

The Size and Value of the Public Domain

Dr Rufus Pollock

University of Cambridge
The EU Public Domain Project

COMMUNIA, 2010-02-01

Background

DG InfoSoc Funded Project

Led by Rightscom - Many Contributing
Researchers

I'm the Lead Economist

My Focus

Several Strands, My Focus:

Quantitative Research

on the

'Formal' Public Domain

Size

What We Do (and Why It's Hard)

- Qu: How many items/works in the PD?
 - Item/work distinction: editions of Hamlet vs. original
- Get a list of works/items – e.g. library catalogues
- Unfortunately: often lack death dates
- Need to merge with other DBs, do extrapolation work
 - Not simple: authors don't have standard names
 - These are big DBs – millions of items

Books

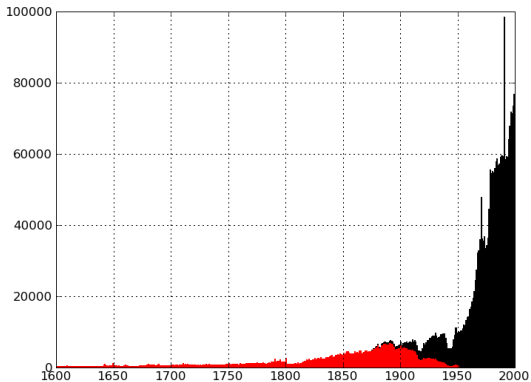


Figure: UK Published Books and Public Domain (Red) (CUL)

Books

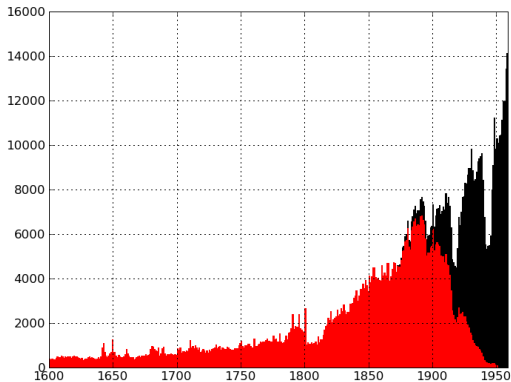


Figure: UK Published Books and Public Domain (Red) (CUL)

Books: Numbers

EU Wide: ~ 10m PD Books

10-20% of all Published Books

Country	Total	Total PD	Percentage PD
Bulgaria	549589	25125	5.0
Czech Rep.	1560000	85542	5.0
Finland	836141	68059	8.0
Italy	6000000	238648	4.0
Ireland	2954000	221116	7.0
Poland	2685169	379800	14.0
Slovakia	338529	17472	5.0
UK (BL)	10561445	2015050	19.0
Total	25484873	3050812	12.0

Table: PD Books Based on National Library Holdings

Recordings

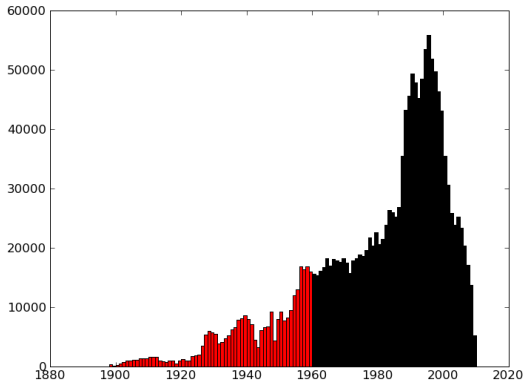


Figure: UK Recordings (BLSA). Red shows PD.

Recordings: Next Decades/Loss from Extension

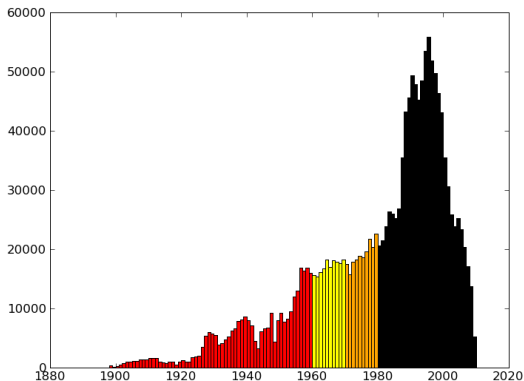


Figure: UK Recordings (BLSA). PD is red, yellow = PD in next decade, orange = PD following decade

Recordings: Next Decades/Extension

	Count	%
To Present	1671370	
In PD	301668	18.0
Entering PD next 10y	171155	10.2
Entering PD next 10y after	191308	11.4

Table: PD Recordings entering PD (no term extension)

Books: Next Decades/Extension

	Current		Next 10		Next 20		Last 10		Last 20	
Items	% PD	No. PD	% PD	No. PD	% PD	No. PD	% PD	No. PD	% PD	No. PD
3.5m	19	0.65m	21	0.73m	23	0.81m	17	0.58m	15	0.52m

Table: Change in Number of Public Domain Books (CUL, UK)

Price and Usage

What's Needed

- We need 'real' price/sales/usage data
- But we also need to classify on PD status
- This is HARD!
 - Price/sales DBs have poor author info (of the kind we need)
 - \Rightarrow classify PD status by hand
 - But need a lot of data to get power ...
- Books: Nielsen
 - 64k titles
 - UK price and sales ...

Usage

Nielsen

Public Domain?	No. of Items	%	No. of Items (Pbk)	%
N	56857	88.95	39605	93.21
Y	3868	6.05	1873	4.41
Y?	235	0.37	56	0.13
YN	12	0.02	0	0.0
YT	2864	4.48	940	2.21
YT?	80	0.13	17	0.04
Unknown	1	0.0	1	0.0
Total	63917	100	42492	100
PD (Broad)	7059	11.04	2886	6.79

Table: PD Status Counts for Nielsen Dataset. N=No, Y=Yes, YT=Translation of PD work (may or may not be PD itself), YN=PD work with notes, '?' indicates there may be some uncertainty.

Summary

Area	Source	PD (%)	Comments
Paperback Book Editions	Nielsen	4.4-6.8	'Pure' PD 4.4, 'broad' PD 6.8. Including unclassified formats raised figure to 6 (11 for broad)
Paperback Book Sales	Nielsen	2-4	'Pure' PD: 2, 'broad' PD 50%-100% greater
Book Borrowing	Slovak National Library	5	
Book Borrowing	Trinity College Dublin Library	3-8	24% relative to pre-1960 books (total to present likely to 4-8 times as much)
Music on Radio and TV	YLE (Finland)	9	
Musical works	SIAE (Italy)	2	Likely to be a partial underestimate (figures up to 8% possible)
Other works (esp. Opera)	SIAE (Italy)	6-27	
Recordings Revenue	SCPP (France)	<1%	Likely that public domain revenues are significantly underestimated

Usage: A Figure

Crudely: \sim 5-10%

Very Approximate!

Big issues with measuring PD usage ...

Issues

- PD items are PD! (So not measured ...)
 - E.g. SCPP
- What is usage?
 - Downloads? E.g. Project Gutenberg has 2.4m monthly
 - Library borrowing?
 - Purchase?
 - Is usage w/o value meaningful?
- What about the wider PD? (Of ideas etc)

Price

Price: Books

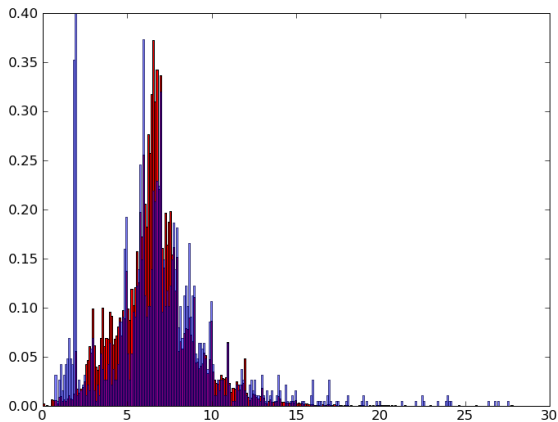


Figure: ASP distbn for PD (blue) and copyright (red) books

Price: Books

£2 PD book versus £6 Harry Potter

66% reduction!

But much diluted in the average: ...

Price: Books

Price difference: 15% RRP, 6.3% ASP (v. stat significant)

Sales weighted: 14% (PD £4.96 v. £5.65 ©)

RRP Limit 10						
	No. Items	Mean	SD	Median	Min	Max
Copyright	31431	7.22	1.57	6.99	0.0	10.0
PD	1618	6.28	2.55	6.99	0.95	10.0
All	33049	7.17	1.65	6.99	0.0	10.0

ASP Limit 10						
	No. Items	Mean	SD	Median	Min	Max
Copyright	36433	6.23	1.79	6.45	0.0	10.0
PD	1683	5.86	2.41	6.31	0.75	9.99
All	38116	6.22	1.82	6.45	0.0	10.0

Table: Price comparison with truncation.

Issues

- Many!
 - Most around unobserved heterogeneity
- Comparing like for like in cross-section
 - Increasing dispersion in PD
 - Works versus items
 - In general \Rightarrow underestimate price impact
- Longitudinal is better (but limited in doing this)
- What about digital
 - Look at ebooks ...

Price: Recordings

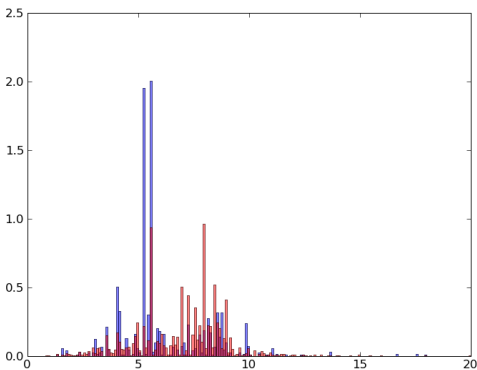


Figure: Price distbn for CDs, PD (blue, released < 1958) and in-copyright (red, rel > 1960)

Price: Recordings

PD 14% cheaper than in-copyright

(t-stat: -12.0, p-value $< e^{-16}$!)

Pre-1959 v. Post-1959						
	No. Items	Mean	SD	Median	Min	Max
PD	1226	6.16347471452	2.28046294735	5.55	1.38	34.99
Not PD	131144	7.032425883	2.46341819619	7.29	0.09	39.99

Table: Prices for PD and in-copyright recordings

Impact Assessment for Term Extension was Wrong!

Price: A Figure

Crudely: \sim 5-15% for means

But could be much more: 50%+ and up to 100%
for ebooks ...

Value

The Demand Curve Revisited

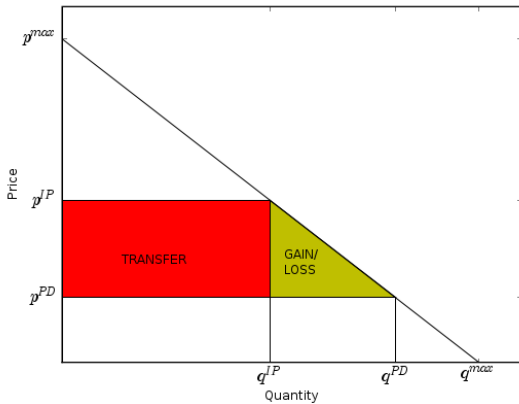


Figure: Demand Curve and Welfare

Summary: Books

Price: -10%, Elasticity: 2-4, PD Usage: 10%

⇒ PD Value = 1-2% of PD sales

(0.1-0.2% total sales)

Total consumer benefits: +10% (direct effect of price reduction) so 11-12%

Sensitive to Price Impact

E.g. Price: -60% (e.g. Harry P)

⇒ benefits 35-70% of rev

Greater price reduction, greater elasticity have
big effect on value ...

eBooks - Conjectural!

Price: -80%, Elasticity: 3-5

⇒ PD Value = 100-160% of current sales

Conclusion

Size

Books: \approx 10m in EU (3m+ for countries studied)
(10-20% of total)

Recordings (UK): 300k and will *double* in next
20y (if no extension!)

Usage

Crudely: $\sim 5-10\%$

(*Per-item* PD usage is 0.5 – 1 of in-copyright –
not bad!)

Excludes pretty much all digital material!

Price

For books + recordings (physical media): 5-15%

Inline with 'copyright' proportion of cost

Much larger for eBooks