

Proposal: Digitization, Copyright Term, and Welfare in Books

Joel Waldfogel, University of Minnesota

How long should copyrights last? And has this changed as the costs of bringing works to market are reduced by digitization?

US copyrights in books traditionally lasted 75 years, which was recently extended to 95 years. While it strains credulity to suggest that the additional 20 years of protection is needed to provide incentives for authors, we don't really know much about how duration would affect revenue.

Digitization has the potential to reduce costs of bringing works to market. Traditionally, retailers received roughly half of retail revenue, while authors themselves received about 8 percent. The remainder went to the publishing houses. With coming digitization, the costs that revenues need to cover to provide creators the incentives to continue creating may be much smaller than they have been. It is therefore possible that consumers can be made better off, while maintaining creative incentives, with shorter copyright duration.

Accordingly, the goal of the proposed project is to ascertain the changes in producer and consumer surplus available with changes in copyright duration for books. That is, I'd like to be able to provide a measures of the gains to consumers and losses in revenue from moving to, say, 5, 10, or 20 year copyright duration. Suppose one can determine that half of current revenue is needed to maintain current creative incentives. Then I could ascertain the potential welfare benefits of the new technologies, given the current book titles in existence.

Descriptively, it would be useful to know the following facts to begin to answer the question of appropriate copyright duration. First, what share of published books continue to generate revenue 1, 2, or N years after publication? Second, even if only a small share of published books continue to generate revenue long after publication, it may be that a high share of revenue is produced by a few books long after publication, so: what share of revenue is produced N years after publication? Third, given the move to electronic books – and the possible circumvention of existing publishing houses – how much revenue is needed to maintain current creative incentives?

What's needed to study the first two of these questions is detailed data on the sales of books. These data exist but are not generally available to researchers. Nielsen has offered me access to their Bookscan data for \$6000 per year. Their data include the entire sales and price histories of essentially every physical book published in the US. The sales data cover the period 2001-present, but the titles include all books in print in this period (including books initially published long ago).

Specifically, I would estimate a model of book demand (using logit, nested logit, or some variant). In the model, demand for each title will depend on the price, its unobserved quality, and its age. Given the model, I can calculate the revenue and consumer surplus generated by each book. I can also calculate the impact of shortening copyright duration on these surplus measures, given the distribution of titles' qualities as well as the book age-demand profile.

As a practical matter, off-copyright books are made available digitally (at Gutenberg or Google, among other sites). I can use data on download volumes at these sites to get a sense of quantity demanded when price falls to zero and the book is available digitally.

One important challenge of the Bookscan data is that their access arrangement will allow me to extract only one book at a time. I am allowed to extract as many books as I wish, but I may not use a scraper. Hence, obtaining the data will entail more costs than simply the access fee. Accordingly, I am requesting the full \$20,000 to allow me to obtain some research assistance in creating the dataset. (I expect I will need a few undergraduates working for a few months each).

I expect that these data will allow me to pursue some different but related research projects. For example, I am interested in replicating my work on the evolution of music quality over time on books.