Friends or foes?

A meta-analysis of the link between online piracy and the sales of cultural goods

Wojciech Hardy - Michal Krawczyk - Joanna Tyrowicz

MAER-Net, Prague

September, 2015

Motivation

Depending on who you ask

- ''Online piracy" is good for sales (activists opposing copyright)
- ''Online piracy" is bad for sales (copyright owners)

The debate is hot, but where is actually science?

Motivation

Most known scientific result? App. 1000 citations on ScholarGoogle

- ''The effect of file sharing on record sales: An empirical analysis'', F. Oberholzer-Gee, K. Strumpf, JPE, 2007 – for foes
- Not a single paper for friends that would exceed 100 citations, including the critique by Liebowitz (2010, could not publish, conflict with J. List, etc).

Original idea

- Maybe the problem is in developing good causal analysis?

Spoilers

- This is not a story about good economics vs. bad economics

Theoretical underpinnings

Friends

- "Pirated" copy is not a perfect substitute
- Consumers are heterogenous in terms of preferences.
- Consumers have "morals", but they also have uncertainty about "quality".
- Network effects.

Foes

- "Pirated" copy is a perfect substitute for the original good.

BUT: Consumption at $price = 0$ cannot be compared to consumption with a downward sloping demand curve at any $price > 0$.

Surveying the field

Starting point: grab everything we can, evaluate, run regressions

The process of data collection

1. EconLit with keywords (digital/online/music/film/-)"piracy" and "displacement/sales/revenue/box office".
2. Same on GoogleScholar (first 300 hits)
3. Inspect their bibliography + Smith and Telang (2012) + Dejean (2009) + Grassmuck (2010)
4. Restricted papers in English : 72 papers.
5. Restricted to empirical: 44 papers
   - 426 estimates
   - 26 published papers and 18 WPs

Overview of the literature

Year of publication	No. of published articles	No. of working papers	No. of estimates	Conclusions Film	Conclusions 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

The literature is far from conclusive

Year of publication	Film industry			Music industry		
	Neg.	Inconc.	Pos.	Neg.	Inconc.	Pos.
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

Caveat 1 - how to measure sales

Table: 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

Caveat 2 - how to measure piracy

Table: Measures and proxies for "piracy" (as independent variables)

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

Caveat 3 - how can we tell if results are not spurious

Follow the author(s)

Best set

Worst set

Results

Highlights

1. Little support for either of the camps, conclusions do not depend on (how robust is the) method
2. There seems to be a negative time trend for films and somewhat positive for music $\rightarrow$ the role of technology? or cohorts?

Putting evidence together: "meta-regression" for film

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 .

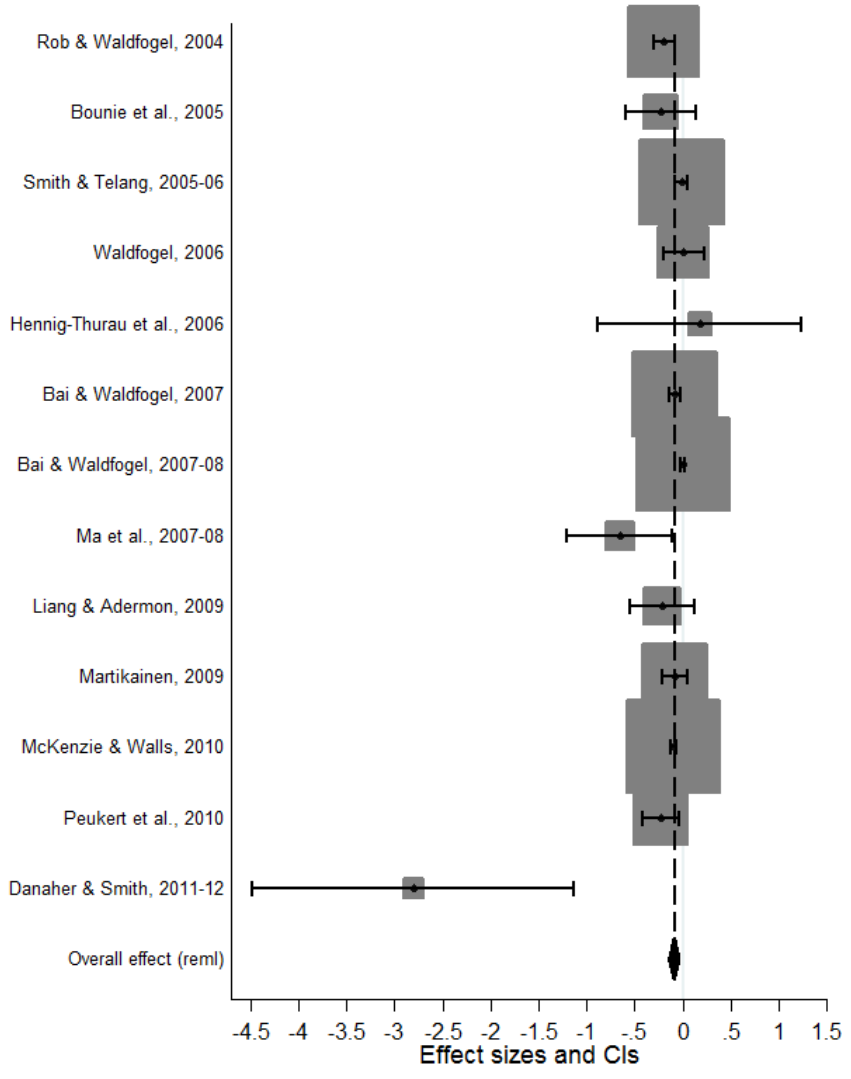
Putting evidence together: "meta-regression" for music

Table 7: Metaregressions for the music industry

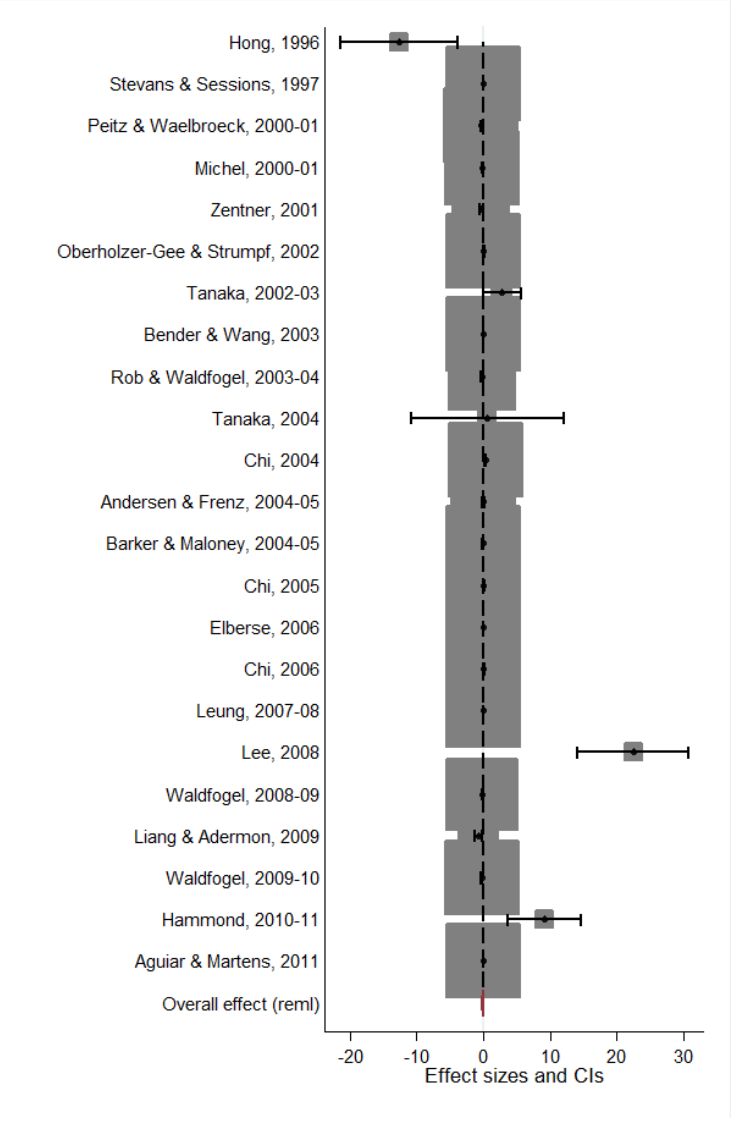
	All regressions			Only not worst sets			Only best sets		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Basic study characteristics									
Year published	0.13***	0.05	0.07	0.13*	0.05	0.08	0.02	0.02	0.05
	(0.05)	(0.05)	(0.07)	(0.06)	(0.07)	(0.10)	(0.03)	(0.03)	(0.05)
Year of study	-0.04	0.09*	0.07	-0.03	0.11*	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 .

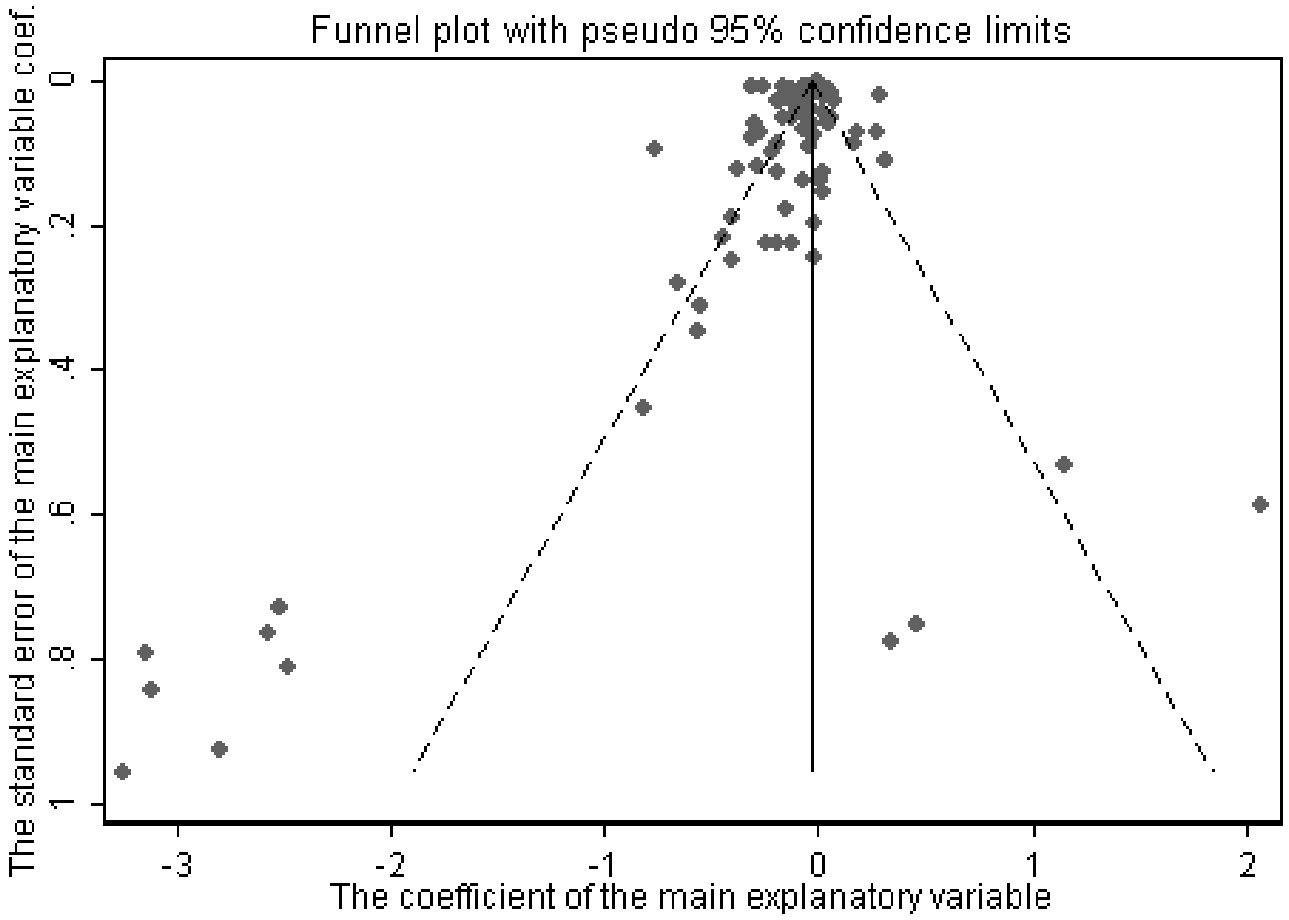
Forest plot for film studies, best set coefficients



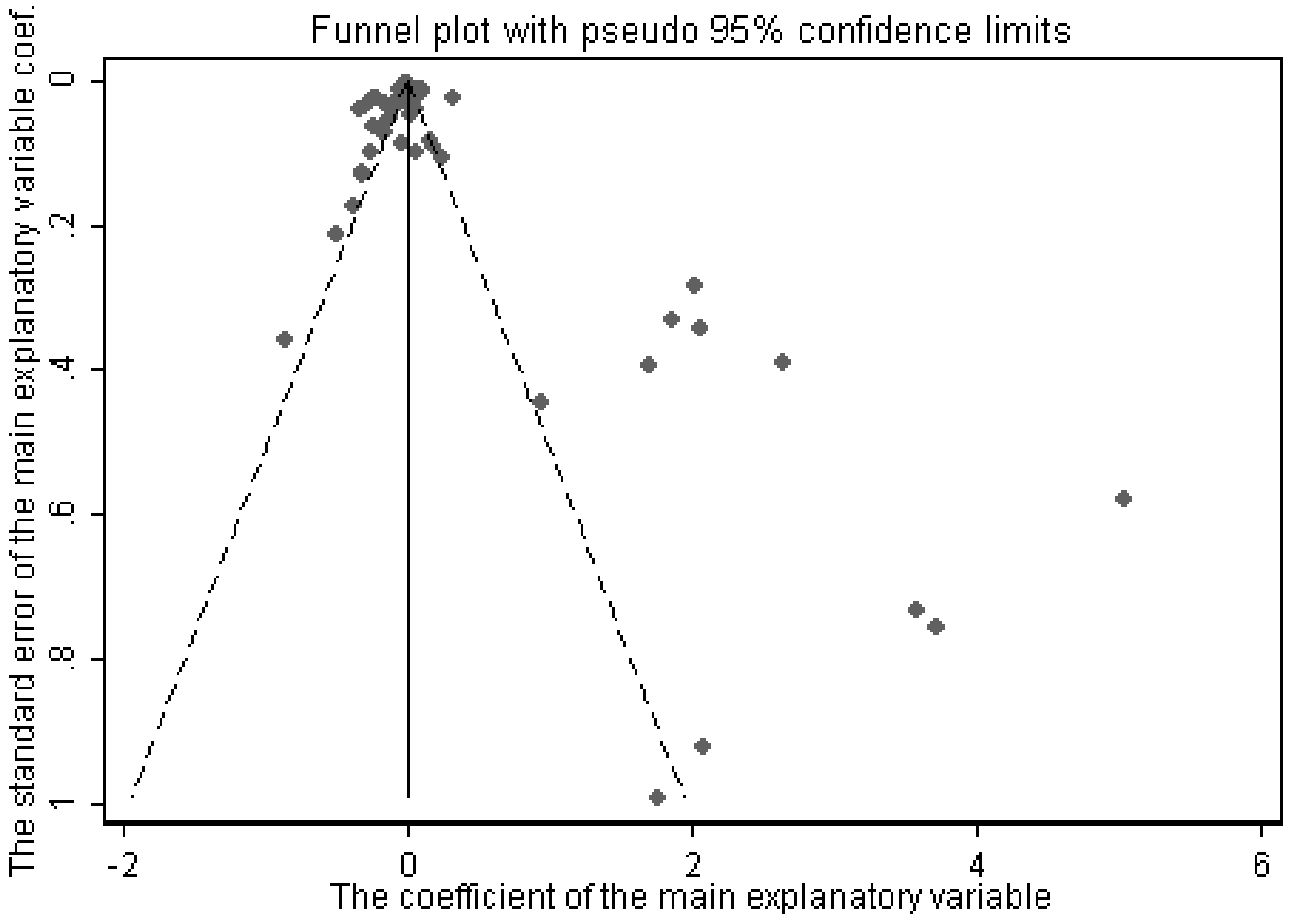
Forest plot for music studies, best set coefficients



Funnel plot for film studies, best set coefficients



Funnel plot for music studies, best set coefficients



Conclusions

1. The field is growing
   - Except for RIAA reports, now nearly everybody takes the effort to actually have an identification strategy, which drives the significance down
   - Literature is very far away from consensus
   - Data (un)availability an important constraint for growth
   - Conceptualization of consumer and consumer choice may be the problem behind the null result in a meta-analysis

2. Music seems to behave differently from film
   - Technology development affected them at different rates
   - Methods of consumption (and thus preferences) differ substantially

Comments or suggestions?