WOJCIECH HARDY
MICHAL KRAWCZYK
JOANNA TYROWICZ

FRIENDS OR FOES?
A META-ANALYSIS OF THE LINK BETWEEN "ONLINE PIRACY" AND SALES OF CULTURAL GOODS

WARSAW 2015

Friends or foes? A meta-analysis of the link between "online piracy" and sales of cultural goods

WOJCIECH HARDY
Faculty of Economic Sciences
University of Warsaw
e-mail: whardy@wne.uw.edu.pl

MICHAL KRAWCZYK
Faculty of Economic Sciences
University of Warsaw
e-mail: mkrawczyk@wne.uw.edu.pl

JOANNA TYROWICZ
Faculty of Economic Sciences
University of Warsaw
National Bank of Poland
e-mail: jtyrowicz@uw.edu.pl

Abstract

Over the past decade or so, the literature has sprung in analyses of the impact the so-called online or digital "piracy" has on sales. Since theory posits both positive and negative effects are possible, the question remains purely empirical. Consequently, there is a variety of published articles and working papers arguing in both ways, many of which attempt to account for the challenge of providing a reliable and causal effect. The objective of this survey is to review and discuss the accomplishments of the field so far. We also provide a tentative meta-analysis. Despite the multiplicity of measures and methods used we argue that the literature as a whole fails to reject the null hypothesis of no effects on sales.

1 Introduction

It is rare that a social phenomenon attracts public attention and also sparks the scientific interest. In some cases, it is because from an academic perspective the topic is obvious and the public debate concerns social consensus rather than scientific facts. In some other cases, data scarcity makes it problematic to design and complete a quality research agenda. Finally, whether or not funding is available for research in some topics is another constraint affecting the quantity and the consensus among the researchers. In this paper we analyze one case that suffers on all these three accounts: the link between activity referred to as "online piracy" and the sales of cultural goods. First, while "online piracy" is a popular phenomenon, its effects on sales have received considerably less attention from the scientific community than from the commercial or civic actors. Second, being largely unobservable, data on "online piracy" are relatively scarce whereas the construction of the reliable counter factual for the sales is indeed challenging. Third, funding for these topics is typically provided by commercial or social actors with clear interests – raising finance for research agenda without prior normative stand may be hard or even impossible.

Given these problems it is not surprising that the field has no clear consensus. On the one hand, careful reader will find at least a few articles arguing in favour of a positive effect of acquiring unauthorized content online. They rely on a a range of arguments such as sampling and exploration, "dipping the toe", increased revenues for complementary goods, etc. On the other hand, some studies find an effect of "online piracy" substituting legitimate sales. Indeed, zero or low price competition from unauthorized distributors can effectively reduce demand for usually higher priced CDs, MP3 files, DVDs, movie theatre tickets, etc. This discrepancy of conclusions is in itself a sufficient justification to engage in a meta-analysis.

In addition to conflicting results, there is also a disparity in measurement. Indeed, "online piracy" is rarely directly observable, which necessitates the use of indirect measures or proxies. Furthermore, the measures of legitimate sales or proceeds from selling authorized content are often unavailable or replaced by other indicators. The very term of "online piracy" – though commonly used – is highly imprecise. Intuitively, most people recognize that acquisition of content from unauthorized distributors constitutes "online piracy", but from a legal perspective it is not as clear. For example, in some countries (e.g. the United States) acquiring copyrighted content from unauthorized distributors is a crime and can result in fines or even imprisonment. While this demarcation line seems fairly clear, the law permits the so-called fair use, i.e. sharing files with family or friends in a way that would be analogous to non-digitalized cultural goods. In fact, the peer-to-peer networks rely exactly on users sharing cultural content one with another rather than some centralized distribution of files, thus taking the concept of a fair use to or even beyond its legal limits. Also in the recent decade more and more unauthorized distributors make content available through streaming, which implies that the "downloader" actually never acquires the file but rather views it in real time from a server located outside the jurisdiction of e.g. American courts. Partly in response to these changes and partly for other reasons the legislation differs from the American in many countries. For example in Germany or Poland it is unauthorized distribution that violates law, whereas acquisition is either entirely legal or an offence pursued on the grounds of civil law and relatively troublesome to prove in court. In most countries, whether or not the client paid for the content is irrelevant for the legal status of the deed, but local unauthorized distributors differ in business models: some specialize in commercials, others request the so-called transfer payment, finally some distributors expect a constant fee for maintaining the account; to avoid criminal charges most unauthorized distributors offer the actual content for free. It is also not relevant if the client had understanding of whether the distributor is authorized to distribute this

content. Given these definitional difficulties, measurement becomes particularly troublesome, especially as the state of law might change (e.g. the case of Netherlands, where until recently downloading was considered fair use)[1].

In addition to the variety of measures/proxies used, there is also a variety of methods used in this literature. Indeed, a positive correlation between (a proxy of) "online piracy" and legitimate sales cannot be meaningfully interpreted as these two products being complementary, because typically more attractive content will be more frequently acquired from both authorized and unauthorized distributors. In the same vein, a consumer with a larger taste for culture may acquire both more authorized and unauthorized versions of content. Typically one would interpret content distributed with and without authorization as substitutes if cross-elasticity of demand was positive. However, as already discussed, unauthorized distribution of content usually has no per-good price: access is either entirely free of charge or with a zero marginal cost. Consequently, even if data was available on per-good per-client basis, cross-elasticity could not be computed. One way to address this problem would be to substitute actual price of content acquired from unauthorized sources with a fine multiplied by the individually perceived probability of being forced to pay this fine. However, fines are either absent or rarely given in most countries, so it would be rational to have near-zero expected probability of fine, bringing the estimates of cross-elasticity again to near-zero price for "pirated" goods.

With these two paramount problems (popularity driving both legitimate sales and "piracy" as well as near-zero price for "pirated" content), the field has offered a variety of methodological solutions. While they remain controversial – see for example Oberholzer-Gee and Strumpf (2007) and the subsequent critique by Liebowitz (2010) – they are designed with the intention to discriminate between the theoretical insights which suggest room for both positive and negative effects of "piracy" on sales of cultural goods. To avoid some of the delineation issues, the articles typically analyze fairly similar cultural goods (e.g. music or movies). Some of the studies also try to narrow the analyzed topic geographically (e.g. analyzing sales within a campus or in shops in certain districts). Methodologically, it seemed relevant to several authors to control for individual tastes of culture or to apply fixed effects estimates to "erase" this and other unobservable effects. A strand of research attempted to instrument for "piracy" with variables seemingly unrelated to legal sales or to identify via event studies (such as the introduction of new laws or the shut down of major unauthorized distributors). This proliferation of methods, subjects and identification strategies constitutes yet another motivation to engage in a meta-analysis. In fact, in total, for the final analysis we have identified as many as 40 studies (with more than 600 estimates) of which 4 argue in favor of a positive effect of "piracy" on sales, 21 demonstrating the opposite, 6 finding no relationship whatsoever and 5 finding the direction of the link dependent on the type of content or analyzed sample. In addition, in most of the papers, at least some of the specifications were insignificant.

Our objective in this paper is to provide an (systematic and fairly synthetic) evaluation of the achievements in this field of economics. We analyze the available research papers with two major objectives. First, we aim to shed some light on the developing standards in this strand of the literature, emphasizing the relevance of endogeneity bias and the need for the strong identification strategy. We review measurements and methods used with the objective to shed some light on the external validity of some positive and some negative results. Second, we implicitly intend to test the contention that with more robust methods and more comprehensive measures the actual substitution effect becomes fairly small –

---

[1]It is popular to use "piracy" to express both acquisition from unauthorized distributors and unauthorized distribution. In the reminder of this paper, for clarity, if we use the term "piracy" we refer to acquisition or "downloaders". The term for uploading content without authorization will not abbreviated. The verb "to pirate" will be used to refer to downloading from unauthorized distributors.

possibly even negligible. This result would be striking given the centrality of this topic to economics. Popularity of acquiring content from unauthorized distributors is after all yet another example of a concept referred to in the literature as enforcement problem. The ability to protect one's property from unauthorized use is not a new topic in economics – its traces may be found in numerous applications of empirical institutional economics, conflict economics, endogenous growth theories to name just a few.

The paper is structured as follows. In the next section we discuss the insights from the available literature in a broader context of institutional economics and property rights protection. In section 3 we discuss the literature sampling procedure and describe quantitatively the literature of the field. The discussion of the estimates, including a deeper insight into the strengths and weaknesses of proxies and methods used is presented in section 4. After the exploratory analysis, in section 5 we move to the meta-regressions, assessing quantitatively the role of robustness for the conclusions from the literature. The policy and research implications of our study are discussed in the conclusions.

2   Theoretical insights

Over the last two decades or so several theoretical advances have been made as to the impact of unauthorized access ("piracy") on legal sales. (Peitz and Waelbroeck, 2006a) provide an excellent review of models of piracy. Some of these, however have been made less relevant by technical progress (e.g. sharing in non-overlapping clubs (Bakos et al., 1999)) and others focus on other issues than substitution between original and pirated versions. Furthermore, the scope of that review is obviously limited by the year of publication.

In the simplest model, assuming that the pirated copy is a perfect substitute of the original for all customers and is available at no direct or indirect cost, all sales are necessarily cannibalized. Even in this extreme case, it should be noted, calculating lost revenues on the one-to-one basis is clearly incorrect–under downward sloping demand curve faced by an individual company (which will generally be the case in any imperfectly competitive market), consumption at zero price will be larger than under optimal pricing absent piracy.

Several more complex models allow for some sort of quality gap between the original and the unauthorized copy, typically with customer heterogeneity in terms of valuation of this difference. Indeed, pirated copies may have lower technical quality; they may come without physical complements such as printed manuals; they will typically allow no access to customer support etc. Even more fundamentally, while e.g. digital files downloaded to a computer might be a good substitute for a DVD, this will not be true for a movie screened in a theater (especially if its appeal largely lies in special effects). If the pirated version is not a perfect substitute, positive legal sales are still possible at non-zero price. An optimal response on part of producers in a non-competitive market will typically involve lowering of prices. This can even lead to increased legitimate usage, although not higher profits, see Bae and Choi (2006).

Yet models consider a broader notion of utility loss, in particular associated with moral cost of unauthorized acquisition of the product, cfr. Chen and Png (2003); Kobus et al. (2013). Likewise, Balestrino (2008) investigates the possible social norm of no-piracy; basing on functionalist view of norms, however, he argues that no social stigma will be attached to digital piracy, because it involves no perceived social cost.

Other authors account for possible positive impacts of piracy on sales. First, many digital products show substantial network externalities. If consumption value indeed increases in the number of users (because a movie enters a "must-see" category, because user-made additions to software are more easily available etc.), also unauthorized access may benefit legitimate users (Reavis Conner and Rumelt, 1991). For some distributions of taste for

the original and pricing options available to the producer, this effect may even boost legal consumption. Another potential benefit of piracy involves sampling. Consumers who are uncertain about the value of the product can download an unauthorized version and, provided it turns out to match their needs, make a legitimate purchase, see Gopal et al. (2006a) for a single-product and Peitz and Waelbroeck (2006b) for a multi-product setting. It must be noted, however, that both of these benefits of the low-cost version could in principle be provided by the producers themselves (King and Lampe, 2003; Halmenschlager and Waelbroeck, 2012), as in the popular "freemium" business models in services such as Spotify.

It is safe to assume that rapid technological changes will continue to shape development of theoretical models of piracy in near future. For example, ten years ago slower Internet connection, lower file compression and computer screens of lesser quality, accompanied by fewer means of professionally recording and uploading movies – all led to a much smaller focus on the movie industry. Simply downloading a video quickly and in high quality was much harder than it is nowadays. With faster bandwidth and more organized 'pirate' networks like the Warez groups, it may now be a case of minutes to acquire virus-proofed content, even in BluRay quality. On the other hand, availability of video streaming platforms for PCs is substantially larger than for mobile devices, which is likely to contain acquiring unauthorized content to computers. This example is far from unique – the need for unauthorized content may be limited both in scope and in scale by the availability of the alternatives, not necessarily costless, such as streaming services, VOD, etc.

3   Literature Sampling

Given the goal of this paper, we searched for empirical studies quantifying the relationship between "online piracy" and sales of cultural content (or proxies thereof). In line with standard methodology of meta-analyses we spared no effort to ensure that our list of studies is as comprehensive as possible. We wanted to include both published and unpublished papers, irrespective of their quality, notion of "piracy" involved or details of the estimation procedure. For each paper under consideration we read the abstract (and the body of the paper if necessary), until we were able to determine whether the paper indeed contains original empirical work quantifying the link between some measure of "piracy" and some measure of legal sales of the same digital product category. We could only include papers that contained coefficients on explanatory variables identified as (in)direct proxies for piracy and explained variables identified as (in)direct proxies for sales.

Our starting point were the literature reviews by Smith and Telang (2012) and Dejean (2009), which cite 27 relevant articles.[2] Additionally, a systematic search using EconLit data base performed in January 2014. Specifically, in EconLit we searched for (piracy AND sales) (piracy AND displacement) (piracy AND effect), ("digital piracy"), ("online piracy"), ("music piracy"), ("film piracy"). We inspected all of the studies returned in response to this search. The studies found *via* published papers and registered working papers series available in EconLit were relatively few and were also missing a number of references reviewed by Smith and Telang (2012) and Dejean (2009). To make our sample more comprehensive, we also employed Google Scholar service in January 2014.[3].

We used the same search words as in the case of EconLit, which produced a much

---

[2]After finishing our search we have also found the literature review by Grassmuck (2010). However, no additional relevant papers were found in this work.

[3]We are aware of the criticism raised against employing Google Scholar service for the search of material for meta-analysis. We made sure that some of the hazards associated with this method are minimized. First, we sought a published version, if it were available online but missing from EconLit. Second, we referred to the newest available working paper if the published paper was non-existent.

larger number of hits, but also a much smaller fraction of relevant scientific papers. We thus placed queries where each of the terms (piracy) (piracy AND displacement), (piracy AND effect), ("digital piracy"), ("online piracy"), ("music piracy"), ("film piracy")[4], was coupled with (sales OR revenue OR "box office"). Relevant articles were generally very scarce beyond the first 100 hits, we restricted our attention to the first 200 hits per search in the case of Google Scholar. As a final stage of collecting the sample, we have inspected lists of references of all the papers included in the sample.

Table 1: Number of studies per year and numbers of derived coefficients

Year of publication	No. of published articles	No. of working papers	No. of estimates	Conclusions Film	Music
2004	0	3	40	0	7
2005	2	0	2	0	2
2006	5	1	36	3	7
2007	5	0	85	5	3
2008	2	1	29	0	3
2009	4	1	41	2	4
2010	4	0	29	2	4
2011	0	3	16	4	2
2012	3	2	60	6	4
2013	1	7	88	6	7
Total	26	18	426	28	43

Note: whenever a published paper is available, the working paper is disregarded.

In total our search procedure returned 72 research papers on the relationship between "online piracy" and sales. We then searched the sample and eilminated from the further analysis these papers that did not in fact establish this link empirically e.g. were theoretical models, or were in fact focused on downloading behavior or its effect on supply of culture. A share of articles were legal papers and philosophical essays, containing no empirical estimates of the relationship in question. Also, some papers were economic, but purely theoretical or descriptive, which too prevented empirical conclusions. In addition, papers or articles that did not *demonstrate* the analysis of "online piracy" and sales could neither be considered. This second group of articles and working papers includes some relatively highly cited examples, where authors e.g. assumed a degree of harm done and calculated the 'actual' losses. In the third step, after careful inspection, the analysis was narrowed to the papers that actually estimated the effects of "online piracy" on sales, rather than other relationships (e.g. propensity to download content from unauthorized distribution with no link to legitimate sales or "piracy" effect on the supply of content). This procedure reduced the final sample to be analysed from 72 to 44 papers.

The final sample of 26 published articles and 18 working papers yields 71 general conclusions about piracy's effect on sales (some studies analyzed more than one type of good, or considered different samples) and 426 estimates that could be studied through a meta-regression, Table 1.[5] The sample is fairly balanced between studies on movie and music piracy. There are also fewer studies of movie industry in the first years of our sample and more of music industry. The opposite can be said about more recent years. There seems to be a much stronger consensus in the film industry than in the music industry as well, although in both cases there is support to both positive and negative effects of "online

---

[4]It should be noted that we have searched for studies in industries other than movie and music, although with poor outcomes - we have only identified one study on "piracy's" impact on video games' sales (Fukugawa, 2011) and therefore we do not include it in any of the following tables and analyses.

[5]The list of all papers used in our study is available in Table 8 in Appendix A. Table 1 contains the numbers of studies published or disseminated in subsequent years.

piracy" on sales, see Table 2. One interpretation is that in the first years of the 21st century the Internet "piracy" was considered a threat mainly to the music industry and was therefore studied more vigorously. Indeed, it is the music-sharing platform Napster that is often referred to as the starting point of the "piracy age". One should also remember that files with films are much heavier and therefore more difficult to transfer – leading to the conclusion that film piracy might have reached a higher scale only lsubsequent to the increasing coverge of broadband internet and reduced fares for access.

3.1 Descriptive statistics

Most of the studies in our sample find that "piracy" displaces legal consumption of cultural goods, see Table 2. While many of the papers comprise more than one single estimate, authors interpret the findings of their paper in abstract, introduction or concluding sections. We classify these interpretations of the authors to positive, negative or inconclusive, depending on the wording. This procedure entails as little discretion from our side as possible.

Table 2: Relationship between piracy and sales, by the publication year and industry.

Year of publication	Film industry			Music industry		
	Negative	Inconclusive	Positive	Negative	Inconclusive	Positive
2004	-	-	-	2	2	3
2005	-	-	-	2	0	0
2006	2	1	0	5	0	2
2007	5	0	0	1	2	0
2008	-	-	-	2	0	1
2009	1	1	0	2	1	1
2010	1	1	0	3	1	0
2011	1	3	0	2	0	0
2012	2	3	1	2	0	2
2013	3	1	2	5	1	1
Total	15	10	3	26	7	10

Negative effect is far from universal, since there is also strong evidence of the contrary. In fact, as much as a third of estimates concerning the film industry find no effect, whereas approximately one in four studies on music finds a positive estimate of the link between "piracy" and sales. These conflicting results may stem from a number of effects. First, in principle there are two theoretical effects: dipping the toe (positive) and substitution (negative). If they have a fairly similar size, the failure to reject the null hypothesis may actually imply *strong opposing effects* rather than *no effect*. Second, both positive and negative estimates are likely to be biased due to the omitted variable problem, sample selection as well as reverse causality. Not all studies attempt to account for this bias, which necessitates caution in deriving conclusions from this table. Third, while majority of the studies concerns the United States, the 44 analyzed research papers comprise also studies for Australia, Canada, China, France, Germany, Japan and Sweden, along with several cross-country studies.

The final and possible most important explanation for the differentiated results summarized in Table 2 consists of the methodological differences in how both "online piracy" and sales are measured or proxied. The studies of the film industry include analyses of "piracy's" effect on box office revenues, DVD sales, TV watching, video rentals etc., while the ones of music industry distinguish between digital and physical formats of music. To simplify the categories we have constructed eight groups based on essential differences between the explained variables. Five of the groups are clearly related to consumer-level data on individual behaviour. These usually came from surveys or some forms of tracking

software. Three of the groups related to producer or industry-level data, like general sales volumes, revenues or chart ranks, Table 3.

Table 3: Measures and proxies for sales (as dependent variable)

Proxy		No. of papers	No. of estimates
Consumer side	Viewings	6	107
	Purchases	11	98
	Clicks on authorized websites	1	36
	Expenditure	3	7
Producer side	Sales	8	118
	Revenues	7	57
	Rank	1	3

Similarly large variety concerns the way "online piracy" was proxied or measured in the analyzed papers. Indeed, this variable remains largely unobservable. Indicators such as downloads or page visits may be informative for some sources only, while unauthorized distributors rarely make such data (e.g. popularity rankings or downloads) available even for research. Some studies thus assume that rather than actual act of "online piracy", a proxy should focus on the 'window of opportunity' which resulted in a large number of papers relying on broadband/internet connection, communicational proficiency of the user or access to means that can be used for "online pirating". These studies should be clearly separated from those, which used as proxies an institutional change, such as a change in law or a shutdown of a major source of unauthorized content, see Table 4. Based on a careful examination, we have constructed thirteen classifying these predicting variables. Means for "online piracy" signifies studies which used measures such as owning equipment (e.g. CD burners). Tech/Law/State change refers to studies of impact of a huge dimension in the market for pirated products. Those include: introductions of new laws or shutdowns of big piracy networks like Megaupload or Napster. Usage of piracy refers to person's tendency to use software associated with "piracy" (e.g. P2P software), or the general penetration rate of "piracy" associated networks (e.g. Megaupload). Finally, mixed refers to a combination of several proxies (e.g. interactions between other categories).

Table 4: Measures and proxies for "piracy" (as explaining variable)

Proxy	No. of papers	No. of estimates
Downloads	18	267
Spread of piracy	4	39
Clicks on unauthorized websites	1	36
Tech/Law/State change	5	32
Internet/tech proficiency	1	29
Supply	5	23

Summarizing, already from the literature sampling it is clear that the literature in the field has grown substantially both in numbers and in variety of implemented empirical strategies. The general summary of the results is far from conclusive - arguments in favor of both positive and negative effects of "online piracy" on sales can be found, on top of which a fairly large share of papers is inconclusive. The most research areas are film and music industry. In the reminder of this paper we carefully examine quantitatively the link between the characteristics of the studies and their conclusions.

4 Exploratory Meta-Analysis

This section provides an overview of the conclusions provided in the literature. While we discuss extensively the role of methods employed in section 5, we first show that even the bird view of the main findings is inconclusive. Arguably, the number of estimates and papers with negative results is larger than those with positive findings, but both are fairly common. We show that quantitatively a dominance of either of the outcomes is labile both over time and between industries. First, however, we discuss which papers could not be included in the meta-analysis and provide motivation for this selection.

4.1 Excluded papers and observations

During the course of our search we have encountered a number of studies related to our topic that could not, however, be included in our metaregression. There were also cases of studies that could be used but with some omissions. Those exclusions can be divided into five groups and are based on rules that we followed consistently.

First, we did not include any study where the authors did not interpret the results as the effect of file-sharing, or when the variables reflected more than just sales or "piracy". For example, while some authors used Internet penetration as a proxy or instrument for "piracy" (and interpreted it this way), some others specifically mentioned that they cannot and do not interpret the effect of Internet as a proxy for "piracy". Thus, we only included the papers where the authors themselves believed that their results reflected the effect of "piracy", even if the ones without such interpretation shared the methodology As another example, Zentner (2008) study the Internet's penetration effect on the survivability of music stores. However, this proxy for sales catches not only the demand but also the supply side. Moreover, the number of stores is a very different type of proxy to the others used in the literature (which all describe some form of direct consumption). [6]

Second, we always included the newest version of the study that we were able to find (if we found a working paper, we would then check if there were any newer – or published – versions of it). If there were more coefficients in an older version (consider Blackburn (2004) that had several more, intermediate, specifications than the newer Blackburn (2006) version), we did not take them into account.

Third, we did not include any article that had its results based on non-regression calculations. This led to exclusion of, mostly, reports prepared for specific institutions and industry representatives. Among those were e.g. Siwek (2006, 2007a,b).

Fourth, we did not include studies (or some of the coefficients from studies in our sample), where the model was too atypical to be included within our methodology. Whenever possible, we reported all the coefficients (when there were two or more piracy variables in one regression, we would treat each of them as a separate observation). However, in some cases, when the model's interpretation was based on e.g. two or more variables *simultaneously* (e.g. regressions with piracy interacted with popularity in Blackburn (2006)) or on the comparison between two specifications, there was no natural way to include those results within our framework, while keeping the authors' interpretation of the results.

Fifth, as mentioned earlier, we have found only one study related to video games' online piracy's effect on sales, see Fukugawa (2011). Since it was too little to include it in the analysis, we concentrated on the music and movie industries and left it out of our sample.

---

[6]Another issue in Zentner (2008) is the interpretation of the piracy proxies (internet penetration AND broadband penetration), which relies heavily on the study design, making it incomparable to other studies).

4.2 General tendencies in the literature

As discussed earlier, there is a number of reasons why the overall conclusion on the link between "online piracy" and sales could be perfectly consistent and yet differentiated across papers. First, music and film industry could react differently to "online piracy" while the fact that due to the technological limitations they have expanded in different periods could provide additional source of heterogeneity. Second, it is a fairly general phenomenon in economics that earlier studies are typically less concerned with the sample selection, omitted variable and endogeneity biases, whereas the later studies more often seek a clean causal identification strategy, which may happen at the expense of precision too. Finally, if positive and negative effects are indeed strong, it is possible that one of them accidentally dominates in a particular context. Thus, accounting for industry specificity, method and context of the study should reveal a pattern in the results. This is summarized in Table 5.

Table 5: Results by applied methodology, analyzed industry and paper status.

	No. of papers	No. of estimates	% Positive		% Negative	
			All	Best	All	Best
Methods						
OLS/probit/tobit	22	220	40%	34%	43%	44%
Panel OLS/probit/tobit	15	86	21%	23%	55%	55%
IV	12	110	46%	48%	23%	42%
PSM and others	3	10	10%	0%	70%	50%
Industries						
Movie (DVD or VHS sales)	6	19	11%	14%	47%	43%
Movie (Digital)	1	12	0%	0%	100%	100%
Movie (TV watching)	2	42	41%	29%	29%	21%
Movie (Theatre)	8	63	19%	8%	59%	77%
Movie (DVD or VHS rental)	3	13	8%	0%	39%	33%
Movie (Any)	2	34	21%	30%	50%	55%
Music (Album sales)	15	148	35%	56%	36%	30%
Music (Song sales)	6	59	61%	48%	39%	57%
Music (Any)	3	35	83%	43%	14%	52%
Paper Status						
Published	17	219	21%	20%	41%	50%
Working Paper	15	207	54%	51%	41%	44%

The image from Table 5 is partly blurred by the fact, that some of the papers present multiple specifications, of which only some are considered valid by the authors (e.g. authors first show a biased OLS estimator to subsequently move to interpreting the unbiased estimator from an IV specification). In response to this tendency, for all the studies in our metaregression sample we have identified regressions that could be considered 'best sets'. Some of the authors specifically pointed at regressions they considered to be appropriate for interpretation or that they favoured. For these studies the choice of best sets was straightforward. For studies that reported only one result (or several but for different samples/types of good), we marked all the coefficients as 'best'. For the reminder of the studies we employed the following procedure. We mark an estimate as 'best set' if:

- the author(s) followed a particular specification for the robustness checks (i.e. referred to a previous specification in the main study)

- the author(s) quoted, in the abstract or the conclusions section, the result(s) from particular specification(s),

- the author(s) quoted a particular result to compare effect sizes between robustness checks and previous specifications

- the author(s) performed a regression that they indicated will control for potential biases in previous regressions.

If the author(s) reported several regressions but made no indication on which they considered better, we have included them all as best sets. This procedure yielded 177 best set estimates: 91 for the film industry and 86 for the music.

Given how capital the endogeneity bias can be in the study of the effects of "online piracy" on sales, we place great emphasis on isolating studies which made provisions to address this issue. Table 5 reveals however, that it is not only the estimation method but also the design that may matter substantially for the final conclusions. While tenativey it seems to be a tendency that the results are more often negative if one employs propensity score matching, this conclusion no longer holds for studies which rely in instrumenting the endogenous variable. With the instrumental approach, the estimates are only as reliable as the instrument is powerful in providing the identification. Indeed, a large fraction of the papers proxies for "online piracy" with indicators such as broadband access, while some others dispose of a tentative measure of "online piracy" and instrument for it with broadband access. Comparing estimates from such two studies involves acconting not only for the measures used, but also for the method. In addition, in some cases of the panel estimates, one could say that the individual specificity may be exluded and if it were the main source of bias, fixed effect estimators could be considered more reliable.

5   The meta-regression

As argued above, the literature is abundant but far from conclusive. The objective of our analysis is to provide insights on the which data and methodological characteristics are more likely to yield a statistically significant result and which of which sign. Unfortunately, it is also fairly diversified, which necessitates caution in interpreting the results of any meta-regressions. To provide results we separate in the analyses the movie and music industries for two major reasons. First, it is likely that the technological progress in the online exchange of files has affected these two industries differently, which could imply that the estimated relationship would not be similar. Second, there is a substantial dispersion in how "online piracy" is proxied and analyzed for these two industries.

Intuitively, techniques robust to the endogeneity bias should be treated as more reliable. On the other hand, it is fairly typical that robust methods yield higher standard errors, less frequently giving statistically significant estimates. Thus, for the purpose of the meta-regression we identify if authors have acknowledged the hazard of the endogeneity bias (be it reverse causality, omitted variable problem or sample selection bias). Second, we also acknowledge if it was tackled and how. Finally, as suggested by the exploratory analysis, we also pay close attention to the proxies for "online piracy" and proxies for sales. We separate the sale data into revenue streams, number of units sold, declared purchases and direct consumption with the following premises. If "online piracy" serves the purpose of testing the product, there could be strong effects on revenues and sales, but much weaker on self-declared purchases and direct consumption. Conversely, if "online piracy" was directly substituting the legitimate distribution, there should be no difference between these measures: all statistically significant and negative. Similar reasoning holds for the proxies of "online piracy".

We also include all the standard controls. The explanatory variables include the year of publication as well as the period covered by the study. While these two variables are likely to be fairly strongly correlated, we include both because many of the studies focused on particular phenomena (such as closing of Megaupload), but were drafted, circulated and published at different distance from that event. We also control for the number of

variables used, number of observations in the study and the number of regressions in the study. Larger samples are more likely to provide more statistical power to the estimates, the opposite holds for the number of variable used. High number of regressions may be indicative of a large number of robustness checks and possibly also heterogeneity of the analyzed phenomena. We also distinguish explicitly between the cross-country or cross-product data, as opposed to individual level data.

The results for the movie industry are reported in Table 6, whereas the results for the music are reported in Table 7. We start from analyzing the time effects – separating the period of the data and the actual date of publication. We find no time effects for the music industry and weakly positive time effects for the film industry in both measures. Consequently, contrary to the intuition, it seems that studies exhibit no clear time pattern for music and somewhat more time pattern in the case of film. This result is confirmed by the time plots and the funnel plots, see Figures B.1-B.6 and Figures C.1-C.6, respectively, in the Appendix.

When it comes to the characteristics of the studies, the results obtained from larger samples and wider set of controls do not seem to be more powerful. This conclusion may come from the fact that there can be confusion between medium sample size studies, which may relate both to cross-country comparisons and small-scale experimental data. While the two types of the studies have similar sample size, they differ substantially in the ability to robustly address the causal link between "piracy" and sales. This is why we focus mostly in this study on the consequences of controlling for reverse causality and sample selection. In the case of music, the estimates from the studies which claim accounting for either of the two are likely to yield negative and significant coefficients, but this conclusion is not robust to the inclusion of outliers. Namely, if we control for the types of proxies for sales and "piracy" (our preferred specifications for methodological reasons discussed earlier), this effect disappears, while the estimates actually even change signs (admittedly, they are insignificant). In the case of the film industry, studies which claim controlling for sample selection are more likely to find a significant negative coefficient and this result seems fairly robust across specifications. While lower power of robust estimators is something to be expected, as we argued in Table 2, in fact also robust estimators give sometimes negative and sometimes positive effects even within film industry alone. Indeed, the effects for music must be driven by the outliers, as there is no direct effect on the estimates stemming from the measurement of sales nor "piracy". This is confirmed in the forest plots, see Figures D.7-D.12 in the Appendix.

Summarizing, even when we account for a variety of characteristics, the results do not seem to take consistently any particular side in the debate between the creative industries and the sharing communities. One would be tempted to expect the results which account for reverse causality to show a less negative effect, because the omitted variable effect (and the reinforcing pattern) stemming from popularity should bias the coefficients upwards. However, this is only weekly confirmed in the case of music industry and not at all the case for the movie industry.

Table 6: Metaregressions for the film industry

	All regressions			Only not worst sets			Only best sets		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Basic study characteristics									
Year of study	-0.07**	-0.18***	0.06	-0.07**	-0.20***	0.15**	-0.06*	-0.18***	0.16***
	(0.03)	(0.04)	(0.05)	(0.03)	(0.05)	(0.06)	(0.03)	(0.05)	(0.06)
Year published	0.04*	0.10***	-0.04	0.06**	0.14***	-0.07	0.06**	0.13***	-0.08**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
# variables	-0.00	-0.00	-0.01***	0.00	-0.00	0.00	0.00	0.00	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.02)
# observations	0.001*	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00
(in 1000s)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cinema based study			-0.02			0.01			0.01
			(0.08)			(0.10)			(0.10)
Macro data			-2.73***			-3.22***			-3.03***
			(0.42)			(0.51)			(0.51)
Reverse causality		0.14***	0.04		0.26	-0.47		0.16	-1.35
		(0.05)	(0.04)		(1.08)	(1.06)		(1.16)	(1.10)
Sample selection		-0.31***	-2.53***		-0.29***	-3.18***		-0.28***	-2.96***
		(0.09)	(0.41)		(0.10)	(0.50)		(0.10)	(0.50)
Sales measure:	*Other* proxies as reference level								
Purchases			-0.08			0.30			0.30
			(0.49)			(0.59)			(0.59)
Revenues			-0.01			-0.16			-0.08
			(0.17)			(0.19)			(0.18)
Consumption			-0.03			0.23			0.23
			(0.48)			(0.59)			(0.58)
Piracy measure:	*Other* proxies as reference level								
Downloads			-0.18			-0.32**			-0.33**
			(0.14)			(0.15)			(0.14)
Market change			-0.22			-0.32			-0.56**
			(0.23)			(0.26)			(0.26)
Constant	52.56**	160.64***	-41.37	19.00	113.43***	-160.80***	8.36	100.47**	-160.73***
	(22.29)	(35.18)	(43.06)	(27.11)	(41.35)	(53.30)	(28.73)	(42.36)	(50.88)
# coefficients	179	179	179	103	103	103	90	90	90

Notes: Estimates for the meta regression for all the estimates within all the reported coefficients of the effect of digital piracy on sales (obtained from `-metareg-` syntax in STATA). We drop one study with analysis based on 5 observations as it comprises an obvious outlier (both in terms of the data characteristics and its results). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 .

Table 7: Metaregressions for the music industry

	All regressions			Only not worst sets			Only best sets		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Basic study characteristics								
Year published	0.13***	0.05	0.07	0.13*	0.05	0.08	0.02	0.02	0.05
	(0.05)	(0.05)	(0.07)	(0.06)	(0.07)	(0.10)	(0.03)	(0.03)	(0.05)
Year of study	-0.04	0.09*	0.07	-0.03	0.11*	0.19**	-0.02	0.01	0.01
	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.09)	(0.02)	(0.03)	(0.06)
# variables	0.01	0.00	0.03*	0.02	-0.00	0.04*	0.01	0.00	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
# observations	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00
(in 1000s)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Macro data			-0.70			-1.75**			-0.25
			(0.62)			(0.88)			(0.67)
Reverse causality		-0.30**	0.09		-0.38**	-0.06		0.09	0.13
		(0.14)	(0.16)		(0.19)	(0.25)		(0.14)	(0.23)
Sample selection		-0.60***	-0.22		-0.68**	-0.56		-0.28	-0.24
		(0.22)	(0.32)		(0.34)	(0.47)		(0.17)	(0.28)
Sales measure:	*Other* proxies as reference levels								
Purchases			-0.78*			-1.52**			-0.14
			(0.43)			(0.63)			(0.39)
Clicks			-1.44**			-2.51***			
			(0.60)			(0.86)			
Sales			0.60			1.12*			0.47
			(0.51)			(0.67)			(0.52)
Piracy measure:	*Other* proxies as reference levels								
Downloads			0.02			0.24			-0.09
			(0.25)			(0.41)			(0.29)
Piracy spread			-0.65**			-0.17			-0.08
			(0.29)			(0.49)			(0.51)
Clicks									-0.41
									(0.53)
Constant	-180.12***	-280.75***	-281.58***	-197.54***	-328.93***	-534.00***	-16.99	-48.78	-126.17
	(49.74)	(63.61)	(86.36)	(62.13)	(91.99)	(148.57)	(43.98)	(43.34)	(108.63)
# coefficients	243	243	243	182	182	182	86	86	86

Notes: Estimates for the meta regression for all the estimates within all the reported coefficients of the effect of digital piracy on sales (obtained from `-metareg-` syntax in STATA). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 .

As discussed earlier, the results of different studies are not exactly comparable because diverse proxies, measures and identification strategies were used. The estimates of the effect of "online piracy" on sales show thus a considerable dispersion, which raises the hazard that estimates of the meta-regression are spurious. Meta-regression, being in fact a modification of plain OLS, automatically yields large weight to extreme values of the estimated (significant) coefficients reported by the studies. This necessitates caution in interpreting the results of our findings.

Another diversity in the literature concerns the fact that not all measures of the "piracy" intensity actually measure unauthorized distribution. For example, Peitz and Waelbroeck (2004); Rob and Waldfogel (2007); Hennig-Thurau et al. (2007); Bai and Waldfogel (2012) analyze self-reported information on downloading behavior, without specifying explicitly that they should come from an unauthorized distribution. The problem is probably less severe in the case of earlier works, but with the increasing availability of authorized streaming and sales service providers, differentiating between the types of the downloads seems crucial.

Secondly, the quality of intensity measurement may be highly country specific. For example, Danaher and Smith (2013) and Peukert et al. (2013) find the opposite effects for the shutdown of an unauthorized Megaupload service provider, but this disparity does not stem from the reliability of the identification. In fact, one of the paper uses data for 12 countries, and the other for as much as 50. The country specificity may stem from phenomena related to "piracy" – such as market maturity, quality of domestic cultural goods, knowledge of foreign languages, presence of international commercial authorized service providers, etc – but also from phenomena related to general consumption of cultural goods (such as average number of hours worked, cultural norms concerning leisure time, etc.). Identifying causal effects in the international context – even if econometrically the identification strategy is sound – will remain a challenge.

Finally, the third dimension of heterogeneity may come from the cultural products themselves. The diversification of the cultural goods encompasses dimensions such as highly imperfect substitution, preferences instable over time, numerous grounds for complementarity, presence of phenomena such as addiction consumption, etc. Last but not least, there is also a notion of popularity – relative or absolute. For example, Blackburn (2006) argues that unauthorized distribution helps the sales of less popular goods and does harm only to the highly popular ones.

6   Conclusions

There is a number of theoretical insights on how the authorized sales of cultural goods may be affected by the unauthorized distribution of their digital versions online. The representatives of the creative industries race in providing the dollar estimates of the losses due to "digital piracy", while the proponents of sharing argue that the decision to acquire and decision to purchase may be driven by different intrinsic motivations, thus breaking the link between the number of files exchanged (admittedly illegitimately) and the sales. The economic profession has taken a significant number of attempts to tackle this empirical issue, developing a variety of identification strategies. The objective of this study was to summarize these achievements and provide a synthetic overview of the overall conclusions from the studies so far.

Our focus in this study was on the reliable causal estimates of the consequences that "digital piracy" has on sales. Notably diverging estimates to be found in the literature were likely to be driven by the changing time patterns as well as the differentiated ability of respective studies to handle adequately the problem of causality. Our results suggest that there is no clear conclusion so far in the literature – even if we narrow the analysis

to what the authors considered the most reliable set of estimates (or least unreliable), the results reveal a weakly positive time trend for the film industry and no trend for the music industry. In the case of the former, the studies published later are somewhat more likely to show negative effects on sales, but this effect is neither large, nor robust to the set of controls included in the meta-regressions. The only clear pattern that emerges from these studies is that cross-country comparisons suggest a fairly robust negative link, but clearly causal identification is the weakest in this context.

There is a number of reasons why the literature may be so unequivocal after two decades of research. First, both sales and "digital piracy" belong to a group of fairly poorly measurable phenomena. While in the case of "piracy" this requires no further explanation, in the case of sales that may come as a surprise, since in most other fields of studies (i.e. for other industries) this type of data seems readily available. In the case of the creative industries, however, there is a variety of indicators – such as sales revenues, number of items sold or number of viewers are not necessarily similar to one another. Also, they would need to be available at product level for meaningful comparisons, which is rarely the case.

In addition to low data availability, the identification of the causal link remains a challenge. The literature has invoked either controlled experiments or quasi-natural experiments as the means to overcome this problem, due to the lack of proper instruments. Yet, in the case of controlled experiments most of the results rely on self-reported purchasing and downloading behavior, whereas the quasi-natural experiments usually suffer from the insufficient data breadth or quality.

The third – and possibly the most important reason – concerns the underlying intuitions concerning the links between unauthorized distribution and legitimate sales. The phenomenon of "digital piracy" seems to be highly complex in nature. Except for the obvious download piracy, there is also the case of upload piracy, leak piracy, potential piracy and the lag between the piracy and the observed sales. Especially this last point requires more specific treatment in the literature due to the specific attributes of the cultural goods. On the one hand, these are durable goods in a sense that a song or a movie usually does not need to be repurchased. Yet, if the first-time consumption occurred via television or radio, consumers may be actually more inclined to purchase an own copy. On the other hand, the arrival rate of new cultural goods may be compared to the FMCG industry, as the song or a movie purchased last week or last month may no longer be of interest for the consumer if a new one arrives (possibly by a different artist and a different producer). Thus, the cultural goods exhibit a joint set of difficulties of durable and non-durable goods, which all affect the consumer decision process and which all exhibit different reaction to the presence of unauthorized, virtually free distribution. Until research is able to disentangle these effects – possibly also by identifying the type of consumers or, better yet, the type of consumption – the results of the studies will naturally confound a variety of effects.

References

Aguiar, Luis and Bertin Martens, "Digital Music Consumption on the Internet: Evidence from Clickstream Data," 2013.

Andersen, Birgitte and Marion Frenz, "Dont blame the P2P file-sharers: the impact of free music downloads on the purchase of music CDs in Canada," *Journal of Evolutionary Economics*, October 2010, *20* (5), 715–740.

Bae, Sang Hoo and Jay Pil Choi, "A model of piracy," *Information Economics and Policy*, 2006, *18* (3), 303–320.

Bai, Jie and Joel Waldfogel, "Movie piracy and sales displacement in two samples of Chinese consumers," *Information Economics and Policy*, 2012, *24* (3–4), 187–196.

Bakos, Yannis, Erik Brynjolfsson, and Douglas Lichtman, "Shared Information Goods*," *The Journal of Law and Economics*, 1999, *42* (1), 117–156.

Balestrino, Alessandro, "It is a theft but not a crime," *European Journal of Political Economy*, 2008, *24* (2), 455–469.

Barker, George and Tim Maloney, "The Impact of Free Music Downloads on the Purchase of Music CDs in Canada," SSRN Scholarly Paper ID 2128054, Social Science Research Network, Rochester, NY August 2012.

Bender, Mark T. and Yongsheng Wang, "The Impact of Digital Piracy on Music Sales: A Cross-Country Analysis," *International Social Science Review*, June 2009, *84* (3/4), 157–170.

Bhattacharjee, Sudip, Ram D. Gopal, Kaveepan Lertwachara, James R. Marsden, and Rahul Telang, "The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts," *Management Science*, January 2007, *53* (9), 1359–1374.

Blackburn, David, "On-line piracy and recorded music sales," *Unpublished manuscript*, 2004.

_ , "The Heterogeneous Effects of Copying: The Case of Recorded Music," *Job Market Paper (Harvard Ph. D. Programme)*, 2006.

Boorstin, Eric S., "Music sales in the age of file sharing." PhD dissertation, Princeton University 2004.

Bounie, David, Marc Bourreau, and Patrick Waelbroeck, "Pirates or Explorers? Analysis of Music Consumption in French Graduate Schools," in "Annual Congress of the Society for Economic Research on Copyright Issues, Montreal. Available on http://www.serci.org/documents.html" 2007.

_ , **Patrick Waelbroeck, and Marc Bourreau**, "Piracy and the Demand for Films: Analysis of Piracy Behavior in French Universities," *SSRN eLibrary*, 2006.

chen Wang, Chia, Chin ta Chen, Shu chen Yang, and Cheng kiang Farn, "Pirate or Buy? The Moderating Effect of Idolatry," *Journal of Business Ethics*, November 2009, *90* (1), 81–93.

Chi, Wendy, *Does file sharing crowd out copyrighted goods? Evidence from the music recording industry*, November, 2008.

Conner, Kathleen Reavis and Richard P Rumelt, "Software piracy: an analysis of protection strategies," *Management Science*, 1991, *37* (2), 125–139.

Danaher, Brett and Joel Waldfogel, "Reel Piracy: The Effect of Online Film Piracy on International Box Office Sales," *SSRN eLibrary*, January 2012.

_ **and Michael D. Smith**, "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales," Technical Report ID 2229349, Social Science Research Network, Rochester, NY 2013.

_ **, Michael Smith, Rahul Telang, and Siwen Chen**, "The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France," SSRN Scholarly Paper ID 1989240, Social Science Research Network, Rochester, NY January 2012.

_ **, Samita Dhanasobhon, Michael D. Smith, and Rahul Telang**, "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," November 2010.

Dejean, Sylvain, "What Can We Learn from Empirical Studies About Piracy?," *CESifo Economic Studies*, 2009, *55* (2), 326–352.

Elberse, Anita, "Bye-bye bundles: The unbundling of music in digital channels," *Journal of Marketing*, 2010, *74* (3), 107–123.

Fukugawa, Nobuya, "How Serious is Piracy in the Videogame Industry?," SSRN Scholarly Paper ID 1828642, Social Science Research Network, Rochester, NY May 2011.

Gopal, Ram D, Sudip Bhattacharjee, and G Lawrence Sanders, "Do Artists Benefit from Online Music Sharing?*," *The Journal of Business*, 2006, *79* (3), 1503–1533.

Gopal, Ram D., Sudip Bhattacharjee, and G. Lawrence Sanders, "Do Artists Benefit from Online Music Sharing?," *The Journal of Business*, May 2006, *79* (3), 1503–1533.

Grassmuck, Volker Ralf, "Academic Studies on the Effect of File-Sharing on the Recorded Music Industry: A Literature Review," SSRN Scholarly Paper ID 1749579, Social Science Research Network 2010.

Halmenschlager, Christine and Patrick Waelbroeck, "Fighting Free with Free: Streaming vs. Piracy," Technical Report, Working paper 2012.

Hammond, Robert, "Profit Leak? Pre-Release File Sharing and the Music Industry," SSRN Scholarly Paper ID 2059356, Social Science Research Network, Rochester, NY May 2012.

Hennig-Thurau, T., V. Henning, and H. Sattler, "Consumer file sharing of motion pictures," *Journal of Marketing*, 2007, *71* (4), 1–18.

Hong, Seung-Hyun, "Measuring the Effect of Napster on Recorded Music Sales: Difference-in-Differences Estimates Under Compositional Changes," *Journal of Applied Econometrics*, 2013, *28* (2), 297324.

King, Stephen P and Ryan Lampe, "Network externalities, price discrimination and profitable piracy," *Information Economics and Policy*, 2003, *15* (3), 271–290.

Kobus, Martyna, Michal Krawczyk, and Piotr Milos, "Piracy as an ethical decision," *WNE UW Working Paper*, 2013.

Koh, Byungwan, B. P. S. Murthi, and Srinivasan Raghunathan, "Shifting Demand: Online Music Piracy, Physical Music Sales, and Digital Music Sales (Shifting Demand: Piracy, Physical & Digital Music Sales)," 2013.

Lee, Jonathan, "Private Filesharing Markets and Album Sales," 2009.

Leung, Tin Cheuk, "Music Piracy: Bad for Record Sales But Good for the iPod?," SSRN Scholarly Paper ID 2244111, Social Science Research Network, Rochester, NY June 2013.

Liang, Che-Yuan and Adrian Adermon, "Piracy, Music, and Movies: A Natural Experiment," SSRN Scholarly Paper ID 1752224, Social Science Research Network, Rochester, NY January 2011.

Liebowitz, Stan J., "Research Note: Testing File Sharing's Impact on Music Album Sales in Cities," *Management Science*, January 2008, *54* (4), 852–859.

Liebowitz, Stan J, "The Oberholzer-Gee/Strumpf File-Sharing Instrument Fails the Laugh Test," Working paper 576, Regulation 2.0 2010.

Liebowitz, Stan J. and Alejandro Zentner, "Clash of the Titans: Does Internet use reduce television viewing?," *Review of Economics and Statistics*, 2012, *94* (1), 234245.

Ma, Liye, Alan Montgomery, Param Singh, and Michael Smith, "The Effect of Pre-Release Movie Piracy on Box-Office Revenue," SSRN Scholarly Paper ID 1782924, Social Science Research Network, Rochester, NY March 2011.

Martikainen, Emmi, "Does File-Sharing Reduce DVD Sales?," SSRN Scholarly Paper ID 1742443, Social Science Research Network, Rochester, NY January 2011.

McKenzie, Jordi and W. D. Walls, "File-Sharing and Film Revenues: An Empirical Analysis," Working Paper 2013-14, University of Sydney, School of Economics 2013.

Michel, Norbert J., "The impact of digital file sharing on the music industry: An empirical analysis," *Topics in Economic Analysis & Policy*, 2006, *6* (1).

Mortimer, Julie Holland, Chris Nosko, and Alan Sorensen, "Supply responses to digital distribution: Recorded music and live performances," *Information Economics and Policy*, March 2012, *24* (1), 3–14.

ning Chen, Yeh and Ivan Png, "Information goods pricing and copyright enforcement: Welfare analysis," *Information Systems Research*, 2003, *14* (1), 107–123.

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

Peitz, M. and P. Waelbroeck, "The effect of internet piracy on CD sales: Cross-section evidence," 2004.

Peitz, Martin and Patrick Waelbroeck, "Piracy of digital products: A critical review of the theoretical literature," *Information Economics and Policy*, 2006, *18* (4), 449–476.

_ **and** _ , "Why the music industry may gain from free downloadingthe role of sampling," *International Journal of Industrial Organization*, 2006, *24* (5), 907–913.

Peukert, Christian, Jeorg Claussen, and Tobias Kretschmer, "Piracy and Movie Revenues: Evidence from Megaupload: A Tale of the Long Tail?," SSRN Scholarly Paper ID 2176246, Social Science Research Network 2013.

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

_ **and** _ , "Piracy on the Silver Screen," *The Journal of Industrial Economics*, 2007, *55* (3), 379–395.

Rochelandet, Fabrice and Fabrice Le Guel, "P2P music sharing networks: why the legal fight against copiers may be inefficient," *Review of Economic Research on Copyright Issues*, 2005, *2* (2), 6982.

Safner, Ryan, "Americas Most Wanted Movies: The Effect of the Top Pirated Movies on Box Office Revenues," 2013.

Siwek, Stephen, "The True Cost of Motion Picture Piracy to the U.S. Economy," Policy Report 186, Institute for Policy Innovation 2006.

Siwek, Stephen E., *The true cost of Copyright Industry Piracy to the US economy*, Institute for Policy Innovation, 2007.

_ , *The true cost of sound recording piracy to the US economy*, Institute for Policy Innovation, 2007.

Smith, Michael and Rahul Telang, "Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy," *MIS Quarterly*, 2009, *33* (2), 321–338.

_ **and** _ , "Assessing the Academic Literature Regarding the Impact of Media Piracy on Sales," SSRN Scholarly Paper ID 2132153, Social Science Research Network, Rochester, NY 2012.

Stevans, Lonnie K. and David N. Sessions, "An Empirical Investigation Into the Effect of Music Downloading on the Consumer Expenditure of Recorded Music: A Time Series Approach," *Journal of Consumer Policy*, September 2005, *28* (3), 311–324.

Tanaka, Tatsuo, "Does file sharing reduce music CD sales?: A case of Japan," IIR Working Paper 05-08, Institute of Innovation Research, Hitotsubashi University 2004.

Vany, Arthur De and W. Walls, "Estimating the Effects of Movie Piracy on Box-office Revenue," *Review of Industrial Organization*, 2007, *30* (4), 291–301.

Waldfogel, Joel, "Lost on the web: Does web distribution stimulate or depress television viewing?," *Information Economics and Policy*, June 2009, *21* (2), 158–168.

_ , "Music file sharing and sales displacement in the iTunes era," *Information Economics and Policy*, December 2010, *22* (4), 306–314.

Zentner, Alejandro, "Measuring the Effect of File Sharing on Music Purchases*," *The Journal of Law and Economics*, April 2006, *49* (1), 63–90.

_ , "Online sales, Internet use, file sharing, and the decline of retail music specialty stores," *Information Economics and Policy*, September 2008, *20* (3), 288–300.

_ , "Ten Years of File Sharing and Its Effect on International Physical and Digital Music Sales," SSRN Scholarly Paper ID 1724444, Social Science Research Network, Rochester, NY December 2010.

_ , "Measuring the Impact of File Sharing on the Movie Industry: An Empirical Analysis Using a Panel of Countries," SSRN Scholarly Paper ID 1792615, Social Science Research Network, Rochester, NY March 2011.

Appendix A. List of papers

Table 8: List of papers with regression estimates on the link between pirates and sales

Paper	No. of estimates	Relationship
Boorstin (2004)	5	Varying
Peitz and Waelbroeck (2004)	32	Negative
Tanaka (2004)	8	Neutral
Rochelandet and Le Guel (2005)	Not in metaregression	Negative
Stevans and Sessions (2005)	2	Negative
Blackburn (2006)	9	Varying
Bounie et al. (2006)	11	Varying
Gopal et al. (2006b)	Not in metaregression	Varying
Michel (2006)	3	Negative
Rob and Waldfogel (2006)	14	Negative
Zentner (2006)	5	Negative
Bhattacharjee et al. (2007)	Not in metaregression	Varying
Bounie et al. (2007)	41	Varying
De Vany and Walls (2007)	4	Negative
Hennig-Thurau et al. (2007)	18	Negative
Oberholzer-Gee and Strumpf (2007)	42	Neutral
Rob and Waldfogel (2007)	30	Negative
Chi (2008)	29	positive
Liebowitz (2008)	7	Negative
Zentner (2008)	29	Negative
Bender and Wang (2009)	1	Negative
Lee (2009)	1	Positive
Smith and Telang (2009)	3	Neutral
Waldfogel (2009)	54	Negative
Wang et al. (2009)	Not in metaregression	Varying
Andersen and Frenz (2010)	7	Neutral
Danaher et al. (2010)	Not in metaregression	Varying
Elberse (2010)	2	Negative
Waldfogel (2010)	20	Negative
Zentner (2010)	20	Negative
Zentner (2011)	92	Varying
Fukugawa (2011)	Not in metaregression	Neutral
Liang and Adermon (2011)	4	Varying
Ma et al. (2011)	10	Negative
Martikainen (2011)	2	Neutral
Bai and Waldfogel (2012)	28	Varying
Barker and Maloney (2012)	4	Negative
Danaher and Waldfogel (2012)	54	Negative
Hammond (2012)	28	Positive
Liebowitz and Zentner (2012)	13	Varying
Mortimer et al. (2012)	Not in metaregression	Varying
Aguiar and Martens (2013)	36	Positive
Danaher and Smith (2013)	12	Negative
Danaher et al. (2012)	Not in metaregression	Negative
Hong (2013)	2	Negative
Koh et al. (2013)	Not in metaregression	Varying
Leung (2013)	3	Varying
McKenzie and Walls (2013)	25	Negative
Peukert et al. (2013)	11	Varying
Safner (2013)	16	Negative

Appendix B. Coefficients over time

Figure B.1: Number of significant coefficients, all reported coefficients

Graphs by industry
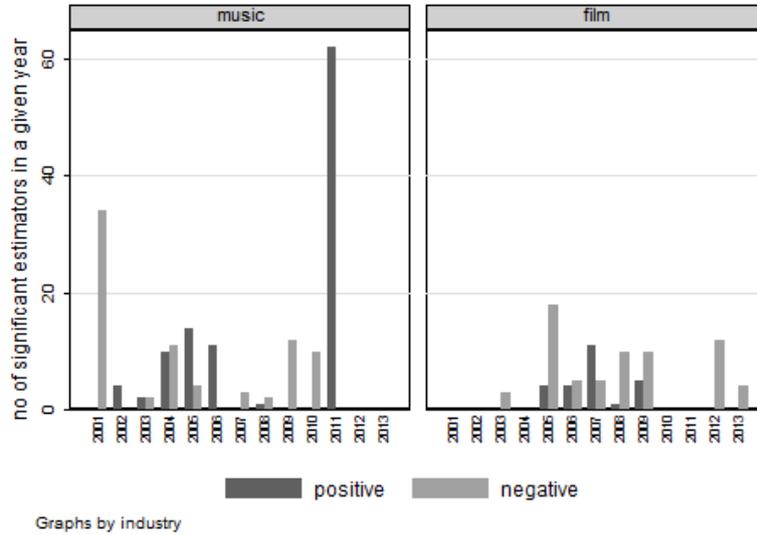
Figure B.2: Share of significant coefficients, all reported coefficients
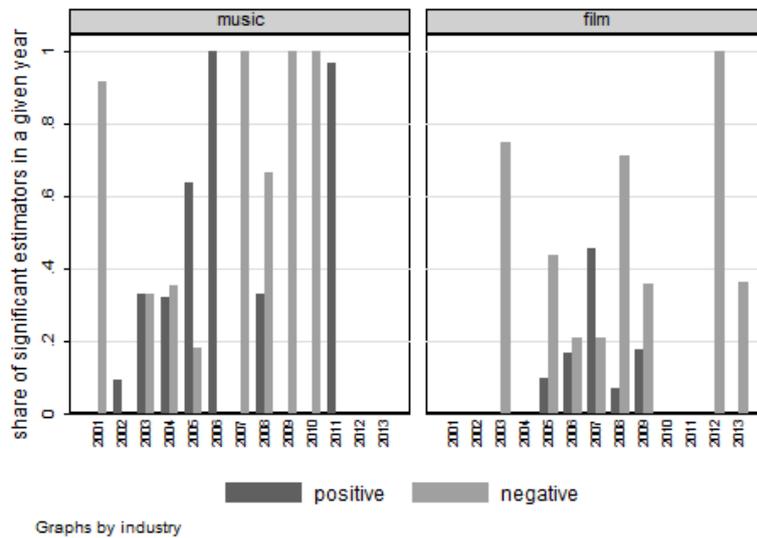
Graphs by industry

Note: Yearly values calculated for all studies, time defined by year encompassed the particular study, unweighted average. Positive coefficient signifies complementarity between "digital piracy" and sales, negative coefficient is consistent with substitution.

Figure B.3: Number of significant coefficients, 'worst set' coefficients excluded
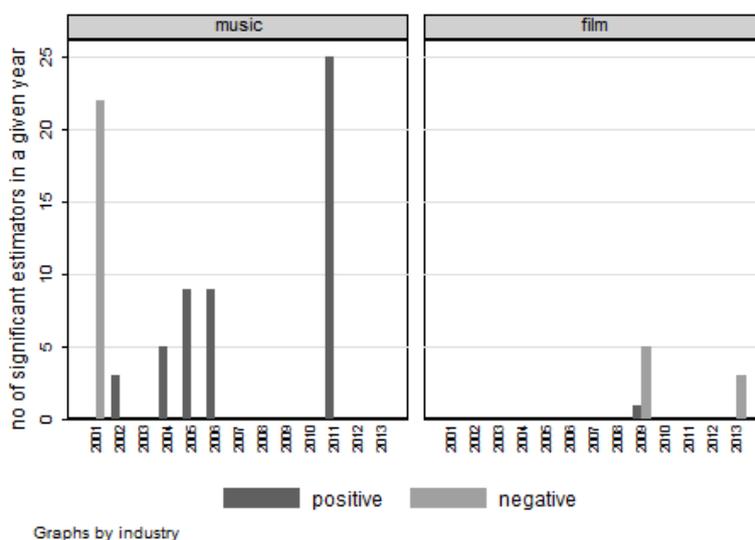
Figure B.4: Share of significant coefficients, 'worst set' coefficients excluded
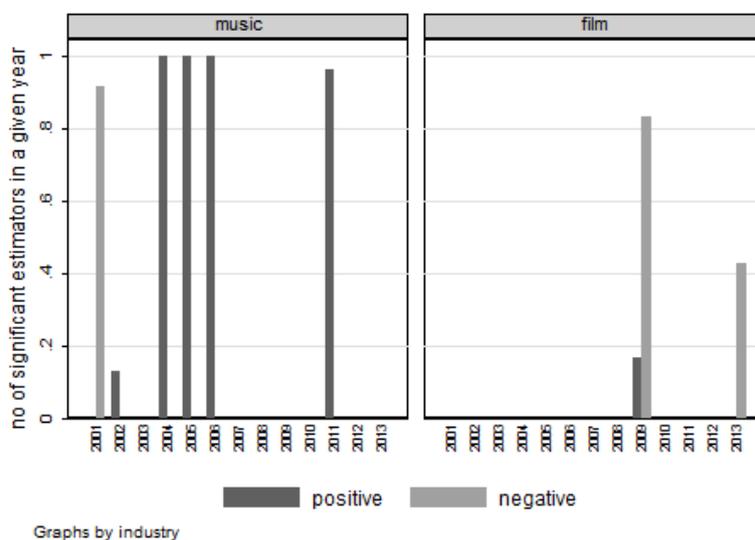
Note: Yearly values calculated for all studies, time defined by year encompassed the particular study, unweighted average. Author's evaluation used to define 'worst set' coefficients to be excluded. Positive coefficient signifies complementarity between "digital piracy" and sales, negative coefficient is consistent with substitution.

Figure B.5: Number of significant coefficients, only 'best set' coefficients

Graphs by industry
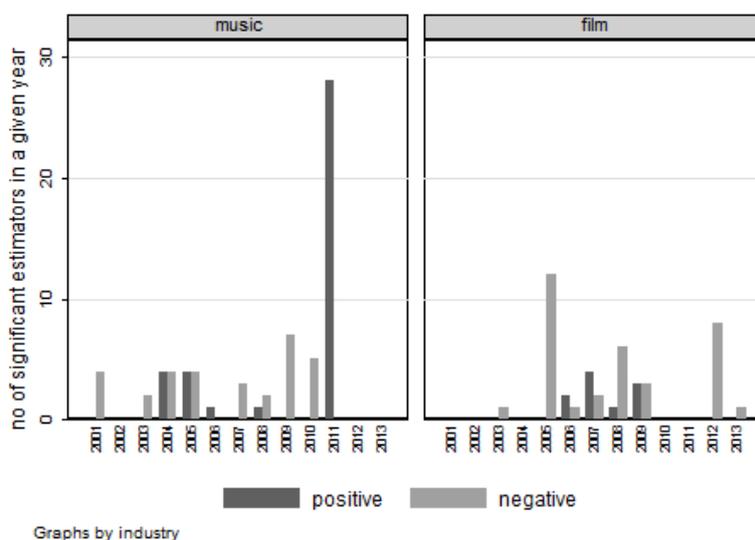
Figure B.6: Share of significant coefficients, only 'best set' coefficients excluded
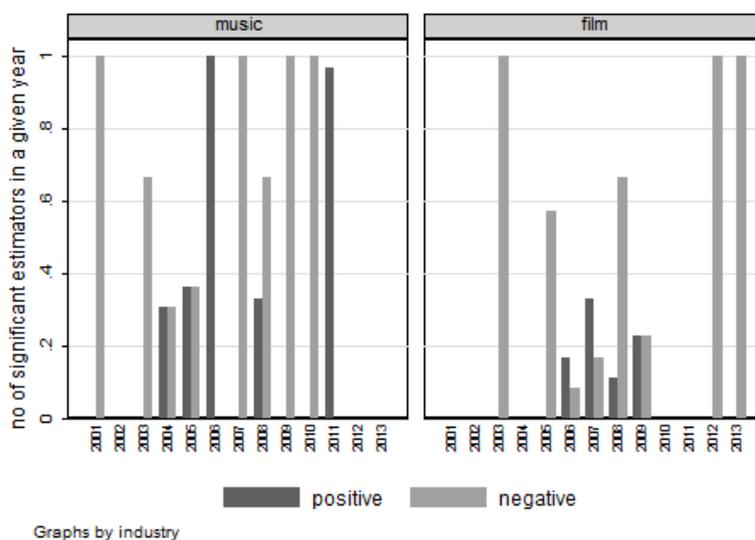
Graphs by industry

Note: Yearly values calculated for all studies, time defined by year encompassed the particular study, unweighted average. Author's evaluation used to define 'best set' coefficients to be included. Positive coefficient signifies complementarity between "digital piracy" and sales, negative coefficient is consistent with substitution.

Appendix C. Funnel plots for film and music studies

Figure C.1: Funnel plot for film studies, all coefficients
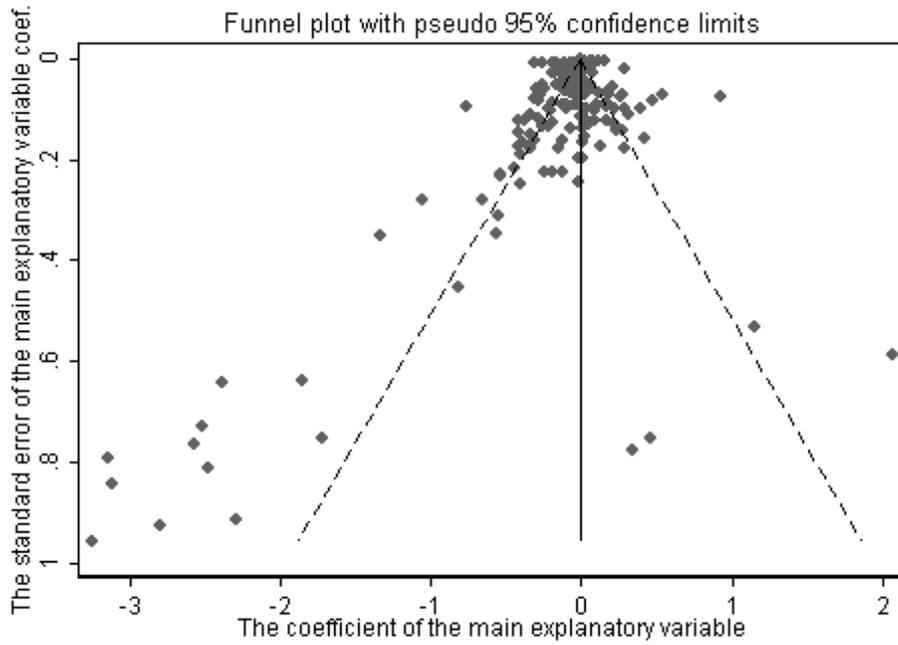
Figure C.2: Funnel plot for film studies, 'worst' coefficients excluded
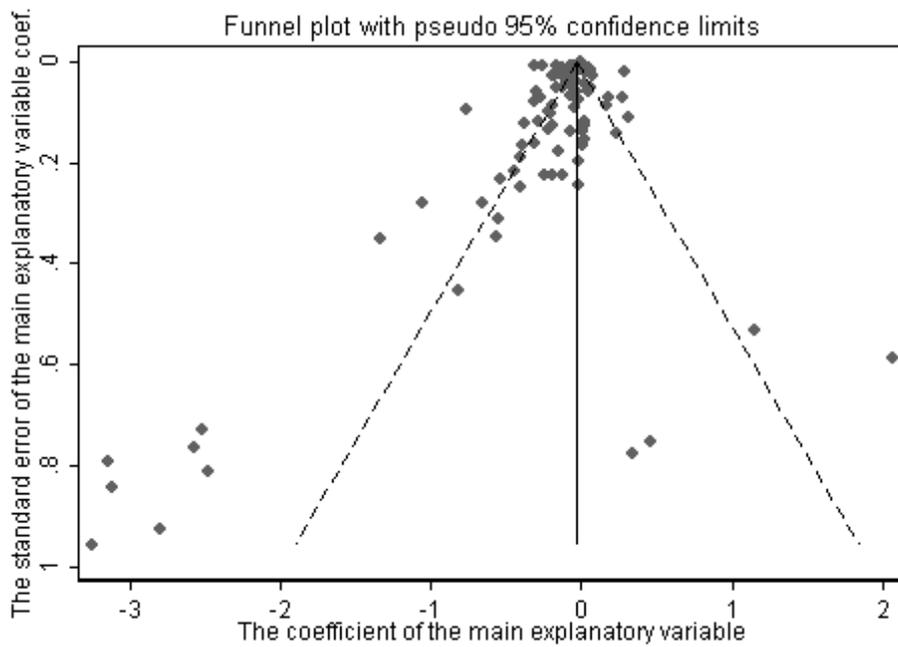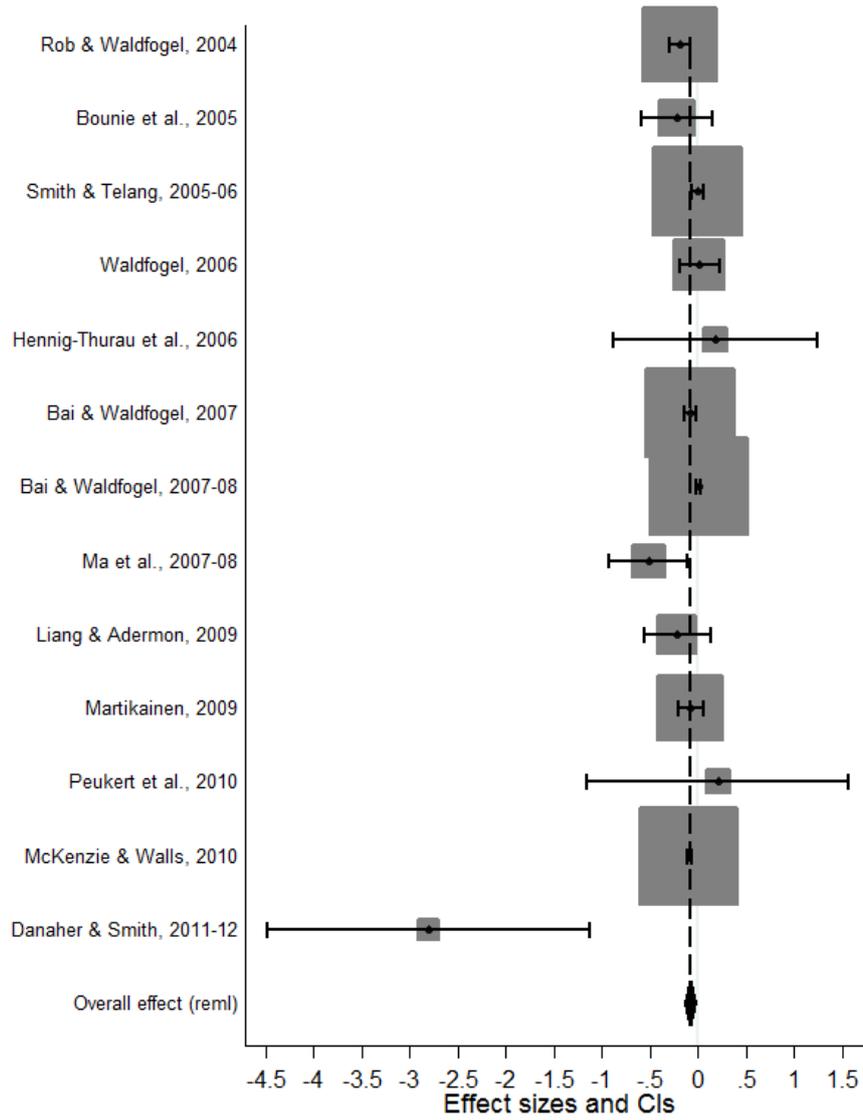
Figure C.3: Funnel plot for film studies, best set coefficients

Funnel plot with pseudo 95% confidence limits
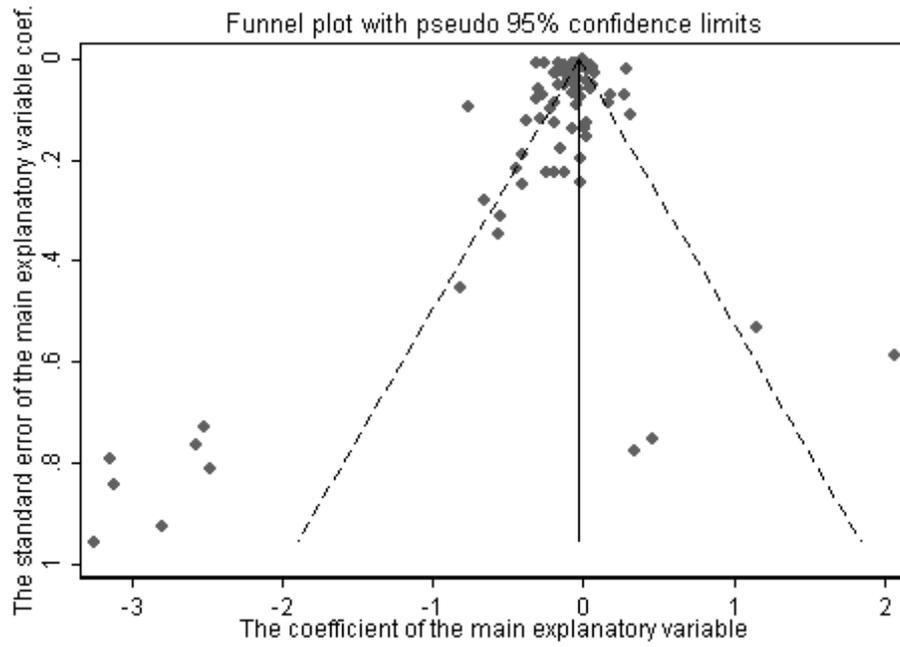
Figure C.4: Funnel plot for music studies, all coefficients

Funnel plot with pseudo 95% confidence limits
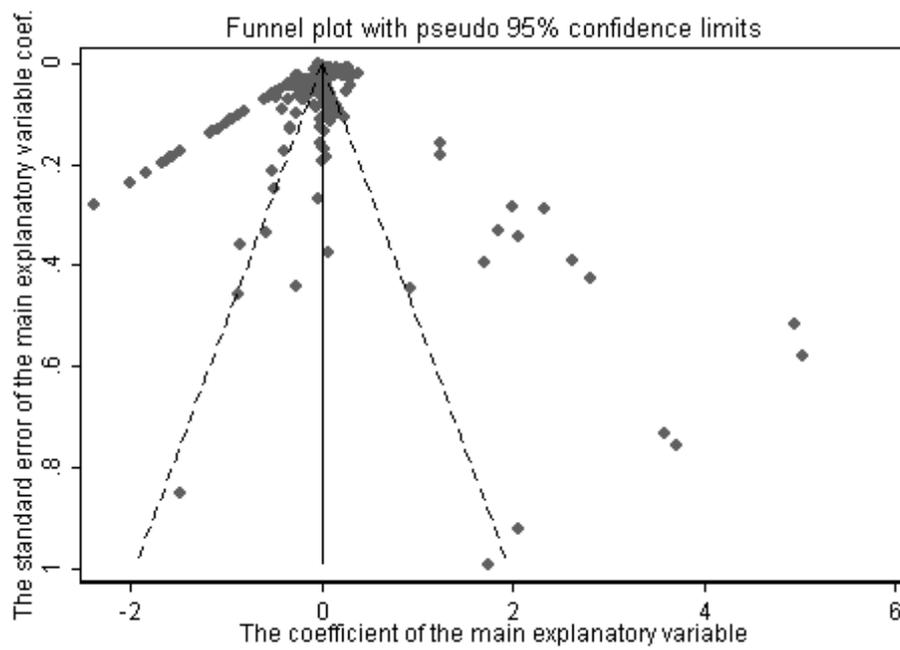
Figure C.5: Funnel plot for music studies, 'worst' coefficients excluded
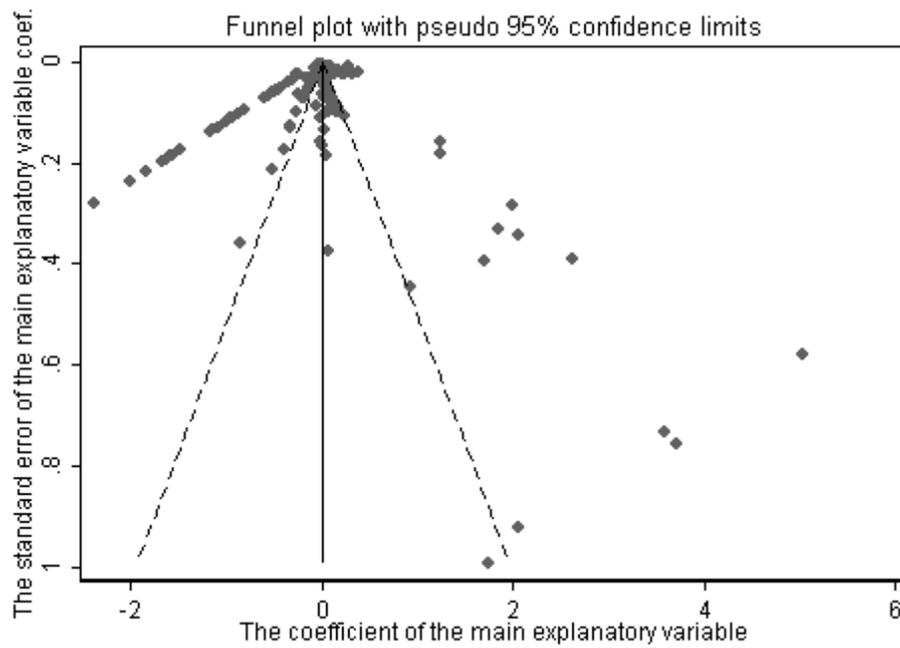
Figure C.6: Funnel plot for music studies, best set coefficients
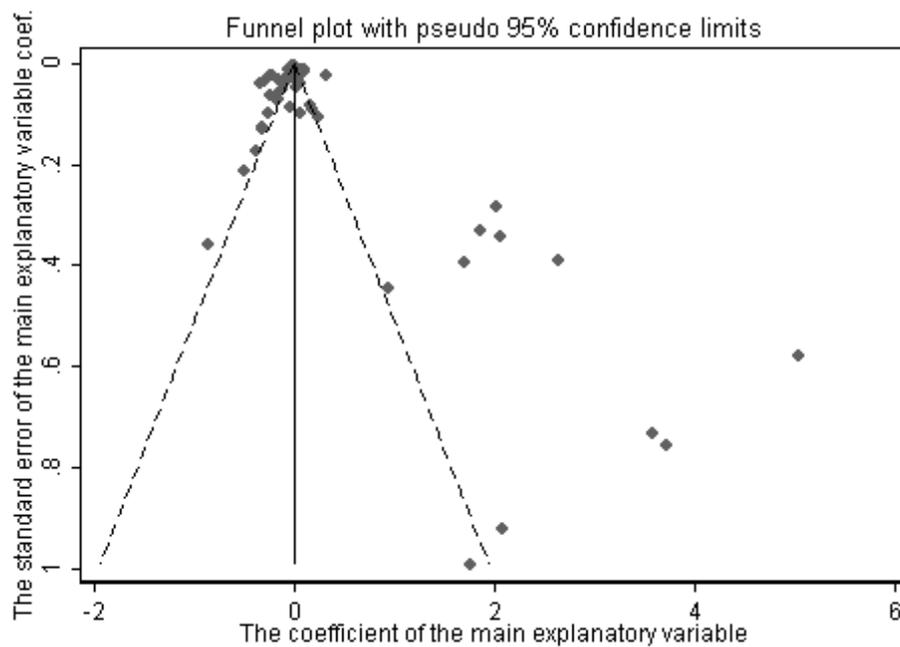
Appendix D. Forest plots for film and music studies

Figure D.7: Forest plot for film studies, all coefficients
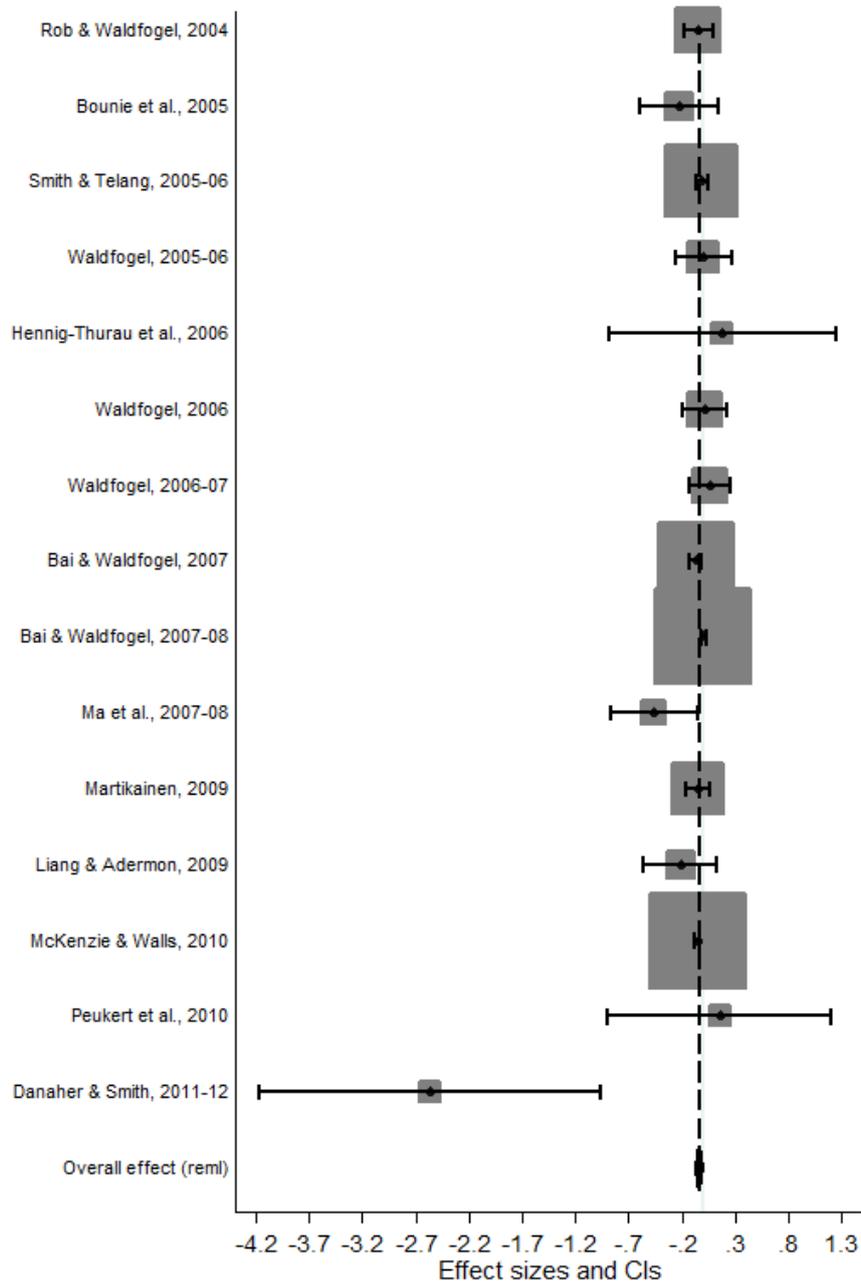
Note: All coefficients, unweight average.

Figure D.8: Forest plot for film studies, 'worst' coefficients excluded

Note: 'Worst' coefficients excluded, unweight average.

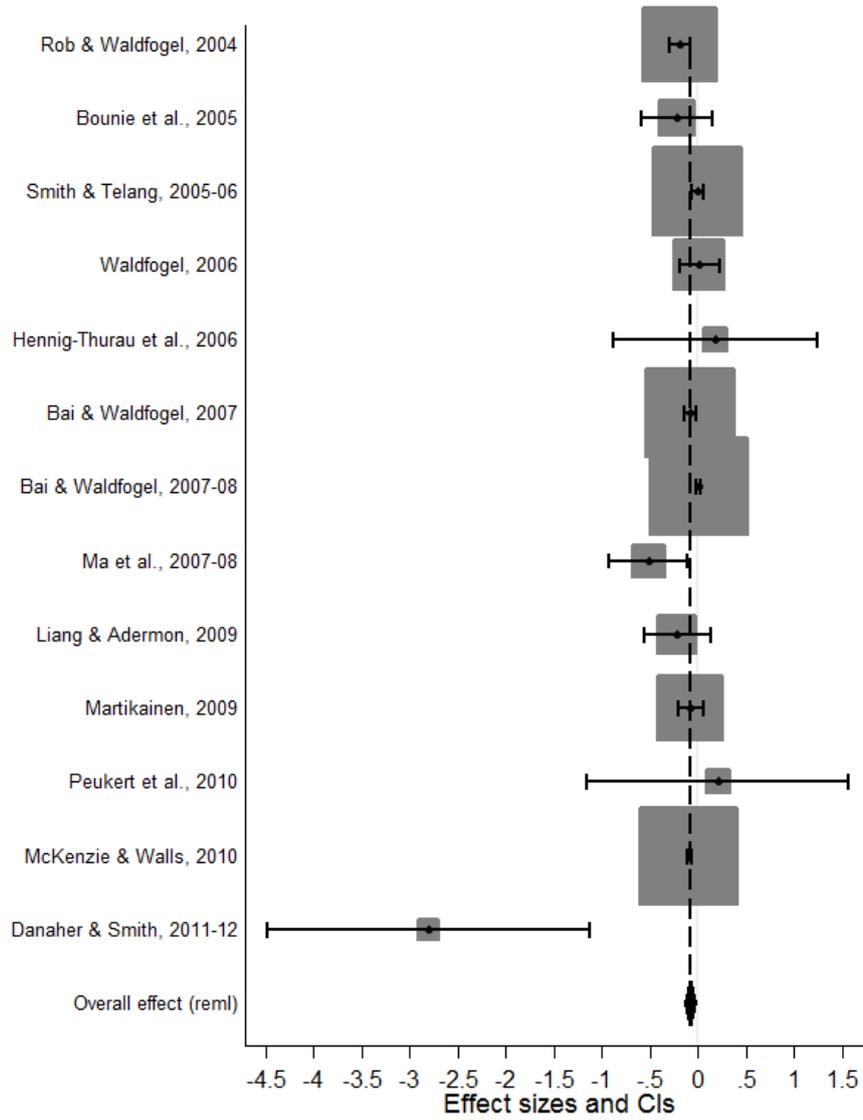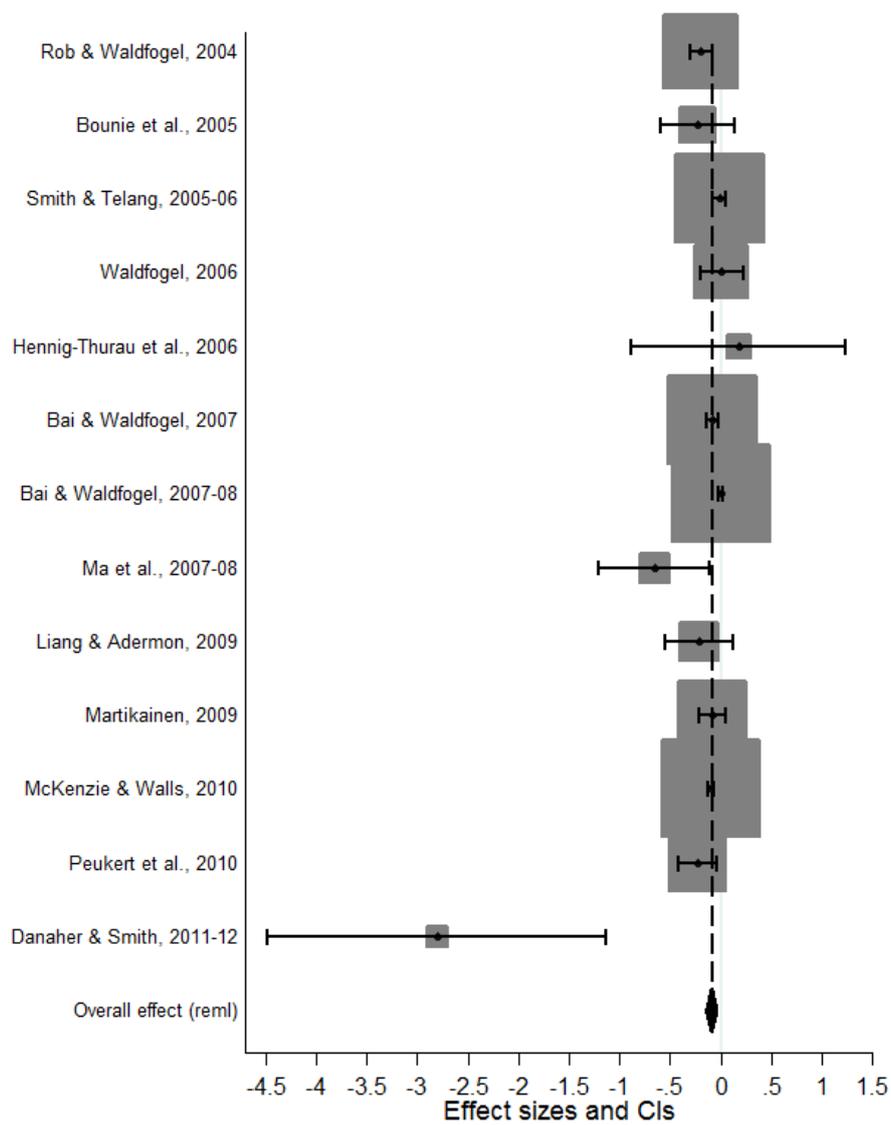Figure D.9: Forest plot for film studies, best set coefficients

Figure D.10: Forest plot for music studies, all coefficients
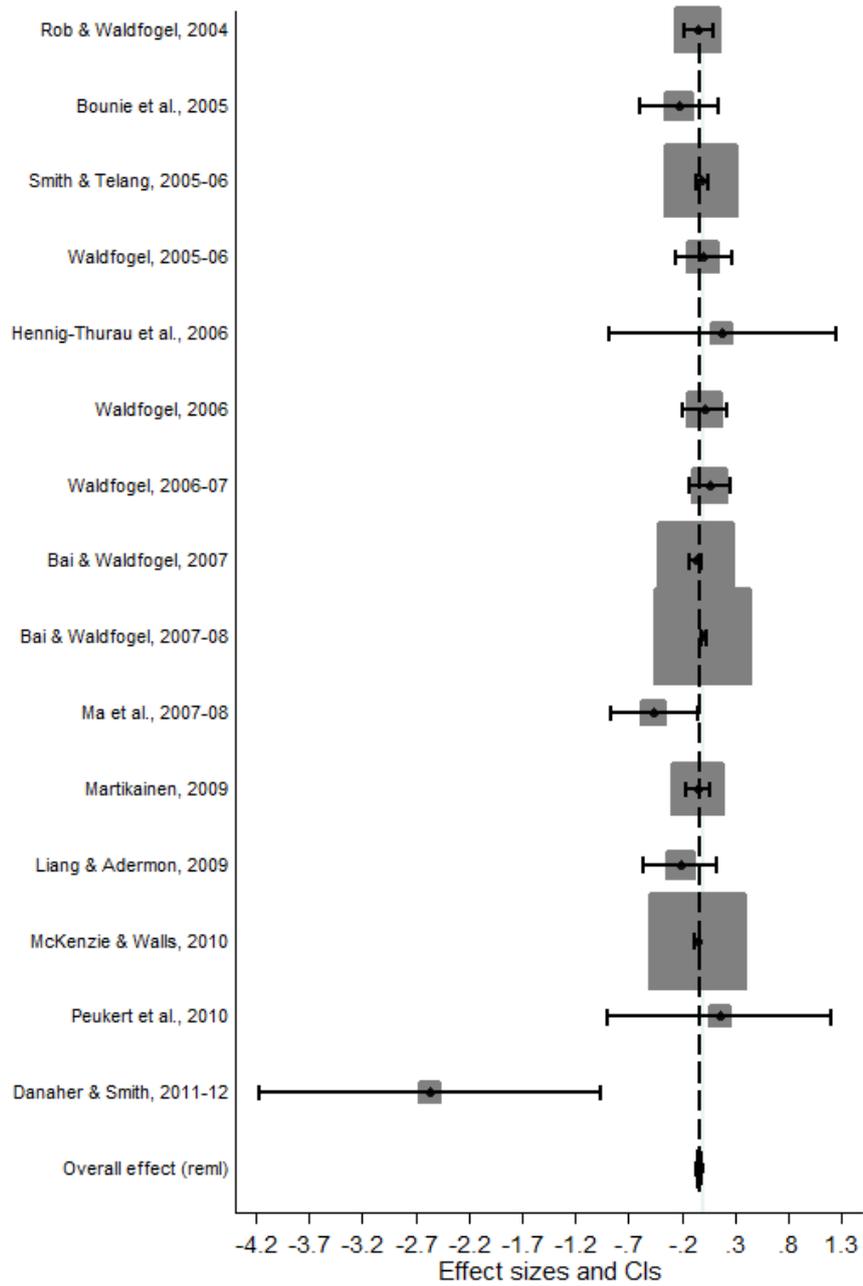
Note: All coefficients, unweight average.

Figure D.11: Forest plot for music studies, 'worst' coefficients excluded

Note: 'Worst' coefficients excluded, unweight average.

Figure D.12: Forest plot for music studies, best set coefficients

Note: Only best set coefficients, unweight average.