How do copyrights influence science?

Student's name: **Bihong Pei**

1. Introduction

1.1 Context

Typically, copyrights prevent publishers from copying and selling the content of scientific texts that they do not own (unless they pay for a license). In 1942, the US Book Republication Program (BRP) allowed US publishers to reprint the exact content of German-owned science books. This program broke publishers' monopoly for these books and caused German-made books to become available in the market, leading the price of the average BRP book to decline by 25 percent.

1.2 Research Question

In this lab, we will use data including citations to BRP and SWISS books, to investigate an exogenous change in copyrights to examine the potential costs of copyright on science.

2. Data Summary

2.1 Data Description

The lab's data include BRP books in chemistry and mathematics, a control group of Swiss books in the same fields, and all new articles and books that cite BRP and Swiss books between 1920 and 1970.

This data lists id, years of citation, book author, book title, the city where the book is published, publisher, year or publication, whether the book is English, and the original price and the reproduction price.

From Table (1), we can observe that there are 291 BRP books with at least one citation, and 486 Swiss books with at least one citation, the total number of citations to BRP before 1942 is 935, and the total number of citations to BRP after 1942 is 7915. Because of BPR, the total number of citations to BRP increase a lot.

Table 1

	The total number of BRP	The total number of Swiss
Number	**291**	**486**
	The total number of citations to BRP	The total number of citations to Swiss
Number	**8850**	**1526**
	The total number of citations to BRP	The total number of citations to Swiss
N (Before 1942)	**935**	**282**
	The total number of citations to BRP	The total number of citations to Swiss

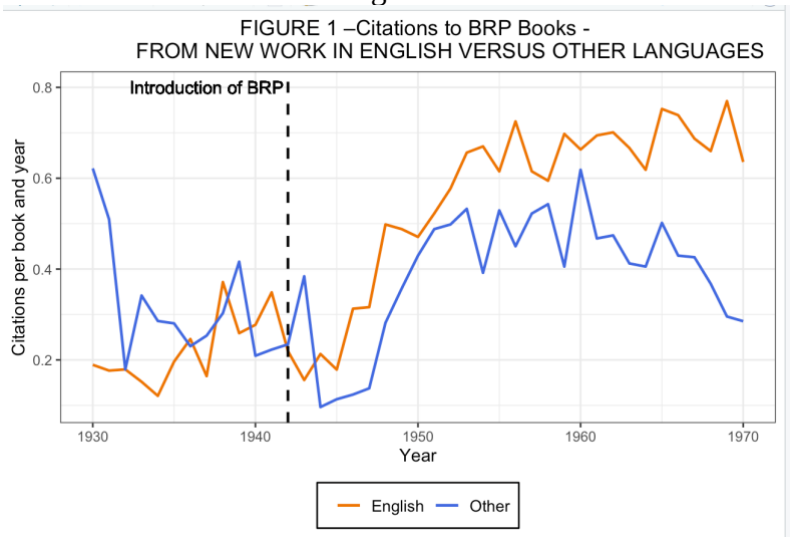
N (After 1942)		**7915**	**1244**
	The total number of citations to BRP		The total number of citations to Swiss
N (By English-language authors)		**5120**	**873**
	The total number of citations to BRP		The total number of citations to Swiss
N (publishing in other language)		**3755**	**645**

2.2 Data Analysis

To identify the causal effects of copyrights, we use the first strategy by comparing changes in citations to the same BRP book from English-language authors and from authors publishing in other languages.

Before 1942, we can observe that counts of new publications that cite BRP books in English and other languages are a similar trend. Moreover, After BRP, we can find that English-language publications increase more than the citations to the same books by authors writing in other languages. But trends in the English language and other citations were nearly identical until 1942 from Figure 1.

Figure 1



To estimate the effects of the BRP, we first estimate OLS regression:

$$cite_{ilt} = \partial English_l + \beta English_l * post_t + book_i + \tau_t + \varepsilon_{ilt}$$

The identifying assumption for this test is that changes in English-language and other citations to BRP books would have been similar in the absence of the BRP.

Table (2)

	(1)	(2)
english	-0.038	-0.038
	(0.036)	(0.041)

english_post	0.208 ***	0.208 ***
	(0.039)	(0.064)
State FE	Yes	Yes
Year FE	Yes	Yes
Pre-1942 mean	0.269	0.269
N	20232	20232
R^2	0.357	0.357

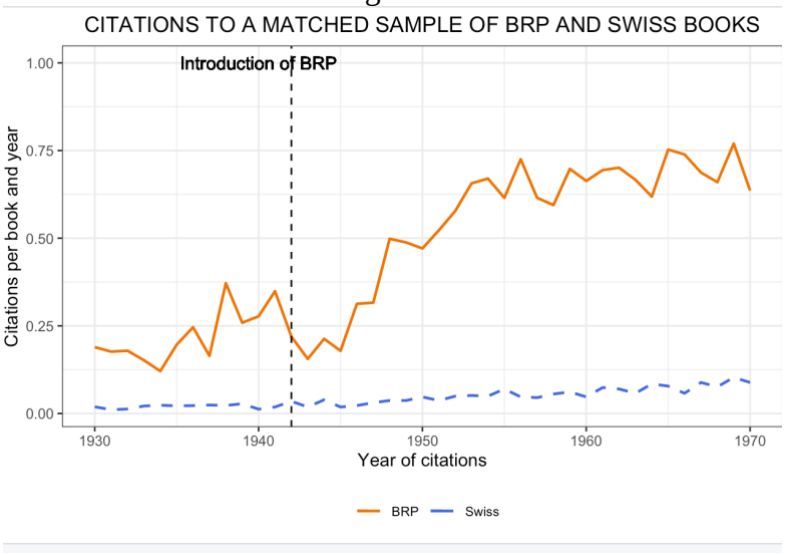
From Table (2), we can see that the coefficient β estimates the causal effect of the BRP, and we can find that coefficient β is 0.208, which means that citations to BRP books increased by an additional 0.208 per book and year after 1941 compared with citations from other languages, significant at 1 percent. Furthermore, we can find that the pre-1942 mean of the outcome variable is 0.269, this implies an 77.32% percent increase in citation in response to the BRP. The pre-1942 mean of the outcome variable helps us to calculate the increasing rate, so we use it.

However, this assumption may have problems, English-language citations may have increased mechanically after World War II if the research output of US scholars increased after World War II, relative to other authors. The reason why the English-language citations increase is World War II instead of BRP, which is independent of the BRP, in this way, the difference-in-differences test would overstate the effect of the BRP.

To solve this potential problem, we will use the second strategy, comparing changes in citations by English-language authors to BRP books with changes in citations by English-language authors to Swiss books.

Firstly, we plot the number of citations for BRP books and Swiss books per year, and we find that Swiss books in this sample receive fewer citations than BRP books (Figure 2). After 1942, citations to BRP books grow to 0.75 in 1965, while citations to Swiss books increase much less, the trend is basically flat. Citations to BRP books remain high around 0.75 per book year after 1965, while citations to Swiss books remain below 0.125.

Figure 2



Secondly, in order to estimate the effects of the BRP, we estimate the OLS regression:

ilt=l* postt + booki + t+ ilt

where the dependent variable cites it measures citations to BRP and Swiss books by new English-language publications to book I per year t between 1920 and 1970.

Table (3)

	(1)	(2)
brp_post	0.393 ***	0.393 ***
	(0.021)	(0.082)
State FE	Yes	Yes
Year FE	Yes	Yes
N	29879	29879
R^2	0.549	0.549

From Table (3), we can observe that an additional 0.393 new publications by English-language authors per year cite BRP books after 1941 compared with Swiss books, significant at 1 percent.

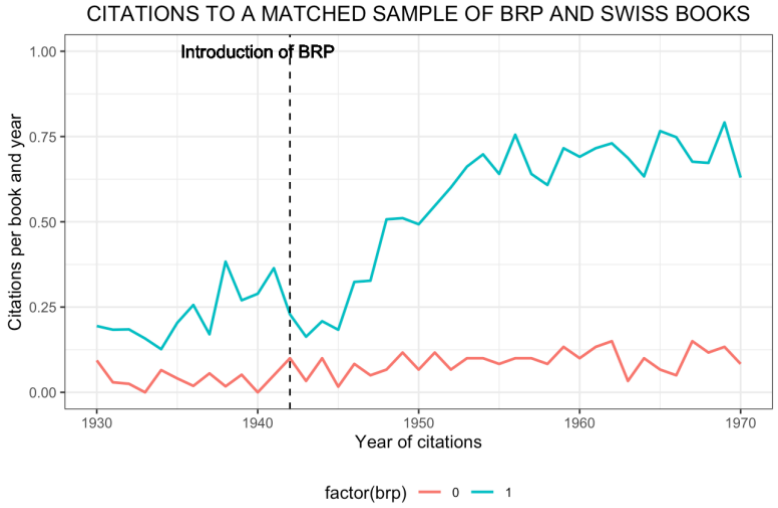
The reason why changes in citations to Swiss science books are a suitable control for changes in citations to BRP books is that books with Swiss-owned copyrights were not eligible for the BRP due to Switzerland's neutrality during the war, so citations to Swiss will be a good control variable.

The identifying assumption is that people would have no selection bias in choosing the book or BRP and Swiss, which means people randomly selected books, and the trends of citations

between BRP and Swiss books would be the same. However, publishers would pick books with high expected demand, so if they chose books for the BRP, in this way, estimates would overstate the effects that the BRP.

We need to choose the Swiss books and the BRP books in the same fields to keep them on the same level.

Figure 3



When I use the new subsample to do the figure and regression, the outcome is in figure 3, the trend of Swiss books is similar to the trend of BRP, and the trend of Swiss books is different from figure 2. And the regression outcome is in the Table4, in question(3), we have the selection bias, since BRP books were picked by publishers so the estimates would overestimate, but in this question, we solve the problem of overestimating, so this is the difference of the outcome in Table 4 regression with Table (3), and the coefficient estimate is 0.388, at the significant level 1, which means that an additional 0.388 new publications by English-language authors per year cite BRP book after 1941 compared with Swiss books. Therefore, our preferred estimates combine the two strategies by comparing the differential change in citations to BRP books from English-language and other authors with the same differential change for Swiss books.

Table (4)

	(1)	(2)
brp_post	0.388 ***	0.388 ***
	(0.052)	(0.089)
State FE	Yes	Yes
Year FE	Yes	Yes

N	12118	12118
R^2	0.553	0.553

*** p < 0.01; ** p < 0.05; * p < 0.1.

The triple-differences OLS regressions estimate:

$cite_{ilt}=\beta_1 English_l+ \beta_2 English_l$*BRP+ $\beta_3 English_l$* $post_t + \beta_4 BRP_i$ * $post_t + \beta_5 English_l$* BRP_i * $post_t$ + $book_i$ + τ_t+ ε_{ilt}

β 5 helps us calculate the differential increase in citations from English-language authors compared with other authors. The identifying assumption behind this triple-differences OLS regression estimate is that the difference between citations by English-language authors and citations by other authors would have been similar for BRP and the comparable set of Swiss books in the absence of the BRP.

The regression outcome is in Table (5), column 1 is iid, and column 2 is the cluster estimate. From column 1, we can find the coefficient estimate is 0.175 which means that BRP books receive an additional 0.175 English-language citations per year after the BRP compared with Swiss books.

Table (5)

	(1)	(2)
english	-0.012	-0.012 ***
	(0.012)	(0.004)
english_brp	-0.026	-0.026
	(0.025)	(0.041)
english_post	0.033 **	0.033 ***
	(0.014)	(0.007)
brp_post	0.136 ***	0.136 ***
	(0.020)	(0.052)
english_brp_post	0.175 ***	0.175 ***
	(0.028)	(0.064)
State FE	Yes	Yes
Year FE	Yes	Yes

N	59758	59758
R^2	0.375	0.375

*** p < 0.01; ** p < 0.05; * p < 0.1.

3. Conclusion

When the US government broke the copyright monopoly for German-owned science books, BRP books became popular across the United States, allowing researchers to use these books in their research.

From the above analysis, we can know that English language translations were the closest substitute for a BRP book, in order to further investigate whether substitutes for BRP books were available in 1942, we would use Amazon's sales algorithm to do the regression.

But we also need to know the effect of BRP on the prices, and the influence of the changes in prices on the increase of citations. Besides, we need to know the interaction of human capital with the BRP. Moreover, we need to extend our research into different disciplines, like Mathematics.