-share purchases music that his file-sharing friends can enjoy when they come over to visit. These do not appear to be powerful factors, so as a practical matter network effects from file-sharing would seem likely to be weak.

Although it is conceivable that network effects could play a role in the overall sale of CDs, and it is conceivable that file-sharing could increase overall usage sufficiently that CD sale might increase, it seems unlikely to be more than a minor impact.

The only test of network effects on record sales that I am aware of is indirect – looking at the impact of radio on record sales. Again, network effects from radio listening need to be separated from the sampling effects from radio, which is a more obvious impact.

One might argue that radio play is more likely to exhibit global network effects. Global network effects are possible with radio since radio play is one form of communicating what ‘everybody’ is listening to. If individuals derive value from being part of the crowd, then there would be global network effects from radio play. Second, local network effects are also possible since groups can sit around listening to the radio and commenting on the songs. This might

increase sales for those individuals wishing to make their friends happy in exactly the same manner as it might for file-sharing.

Undoubtedly, radio play influences the sales of records. Its main impact, however, seems to be to shift sales from one recording not receiving airplay to other recordings that do receive airplay. This conclusion is reached in Liebowitz (2004a) who examined two instances where radio broadcast did not increase overall record sales. In the first instance, the introduction of radio in the US in the 1920s, the sale of sound recordings fell dramatically. In the second instance, the introduction of private radio stations in the UK during the last three decades of the twentieth century led to no increase in the sales of records. Although these results were not meant to be tests of network effects *per se*, they are certainly inconsistent with the possibility of strong network effects.

2.4 Indirect appropriability

The final impact of copying that might apply to file-sharing is indirect appropriability. This is a concept coined in Liebowitz (1985) and analyzed for the case of file sharing in Liebowitz (2002) which I summarize below. It has recently been brought up by Boldrin and Levine (2003) whose work was then critiqued by Klein and Murphy (2003). The basic idea is that originals from which copies are made might undergo an increase in demand as those making copies of originals capture some of the value from those receiving the copies and transfer this value into their demand for the originals that they purchase.

If, for example, everyone who purchased a CD made one cassette to play in their automobile, then the demand for the original CD would increase by the value of being able to make the tape and the sellers could capture some of this higher value by increasing the market price, as would happen when demand increased. This value is captured indirectly since there is no direct payment made for the copy.

In order for indirect appropriability to work, however, one of two conditions must hold. First, the variability in the number of copies made must be small, as in the example above. Or else the seller needs to be able to charge higher prices for those originals from which the most copies are made as in the real world example of photocopies. The most heavily photocopied copyright materials are journals, and most photocopying of journals takes place in libraries. Thus publishers of journals were/are able to charge higher prices to libraries than they charge to individual subscribers to take account of photocopying and indirectly appropriate some of the value from copying. Support for this thesis came from empirical work revealing that such price discrimination was practically unheard

of prior to the advent of the photocopier, and the most heavily copied journals were also those with the greatest price differential.

Contrary to the claims of Boldrin and Levine, indirect appropriability will not work for file-sharing, however. Because there is great variability in the copies made from each original, and sellers of originals cannot identify which originals are going to be used on file-sharing systems, the mechanisms that allow indirect appropriability to function will not work.

3 The prima facie case against file-sharing

Data on the sales of recorded music tend to be available on a yearly basis from organizations of record producers, such as the Recording Industry Association of America (RIAA) and the International Federation of the Phonographic Industry (IFPI). The United States is by far the largest single market.

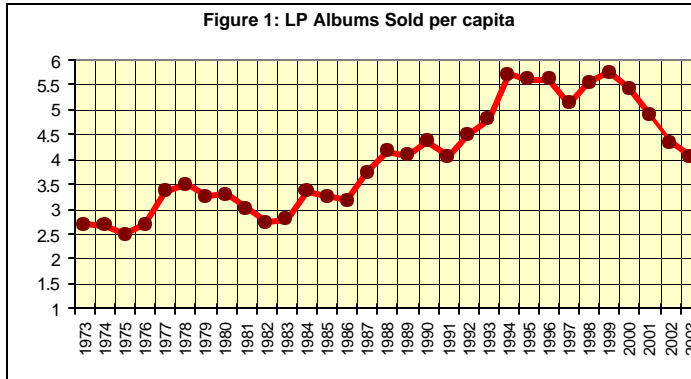
Figure 1 represents the per capita sale of full-length albums sold in the United States since 1972, using data from the RIAA.¹⁶ The recent decline in sales is readily apparent. Clearly, there are yearly fluctuations in this series. These statistics indicate a somewhat smaller decline than usually suggested by the industry since I have stripped out the impact of 'singles' because their decline appears to be part of a much longer secular decline quite separate from file-sharing. Nevertheless, the current decline stands at thirty percent.

This recent decline is clearly the largest that has occurred in the last thirty years. In absolute terms (units per capita) it is more than twice as large as the next closest decline and in percentage terms it is nearly fifty percent larger than the next largest decline, which may also have been impacted by copying.¹⁷ Another way to gauge the size of the drop is to count how many years back one would need to go to equal the new lower sales level. In the case of the current drop, one needs to go back to 1987 – 17 years – before encountering a year with lower sales per capita.¹⁸ The second largest decline ended in 1982 and erased the (much smaller) gains of the prior 7 years.

¹⁶ Data on quantities were reported beginning in 1973. For prior years only the industry revenues were reported. The data on revenues tend to be estimates based on the list price of recordings. I use full-length albums to avoid measurement problems as, for example, with singles, as they have been in a twenty year secular decline.

¹⁷ The second largest decline occurred in 1978-82. Long playing vinyl records were in decline at that time, being replaced by prerecorded cassettes. This decline coincided with a fairly deep recession and the possibility that blank cassettes were being used to copy records. As I explain in Libowitz (2004), cassettes ultimately allowed, for the first time, portability of prerecorded music, leading to a large increase in sales that overcame any negative impacts of copying.

¹⁸ 1991 is almost as low, but this still would be 13 years.



This recent decline is sufficiently striking that it would appear that something unusual has occurred in the last few years. Such a large change would be caused by either an unusually powerful but established factor, an unusual confluence of established factors, or some completely new factor.

Napster, which was the forerunner of modern file-sharing, came into existence in 1999, which also happens to be the peak year in sales. At the time, CD burners were still relatively expensive and did not yet have a large market penetration, making MP3 files less than perfect substitutes.¹⁹ Although Napster was closed down in 2001, file-sharers soon migrated to other file-sharing services.

The question, in a nutshell, is whether the decline in sales that began in 2000 is due to file-sharing. This file-sharing hypothesis is at the center of a political firestorm.

It is incumbent on us to respect the facts that we have. The confluence of file-sharing's birth and the decline in record sales, the rapid growth to an immense size of file-sharing activities, and the unusually large decline in the sound recording market all support a claim that file-sharing is responsible for the decline in sales. Add to this the predictions of economic theory that file-

¹⁹ The growth in CD burners impacts these arguments in two ways. First, the decline in sales would be expected to continue even if file-sharing were constant if the addition of CD burners make MP3 files better substitutes for purchase. Second, it is possible that individuals exchange copied CD with one another quite independent of file-sharing, thus tending to overstate any sales decline attributed to file-sharing as suggested in Peitz and Waelbroeck (2004a). This last possibility is weakened, however, by the fact that person-to-person copying has been popular for several decades (note the popularity of dual cassette audiotape player/recorders) and except for the fact that the quality of digital copies is somewhat higher, it isn't clear that digital personal copying will have an impact much different than analog personal copying had and thus the rise of CD burners might not have a noticeable impact on CD sales.

sharing should lead to a decline in sales and we have what appears to be a very strong case that file-sharing is the cause of the decline.

Although the 2004 RIAA half-year numbers are not yet out, SoundScan which measures the sales of CDs through the use of barcode readers at major retailers, reports an increase in sales of approximately 7 percent for the first half of the year according to news reports.²⁰ This turnaround nicely matches data from ComScore/MediaMetrix which indicates a decline in the number of file-sharing users of approximately 30 percent although it is likely that a decline in CD prices, brought about by a reduction in list price by one of the five major labels might also have played a role.²¹ This continued linkage between the size of file-sharing and the sales of sound recordings is additional evidence indicating a causal negative relationship between these two variables.²²

Economists, however, are trained to eschew straightforward answers. Nevertheless, the file-sharing explanation for the decline in sales is sufficiently compelling that only extremely powerful empirical evidence to the contrary should allow us to dismiss the file-sharing explanation. As the following review of the evidence shows, compelling and consistent evidence against the file-sharing hypothesis has not been forthcoming.

4 Economists to the rescue?

If we could run controlled experiments it would be easy to determine the impact of file-sharing. Because we cannot run controlled experiments and must use the less than ideal data at hand, we might not be able to answer this question in a convincing manner. Various analyses have been undertaken using different data and different approaches, and more are sure to come.

There are some preliminary points that I should mention before discussing these approaches in detail, however.

²⁰ For example, see <http://uk.news.yahoo.com/040702/325/ex5yb.html>.

²¹ Although the decline in users seems very large relative to the increase in sales, it is likely that the users most likely to stop using file-sharing in response to the RIAA lawsuits were the marginal users whose numbers probably overstate their actual impact on filesharing. This decline in users is mirrored in Nielsen/NetRatings numbers as well as in reported measurements by the NPD group of the number of files shared, although the numbers from Big Champagne reflect a much smaller decline in Internet usage. I am currently trying to evaluate the claims made by these organizations.

²² This recent rise in record sales has been taken by some as indicating that there is no linkage between sales and filesharing. For example, Kolemian Strumpf is quoted in the Guardian as saying: "If file sharing is killing record sales, why are records starting to sell better?" See Suw Charman "Listen to the flip side" July 22, 2004 at <http://www.guardian.co.uk/print/0,3858,4975370-110837,00.html>.

Although measuring the impact of file-sharing on the sound recording industry provides important information, it is only part of the analysis that would be required to answer the question about file-sharing's impact on social welfare or even its impact on the industry's ability to appropriate value. Even if it were the case that file-sharing had no effect on recording industry revenues, it might still have a large impact on appropriability. In the simplest case, assume that half the population purchases CDs and that these individuals do not change their behavior in the face of file-sharing. Then, due to file-sharing, the other half of the population listens to music downloaded from file-sharing sites. The producers appropriate none of this additional value created by their product, so appropriability, which might be defined as producer revenues relative to total value, has fallen.

The new consumption of music by this latter half of the population is an additional surplus for society, but that doesn't mean that this change is optimal.²³ Total surplus is maximized not when static consumer surplus is maximized for a fixed output, but when the sum of consumer and producer surplus is maximized in a system where producers receive payment so as to induce greater output. Some level of appropriability other than zero is usually optimal and it is likely that efficiency would require that some additional revenues be generated from the file-sharers.

Nonetheless, economists have limited themselves to the more prosaic question of whether or not file-sharing has decreased revenues to the sound recording industry. This is a useful first step, but we need to remember that it is only a first step.

4.1 Examining alternative explanations

One straightforward method to help determine whether the prima facie argument is correct is to examine other possible explanations for the decline in record sales that has occurred. I do not run multiple regression analysis for the simple reason that there are not enough data points. Besides, what we are looking for is some change in other variables that might explain the change in record sales that began in 2000. It is unnecessary to understand the full relationship between these variables when our goal is simply to determine whether some other factor(s) might have led to that change.

Examination of other factors that might impact record sales is the approach taken in Liebowitz (2004). Liebowitz addressed four main alternative factors that might impact record sales :

²³ I hesitate slightly to call it 'consumer's surplus' since these individuals do not pay a positive price and do not fit the normal definition of a paying consumer.

- a. Price or Income Changes
- b. Changes in Substitute/Complement Markets
- c. Changes in the Quality of Music
- d. Changes in the Supply of music.

I address these points in turn.

Price and Income changes

Although the claim is often made that price increases are responsible for the decline in CD sales the evidence does not support this view. Real list prices have been virtually constant for the last decade, and the increasing share of the market generated by discount sellers has probably lowered the average transaction price. Prices, therefore, are ruled out as an explanation of the sales decline.

There are numerous other possible complements/substitutes that might be examined. Although multiple regression analysis is our common methodology for examining the relationship between various economic variables, that is not possible here because there are too few data points. More importantly, there is no need to perform such an analysis. Whether a potential substitute/complement actually lives up to its expectation is not of importance. We have a relatively small list of such variables and only if one or more of them change at approximately the same time that record sales change we can presume that factor to be a possible cause of the decline. You can only explain the variation in record sales (from its trend) that begins in 2000 with variables that also have variations from their trends at about that time. Eyeballing the trendlines should be sufficient, and possibly optimal, for this purpose.

Nevertheless, we can begin by estimating a reduced form regression of the quantity of recordings sold. Regressing first differences in albums per capita on real per capita GDP and movie box office revenues per capita (the only measure of expenditures for movies that was available for the entire period) over the period 1973–1999 gives the results found in Table 1.²⁴ Movie box office receipts appear to have little relationship to album sales. The impact of income on sales over the last thirty years is statistically significant, but a

²⁴ It is obviously inappropriate to use the post 1999 data in these regressions since file-sharing started in 1999 and we wish to determine the impact of income and movies on record sales in order to judge whether the post 1999 sales were negatively affected by file-sharing. If file-sharing had an impact on sales, it would distort regression results ignoring file-sharing. If we use the entire sample period with a filesharing dummy for the years after 1999 the dummy is significant with a value of $-.54$.

change in per capita disposable income of \$1000 would alter per capita sales by only .28 units and the largest sustained decline in per capita GDP (1982) during the last thirty years was approximately \$661. Although the US experienced a recession in 2001, it was far too mild, given the historical impact of income on sales, to be able to explain the decline in sales that occurred.

Table 1

1973–1999

	Coefficients	Std. Error	t	Sig.
(Constant)	-0.0269604	0.089802	-0.3	0.767
Per Capita Change in GDP	0.0002794	0.0001245	2.25	0.035
Change in Movie Box Office	0.0421614	0.059508	0.71	0.486
R Square	Adjusted R Square	Dependent Variable: Change in Albums per Capita		
0.2122	0.1437	N = 26		

In fact, since 1999 real GDP per capita has increased by \$1883, which, according to the regression, should have led to an increase in sales of .52 units per capita. The year-by-year sales figures after 1999, based on changes in income, are provided in Figure 1x as the line with diamond markers. Of course, changes in GDP are not capable of explaining much of the prior declines so there are obviously other factors at work.

The impact of substitutes, such as the DVD market

One claim that is frequently encountered is the claim that record sales are down because alternatives, such as videogames, and DVDs are up.

We have already seen that there doesn't appear to be much of a relationship between movie box office receipts and record sales. Table 2 lists correlations between yearly changes in the sound recording industry and changes in these other industries. These forms of entertainment seem to be impacted by similar factors since yearly changes in each are positively related to one another. In some instances the number of observations is quite small due to the relatively recent creation of these markets, reducing our confidence in the applicability of these measurements.

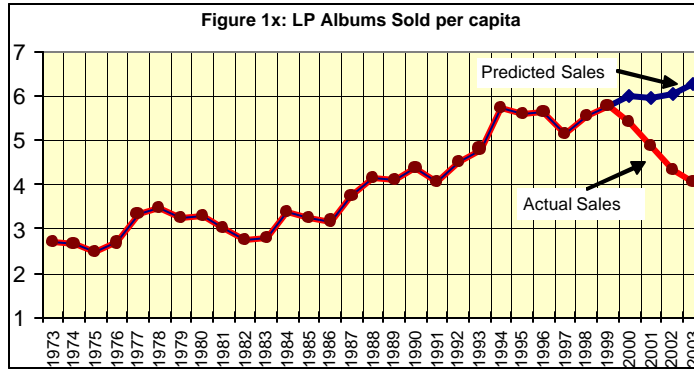


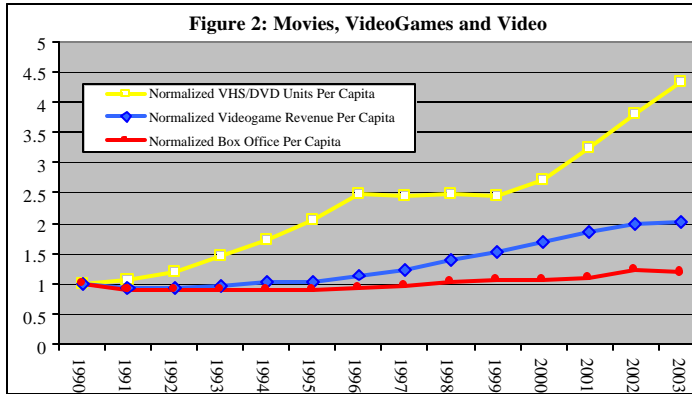
Table 2

Correlation of change in per capita unit record sales with change in ...

movie revenue per capita	0.199	1973–1999
videogame revenue per capita	0.265	1991–1999
units of prerecorded video per capita	0.177	1990–1999

Whether these positive correlations indicate that these products are complements with one another or whether they are due to a left-out variable bias, we cannot be sure, although income appears to have different impacts on these markets.²⁵ Certainly, videogames are often based on movies (and occasionally vice versa) and some soundtracks of movies sell very well, but it is difficult to believe that these complementarities are prevalent. Nevertheless, just as we saw that movies had an insignificant impact on record sales in Table 1, the positive coefficients reported in Table 2 should not be taken to indicate much of relationship between these variables. The point, of course, is that there is no evidence that these products have *negative* impacts on one another.

²⁵ The correlation of income changes with changes in movie revenues is .06, with videogames is .41, and with prerecorded videos is -.37.



More detailed information on the relationship of these markets can be gained by examining Figure 2.

Here we compare the changes in these markets since 1990, with 1990 values normalized to equal 1. We know that record sales began their steep dive in 2000. The only one of these other markets that has a fairly abrupt shift at about that time was the prerecorded movie market. Driven largely by the growth in DVDs, the prerecorded video market experienced a large growth in unit sales beginning in the year 2000. This might be why DVDs are often mentioned as a potential cause of the decline in CD sales.

But there is a fly in this ointment. The VHS market experienced a similar increase in unit sales from 1991 until 1996. Returning to the data in Figure 1 we can see that CD sales during 1991-96 had the largest *increase* of any 5 year period in our data. This largely explains why the correlation between CD sales and video sales was positive over the decade.

If the recent increase in DVD sales was the cause of consumers reducing their purchases of CDs, then the same market forces would be expected to have led to a decline in sound recording sales during the prior large increase in VHS sales. Instead, the exact opposite happened.

Which, naturally, brings us back to the question: why did these other media move generally in tandem with CD sales? One possibility is that interest in electronic gizmos, like other trends in consumer tastes, goes through cycles, and that music and video both tend to benefit from increased interest in home entertainment. Even under such circumstances, where both the video and sound recording markets are impacted by the same outside factor(s), the simple correlation that results between sales in these two markets still provides useful informa-

tion. As long as that third factor continues to be the dominant impact in both markets, video and sound recordings would be expected to continue to move together. Again, the divergence of the fortunes in these markets indicates that something changed in the year 2000. The most viable candidate is file-sharing.

Another somewhat different way of looking at this issue is to examine the average usage of time taken up by these activities. Published estimates indicate that in 1999 listening to recorded music (as opposed to radio) took approximately 48 minutes of a person's time per day, whereas going to movies took 2 minutes, watching prerecorded movies took 9 minutes, and playing videogames took 10 minutes.²⁶ If these measurements are at all accurate then the relatively low time usage of prerecorded movies or videogames does not appear capable of usurping much of the time spent listening to prerecorded music.

Since theatrical movies take up a trivial amount of time, and videogames do not preclude listening to prerecorded music, we can focus on prerecorded videos. We can take the most extreme case possible. Assume that each minute of watching prerecorded videos replaces a minute of listening to prerecorded music (given the much larger potential timetargets of television and radio listening this is quite an extreme assumption). How much of a decline would this have caused to the sound recording industry?

The growth in per capita prerecorded video was 80 percent from 1999 through 2003. An 80 percent increase would raise the time spent watching VHS tapes and DVDs by 7.2 minutes per person per day. Reducing time spent listening to prerecorded music by 7.2 minutes, if it translated one-to-one into lost sales, would reduce sales of sound recordings by 15 percent. Yet per capita sales of sound recordings fell by 30 percent during this period of time. Thus, even in the most extreme case imaginable, a time substitution could have caused no more than half of the actual decline, and this is an implausibly large estimate. Further working against this hypothesis is the fact that the decline in record sales in 2003 was only half that of the prior two years whereas the increase in prerecorded video was as large in 2003 as the prior two years.

We must also remember that record sales might have been expected to increase after 1999, as discussed in the last section. This stretches even further the credibility of these alternative explanations.

²⁶ US Statistical Abstract, Table No. 1102. Media Usage and Consumer Spending: 1996 to 2005. <http://www.census.gov/prod/2003pubs/02statab/info.com.pdf> Original source is Veronis, Suhler, Stephenson.

Table 3**Unit sales of prerecorded video**

year	VHS	DVDs	sum	per capita
1999	561	91.3	652.3	2.390954
2000	565	174.4	739.4	2.685739
2001	539.6	350	889.6	3.196482
2002	407.5	650.6	1058.1	3.761384
2003	240.4	985.3	1225.7	4.3112

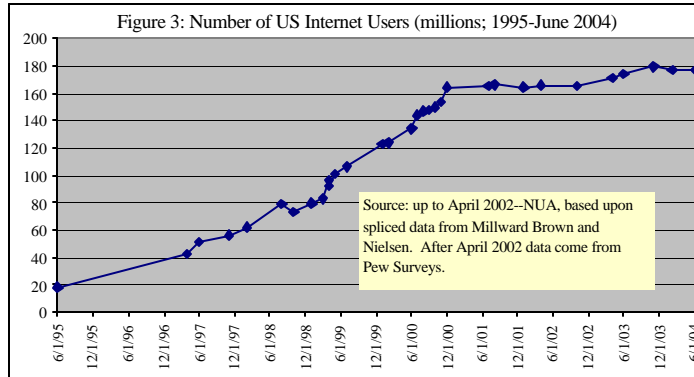
Finally, it has been suggested that Internet use itself, or Broadband penetration might explain the decline in music CDs. Yet it is not the case the Internet usage necessarily substitutes for music listening since both can be undertaken at the same time. Even so, the Internet, as shown in Figure 3, has been growing since its inception and if anything, its growth declined in the period of interest (1999-2003) and thus would not seem to be a viable explanation of the decline in music CDs. Broadband is just another form of Internet usage and although it might make file-sharing more attractive, as would CD burners, file-sharing is still the activity that would be causing the decline in CD sales. In order for broadband to be a factor in explaining the decline in CD sales, one would have to hypothesize that broadband somehow increased time spent on Internet activities that precluded music listening, and I am not aware of any data supporting this hypothesis.

Although this is an area for further research, it seems most unlikely that under more precise measurements these variables could have the type of negative impact required to explain much, if any, of the fall in record sales.

Has the music changed?

When the record industry has a decline, critics of the industry will often claim that sales fell because the music was not interesting. Thus, it is not surprising that this has emerged as an answer to the question: “if MP3s are not hurting the industry, what is?”

I looked at two pieces of evidence related to the ‘quality’ of music. One was the financial success of concerts from 1990–2001. Although there are defects with this measure, the years 2000 and 2001 had the largest real increases in concert revenues – at the same time the record industry was experiencing unusually large decreases in revenues.



A second potential source of data is time spent listening to radio. If music is losing its luster, then radio listening should decline. Data from the last 5 years, shown in Table 4, do reveal a fairly serious decline in listening – 9percent. The decline, however, is across the board, not just for groups that listen to new music. It is implausible that those over the age of 65, for example, would find their attraction to radio diminished because modern musical compositions were not to their liking.

Table 4

Fall 98 – Fall 2003

Group	Decline in listening (in %)
Persons 12+	9.30
Teens 12 – 17	11.67
Persons 18 – 34	13.48
Persons 25 – 54	8.79
Persons 55 – 64	6.74
Persons 65+	5.75

Further doubt about any claimed doldrums in current music composition comes from examining the listening shares of radio formats. If current music were inferior to past music, its share of listening should fall relative to old music and non-music categories. Yet the category of ‘News and Information’ is only up 1.8 percent over this period. The largest decline in radio listening (51.3 percent) is for the ‘Adult Standards’ category which consists of music

from the pre-rock era, hardly a repository of new music. The categories of 'Contemporary Hits' and 'Urban,' both of which have a great deal of new music, each increased their shares (5 percent and 23 percent respectively) and 'Alternative' (modern rock) was down only 2 percent. It is true that 'Country' music had a 9 percent decline, but this was largely offset by an increase of 43 percent in the 'Religious' category (which contains two categories of contemporary music).

The evidence on musical quality is far too weak and scant to explain the very large decrease in sales that has occurred.

Changes in the supply of music

It has been claimed by some that the number of new titles has fallen in recent years and this decline in new titles is responsible for the decline in sales.²⁷ Of course, the number of new releases is not exogenous and we would expect the quantity of releases to fall if demand fell because of file-sharing. Unless we are willing to entertain the possibility that the supply of songwriters and musicians has suddenly dried up at the same time that file-sharing has started, we would expect that changes in demand are more likely to influence the number of new titles than would changes in underlying supply conditions.

Further, this discussion has suffered from an imprecision in the measured numbers of new releases.²⁸ One factor that is clear, however, is that the large majority of new releases, between 65 percent and 80 percent, are from independent record labels which account for merely 10 to 15 percent of industry sales. Looking at industry-wide new releases overweights independent albums at the expense of major labels and is like having the tail wag the dog. Thus the total number of new releases is a misleading statistic as regards the overall sales in the industry and really cannot be a useful barometer of supply even if it were exogenous. It would be like comparing the total footage of home movies shot year-by-year to explain the yearly economic performance of motion pictures. They are not really part of the same market.

²⁷ See for example George Ziemann "RIAA's Statistics Don't Add Up to Piracy" available at <http://www.azoz.com/music/features/0008.html> The analyses contained on these pages are probably best described as 'rants' but they have nevertheless been taken seriously by some. Understanding the differences between retail sales volume measured by SoundScan's barcode reading technology and wholesale shipment numbers reported by major labels would be a useful exercise, but instead the difference tends to be treated in these analyses as some sort of industry conspiracy.

²⁸ Apparently the RIAA at one time provided the number of new releases and then stopped. A general statement about new releases on the RIAA was then taken by Ziemann and used in his analyses, although that was apparently an error. See <http://www.theinquirer.net/?article=9048>.

The perfect storm?

The conclusion from examining these alternative explanations is that they do not appear capable of explaining the decline in sound recordings that has occurred. It would take a remarkable confluence of events, a perfect storm if you will, to explain the large drop that has occurred in the sound recording market. That doesn't mean that it could not have happened. But in a choice between file-sharing as an explanation and the confluence of various disparate factors all perfectly aligned to harm the sound recording industry, Occam's razor requires that we accept the file-sharing hypothesis.

4.2 Using countries as the unit of analysis

There are several papers that take this approach. The idea is straightforward enough: compare changes in sales of sound recording in different countries over time using some measure, such as the share of Internet users, to proxy for the impact of file-sharing across countries.

Peitz and Waelbroeck (2004) use data from 16 countries for the period 2000-2001. They find a rather large decline for this one year period – 11 percent or 12 percent, for the world or US respectively. Using survey data from the US they then conclude that the growth in the number of downloaders couldn't have been responsible for more than a quarter of the decline that occurred in 2002.

Alejandro Zentner (2004) uses international cross section data from 97–98 to 2001–2002. He finds a worldwide decline of 15 percent and a US decline of 30 percent. Using an analyses of European survey data he finds that file-sharing reduces the probability of buying music by 30% and a drop in music sales of 7.8 percent.

Difficulties with the approach used by these authors include the fact that many of these countries are very different from one another. They have different levels of organized piracy, speak different languages, and have very different levels of per capita income, CD sales, Internet use, and stereo equipment. This great heterogeneity across countries makes it less likely that file-sharing will have the same impact in each country and imposes on the researcher a difficult burden, although this does not mean that it cannot be overcome.

4.3 Using cities as the unit of analysis

A paper by Eric Boorstin, a student at Princeton, used cities as the unit of analysis. This paper has received some attention because one of his advisors,

Ed Felton, is a very influential computer scientist who has declared on his widely read web page that Boorstin's analysis provides the best explanation of the impacts of file-sharing.

The paper is quite clever. Boorstin uses Census data on the number of individuals in each city with and without Internet access in his attempt to determine the impact of file-sharing on the sales of CDs within a city. Internet use is taken as a proxy for file-sharing. In this sense his paper is similar to those performing international comparisons except that his comparisons are based across cities instead of countries. He uses SoundScan data for the sales of CDs in each metropolitan area.

This approach has certain advantages over international comparisons. Socio-economic variations, as related to music are likely to be smaller across cities in the US than across the countries used in some other analyses. The language is common throughout the country, and the popularity of songs is fairly uniform within the country.

A further advantage is that Census data on Internet usage is broken down by age group so that, in principal, the analyst can determine the impact of file-sharing for different age cohorts of individuals.

The data used by Boorstin cover the years 1998, 2000, and 2001. By ending in 2001, his data contain only two of the four recent years of declining sales. Additional controls used by Boorstin are two (unusual) measures of income within a metropolitan area and dummy variables for the years 2000 and 2001 (of which I will have more to say below).

Boorstin provides regression results (using a fixed-effects model) with CD sales as the dependent variable. These results are partially reproduced in Table 5 (only the results related to file-sharing are reproduced with other variables not reported). Those results have a certain intuitive appeal. The impact of file-sharing is measured as the difference between the coefficients of the cohort of an age group with Internet access and those in the cohort without Internet access. For youthful individuals, file-sharing is deemed harmful to sales as indicated by the negative coefficient in the Internet effect column, whereas for older individuals file-sharing is found to be beneficial. Boorstin attributes the positive impact of older individuals to the sampling hypothesis, where he followed other writers in accepting that sampling would have a beneficial impact on sales. He concludes that the overall impact of file-sharing is positive

since the positive impacts of the older age cohorts outweighs the negative impacts from the younger age cohorts.²⁹

Table 5

Boorstin results, dependent variable = sales

Total sales	Coefficient	t-statistic	internet effect	significance
Age 0-4	-0.301842	-0.2		
5-14 no internet	4.831504	3.71		
5-14 with internet	1.245786	0.76	-3.585718	0.1196
15-24 no internet	7.134016	5.41		
15-24 with internet	2.897334	2.13	-4.236682	0.0311
25-44 no internet	-2.225841	-2.17		
25-44 with internet	2.921416	2.34	5.147257	0.0016
45+ no internet	-0.597636	-0.63		
45+ with internet	7.614472	6.23	8.2121077	

Upon closer inspection, however, several problems arise with regards to his specification. One important problem has to do with differentials between Internet users and non-Internet users in a city. There are no controls in the regression to account for socioeconomic differences between these Internet users and non-Internet users. Surely, we expect that Internet users are likely to be more affluent and better educated than non-Internet users. Isn't that what the whole debate on the digital divide was about? The data used by Boorstin supports this view since cities with higher incomes also have a greater share of individuals in each age group who have Internet access, as Table 6 makes clear.

Greater income should lead to greater record sales for the groups with Internet access, although I do not know which way education would likely cut. Because Boorstin's regressions do not account for income differentials by Internet access, we should expect that Internet users will have higher coefficients on the purchase of sound recordings merely due to their higher incomes. This will bias upward (in a positive direction) the impact of file-sharing as measured by Boorstin. This is a potentially serious defect. It is also a possible explanation for the positive impact of file-sharing that Boorstin finds for the older age groups when we have seen the economic theory predicts no such positive impact from 'sampling'.

²⁹ Boorstin multiplies these Internet effects by the number of individuals in the cohort. I believe it would have been more appropriate to perform this multiplication using the share of CD purchases made by each cohort, as I do. It would not have changed his conclusion, however.

Table 6

Correlation with city income

Age group	Share of individuals without internet access	Share of individuals with internet access
10–24	–0.3321	0.4063
25–44	–0.2866	0.4875
45–64	–0.3534	0.3797
65+	–0.4483	0.1433

For this reason, one might argue that the groups with the oldest individuals should be the benchmark for zero impact of file-sharing since it seems very unlikely that those over 65, for example, would engage in file-sharing to any serious extent. If we were to use Boorstin's oldest age group as a benchmark, the other groups would then have various negative impacts from file sharing. I come back to this issue later.

The single most glaring problem, however, is his inclusion of dummy variables for the years 2000 and 2001. Since these years are the years when file-sharing was coming into existence, some of the impacts of file-sharing are likely to be included in the dummies, as Boorstin freely admits.³⁰ Boorstin suggests that this was a difficult time for the music industry although the industry has claimed that the difficulty was due to file-sharing. But even if the industry were having difficulty for reasons other than file-sharing, it would be inappropriate to include these dummies which will interfere with the file-sharing variables. The pure file-sharing variables are unlikely to be impacted by factors such as a decline in music quality, so the inclusion of these dummies would bias upward the file-sharing variables and serve little extra purpose.³¹

Making the single change to Boorstin's results of excluding the dummy variables for the years 2000–2001 provides the following results.

³⁰ On page 54 of his paper Boorstin states: "The music industry has been complaining about a decline from its 1999 peak, and these dummy variables will pick up the market factors not included in my model. The increasing popularity of substitutes and theorized decline in overall music quality would be captured by these dummy variables. Also captured would be a change in the level of piracy that cannot be captured by Internet usage statistics."

³¹ Pre-recorded video usage may or may not be related to Internet/income levels, although there is no compelling reason to believe that there is such a link. Even if there were, however, the period 1993–1996 saw a strong upsurge in the sale of prerecorded videos but no decline in the sales of music albums.

Table 7

Removing year dummies

Total sales	coefficient	t-statistic	internet effect	significance
Age 0-4	-4.75724	-2.54		
5-14 no internet	1.630779	0.99		
5-14 with internet	-5.3804	-2.73	-7.011182	0.0187
15-24 no internet	10.41141	6.22		
15-24 with internet	-5.27322	-3.61	-15.684629	0
25-44 no internet	1.485543	1.2		
25-44 with internet	6.496813	4.73	5.01127	0.0065
45+ no internet	0.629566	0.52		
45+ with internet	7.082242	4.96	6.4526759	0

This table now looks to tell quite a different story. The coefficients indicate larger negative impacts for young people and smaller positive impacts for older individuals compared to Boorstin's original results. In fact, we can use the information in Table 8 which lists the age distribution of record purchasers³² to estimate the overall impact of file-sharing from the new analysis.³³

In 2001, for example, 41.7 percent of purchases were made by consumers between 25 and 44 years of age and consumers older than 45 purchased 23.7 percent of the total. Applying the coefficients from the regression to these purchase share numbers implies an overall decline in record sales of .92 units, which is just slightly more than the decline that had occurred as of 2001 (.68 from 1998 and .88 from 1999). Given that this is likely to be an underestimate due to the differential incomes of Internet/non-Internet users, this would imply that the entire decline (plus some forgone increase) in the sales of albums was due to file-sharing.

There is one further change to Boorstin's analysis that I would suggest. In his regressions the number of individuals in various groups in each city was regressed on the number of CDs sold. A more natural and more exacting specification is to use the relative size of these groups (measured as a share of the population) in each city to explain per capita sales in each city. This removes the strong correlation that would be expected merely from the large size

³² Boorstin uses the numbers of individuals in each age group.

³³ This information comes from the RIAA and is based on polls taken by Peter Hart Research. More information is available at the RIAA website www.riaa.com. I have not seen the surveys used to generate this information so I can not form an opinion about their likely veracity. For our purposes, the implications of the results in the table will hold as long as any imperfections in the survey remain consistent over time.

differentials between cities. The results are reported below. I also rearranged the groupings into more even natural and ones.

Table 8**Consumer information 2001**

Age group	Share of purchases	Purchase intensity (in %)
10–14	8.5	115.93
15–19	13	182.68
20–24	12.2	176.54
25–29	10.9	164.02
30–34	10.3	141.84
35–39	10.2	130.60
40–44	10.3	128.80
45+	23.7	64.77

Table 9**Age groups based on share of population**

Per capita units	Coeff.	Std. Err.	t	internet effect	significance
10–29 no internet	2.863291	1.82591	1.58		
10–29 internet	0.2094592	1.570257	0.13	–2.6538318	0.1024
30–44 no internet	1.751894	1.8017	0.97		
30–44 internet	2.658693	2.121496	1.25	0.906799	0.606
45–64 no internet	3.471715	1.528949	2.27		
45–64 internet	5.396905	1.691543	3.19	1.92519	0.2006
65+ no internet	4.859013	1.651844	2.94		
65+ internet	5.962898	2.43968	2.44	1.103885	0.6637

Under this specification the only Internet effect that is significant is the negative impact for the 10–29 age group. The coefficients are somewhat more reasonable since the Internet effect for the 10–29 age group now implies that a complete shift from no Internet usage to full Internet usage by this group would lower sales of sound recordings by 2.7 units per person, which is roughly a drop of one third for this age group which has average per capita purchases in the range of 9 units.

Of course, the Internet effect is still understated due to the income differential between groups. The 65+ group seems very unlikely to be impacted by file-sharing and thus this may be taken as a benchmark for the impact of higher income on sales of sound recordings. If so, the Internet effect changes to -3.75 for the 10–29 age group, implying that Internet usage would lower sales to this group by approximately 40 percent if everyone switched or about 24 percent at the average Internet penetration of about 57 in 2001.³⁴ With this age group responsible for almost one half of all sales, this would lead to a decline of about 12 percent due to file sharing, compared to the actual decline of about 15 percent by 2001. These two values are so close that we can say that this evidence is consistent with a view that all of the decline is due to file sharing.

This is a charming story, but it isn't clear how reliable these results are.

First of all, the magnitudes of the coefficients in the initial specification are troubling. Actual per capita sales are in the range of 5 or 6 CDs per year, yet according to the initial regression the impacts of Internet access are in the vicinity of 5 to 15 CDs. These large magnitudes are difficult to believe although the specification in Table 9 seems to address some of these concerns, although the high coefficients for older age groups is still troubling.

Second, if this regression is capable of measuring the impact of Internet use, it should also be capable of using population differentials across cities to determine relative purchase rates of different age groups, for which we have statistics from other sources. We have already noticed the disturbing results that older age groups appear to have larger coefficients than younger age groups when broken down by Internet access. Running a regression without Internet groupings provides the results found in Table 10.

Based on these regressions, individuals over 45 would appear to purchase records with a greater frequency than younger individuals. We know, however, that individuals in the 45+ age category actually purchase records with the slightly less than half the intensity of younger individuals according to purchase statistics reported by the RIAA. Unfortunately, the inability of these regression results to generate results consistent with known behavior reduces any confidence that we can have in the reliability of this approach.

³⁴ If we apply this logic to Boorstin's specification we find that when consumers over 65 are put in a separate category the other groups all indicate various negative impacts from file sharing. This would then imply a considerably larger (and implausible) decline in sales than has actually occurred (it would essentially imply that filesharing would have reduced sales by 8 units per person, which is larger than sales per person before filesharing occurred).

Table 10**Predicted versus actual purchase intensity**

group	Predicted purchase coef.	Sd. Err.	t	Actual purchase intensity (in %)
Age 10–24	0.8066711	1.510765	0.53	157.78
Age 25–44	2.022539	1.513401	1.34	140.34
Age 45+	4.179985	1.192174	3.51	64.77

It is not clear how much confidence we can place in these results. But they certainly do not show file-sharing to increase sales, or even to have a neutral result once the year dummies are removed. The strong negative impact for younger individuals appears quite robust, however, providing some possibility that this technique provides some useful information.

4.4 Using records as the unit of analysis

A recent paper by Oberholzer and Strumpf (O&S) uses records as the unit of analysis. Unlike almost everyone else, Oberholzer and Strumpf were allowed access to actual downloads logs on a server that was part of a file sharing system. They then matched music downloads to the recordings where the songs appeared. Using data from SoundScan on recordings they then compared the downloads of songs to the sales of songs over a seventeen week period. The creation of this data set was a substantial effort requiring much work and creativity. Whether it can be used to determine the impacts of file-sharing, however, is another question.

There are two potential problems. First, can an analysis using records as the units of observation provide information about the impacts on the entire industry, as opposed to the impacts on individual recordings? Assuming that this first problem is not an issue, a second obstacle exists – there is a serious simultaneity problem in the data because popular songs are going to be both heavily purchased and heavily downloaded. Overcoming such a simultaneity problem is no easy task.

First, however, I discuss their results.

Their conclusions versus their results

O&S suggest that their results indicate that file-sharing has no impact on the sales of sound recordings because they emphasize their most 'pessimistic' results. I read their paper differently and believe that a more appropriate representation of their results would be to conclude that file-sharing has a positive impact on sound recordings.

O&S begin their regression analysis by looking at their universe of approximately 650 recordings and the relationship of sales to downloads (their Table 11). As a benchmark they run a regression with sales as the dependent variable and downloads as an independent variable (using musical genres as additional independent variables). The coefficient on downloads is 1.07 and statistically significant. Of course, this coefficient is infected by the simultaneity problem mentioned above, implying that the coefficient is biased upward.

In an attempt to solve the simultaneity problem, O&S use several instrumental variables related to the length of songs on the albums, which they argue are appropriate instruments in a first stage regression with downloads as the dependent variable. In the second stage of their two stage regression the coefficient on fitted downloads (with sales as the dependent variable) is significantly positive, at 1.467. Strangely, the use of instruments that were supposed to control for the upward bias in the coefficient on downloads actually leads to an increase in the coefficient. It would appear that the instruments are completely failing to control for the simultaneity, yet the diagnostics relied upon by Oberholzer and Strumpf do not warn of this failure.

Given parameters used elsewhere in their paper, the coefficient in this instrumented regression translates into additional sales of 198 million albums caused by file-sharing. This result would imply, if taken seriously, that file-sharing has led to an increase in sales on the order of 25% of year 2003 sales. Given the actual decline since year 2000 of almost 300 million units, this positive impact of file-sharing is rather difficult to accept as being even remotely plausible. O&S tend to skip quickly by this result, however.

Their Table 12 adds weekly information into the mix, allowing O&S to use a fixed effects model. Since the fixed effects model includes a dummy for each sound recording, some of the simultaneity is taken care of merely by the use of the fixed effects model. The coefficient on downloads drops from 1.193 to .28. However, there is still simultaneity going on week to week since records are more and less popular over time. Again, O&S introduce several instruments in

an attempt to solve the simultaneity problem.³⁵ The second stage coefficient on downloads, in their two preferred models, are close to zero, taking on values of $-.014$ and $+.088$. The $-.014$ implies that it would take 5000 downloads to decrease sales by 1 album, a number they have repeated several times when describing their results. The “preferred” measurement ($.088$) would translate into a gain due to downloading of 12 million albums, which is quite small.

There is an important conceptual error in using these regressions to calculate the overall impacts, however. These regressions give each record album equal weight in the analysis. This would not be appropriate to answer a question about the overall impact of downloads on sales unless the impact of downloads on sales was the same for high selling records as it was for low selling records. O&S accidentally test this assumption in their Table 13 where they find that the impact of downloads is higher for albums in the top quartile of sales than it is for less popular albums. Since the top quartile of albums represent the large majority of total sales (85%), giving equal weight to all albums causes O&S to come to erroneous conclusions regarding the overall impact of downloads.

In their table 13, the quartile of largest sellers has a coefficient of $.468$ ($t=1.52$). This coefficient implies an increase of 63 million albums due to downloading. Although the level of statistical significance is slightly below traditional norms, this is not the ‘zero’ impact that O&S emphasize in their conclusions.

The unit of analysis: Potential fallacy of composition

To make this concern concrete, imagine that we wished to learn what the elasticity was for the wheat industry. What would happen if we estimated the elasticity of demand for each of the firms within the industry and then assumed that the elasticity for the industry was the average of the elasticities of its constituent parts? We all know, and teach our students, that industry elasticity is much lower than the elasticity facing the firms because there are fewer close substitutes for consumers at an industry level than there are at the firm level. This would be an example of the fallacy of composition that sometimes occurs when going from constituent parts to the whole.

A similar compositional problem would occur with respect to advertising at the firm level versus advertising at the industry level. The advertising analogy is

³⁵ One of these variables is the number of German schoolchildren on holiday, which can be viewed as either a choice of subtle inspiration or one bordering on the ridiculous. I tend toward the latter view and the reader can see my webpage for more details <http://www.utdallas.edu/~liebowit/intprop/germankids.htm>.

particularly useful apropos record sales due to the fact that downloading, sales, and popularity are all related in some sense to the 'buzz' that surrounds a recording, part of which is directly due to advertising.

Assumethat downloading increases the sales of a record. Perhaps downloading creates a buzz for record x and people buying records buy those records with buzz. Or perhaps downloading allows consumers to sample a high quality recording such as x leading to an increase its sales. In this example, heavy downloading makes recording x more successful than lightly downloaded recording y .

A regression using downloads to explain sales would return a positive coefficient, assuming that all other simultaneity problems were overcome. After all, increases in downloads, by assumption, lead to an increase in the sales of the downloaded recordings in this example. But that does not mean that downloads increase overall record sales. A positive coefficient could be entirely consistent with record sales being severely harmed by downloads and thus couldn't answer the question about the overall impact of downloading.

The sampling story, discussed above, is a case in point. Sampling might lead to greater sales for those recordings providing higher quality, but it is likely to lead to an overall decrease in record sales as discussed above. Any informational aspect of downloading separate from sampling, if there were one, has a similar result. Political advertising clearly increases the market share of individual candidates (increases their 'buzz'), but it need not, and probably has not historically increased total voter turnout. Yet a regression on individual advertising campaigns would show that when candidates increase their spending they generate a higher share of votes (or higher poll numbers).

Liebowitz (2004a) shows a similar result for the impact of radio broadcast on record sales. Although radio play obviously has a positive impact on market share of individual records, and would be expected to have positive coefficient in a regression of radio broadcast impact on the sales of individual records, its impact on overall record sales appears to be zero or negative.

Still, the existence of compositional difficulties is not proven but merely suggested. Nevertheless, the methodology of O&S is suspect and should be treated with caution.

4.5 Using households as the unit of analysis

Michel (2004) runs a regression analysis using households as the unit of analysis. The data come from the Consumer Expenditure survey. He uses data for 1998 and 2001 to calculate his estimates. He does not have identical

households in the two years, but instead uses households drawn from stratified samples. He uses computer ownership as a proxy for file sharing. He performs his analysis for expenditures on theatrical movie and, separately, for expenditures on sound recordings.

He finds that computer ownership caused a 4.5 percent decline in sound recording expenditures over this four year period, although the coefficient is not quite statistically significant. Interestingly, real sound recording revenues *rose* by just under one percent during the four year period from 1998 until 2001, so the decline in CD sales due to file-sharing that he measures indicates that without file-sharing CD sales would have increased by a larger amount over this period.

As is true for all studies, there are potential problems. One can question whether the ownership of a computer is a close enough proxy for downloading. The measured values of changes in sales of records and movie box office revenues also do not seem to reflect the changes that occurred in the overall market. This might be a weakness of comparing different households at two different periods of time, or it might be due to imprecision that is inherent in data based on surveys and diaries. Any analysis of these data is only as good as the reliability of the respondents, which is at least a somewhat questionable assumption.

4.6 Using genres as the unit of analysis

I had high hopes for this methodology, although no one has attempted to use it to my knowledge. With data such as that created by Strumpf and Oberholzer, one could determine the variation in file-sharing proclivity by genre, measured as the number of files downloaded as a share of total sales. It is likely that certain groups of users (rock, hip hop?) would be far more likely to engage in file-sharing than other groups (classical).

If yearly sales by genres was reasonably smooth, then we could look to see if those genres with the greatest incidence of file-sharing suffered relative to other genres. As a control, we could use the share of radio listenership by genre.

My high hopes for this methodology were quashed when I purchased data on genre sales from SoundScan. First, they had only seven genres plus a catchall genre called "current." Second, and more important, sales by genre had extremely large changes over very short periods of time. The chart below, showing two genres, has a 170 percent increase in sales in the Metal genre in 1999 and a 100 percent increase in the R&B genre in 1997. Without more information that might help explain such sudden jumps, this data seem unlikely to prove reliable enough to use for any important empirical tests.

	Metal	R&B
1994	38,739	80,819
1995	31,101	80,718
1996	26,409	74,035
1997	28,983	141,613
1998	30,086	166,379
1999	82,698	175,339
2000	89,924	197,141
2001	88,158	195,498
2002	74,677	160,183
2003	74,629	149,972

5 Conclusions

The analysis of file-sharing is a relatively new phenomenon. Economists are just getting started. Nevertheless, progress is being made.

The theory underlying the analysis of file-sharing has not received the attention that it deserves. It has always been clear that some possible aspects of file-sharing would harm copyright owners, such as the substitution of copies for the purchase of originals. What has not been understood is that the use of file-sharing to sample products is also likely to lead to harm to copyright owners. Although one can still construct theoretical conditions under which file-sharing might benefit copyright holders, these conditions seem quite far fetched. A broad analysis of the various theoretical factors at work supports a view that file-sharing is likely to cause serious damage to the owners of copyright materials that are so shared.

The last few years have provided economists a laboratory of sorts to analyze the impacts of file sharing. The basic evidence in the United States over the last few years – the birth of file-sharing and the subsequent decline in CD sales – makes for an extremely compelling and simple explanation in spite of the protestations to the contrary from a large and vocal group of individuals supportive of file-sharing.

Empirical examinations by economists must be undertaken against the backdrop of this simple explanation. That these basic facts fit nicely with the economic theory only adds support to this explanation. With such a strong prior, empirical studies need to meet a higher hurdle than normal before they might be considered to overturn this expectation.

All of the empirical works to date suffer from various imperfections. Nevertheless, the majority of studies find results supportive of the thesis that file-

sharing is causing harm. As more evidence accumulates, both in the world at large and in the pages of academic publications we can expect to learn more about the impacts of file-sharing. The political arena is calling for an answer now, however, and is impatient to wait for academics to reach unanimity, which academics almost never do anyway. If an answer is needed, the answer that would appear to have the greatest likelihood of being correct given our current state of knowledge is that file-sharing hurts copyright owners and that it is responsible for most, if not all, of the recent decline in sales.

References

- Boldrin, M. and D. Levine (2002), "The Case Against Intellectual Property", *American Economic Review*, 209–212.
- Boorstin, E (2004), *Music Sales in the Age of File Sharing*, Senior Thesis, Princeton University.
- Bouillin-Dartevelle R (1989), "Belgium: Language Division Internationalized.", in: L. Becker and K. Schoenbach (eds.), *Audience Responses to Media Diversification*, LEA, Hillsdale, New Jersey.
- Conner K.R. and R.P. Rumelt (1991), "Software Piracy-An Analysis of Protection Strategies", *Management Science* 37(2), 125–139.
- Gopal, R.D., S. Bhattacharjee and G.L. Sanders (2005), "Do Artists Benefit From Online Music Sharing?", *Journal of Business*, forthcoming.
- Hirshleifer, J. (1971), "Suppression of Inventions", *Journal of Political Economy* 79, 382–383.
- Klein, B., A.V. Lerner and K.M. Murphy (2002), "Intellectual Property: Do We Need It? The Economics of Copyright 'Fair Use' in a Networked World", *American Economic Review*, 206–208.
- Lessig, L. (2004), *Free Culture*, The Penguin Press, New York.
- Liebowitz, S.J. (2004), "Will MP3 downloads Annihilate the Record Industry? The Evidence so Far", *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth* 15, 229–260.
- Liebowitz, S.J. (2004a), "The Elusive Symbiosis: The Impact of Radio on the Record Industry", *Review of Economic Research on Copyright Issues* 1, 20–45.
- Liebowitz, S.J. (2002), *Rethinking the Network Economy*, Amacom, New York.

- Liebowitz, S.J. (1985), "Copying and Indirect Appropriability: Photocopying of Journals", *Journal of Political Economy* 93, 945–957.
- Liebowitz, S.J. (1982), "The Impacts of Cable Retransmission on Television Broadcasters", *Canadian Journal of Economics* 15, 503–524.
- Liebowitz, S.J. (1982a), "Durability, Market Structure And New -used Goods Models", *American Economic Review* 72(4), 816–824.
- Michel, N.J. (2004), "The Impact of the Digital Age on the Music Industry: A Theoretical and Empirical analysis", working paper.
- Oberholzer, F. and K. Strumpf (2004), "The Effect of File Sharing on Record Sales An Empirical Analysis", working paper.
- Olderaan F. and N. Jankowski (1989), "The Netherlands: The Cable Replaces the Antenna", in: L. Becker and K. Schoenbach (eds.), *Audience Responses to Media Diversification*, LEA, Hillsdale, New Jersey.
- Peitz, M. and P. Waelbroeck (2004), "The Effect of Internet Piracy on CD Sales: Cross-Section Evidence", *CESifo Working Paper* No. 1122.
- Peitz, M. and P. Waelbroeck (2004a), "An Economist's Guide to Digital Music" working paper.
- Romer, P. (2002), "When Should We Use Intellectual Property Rights?", *American Economic Review* , 213–216.
- Shy O. and J-F. Thisse (1999), "A Strategic Approach to Software Protection", *Journal of Economics and Management Strategy* 8, 163–190.
- Takeyama, LN. (1994), "The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities", *Journal of Industrial Economics* 42, 155–166.
- Varian, H. (2005), "Copying and Copyright", *Journal of Economic Perspectives*, forthcoming.
- Watt, R. (2004), "The Past and the Future of the Economics of Copyright", *Review of Economic Research on Copyright Issues* 1(1), 1–11.
- Weimann, G. (1996), "Cable Comes to the Holy Land: The Impact of Cable TV on Israeli Viewers", *Journal of Broadcasting & Electronic Media* 40, 243–257.
- Zentner, A. (2003), "Measuring the Effect of Online Music Piracy on Music Sales", working paper.
- Zentner, A. (2004), "Measuring the Effect of Music Downloads on Music Purchases", working paper.