The Pecuniary Motivation to Create

Stan J. Liebowitz

liebowit@utdallas.edu

Alejandro Zentner

azentner@utdallas.edu

University of Texas at Dallas

Abstract

Economists have long debated the degree to which inventive and artistic activities respond to expected pecuniary rewards. Copyright and patent laws are based on a view that monetary rewards affect creativity. In this paper we attempt to provide an empirical analysis of this assumption by using a recent panel data set covering the book production industry. We find evidence that authors publish more new books after having earned higher payments for previous works, although at high enough levels of payment their production of new works begins to decline, consistent with a backward bending supply function. We also find that few authors are in the backward bending region and that the overall effect of payments to authors is to increase their output. Nevertheless, our results are also consistent with the belief that non-monetary factors can play an important role in the creation of books.

"No man but a blockhead ever wrote except for money." Samuel Johnson

Introduction

Why do humans create new books, new songs, or new inventions? To what extent is creation an expression of an innate imaginative impulse, an attempt to acquire monetary reward, or a response to other motivations? Although these questions have long beguiled philosophers and psychologists, our concern is a more prosaic one. We wish to understand how economic systems might promote creation, or at least promote creations where the benefits are greater than the costs. This topic has only become more important in recent decades as the Internet's 'creative destruction' of various copyright-based industries has focused attention on the nature of creation and production in these industries. The Internet has lowered the costs of production and distribution of creative works, while simultaneously increasing the magnitude of unauthorized usage. These conflating factors limit researchers' ability to use the advent of the Internet to analyze the impact of monetary rewards on the production of creative works (Smith and Zentner, 2015).

The changing landscape for producing and consuming creative works has led to a reexamination of just how large a role, if any, expected revenues play in inducing creation. Answering this question would inform the general analysis of the justification for intellectual property. If the supply of creative works was unresponsive to expected revenues, for example, then providing ownership of these works, as copyright (or patent) law does, would allow authors and publishers to charge prices above the cost of reproduction, leading to an economically inefficient outcome since these are non-rivalrous goods. Answering this question also would allow us to better understand how the reduced revenues due to piracy (Liebowitz, 2016, Danaher et al., 2015) might impact creativity.

The belief among economists that the supply of creative works might not be positively related to the revenues provided by our intellectual property system is not new (Taussig, 1915, Plant, 1934, Pigou, 1952).[1] Anecdotes have generally been the empirical basis for this claim. There have even been suggestions that overpayments to successful creators would decrease their creative output due a backward bending supply, or the monetarily induced effort needed to enforce their rights on old works (Scherer, 2004, Boldrin and Levine, 2008, Lunney, 2018).

Our purpose here is to examine this issue by measuring relationships between the payments to book authors and the number of books those authors then write. We use a dataset containing about half a million authors over a thirteen-year period, which, to our knowledge, is a fuller population of creators than has previously been examined. Using this panel data, we hope to shed some light on revenue's role in the production decisions of authors.

I. Previous Literature

Although often overlooked in the literature on innovation and intellectual property, the question of how inventive and artistic activities respond to expected pecuniary rewards is related to the question of how workers change their willingness to work in response to changes in payment, more commonly referred to as the supply curve of labor. Although upward sloping supply curves are generally considered the norm, one of the central exceptions to that norm is the infamous 'backward bending' supply of labor. The backward bend occurs when leisure is a normal good and the income effect from

[1] There have been critics of copyright since copyright has come into existence. These critics have often made the claim that creation is, at least to a large extent, an innate personal characteristic having little to do with remuneration. Plant (1934), for example, argues that many creators are motivated by nonpecuniary goals, although he admits that some creators are motivated by expectations of pecuniary rewards.

an increase in wages overpowers the substitution effect. Many estimates of the elasticity of labor supply imply that employed male workers are on the vertical or slightly backward bending portion of their supply curve (Borjas, 2013, Goolsbee, 2000). Although that has generally not been the finding for the supply of female workers,[2] recent evidence indicates that female labor elasticities have been declining and are approaching the levels of men.

Those estimates are generally based on how marginal hours of work change as the wage rate changes, or how labor supply changes in response to changes in the income tax rates that are used as exogenous shifters of the wage rate that are used as an instrument or natural experiment. However, it is important to note that using an exogenous shifter would have its own shortcomings since more than one instrument would be required to estimate a backward bending labor supply function (the demand function would intersect the supply function at two points). Moreover, empirical fluctuations in average wage rates are usually modest and changes in tax rates are infrequent, possibly limiting our knowledge of how hours worked would respond to major changes in income. Nor do these estimates usually measure whether actual production (as opposed to hours worked) changes with wage rates, as would be the case with, say, piece-rate payments.

Our analysis, on the other hand, is based on panel data tracking the creations of authors who may experience very large fluctuations in payments over time, whose effort can be measured by simply counting the output of works, and where payments largely depend on the number of works produced and the sales of those works. Using panel data allows us to control for author-level baseline productivity, which provides some support for a causal interpretation of the correlations that we find. Because authoring books is usually an activity undertaken by individuals, our approach is not complicated by measurement of effort issues that occur in team production. On the other hand, our analysis is complicated by the possibility that some authors, probably a majority, appear to have other sources of income, and that sometimes such income is complementary to their book output.

In shedding light on the importance of financial rewards in the creative activities of authors, we can exploit the considerable heterogeneity in the types of authors, types of books, and types of business models used by authors, to gauge how 'artistic' creators do or do not differ in terms of their response to pecuniary payment. We believe that ours is the first attempt to examine the modern production of creative works using a data set that covers virtually an entire industry, with some caveats that we discuss later.

A precursor to our work can be found in F. M. Scherer's (2004, 2008) comprehensive examination of the production of classical music from the seventeenth to the nineteenth century, as the forms of payment shifted from patronage-based payments to markets based on copyright. Scherer failed to find any increase in overall productivity after the emergence of copyright, and he provided some specific examples where productivity appeared to decrease when payments increased, although he also provided many examples where money was a clear motive for creation of musical compositions.

Creation undertaken without the expectation of external rewards (either income or fame) can be treated as a taste or instinct, in which case it is not itself an instance of the typical economic model where rewards must overcome the disutility of work. For example, Pigou (1952) claimed that "[patent laws] do not, indeed, appreciably stimulate inventive activity, which is, for the most part, spontaneous." F. W. Taussig (1915), a former president of the American Economic Association, suggested that the behavior of inventors and artists is instinctual and innate in the same way that birds work on their nests

[2] In his textbook, Borjas (2013) states (p. 46) "These surveys conclude that the elasticity of the male labor supply is roughly around -0.1…one would not be stretching the truth too far if one were to claim that the male labor supply elasticity is essentially zero." He reports (p. 53) that "most studies of female labor supply find a positive relationship between a woman's hours of work and her wage rate."

or beavers work on their dams. Taussig states "one thing stands out conspicuously: the race of contrivers and inventors does obey an inborn and irresistible impulse. Schemes and experiments begin in childhood, and persist so long as life and strength hold."

He contrasts this view with more standard economic analyses:

> This had been the view of the older utilitarians: men contrived simply because this was conducive to gain, and would not contrive unless prompted by the experience and prospect of gain. Hence there must be premiums and prizes, patent laws, protected trade-marks, the bait of profit. But if there is a spontaneous impulse, — spontaneous in the sense of not being dependent for its initiation on a calculated gain, — we may be led to conclude that the patent system, for example, is a huge mistake. Men would invent anyhow: they obey the instinct and therein take joy. So poets are actuated, musicians, men of science. In their activities we have long recognized the intrinsic satisfaction from the exercise of inborn impulse. [p. 18-19]

Claims about the innate desire to produce creative works, whether music, books, or inventions, persist in the more recent literature, usually relying on anecdotal evidence. Boldrin and Levine (2004, 2008), for example, repeat several stories appearing to demonstrate a decline in creative activity due to the introduction of patent or copyright, including a history reported by Scherer about the Opera composer Giuseppe Verdi reducing his output of operas after gaining access to copyright protection on his old works. Boldrin and Levine discuss, in the same manner, the supposed slowdown in steam engine innovation because, its patent-holding inventor, James Watt, was supposedly too busy protecting his patent to continue improving the engine's design, although this claim has been disputed by Selgin and Turner (2006, 2009, 2011). In the Internet era, claims that many creators, or that entire classes of creators, are not motivated by financial gain can be found in Benkler (2006) or Boyle (2003). Moser (2013) suggests that patents are as likely to be harmful as helpful to innovation.

The hope of this paper is to avoid a reliance on anecdotes and to provide an empirical first-pass answer to the question of how much creation is due to the prospect of monetary gain, at least among book authors in the early decades of the twenty-first century. We attempt to answer this question by using an extensive data set containing a previously unavailable variable—a measure of revenues from creative works-–and covering virtually the full population of creators in the industry. Our panel data allows us to examine how past revenues affect the production of new works controlling for time invariant author-level characteristics (e.g., baseline productivity, innate ability, utility or disutility from writing books) as well as several other control variables (e.g., genre, nature of the publisher).

II. Authors' and Publishers' Decisions

In the standard economic model of publishing, authors create works with the expectation or hope that the market will reward their efforts. It is common for moderately successful authors to receive a guaranteed lump-sum advance against royalties, which converts into a pure royalty scheme if sales are large enough for the royalties to more than cover the lump sum.[3] But many different variations in payment are possible (see Wu and Zhu 2018 for an analysis of how contract variations and competition affect creativity), including having the author pay for publication with the possibility of receiving future royalties.

[3] Liebowitz (2016) notes that during the nineteenth century American publishers paid American authors straight royalties, British publishers paid British authors primarily lump sums, and American publishers paid British Authors mainly lump sums although some received royalties. The current model of a guaranteed lump-sum advance against royalties is a hybrid of the older models.

Authors may have different motivations for creating, including various intrinsic motives as well as monetary motivations. Some authors may have only a single motivation while other authors may have several. Authors who are motivated to create without the allure of payment may nevertheless produce more works when their payment is greater for the simple reason that they can devote themselves full-time to their creative efforts, eliminating the need to work another job to pay their bills. While we cannot determine what fraction of the creative output is due to every possible motive, we nevertheless hope to be able to provide some first-pass evidence of the fraction of the creative output that is due to monetary rewards versus the fraction that is due to other motives. Authors face a far less certain set of choices than do hourly workers or piece rate workers (e.g., plumbers) when they make decisions about increasing or decreasing their efforts. Hourly workers know with virtual certainty that an increase in hours leads to an increase in pay. Authors are unlikely to be able to accurately predict the future sales of a next book, to say nothing of the sales of a portfolio of future books, since market success for creative works is very unpredictable. Some authors, particularly those with a proven track record, may have far less of this uncertainty because they receive an advance from a publisher guaranteeing them a minimum amount of payment independent of how well the book sells, although those advances will be reduced for future books if current book sales underperform expectations.[4]

The decision to publish a book is a joint decision made by authors and publishers. While it is true that publishers have somewhat different pecuniary motivations than authors (e.g., the well-known textbook example of authors wishing to maximize the revenues on which their royalties are based, whereas publishers wish to maximize profits) we believe this is a second order effect. The more serious distinction between authors and publishers is that the latter often do not wish to publish books from the former if the revenues from those books are expected to be insufficient to cover costs. However, since authors always have the option of self-publishing, (and as we show below, the fraction of authors who do self-publish is between one-third and one-half, even in the early years of our data), we believe that treating authors and publishers as a single unit is not likely to seriously skew our analysis, although it is important to acknowledge that our estimates would not represent the supply of labor for authors alone. Importantly, we note that the distinction between authors and publishers is mostly irrelevant if the focus of the analysis is on how monetary compensation affects creative output available to consumers in the market, independent of whether this output is the result of authors' or publishers' decisions.[5]

Our assumption will be that authors (and their publishers, which from this point on we will assume to be jointly included when we use the term 'authors') take their previous success as a harbinger of future success. For example, by publishing a book, authors can learn their own earnings potential. Clearly, authors might focus on the success of their last book, or last several books, along with other factors such as comments from agents, the quality of the publisher, the marketing plan, and publishers' advances, when forming their expectations. We do not specify how they might weight the performance of their recent works relative to other data, or whether they might be overly optimistic or pessimistic. We merely assume that if authors are motivated by pecuniary factors, then the pay they received for their past works will influence the efforts to produce a larger or smaller future output. Specifically, we assume that authors look to the revenues generated by their

[4] Some contracts have cross-collateralization clauses which require profits above advances from one book to be used to cover advances from other books that were not recouped by the publisher.

[5] We note, however, that the introduction of ebooks reduces the cost of publishing books and may have changed the relationship between monetary compensation and the number of books in the market. Our data set does not, unfortunately, include the titles published in ebook format exclusively, although it includes the titles published in both ebook and print format, and most ebook publishers offer print versions. Our estimates, thus, more accurately represent the relationship between monetary compensation and the amount of book publication available to consumers in the market prior to the introduction of ebook. We discuss below how examining a similar question for other creative industries (e.g., movies or pharmaceuticals) is substantially more complicated (e.g., creating a new drug or a movie may involve decisions made by a large number of players and a substantial investment) and can also change over time with the evolution of the organization of these industries.

books written in period *t-1* to form an income expectation when deciding how many books to write in period *t*, although they may look at other factors as well.

III. The Book Market and our Data

We examine the impact of monetary payment on the productivity of authors by using data on book sales from NPD BookScan. BookScan collects point-of-sale data from approximately 16,000 retail outlets, including all the major book retailers and mass merchandisers (e.g., Barnes & Noble, Amazon, Target, Costco, Walmart), smaller retail channels, and hundreds of independent book stores. BookScan claims that the data cover roughly 85% of the entire US print book market.[6] While the data include online sales of print books (including self-published books), excluded are sales of e-books (e.g., the data include sales of print books from Amazon.com but excludes sales from the Amazon Kindle Store) which became an important part of the market after 2010.[7]

The purchased database access provided weekly and yearly information on book sales for 13 years (2004 through 2016) with some limitations as described below. For each year and for each of the two broad book categories (fiction and nonfiction) under examination, our database includes observations listing information from the 250,000 top-selling book titles. Our data set contains almost the entire population of books for the fiction category but leaves out a substantial number of observations for the nonfiction category because the number of nonfiction works exceeded the 250,000 limit in every year.[8] The fact that our data are not fully complete for nonfiction will require extra care in the analysis that follows. Another problem with the data is that sometimes the number of book titles is overstated because some titles are listed multiple times.[9] Note that most book titles sold in a year were initially published in an earlier year because sales of a particular title often continue for several years, as described in Appendix 1.

The unit of observation for our analysis, however, is the author, not the book. This necessitated aggregating the books written by each author and the yearly sales of those books. The raw data, after eliminating some obvious errors, provided information on 357,035 nonfiction authors and 257,211 fiction authors. The variables included in the yearly database are the list price, the number of copies sold (for the year), the International Standard Book Number (ISBN) 13-digit identifier, the book title, the author, the publisher, the imprint, the publication date (of the edition),[10] the genre (Book Industry Standards and Communications subject code, or BISAC), and the format (e.g., paperback, hardcover, audio). We can

[6] There have been some changes in coverage during our study period. For example, Nielsen BookScan began including book sales from Walmart in 2013.

[7] In Appendix 2 we perform an estimation similar to our overall estimation but for the period 2004-2009, which precedes the period of large e-book sales. The results are similar to those we find for our entire 2004-2016 timeframe, leading us to conclude that the growth of the e-book market since 2010 does not greatly affect our results or conclusions.

[8] For works of fiction, the limit of 250,000 begins to bind starting in 2010. Even when it binds, the restriction still allows for books which sell as few as two copies. The effect of this restriction after 2009 seems to be fairly small because, as we find in Appendix 2, the regression results using two three-year periods prior to 2010 are very similar to what we find in the main text using two periods spanning the entire 2004-2016 period.

[9] Some popular book titles are listed more than once due to translations, typos, or slightly different versions. Stephen King, for example, has six similar titles in the database for his single title 'The Waste Lands.' Charles Dickens has the largest number of titles published between 2004-2016, at 368, but this is a highly inflated measure of the number of his unique published titles because, for example, there are 59 distinct variations of 'A Christmas Carol.' Because that work, along with many other classic works, is not protected by copyright, it is published by many different publishers, all of whom are able to add words to the title such as "illustrated," "vintage," "original," "iconic" and so forth, as they see fit, and to package various combinations of works.

[10] We found that it was often the case that listed publication dates did not match up with the dates when the generation of revenues began, which frequently occurred a few years later. Because revenue generation was central to our analysis, we constructed our own 'publication date' to coincide with the first revenue generation for a title and that is what we use throughout the analysis. Using the listed publication dates does not substantially affect the estimates.

compute the nominal revenue by multiplying the list price of a title by the number of copies sold. We do not directly measure the payments to authors, but we take the relative nominal sales revenues of the book to be a proxy for the relative royalties received by authors.[11] We measure productivity by counting the number of titles published by each author. In contrast to Waldfogel (2012) who uses clever proxies to measure the evolution of music quality over time, we directly adjust for quality by using sales figures since they are observable to us. We, unfortunately, do not observe the number of pages for each book (assuming similar font and page size) which would allow for the use of an alternative measure of output (although controlling for genre takes account of some large page differences, such as in comics which are typically much shorter than other books). Using number of titles (controlling for genre) appears to be a sensible measure of productivity since the size of books in a genre are often quite standard, and this follows how productivity—the number of articles—is typically measured in one market with which we are most familiar, the academic market.

Using the book industry to examine the relationship between reward and creation has several advantages relative to other possible targets of analysis. One major advantage is that books are usually written by a single author, eliminating problems of allocating effort in team productions, and easing productivity measurement. Most similar industries, such as popular music, movies, or television, have products often created by group efforts, with the added complication that individual artists often move from one group to another during their careers. Most authors, by way of comparison, work alone and write under a single name during their careers, allowing their efforts to be accurately tracked over time. Of course, there are authors who write under multiple 'pen' names, authors who are really fronts for company efforts, and authors who work as part of a team, but these represent a small minority of cases. Another advantage of using the book industry is that most books are single, non-divisible creations. Thus, successful authors are less likely to have "best of" compilations of greatest hits, mixing portions of various prior books together in order to create a new commercial product out of old artistic works, as is the case with music. Although we believe there are advantages in focusing on the book industry, we note that the relationship between monetary incentives and creativity may differ under other circumstances and it would be valuable to examine this relationship in other industries (e.g., pharmaceuticals or movies) with different structures.

There is a piece-rate element to this market because the revenues an author is likely to earn should be related to the number of titles that author creates. The rather short 'shelf-life' of many successful books, and the ability of some authors to produce one successful book after another, are two factors that appear to support the view that some authors can increase their overall payments by producing, at fairly regular intervals, products that appeal to the same audience over and over, making their product rather like a form of piece-rate payments. In its extreme form, the 'author' becomes a business, producing books like clockwork, often hiring 'employee' authors to write these books, using the characters, techniques, and style of the original author, and sometimes continuing to do this even after the original author has died.[12]

There is also an element of a tournament in some book markets. That is because authors often compete with one another for a somewhat fixed number of readers. Consumers in these markets are likely to allocate their purchases among the competitors by their expectation of relative quality, and a relative increase or decrease in the overall quality of works in such a market is not likely to change the financial size of the market (e.g., if all elementary economics texts increase or decrease in quality, the size of the market is unlikely to change very much).

---

[11] The relationship between book sales revenue and author payments is generally not linear. The more successful authors usually receive higher royalty percentages than less successful authors, and many authors have contracts where their royalty rate rises as sales increase. Also, royalty rates usually differ between hardbound books and paper or mass market paperbacks.

[12] This is how the new 'Tom Clancy' books, for example, continue being produced, although he died in 2013.

Because our concern is with the impact of past revenues on the production of new works, we attempted to eliminate from our sample books written by authors who are dead. It was impractical to go through obituaries or biographies to find such authors. Instead, we limited ourselves to removing classic works that are still being published, many of which had lost their copyrights. We found that many of these works were removed when we eliminated titles listed as being published prior to 1980, and we removed additional titles by hand when we recognized them as being classics. This seemed to remove most of the old classics from the sample. Thus, Charles Dickens and his 'A Christmas Carol' discussed in footnote 9, is not part of the sample used in our analysis.

The major categories of books that enter our analysis are listed below.

III. A Fiction versus Nonfiction

The two major book categories we explore are adult fiction and adult nonfiction. The works in one are usually poor substitutes for the works in the other. We also expected that there might be important economic differences between authors of these two types of works.

Nonfiction authors, for example, often have jobs complementary to their writings and routinely use their nonfiction writings to enhance their employment opportunities, and vice-versa. Politicians write (or have ghostwritten) supposedly nonfiction works in order to enhance their political opportunities. Many self-help authors claim expertise in particular fields (e.g., medicine, real estate) and write books to help market these other services. Professors write scholarly tracts with the hope of enhancing their academic careers. Thus, even for authors motivated by pecuniary rewards, we might expect that revenues from nonfiction books sales to be only one component of a revenue stream consisting of several complimentary products, and that this might reduce the need to write new titles as frequently as would be the case of book revenues were the only income source. If so, the linkage between book revenues and the incentive to produce new books is likely to be weakened for works of nonfiction.[13]

Fiction authors, on the other hand, even when they have other jobs, are unlikely to have jobs related to their activities as fiction authors. Because revenues from selling fiction books likely represents the entirety of revenues from this activity (except for the most successful authors who can earn movie rights), book revenues might be expected to induce greater publication of new books for fiction than non-fiction works.

A countervailing factor is that we might expect that authors motivated by innate urges (and thus not influenced by monetary payment) to be mainly fiction authors. Authors may want to follow in the footsteps of writers thought to lead exciting lives, such as Kerouac, Hemingway, or London. Such romantic goals seem less likely for works of non-fiction than for fiction.

Thus, although the motivations of authors in these two groups may differ, it is not clear which group would be expected to have a greater or lesser impact of pecuniary success on the publication of new works.

[13] We assume that being a published author of nonfiction enhances one's prestige at being a professional (doctor, lawyer, etc.) independent of how well the book sells. Most customers in the market for the ancillary activity are unlikely to know how well a book sold even though the author may well make them aware of a book's existence.

III.B Genres

There are at least several dozen genres of books within each category of fiction and nonfiction. Major fiction genres include crime, romance, thrillers, fantasy, and so forth. Major nonfiction genres include biography, self-help, travel, and cooking. We note that there is a much larger number of non-fiction than fiction genres in Nielsen's data due to the larger number of BISAC subject headings for non-fiction (e.g., history, religion, business and economics, medical). Some genres allow for much shorter books (e.g., comics) that can be produced relatively quickly and thus the number of books produced over a given time period could easily be a function of the book's genre.

Although we can surmise that the motivations of fiction and non-fiction authors may differ, we are unable to similarly expect, a priori, differences across authors' financial motivations in different genres (e.g., why would a fantasy writer have greater or lesser pecuniary motive than an author of romance novels?). Thus, we will limit our use of genres to controlling for the number of titles written.

III.C Company versus Personal Authors

The large majority of authors in our data are individuals. But for nonfiction works in particular and a limited number of works of fiction there are entities, corporate or otherwise, listed as authors. Examples include restaurant guides, study guides, comics, compendiums of puzzles, and religious works. The motivations of the entities responsible for publishing these works are varied, but it seems likely that the number of titles produced by individuals would differ from the number of titles produced by entities, as might the response of output to revenues.

In order to classify books authored by corporations or organizations, rather than by individuals, we conducted a text-mining analysis of the authors' 'names' in our data. By searching for key words in authors' names (e.g., "Inc", "Corp", "LLC"), names that were single words, and also by manually classifying the most important and obvious corporate or organization authors (e.g., "National Geographic," "Merriam-Webster," "Lonely Planet") we were able to classify authors, albeit imperfectly, into the categories of "company authors" and "individual authors" even though the category of "company" includes various types of organizations.

Because the number of company authors is relatively small, and our ability to discern which books are written by such entities is limited, we should acknowledge our limitations in drawing conclusions about the productivity of companies versus individual authors.

III.D Type of Publisher

We also have categorized publishers into the categories of "major" and "minor" due to the possibility that authors with major publishers might have different productivity from authors with smaller publishers. There were over eight thousand publishers of about half a million titles in fiction and nonfiction.[14] To determine leading publishers, we began by looking at those publishers with the largest number of titles. Unfortunately, some of those 'publishers' were really distributors for smaller publishers and/or working mainly with self-publishing authors.[15] We thus took the ten largest publishers who were not merely distributors and who tended to have strong literary reputations, and counted those as the major publishers,

[14] There were 481,906 titles with listed publishers in the fiction market, and 613,326 nonfiction titles in the raw data. There were 8,426 publishers of fiction and 6,241 publishers of nonfiction.

[15] The largest of these publishers was Lightning Source, but there were numerous others such as Ingram and National Book Network which were about one tenth the size of Lightning.

although some of these publishers tended to specialize in fiction or nonfiction.[16] These publishers were responsible for about one fourth of all titles in each sample.[17] The share of book revenues represented by these publishers was 81% in fiction and 47% for nonfiction, and since their share of revenues is greater than their share of titles, these publishers represented the more successful books, particularly in fiction.

III.E Self-Publishing

The last few decades have seen an explosion in self-publishing, an activity formerly known as "vanity publishing" but recent self-publication has much less of a negative connotation. Perhaps somewhat surprisingly, the extent of self-publishing does not appear to be much different for works of fiction or nonfiction, although limitations of the data make this comparison less compelling than we would like.

We categorized a book as being self-published if the imprint was known as one that specializes in allowing authors to pay to have their titles published, often through print-on-demand (POD) services. Wikipedia provides a listing of such publishers[18] but we found that many of those publishers were not listed in our data as publishers, but instead were listed as 'imprints,' so we used the imprint variable instead of the publisher variable.[19]

There are tens of thousands of imprints, which are far greater in number than is the case for publishers,[20] and we examined all imprints that published more than 25 titles in our data. Self-published books represented 36.8% of all fiction titles, and a seemingly similar if slightly smaller share of nonfiction titles.[21] From 2013 through 2016 this share was approximately 50% but even during the early 2004-2006 period self-published works of fiction made up about 33% of all titles. Because self-published books sell considerably fewer copies than other books, and their prices tend to be considerably lower, their cumulative share of revenues is often less than 1% in each market. We should note that there is other evidence that self-published books generate a larger share of their revenue from electronic versions (e-books), which are not reported in our data, so that our revenue estimates are likely to understate the sales of self-published books (in Appendix 2 we replicate our analysis using a period of time when e-books would not have had a distorting effect on our results).

III.F Cohorts of Authors

Our data cover thirteen years, 2004-2016. In our analysis in the main text we group books into two periods, a six-year first-half consisting of years 2004-2009 and a seven-year second-half consisting of years 2010-2016, referred to as period 1 (P1) and period 2 (P2) respectively (we use other time periods in Appendix 2 as a robustness check).

To avoid complicated life cycle considerations discussed in the labor economics literature (MaCurdy, 1981), and to make P1 authors as similar as possible to one another, we conduct our analysis using only authors who published their first book in

---

[16] This set of publishers were (in descending order by number of titles): Random House, Macmillan, Penguin Group, Simon & Schuster, Harlequin Books, John Wiley and Sons, Harpercollins, Hachette Book Group, Pearson Higher Education, and McGraw Hill Trade.

[17] For fiction is was 25.3% and 21.8% for nonfiction.

[18] That list is available here: https://en.wikipedia.org/wiki/List_of_self-publishing_companies

[19] This was particularly the case with publishers such as Lightning Source, since this 'publisher' (more of a printer and distributor, really) provided POD services for many books not self-published, as well as those that were self-published.

[20] There were 39,011 fiction imprints and 37,514 nonfiction imprints, compared to the much smaller number of publishers mentioned in footnote 14.

[21] We cannot directly measure the self-published share of nonfiction titles because our nonfiction data excludes titles selling less than 65 units, which excludes most self-published titles in our data. Limiting our sample to titles selling more than 65 units, the share of self-published nonfiction titles is about 80% of the share that self-published works of fiction represent.

P1.[22] Restricting the analysis to debuting authors also simplifies the analysis because it allows us to circumvent the need to speculate about authors' histories predating our study period and because it provides a more homogenous and more comparable group of authors.

We examine several samples, but in the main text we only report the results of one sample that includes all authors debuting in P1, independently of whether they also publish a positive or zero number of books in P2. In Appendix 3 we show results for a cohort of "long-term" authors that includes those who published prior to P1, during P1, and also during P2, a group that generally can be thought of as representing the most successful authors.

IV. Summary Statistics

To provide a feel for the data, we first provide summary statistics for revenues and titles. Because our nonfiction sample does not contain observations with fewer than approximately 65 units sold in a year (due to the data restriction of 250k yearly observations), whereas our fiction sample does not have this limitation (see footnote 8),[23] a simple comparison of summary statistics for the two categories of works would significantly overstate the average revenues of nonfiction relative to fiction. To ameliorate this problem, we put the fiction sample on a similar footing, restricting its observations to those authors with more than 65 units sold in a year which has a very large impact on the summary statistic values for the sample of works of fiction.[24] This restriction is only performed for the summary statistics and is not needed for the regression analysis.

We show summary statistics for a sample containing all authors publishing their first works in P1. The authors excluded from this sample are those who first publish in P2 or those who publish prior to P1. Notice that our summary data have more than six times as many nonfiction authors as fiction authors due mainly to our putting the fiction sample on a comparable basis to the nonfiction sample.

Table 1a: List-Price Sales Revenues for Authors

	variable	N	Mean	Median	min	max
Nonfiction	Yearly P1 rev from P1 books	109,808	$26,694	$5,220	$1	$37,400,000
	Yearly P2 rev from P2 books	109,808	$9,986	$0	$0	$58,000,000
	Yearly P2 rev from P2 books (only authors with new P2 books)	23,264	$47,136	$8,127	$0	$58,000,000
Fiction	Yearly P1 rev from P1 books	14,131	$42,095	$4,457	$56	$24,900,000
	Yearly P2 rev from P2 books	14,131	$15,525	$0	$0	$9,767,030
	Yearly P2 rev from P2 books (only authors with new P2 books)	3,821	$57,415	$11,404	$132	$9,767,030

[22] We can only imperfectly determine whether an author publishes prior to P1 by removing authors who have books that sell in P1 or P2 with listed publication dates prior to 2004. This methodology would fail to identify authors who published a title prior to 2004 but sold zero copies (or less than 65 nonfiction copies) of those titles after 2003.

[23] Each year the nonfiction works have somewhat different minimum levels of sales with the lowest value of 45 in 2004 and the highest value of 73 in 2011, leading to an average value of 64.6 per year. A very large share of raw fiction observations (sales of a title in a year) had sales of less than 65 units (2,216,011 out of 2,965,979) so that three quarters of the raw fiction observations are removed when we put the two samples on an even footing where each observation has a minimum value of 65.

[24] The average and median values are much lower for our complete fiction distribution as opposed to the distribution truncated at 65 units sold per year. Mean revenue values for the full nonfiction sample are about fifteen percent of those in Table 1a whereas median revenue values are about one or two percent of those in the Table. More information about this can be found in Appendix 7.

Table 1a shows revenue statistics for authors debuting in P1. The first row for fiction and nonfiction lists revenue in P1 from books first published in P1 and the second row lists revenues in P2 for books first published in P2 including the possibility that authors fail to publish and generate no revenue in P2 (includes authors with zero publications in P2). The third row performs the P2 calculation but only for authors who published works in P2 (excludes authors with zero publications in P2).

The large difference between the mean and median values is due to the very skewed distribution of book sales that is common throughout our data and is to be expected given the long tail nature of the book industry, for both fiction and nonfiction works. The average and median revenue values in P1 in the first row for each category, when each author publishes at least one work, is considerably greater than the P2 values in the second row, when including authors who did not publish any new works in P2. The third row in each category probably gives a better comparison to the first row since it only includes authors who publish in P2, and these authors generate more revenues in P2 than they did in P1, even after accounting for inflation.[25] If the most successful P1 authors are also the most successful P2 authors, this result might suggest that monetary incentives induce productivity, since it is consistent with more successful authors in P1 deciding to continue writing books in P2. Although fiction authors earn considerably higher mean revenues than nonfiction authors, that relationship does not hold using median values.

Of more direct concern is the evidence that average authors cannot support themselves from book sales alone, since typical royalties (10% or 15%) applied to annualized mean sales are not sufficient to support a single person at the poverty level. This result is consistent with the common understanding that creative industries have a large number of "starving" members and a small number of very successful stars. The fact that some of these authors will continue to earn revenues from these books in future years beyond a particular period's end is not likely to alter this general result.

Finally, superstar authors, as indicated by the maximum values, can earn a substantial living, with yearly royalties in the millions of dollars per year. This is true for both fiction and nonfiction authors.

Table 1b: New Titles

	variable	N	mean	p50	min	max
Nonfiction	Titles per year published P1	109,808	0.48	0.33	0.17	27.00
	Titles per year published P2	109,808	0.11	0.00	0.00	30.17
	Titles per year published P2 (excludes zero)	23,264	0.54	0.40	0.14	30.17
Fiction	Titles per year published P1	14,131	0.69	0.50	0.17	14.33
	Titles per year published P2	14,131	0.28	0.00	0.00	17.86
	Titles per year published P2 (excludes zero)	3,821	1.05	0.83	0.14	17.86

Table 1b compares the number of fiction and nonfiction works published in P1 and P2. The results indicate that fiction authors write more books than nonfiction authors, which may explain the greater average revenue for fiction writers. The extreme maximum values are due in part to 'company' authors, such as the Puzzle Society, which published 181 nonfiction titles in P2, but also to individual authors, such as the author of erotic werewolve stories, who wrote 125 works of fiction in P2 (note that the numbers in the table are per year). The table also indicates that, for the entire sample, there are fewer titles

[25] The average CPI in P1 was 203.8 compared to 231.3 in P2, providing an increase of 13.5% between P1 and P2.

published in P2 than in P1, which is to be expected given that many P1 authors do not publish in P2. When we only include authors publishing in P2, productivity is greater in P2 than in P1.[26]

V. The Role of Past Revenues on New Production

Our purpose in dividing the thirteen-year period into two separate periods, a six-year first-half and a seven-year second-half, is to use information from P1, particularly revenues in P1 from titles created in P1, to explain the number of new titles created in P2 by a set of authors beginning their careers in P1. Using this methodology, the revenues in the first period cannot be temporally affected by the then-future titles that are produced in the second period, thus eliminating the obvious simultaneity bias that would exist if we were to compare contemporaneous revenues and production. If anticipated revenues affect production of titles, however, the second period titles should be influenced by the revenues generated in the first period since those revenues are a predictor of P2 success for authors deciding whether to create new titles in P2. The reason for dividing our study period into two periods is that writing a book and having it published may take considerably longer than a year. In addition, using more than two periods would lead to a more complicated model if current productivity is a function of the whole history of revenue shocks in each period and not just in period t-1.

We undertake our analysis by using our data aggregated at the author level. We include P1 revenues in quadratic form to allow for nonlinearities such as backward bending labor supply curves discussed in the labor economics literature. We confirmed out choice of quadratic specification by running nonparametric regressions to see whether and what form of nonlinearity might exist between P2 titles and P1 revenues. In Appendix 4 we discuss those results which reinforced our choice of the inclusion of the quadratic term.

We include several additional independent variables likely to affect the production of titles in P2. One of those variables is the number of titles produced in P1. The influence of the number of titles produced in P1 on the number of titles produced in P2 should indicate, at least partially, the impact of the instinctive production of works independent of payment, since we are controlling for P1 revenues and because it seems reasonable to assume that this instinctive production of works is constant across periods.

We also include several author-level variables that may influence the number of books produced in P2. These include the nature of an author's book publishing arrangement (i.e., self-publishing, using a minor publisher, or using a major publisher), the nature of the author (an individual person or an organizational entity), and fixed-effects for the genre typically chosen by an author, since some genres are often associated with unusually high (e.g., comics) or low levels of production.[27] We run regressions for both the fiction and nonfiction book categories.

More formally, we consider the following empirical model for the number of titles authors write in period 2:

$$Number\ of\ titles\ in\ P2_i = \alpha_0 + \alpha_1\ Revenues\ in\ P1_i + \alpha_2\ (Revenues\ in\ P1_i)^2 + \alpha_3\ Number\ of\ titles\ per\ year\ in\ P1_i + \alpha_4 Z_i + u_i$$

[26] The yearly P1 productivity for authors publishing in both P1 and P2 is 0.60 and 0.98 for nonfiction and fiction respectively which is not much different than their productivity in P2.

[27] The book publishing arrangement and the genre are defined at the book level, not the author level. Because our data are aggregated at the author level, in our regressions for each author we use the mode of the nature of the book publisher and genre for that author.

where we use the variable $Revenues\ in\ P1_i$ as a proxy for expected revenues for period 2, control for $Number\ of\ titles\ per\ year\ in\ P1_i$ as a way to capture author-level baseline productivity that might be considered to be constant at the author level (e.g., utility or disutility from writing books, innate ability, instinctual production), and control for a vector of other author-level characteristics, Z_i.[28] We note that our model is similar to a model that includes fixed effects for each author or a model in differences (which are identical when there are two periods). Specifically, redefining the dependent variable into per year values, setting α_3 equal to one, and rearranging terms we obtain a dependent variable in differences $(Number\ of\ titles\ per\ year\ in\ P2_i - Number\ of\ titles\ per\ year\ in\ P1_i)$.[29] Therefore, the interpretation of our estimates is similar to the interpretation of estimates from fixed effects regression models (identification in our regressions does not arise from the cross-section of authors but from within-author changes in productivity). Although author-level productivity may change over time, it is not obvious to us how the speed of productivity change is likely to be distributed across authors with various baseline productivity levels.

We use this model to predict production in P2 due to revenues in P1, including predicting production in P2 for a zero level of revenues in P1 as an instance of the worst-case scenario for authors if copyrights were to be abolished. Such an approach has its limitations.

One possible criticism of our method is that we do not observe fame in our data, and fame is thought to be possibly more important for authors than for most other occupations. This might be a problem because fame is likely to be correlated with the top end of revenues, although the number of titles published in P1 might capture some of the effect of fame. To the extent that moneyless fame drives creation, and if fame and revenues are positively correlated, then our estimates of the coefficients on revenues in P1 would represent an overestimate of the pecuniary motive. Nevertheless, there are reasons to believe that moneyless fame is not an important productivity driver for most authors, particularly those who write nonfiction.[30] Most nonfiction authors write for a limited audience interested in a particular topic, such as (actual titles) "The Mechanics of Reinforced Concrete" or "Linear Algebra with Applications." The possibility of fame will not influence these authors for the simple reason that no one writing books on narrow nonfiction topics such as these ever expects to achieve wide readership and acclaim. Almost by definition, one's work must appeal to a general audience if the readership of the work is to possibly cause the author to become famous.

It is also the case that even best-selling authors do not usually achieve the level of fame awarded to movie stars, rock stars, or leading athletes, and it is misguided to compare authors with far more famous individuals just because they can be lumped into a broadly defined entertainment industry. Unlike movie or rock stars, how many authors are to be likely to be recognized and mobbed by fans whenever they go out in public? The question answers itself. Perhaps that is because public performance, which seems to be an important ingredient to achieve fame, is part of the job of athletes, actors, and musicians, but not so for authors. If we define "fame" more narrowly, to only include becoming known to a small association or guild of individuals, then writing such a moderately successful book is no different than getting a promotion

[28] Our objective is to measure the effect of expected revenues in P2, E(R2), on P2 number of titles, but we use revenues in P1, R1, because E(R2) are unobservable. If E(R2)=R1+random error, our estimates would provide an underestimate of the effect of E(R2) on the number of titles due to attenuation bias.

[29] Estimating rather than imposing that α_3 is equal to one would be preferable to avoid a specification error, although there would be a small efficiency gain from imposing that α_3 is equal to one if we knew that to be true (one fewer parameter to estimate).

[30] We note that the fact that a large number of authors do not earn enough money to make a living does not imply that authors are motivated by fame since authors are motivated by ex-ante and not ex-post earnings.

(with a raise) that becomes known to fellow workers, and labor economists do not usually consider the impacts of such narrowly defined fame when analyzing ordinary changes to employment prospects such as supply elasticity of workers.

Finally, the tradeoff between fame and income seems to lean heavily in favor of income, perhaps because income pays the bills and fills the stomach whereas fame does not. There are industries whose purpose is to increase individuals' income (e.g., higher education and especially business schools; personal finance books and columns) but seemingly no equivalent-sized institutions dedicated to increasing fame. We often find people who have achieved fame trying to monetize their fame by having autobiographies ghostwritten for them, selling autographed objects such as baseballs and photographs, and so forth. But there seem to be a smaller share of rich individuals desirous of fame (with a few notable exceptions). Successful actors, businessmen, and athletes seem to fight for the greatest monetary income they can receive even when it might hurt the nature of their fame, such as when athletes leave their longtime teams in response to a higher offer elsewhere. Few creators put their works into the public domain although it would maximize their fame by increasing consumption of their works and providing a newsworthy novelty. Few if any athletes, actors, or creators increase their fame by giving most of their earnings to charity.[31] When self-published authors become relatively famous they usually switch to a mainstream publisher even though the higher price of their books will decrease the size of their audience and diminish somewhat their fame if fame is based on the number of readers.[32] When the US did not have copyright, some English authors were famous in the US, yet those authors complained bitterly about their very small American remunerations in spite the extra readership and fame that came from having their works sold for a pittance (Liebowitz 2016).[33]

We note that if future research can find a natural experiment affecting most authors, it might be able to more cleanly focus on revenues separately from fame. Unfortunately, the 2018 American tax reform pass-through deduction does not seem to fit this bill since it does not affect all authors in the same way and determining how many and how intensely authors are affected by this tax change would present its own formidable difficulties.[34] In addition, a change in one instrument (the natural experiment) would not be enough to estimate a backward bending labor supply function. Our approach of using the number of titles published in P1 might capture some of the effect of fame.

Although we present regression results for one sample of authors in the main text, we ran these regressions over various samples of authors and we present some of these results in Appendices 2 and 3. In the results below, we look at a cohort of

---

[31] An article discussing the charitable giving of athletes can be found here. A few authors, such as Cory Doctorow (not related to E. L. Doctorow) who are known for their opposition to traditional copyright, have published successful titles forgoing full copyright for a creative commons license which provides less protection from copying although this is not the same as the author putting the work into the public domain. For example, some of Brandon Sanderson's early titles are listed as having a creative commons license although that does not appear to be the case for most of his later works.

[32] Evidence on this point can be found at https://www.dorrancepublishing.com/notable-self-published-authors/

[33] Charles Dickens is most famous for complaining about this during his first US speaking tour in 1842, although many English authors complained about the lack of copyright protection in the US. Dickens figured out how to better monetize his fame during his second, very lucrative, American speaking tour in 1867.

[34] Most authors would not pay any income tax based on just their income as an author, although they or their spouse may earn additional income that would require income tax payments. Almost half of all Americans do not pay any income tax and this would presumably include many authors. Authors whose income is above a certain level (several hundred thousand dollars) have their tax deduction limited. The tax law change for authors is described in this article: https://www.authorsguild.org/industry-advocacy/authors-take-advantage-of-the-new-tax-deduction/.

authors who first published a book in P1. We note that this excludes numerous authors. We also note that approximately two-thirds of our authors publish only one title in this thirteen-year period.[35]

Because authors who published their first book in the earlier years of P1 have more time to publish additional titles and generate larger revenues in P1 than do authors who first publish late in period P1, the number of titles and revenue generated in P1 are converted into per year values, to put these authors on a more even footing. There is no gain in converting the number of titles and revenues generated in P2 into yearly values because all authors in our sample, having published in P1, are assumed to have the same number of years in which to publish books in period two, since we cannot measure exit in P2 for such authors who publish early in P2, whereas we were able to measure entry in P1. In summary, we use data at the author level to examine whether and to what extent higher revenue per year during 2004-2009 influences the total number of titles written during 2010-2016, controlling for author-level baseline productivity and several author-level variables.

V.A Regression Results

Regression results are found in Table 2. In the regressions in Table 2, we measure revenues in million dollars. Interpreting the marginal effect of an increase in P1 revenues on the number of titles written in P2, however, is difficult from a visual inspection of the coefficients due to the quadratic specification of our empirical model. The coefficients on the linear component of the P1 revenues are positive and statistically significant for both fiction and nonfiction works. The negative and statistically significant coefficients on the quadratic component of the P1 revenues suggests a backward bend, although how many authors, if any, might be on this portion of the curve is unclear from the mere observation of the coefficients.

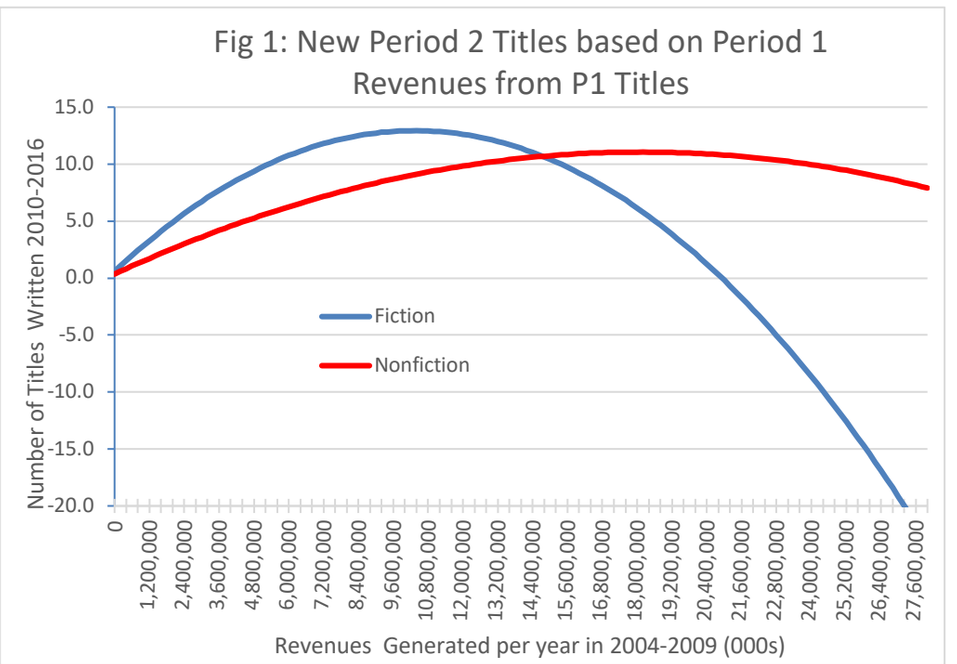
Table 2: # of P2 Titles written (Robust T in brackets)		
	Fiction	Nonfiction
P1 Revenue per year	2.39	1.17
	[-5.608]	[6.655]
P1 Revenue per year ^ 2	-0.11	-0.03
	[-5.020]	[-6.795]
P1 Titles per year	1.63	1.02
	[-13.94]	[10.66]
Company Author	-0.313	0.275
	[-8.372]	[3.516]
Major Publisher	1.1	0.0064693
	[-16.48]	[0.524]
Self-Publish	-0.116	-0.14
	[-6.396]	[-5.644]
Constant	-0.402	-0.135
	[-6.005]	[-3.089]

[35] Although table 1a provided the full number of nonfiction authors in the samples, it did not provide the full number of fiction authors since, in order to make the fiction authors comparable to nonfiction authors, we removed authors who did not sell 65 units in a year. This led to a majority of fiction authors being removed from those samples.

Genres (Fixed Effects)	73	1,983
Observations	62,634	105,781
R-squared	0.178	0.143

Figure 1 provides a graphic representation of the P1-revenue P2-production relationships that facilitates the interpretation of the marginal effect of an increase in revenues in P1 on the number of titles written in P2. The maximum value for new titles peaks at 13 and 11 new titles in P2 for fiction and nonfiction respectively. When the number of new titles begins to fall as revenue increases, authors are on the backward bending portion of their supply curve. To compare this diagram to a supply curve and its possible backward bending component, switch the X and Y axes with each other. The curvature for fiction is greater than the curvature for nonfiction, which is to be expected given the larger absolute coefficients for fiction.[36]

Another concern is the location of the backward bend. Authors of fiction reach their backward bend at lower yearly revenue levels ($10.4 million) than authors of nonfiction ($18.2 million). Remember that these values are revenues generated by sales and not the considerably smaller payments received by authors.



How many authors might be on the backward bending portion of the curve? The answer, for the sample of authors in the regressions, is 4 nonfiction authors who first published in P1, and 5 fiction authors. Thus, there appear to be very few

[36] Note that the intercepts for these curves differ from one another since these intercepts are computed at the mean values of all the independent variables except P1 revenues and P1-revenues squared, and these mean values are different across samples.

authors in the backward bending portion of their estimated supply function. For these 9 authors current copyright protection appears to be inefficiently strong.[37]

Nevertheless, the top nonfiction and fiction authors generate about $37 and $25 million respectively per year during their period of writing, well into the backward bending portion.[38] These results make it possible to examine whether the existence of book revenues, due to say, copyright, increases the number of books written compared to a case where authors received no payment of any sort. As long as the curves in Figure 1 lie above the intercept for zero revenues in P1, the existence of P1 revenues increase P2 titles. Thus, in order for the overall impact of P1 revenue to have a negative net impact on an author's P2 titles, the yearly revenue would need to be $39 million for nonfiction and $21 million for fiction. Since there are zero nonfiction authors generating revenues of this size, author payments clearly have a positive impact in the nonfiction market compared to a regime in which payments would be zero. There are 2 fiction authors who generate sufficient revenues that the existence of payments might reduce their overall book production. But even in this case it would seem very unlikely that the overall impact of a payment regime such as copyright would not increase overall book production across all authors, even though this does not tell us whether copyright provides too much, too little, or just the right amount of protection.

Briefly discussing the other variables included in the regressions, we find that the number of titles that authors produce in P2 is positively related to the number of titles they produce in P1. We take the regression coefficient on P1 to measure authors' baseline productivity, independent of P1 revenues.

Organizations publishing books appear to publish more than average in the case of nonfiction works during P2, which is not surprising given some of the well-known organizations publishing a very large number of titles. Somewhat surprisingly, the small number of organizations publishing works of fiction appear to publish fewer titles than average, but this may be due to imperfection in our method of counting company authors, particularly our treatment of authors with single names as indicating an organization when in fact it might just be an error in the author data field and thus be associated with fewer titles published on average. We cannot definitively explain, however, the conflicting signs of this coefficient. The use of a major publisher is positively related to a greater number of P2 works being published in the case of fiction, but there is little relationship for works of nonfiction. Authors who self-publish do not have P2 publication numbers that differ meaningfully from other authors since the coefficients are very small in size (even when they are statistically significant).

V.B The Revenue Elasticity of Creation

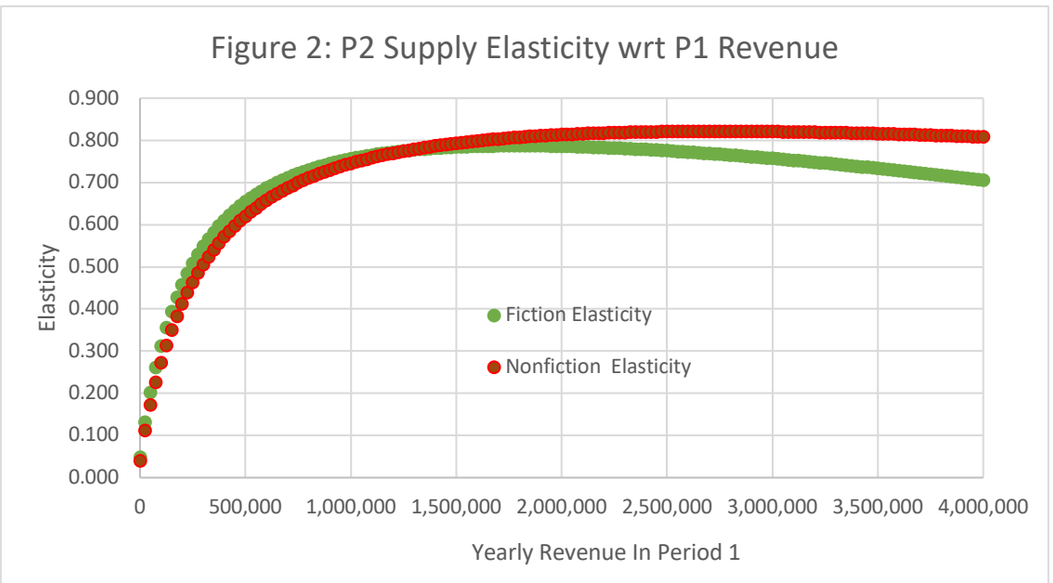
We can dive somewhat more deeply into the results already generated to better understand the responsiveness of creation to financial gain. We begin by estimating the elasticities of supply of P2 titles with respect to changes of P1 revenues. This

[37] We note that copyright may be too strong even when authors are not in the backward bending region of their supply curve. We would need to compare the social benefit of the new work with the author's cost of production for each work to determine optimal copyright. When supply is backward bending, there are fewer new works due to copyright, meaning that the author payment is inefficient since it reduces the production of new works.

[38] Notice that these authors are not necessarily the authors generating the greatest revenues during the 2004-2009 interval because we are only looking at authors first publishing in P1, we are only including revenues from works created in P1, and because we only count the number of years between the first title and the end of the period when we construct the "per year" values. The correlation between revenue per year and overall revenue across authors, however, is .85 for nonfiction and .90 for fiction.

relationship, for both fiction and nonfiction, can be seen in Figure 2 which illustrates the elasticity of new P2 titles with respect to changes in P1 revenue.[39]

Figure 2 was constructed by using the estimates from Table 2. For yearly revenues in the range of $300,000 (which average long-term fiction authors earn), elasticities range between 0.10 and 0.50. Peak elasticities occur when revenues are in the low million dollars, and the peak elasticities are approximately 0.80. These elasticities fall back to zero at the revenue level associated with the backward bend. At very low revenue values such as $5,000 or $10,000, which are below average but close to the median levels, the elasticities are in the range of 0.10, which appear to be relatively high compared to the near-zero values found in the labor literature.



At yearly revenue ranges sufficiently large to support an author in a middle-class lifestyle ($500,000-$1,000,000 of revenue which translate into an income level of $50,000-$100,000 assuming a 10% royalty rate), these elasticities are considerably larger than have been found in the labor literature measuring hourly wage elasticities of male workers, usually measured near the point of full-time employment. Our elasticities are even higher than female hourly wage elasticities which are generally thought to be higher than those of their male counterparts, but there is evidence that female labor elasticities have been in decline, approaching the levels of men (Bishop, Heim, Mihaly, 2009, Kumar and Liang, 2016).

We take these results to indicate that individuals who can support themselves as authors (an income level of $50,000-$100,000, implying revenues of ten times that level), whose sales make up a large portion of total sales,[40] have supply elasticities that are quite high relative to workers that have been examined in the labor literature. These high supply elasticities imply that these authors respond robustly to monetary incentives. It is not clear, however, how comparable our elasticities are to those in the labor economics literature since the decision to publish a book is a joint decision made by authors and publishers, although we reiterate that whether authors or publishers make the decision to publish is mostly

[39] The range of values on the horizontal axis is truncated in order to make it easier to see the relationship at the more common lower revenue values.
[40] Author's generating revenues in this range or above, are responsible for about 38% of all nonfiction revenues and 59% of fiction revenues, although they make up less than 1% of all authors.

irrelevant if the focus of the analysis is on how monetary compensation affects creative output available to consumers in the market.

V.C The Share of Creation Due to Monetary Rewards

In an attempt to gauge the impact of prior revenues on new production, a relatively simple calculation is to use the coefficients from the regressions to generate a predicted number of titles in P2 for each author in our sample, based on the revenues generated in P1 by that author. We then form a predicted revenue for each author in P2 based on the number of predicted titles in P2 multiplied by the revenue per title for each author in P2, or, if they do not sell titles in P2, revenue per title in P1. We sum those predicted revenues and compare the sum to the total actual revenues generated by these authors in P2.

We make this calculation as follows (where PGT stands for *predicted generated titles in P2,* R stands for *revenues per year in P1,* and COEF and QCOEF stand for the linear and squared coefficient estimates on revenue in Table 2). First, we predict the number of new titles in P2 for each author as follows:[41]

$$(1)\quad PGT_{i,P2} = R_{i,P1} \cdot COEF + R_{i,P1}^2 \cdot QCOEF$$

The first row of Table 3 provides the sum of PGT over all individuals in our sample. The results from this calculation show that the additional production of P2 works caused by P1 revenues amounts to only 1,165 fiction and 3,092 nonfiction titles, which represent only a small share (3.08% and 7.69% respectively) of the actual total number of titles produced in P2 by the authors in this sample (see second row of Table 3).[42] These results suggest that most books would exist even without monetary rewards. Specifically, 97% (92%) of all fiction (nonfiction) books in P2 would exist even with zero revenues generated in P1.

However, we also need to consider the quality of these titles (measured by their revenues) that would not have been written had revenues in P1 been zero, which varies by author. This is important due to the long tail nature of the book industry. One way to approximate quality of predicted P2 titles is by using the average revenue per title generated by each of these authors in P2, and then multiplying that value by the predicted number of new works produced by each author due to P1 revenues. For those authors who do not publish in P2 but are predicted to publish in P2, we use the average revenue of P1 titles, a less than ideal solution as we explain below. Summing these values over all authors provides an estimate of the aggregated revenues expected to be generated by new P2 works due to P1 revenues. That is, we make this calculation as follows (where PGR stands for predicted generated revenues in P2, *PGT* stands for predicted P2 titles induced by P1 revenues, and *Rev/Titles* represents authors' average quality either in P2 for authors publishing in both periods or in P1 for P1-only authors):

---

[41] It is unclear whether those few individuals with predicted number of titles less than zero should have those values truncated at zero, since it is not possible to produce a negative number of books. This concern is moot for works of nonfiction in our data since there are no authors with a negative number of predicted titles. There are only two fiction observations with a negative quantity of predicted titles, but one of those authors is the largest fiction seller in our sample and has some influence on the results. We have chosen to allow the negative predicted values for these authors, which lowers our measured impact of revenues on the production of new titles by approximately 2 percentage points.

[42] Note that the new titles produced are the sum of mainly fractional titles calculated over many authors. The number of authors increasing their production of P2 books is therefore much larger than the number of titles predicted to be created in P2. Many authors with fractional predicted increases in written P2 titles will not actually produce any titles in P2. This logic is consistent with the assumption that these authors not producing a work in P2 remain in the market.

$$(2)\ \text{PGR} = \sum_{i=1}^{n} \frac{Rev\ (P2\ or\ P1)_i}{Titles\ (P2\ or\ P1)_i} \cdot PGT_{i,P2} \quad \text{for n authors}$$

Table 3: Impact of Period 1 Revenues on Period 2 Production		
	Fiction	Nonfiction
Generated # of New Titles	1,165	3,092
Share of all P2 Titles	3.08%	7.69%
Generated Revenue from Predicted Induced Titles	$452,000,000	2,650,000,000
Actual P2 Revenue from Sample	$1,260,000,000	$4,020,000,000
Generated Revenue as Share of actual P2 Revenue	35.98%	65.89%

The results from this calculation are presented in the third row of Table 3, which shows that revenues from predicted P2 titles (induced by P1 revenues) are expected to be $452 million for the fiction category and $2.65 billion for the non-fiction category, or, as found in the bottom row, about 36% and 66% of the actual P2 revenues for the fiction and non-fiction markets, respectively. The large difference between the share of predicted P2 titles and the share of predicted P2 revenues generated by P1 revenues indicates that the predicted titles generated by P1 revenues are of much higher quality than the average title, about ten times as great, on average. The P1 revenues, therefore, appear to be a very important inducement for the writing of high-quality books, particularly works of nonfiction.

We can look somewhat more deeply behind the creation of the numbers in Table 3. First, we can examine the composition of the predicted P2 titles induced by P1 revenues. Because approximately 80% of authors in our sample published in P1 but not P2 ("P1-only"), it might naively be thought that most predicted P2 titles would come from P1-only authors. But the predictions from the regressions provide useful information about which authors are most likely to produce titles in P2, so instead of a naïve prediction of P1-only authors being expected to produce 80% of the P2 titles, the regression predicts that P1-only authors will produce 54% of the nonfiction titles and 28% of the fiction titles used in these back of the envelope calculations (the basis for these numbers can be found in Appendix 6). A large fraction of predicted new works, for the summed samples of fiction and nonfiction, therefore, comes from authors who publish in both P1 and P2.

In addition to the composition of P2 titles induced by P1 revenues, there is also the question of revenue generated per predicted title, assumed to be similar to the revenues of an author's actual titles. We do not think that this assumption should be controversial. Obviously, if revenues per title were completely random, this assumption would lead to erroneous results. There is much empirical evidence, however, that that the success of individual titles is not a lottery, and that the odds of a formerly successful author doing well on a new work are better than the odds of a formerly unsuccessful author doing well on a new work. The empirical facts that publishers make advances for unstarted or unfinished manuscripts, and the existence of multibook contracts between authors and publishers, support the view that the success of an author's new titles can be predicted, to some extent, by the success of the author's old titles.

Our use of two different proxies (average revenues per title in either P1 or P2) for predicted P2 revenues per title might seem potentially problematic due to the fact that P1 revenue per title could be different from the level of P2 revenue per

title (e.g., due to inflation or a change in demand) since there is a seven year differential between them.[43] Inflation, as already mentioned, was quite moderate during this period. Looking at average values is somewhat ambiguous. For all titles written in P1 and P2, the P2 revenue per title values are 58% and 14% smaller than P1 values, for fiction and nonfiction respectively. Limiting the sample to authors first publishing in P1, as in our regressions, P2 revenue per title values (for all published titles) are 53% and 40% higher than P1 values, for fiction and nonfiction. The P1 values for P1-only authors, the authors for whom the P1 average revenue per title proxy is used, is 72% and 44% lower than the P2 values used by the two-period authors.

The calculations found in Table 3, however, weight each author by the predicted number of P2 titles, as seen in equation (2). As demonstrated in Appendix 6, the calculated P1 average revenue per title (for P1-only authors) is higher than the calculated P2 revenue per title. As discussed in Appendix 6, the calculations for P1-only authors might bias upward the back of the envelope measures of the importance of past revenues in new works.[44] Appendix 6 also demonstrates that a rough correction of this potential bias using assumptions that we think are likely to excessively reduce the impact of past revenues on new works, would lower the bottom line Table 3 values to 31% and 45% for fiction and nonfiction, respectively.

Another concern that might arise with respect to these back of the envelope calculations is whether the difference between fiction and nonfiction results reported in Tables 2 and 3 are influenced by the fact that nonfiction works selling fewer than 65 units in a year are not included in the data, whereas all works of fiction are included in the data. Appendix 7 reruns the fiction analysis excluding observations with fewer than 65 units sold per year and demonstrates that although the differences between the fiction and nonfiction results get smaller, the differences are not due primarily to truncation of the nonfiction data.

These back of the envelope calculations indicate that, on the whole, increased payments to authors are associated with increases in their output. This is in spite of the finding that the supply is backward bending, but consistent with the fact that very few authors reach the backward bending component. This result is also contrary to some claims that authors of creative works do not respond to monetary reward, although it is compatible with the possibility that some authors, perhaps many, are not influenced by pecuniary motives.

However, the magnitude of generated output due to revenues varies substantially across our samples from a minimum of approximately 14%[45] of the total output value generated by monetary incentives to a maximum of 65%. Overall, all our results indicate that P1 revenues induce a significant fraction, although not necessarily a majority, of the revenues from new titles created in P2.

---

[43] The quality of an author's work, as measured by sales, is positively correlated over time, although the correlation is far from perfect. The unconditional correlations between the per title revenues of an author's P2 works and P1 works (for the sample of authors writing in both periods) are 0.40 for fiction and 0.34 for nonfiction.

[44] Authors with positive shocks in their P1 sales (authors who sell more than their actual quality would imply) will be predicted to produce and sell more works in P2 (i.e., they will have a positive shock in P2 predicted titles and revenues). For P1-only authors, the positive shock in P2 predicted titles is exacerbated by its product with the positive shock in the P1 average revenue per title proxy variable that is used in creating the predicted generated revenues amount (see equation 2). The net effect would be to overestimate the impact of P1 revenues on induced P2 titles and revenues.

[45] In Appendix 3 we present regression results for a 'long-term' sample of authors who published prior to P1 as well as in P1 and P2. This sample includes particularly active and successful authors. For this sample, the share of P2 revenue attributed to titles induced by P1 revenues was only 14% and 21% for fiction and nonfiction respectively. For this sample, the values of P2 titles produced is truncated at 1 unit which makes this sample less representative of the population than the sample we use in the main text.

VI. Conclusions

Copyright and patent systems are economically justified based on the notion that revenues induce creativity, but it is surprising that the evidence for this relationship is only anecdotal or based on data without a revenue component. This evidentiary lacuna is an important part of the controversy over intellectual property systems.

This paper links the literatures on labor economics and innovation. The literature on innovation has not considered the question of how creativity responds to pecuniary rewards as a special case of how workers change their willingness to work in response to changes in payment.

To empirically determine what role, if any, financial payoffs have in the production of creative works, we have used a rich and detailed data set that provides information on virtually the entire population of published authors in the United States over a thirteen-year period. To our knowledge, such a complete data set on creators and their revenues over time has never been the basis for such an examination.

Our results suggest that financial payments are an important motivation, perhaps the most important, for the creation of books. This conforms to our general understanding that writing books is work, and that many authors have an upward sloping supply curve of effort leading to new works, in response to monetary payoffs.

The upward sloping supply curves that we estimated, however, have a backward bend, although only a very small number of authors achieve sufficiently high revenues to enter the backward bending region, and the output from these authors was not large enough to overwhelm increased production from authors on the normal portion of the curve. Further, the relationships we found revealed almost no instances where monetary payments would reduce author output to a level below what would be produced with no payment at all.

We also find that authors successful enough to support themselves through their writings (but not in the backward bend of the supply curve) have supply elasticities with respect to revenue that are much higher than is typically found for workers with respect to wage rates. Supply elasticities for the majority of authors whose revenues are insufficient to support themselves, however, are closer to the more typical elasticities found in the literature.

But our results also leave room for innate urges and other non-monetary factors to play a significant role in the creation of books, conforming to anecdotal empirical evidence of authors who produce works without any expectation of monetary payoffs.

These results are consistent with the possibility that authors are not uniform in their motivations to create and that authors may have multiple reasons for authoring. In other words, some authors are, at least in part, blockheads in the sense of Dr. Johnson's quote with which we began this article, while others fit into the more pragmatic mold he set for himself. We find it unsurprising that authors do not all have the same motivation to create.

There are many questions still unanswered. Do our results hold for other types of creative endeavors and for other time periods? Would the inclusion of demographic characteristics for each author have changed our results? Would our results hold when the works are created by larger teams and not mainly by individuals? Does the relationship between revenue and production change over the life cycle of individual authors?

We view our analysis as only a step in a journey toward compiling useful empirical evidence on the pecuniary motivations to create.

Benkler, Yochai. 2006 "The wealth of networks: How social production transforms markets and freedom" Yale University Press.

Boyle, James. 2003 The Second Enclosure Movement and The Construction Of The Public Domain "Law and Contemporary Problems" 66:33 Winter/Spring.

Bishop, Kelly, Bradley Heim, and Kata Mihaly. "Single women's labor supply elasticities: trends and policy implications." *ILR Review* 63, no. 1 (2009): 146-168.

Blundell, Richard, Antoine Bozio, and Guy Laroque. "Labor supply and the extensive margin." *American Economic Review* 101, no. 3 (2011): 482-86.

Boldrin, Michele, and David K. Levine. 2004. "The Case Against Intellectual Monopoly." *International Economic Review* 45, pp. 327-50.

Boldrin, Michele and David K. Levine, *Against Intellectual Monopoly,* Cambridge University Press, 2008.

Borjas George J., Labor Economics Sixth Edition McGrow-Hill Irwin 2013.

Danaher, Brett, Samita Dhanasobhon, Michael D. Smith, Rahul Telang. 2015. Understanding Media Markets in the Digital Age: Economics and Methodology, Economic Analysis of the Digital Economy. Chapter 13 (pp. 385-406), University of Chicago Press, Chicago, Illinois.

Goolsbee, Austan D. "What Happens When You Tax the Rich? Evidence from Executive Compensation," *Journal of Political Economy*, April 2000, vol. 108(2), 352-378.

Kumar, Anil, and Che-Yuan Liang. "Declining Female Labor Supply Elasticities in the United States and Implications for Tax Policy: Evidence From Panel Data." *National Tax Journal* 69, no. 3 (2016): 481.

Liebowitz, Stan J. "Is efficient copyright a reasonable goal." *Geo. Wash. L. Rev.* 79 (2010): 1692.

Liebowitz, Stan J., "Paradise Lost or Fantasy Island? Voluntary Payments by American Publishers to Authors Not Protected by Copyright," *The Journal of Law and Economics* 59, no. 3 (August 2016): 549-567.

Lunney, Glynn. Copyright's Excess: Money and Music in the US Recording Industry Cambridge University Press 2018.

MaCurdy, Thomas E. 1981. "An Empirical Model of Labor Supply in a Life-Cycle Setting," Journal of Political Economy, vol 89(6), pages 1059-1085.

Moser, Petra. "Patents and innovation: evidence from economic history." Journal of Economic Perspectives 27, no. 1 (2013): 23-44.

Plant, Arnold. 1934. "The economic aspects of copyright in books." Economica 1:167-195.

Scherer, Frederic M. "Quarter notes and banknotes: The Economics of Music Composition in the Eighteenth and Nineteenth Centuries," Princeton, NJ (2004): 142-154.

Scherer, Frederic M. "The Emergence of Musical Copyright in Europe from 1709 To 1850" *Review of Economic Research on Copyright Issues*, 2008, vol. 5(2), pp. 3-18.

Selgin, George, and John Turner. "James watt as intellectual monopolist: Comment on Boldrin and Levine." *International Economic Review* 47, no. 4 (2006): 1341-1348.

Selgin, George, and John L. Turner. "Watt, again? Boldrin and Levine still exaggerate the adverse effect of patents on the progress of steam power." *Review of Law & Economics* 5, no. 3 (2009): 1101-1113.

Selgin, George, and John L. Turner. "Strong Steam, Weak Patents, or the Myth of Watt's Innovation-Blocking Monopoly, Exploded." *The Journal of Law and Economics* 54, no. 4 (2011): 841-861.

Smith, Michael D., and Alejandro Zentner. Internet effects on retail markets. In, Basker, E., (ed.), Handbook on the Economics of Retailing and Distribution, Northampton, MA: Edward Elgar Publishing, 2016.

Taussig, Frank William "Inventors and Money-makers, Lectures on Some Relations Between Economics and Psychology." Delivered at Brown University in Connection with the Celebration of the 150th Anniversary of the Foundation of the University – January 1, 1915 New York: Macmillan.

Waldfogel, Joel, "Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster," *Journal of Law & Economics*, November 2012.

Wu, Yanhui, and Feng Zhu. Competition, Contracts, and Creativity: Evidence from Novel Writing in a Platform Market, Working Paper. August 2018.

Appendices

Appendix 1: Industry Background

Only 14.8% of titles sold in a given year were first published during that year.

The reason that new titles make up such a small share of the market is that book titles last, on average, quite a long time. The full life cycle of titles can be seen in Figure A1, which shows the degree to which sales decay over time relative to the first year of publication, for new works of fiction and nonfiction first published in 2004. Sales of fiction titles decay slightly faster than nonfiction, but the difference is small.[1]

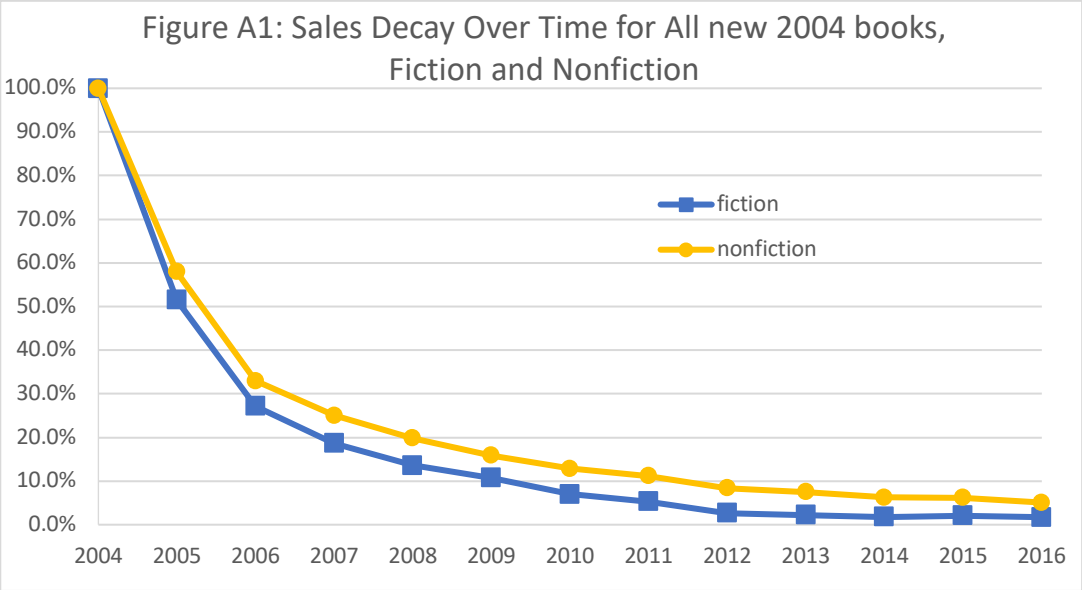
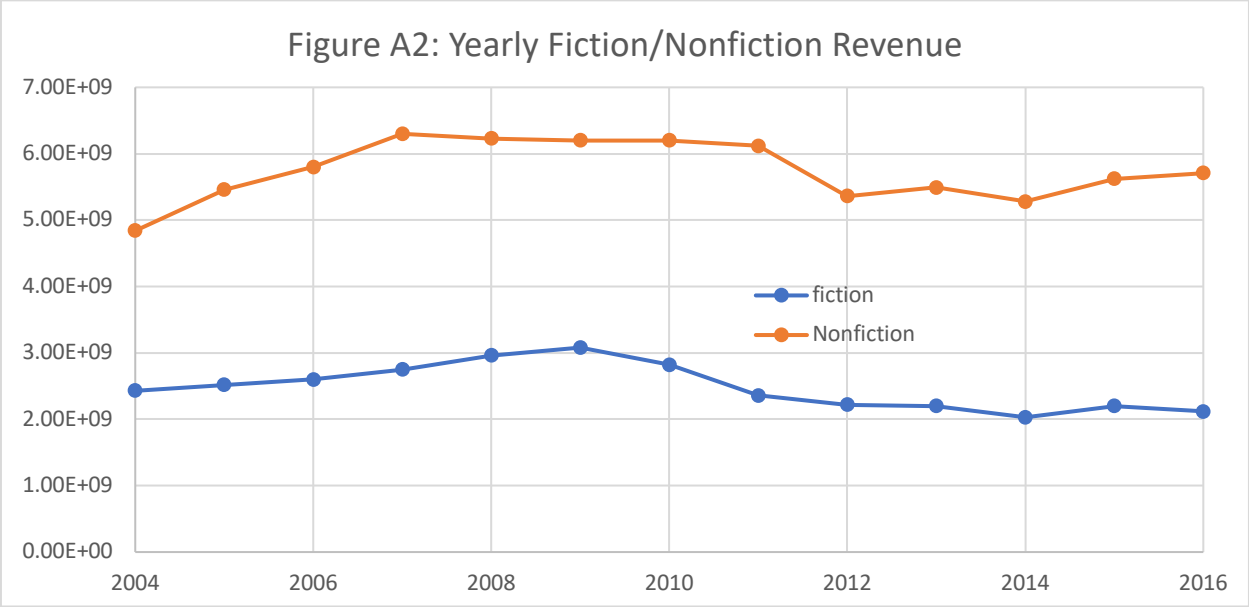


Figure A2 shows the seeming interruption of an upward trend in the revenues of this industry that coincides with the introduction of e-books, which might suggest substitution between e-books and print. Note that sales of e-books are not included in these data, so we can conclude that there likely was an increase in revenues for the entire book industry over this period if we were able to include e-book revenues.

[1] The difference in decay rates between these groups is somewhat smaller when works of fiction selling fewer than 65 copies are removed (as is the case for works of nonfiction due to data availability), making for a more comparable sample.



Appendix 2 – How e-books affect our estimates?

The exclusion of e-book from our data could affect our empirical estimates (e.g., the data excludes sales from the Amazon Kindle Store). For example, due to this data limitation we would underestimate period 2 productivity for authors who published book titles in period 2 exclusively in e-books format (and not in print format).

One way to address this concern is to restrict our examination to a period of time when e-books were not a viable option. While e-books existed even before the beginning of our study period (year 2004), the development of this market was in its infancy before year 2010 (https://files.eric.ed.gov/fulltext/EJ1032678.pdf). We could thus gauge how e-books affect our estimates by restricting our analysis to years 2004-2009.

We replicate the examination in the main paper but redefine periods 1 and 2 to include years 2004-2006 (P1) and years 2007-2009 (P2). Table A1 and A2 replicate Tables 2 and 3 in the main paper. The results are consistent with the results in the main paper. For fiction, Table A3 suggests that revenues in P2 would have been 60.3% lower in P2 than the level observed in P2 had revenues in P1 been zero (compared to 36% in Table 3 in the main paper). For non-fiction, Table A4 suggests that revenues in P2 would have been 61.9% lower in P2 than the level observed in P2 had revenues in P1 been zero (compared to 65.8% in Table 3 in the main paper).

Table A3: # of P2 Titles written (Robust T in brackets)

	Fiction	Nonfiction
P1 Revenue per year	2.15	0.696
	[5.601]	[8.199]
P1 Revenue per year ^ 2	-0.313	0.038
	[-4.147]	[-5.173]
P1 Titles per year	.879	.442
	[16.53]	[11.84]
Company Author	-.233	.139
	[-5.104]	[2.222]
Major Publisher	.587	.02726
	[11.98]	[2.989]
Self Publish	-.19	-.169
	[-16.10]	[-15.75]
Constant	-0.272	-.0816
	[-6.853]	[-3.275]
Genres (Fixed Effects)	59	1,841
Observations	26,549	53,676
R-squared	0.254	0.131

Table A4: Impact of Period 1 Revenues on Period 2 Production

	Fiction	Nonfiction
Generated # of New Titles	491	1,001
Share of all P2 Titles	4.49%	7.64%
Generated Revenue from New Titles	$204,000,000	$689,000,000
Share of P2 Revenue	60.30%	61.95%

Appendix 3: Long-Term Authors

The regressions we presented in the main paper limit the analysis to authors debuting in period 1. While this has the advantage of avoiding life cycle considerations, these samples are not necessarily representative of other swaths of the population of authors. In this appendix we replicate our analysis using a sample of authors who published prior to P1, as well as in both P1 and P2. This is a sample of interest since these authors are particularly active and successful. We label the period prior to P1 as P0.

Conducting the analysis using this sample has a few caveats since we do not observe revenues prior to year 2004, although we do observe revenues in years 2004-2016 for successful books that still sell during our study period that these authors published before 2004.

We adjust the empirical model used in the main text for the number of titles authors write in period 2:

$$Number\ of\ titles\ in\ P2_i = \alpha_0 + \alpha_1\ (Revenues\ in\ P1_i)^* + \alpha_2\ ((Revenues\ in\ P1_i)^2)^* + \alpha_3\ Number\ of\ titles\ in\ P1_i + \alpha_4\ Number\ of\ titles\ in\ P0_i + \alpha_5 Z_i + u_i$$

where the variables $(Revenues\ in\ P1_i)^*$ and $((Revenues\ in\ P1_i)^2)^*$ in the adjusted model represent the revenues and the square of revenues per year in period 1 from books published both during period 1 and before period 1. The model also includes the variable $Number\ of\ titles\ in\ P0_i$ representing the number of titles per active year published prior to 2004 (between years 1980 and 2003), which together with $Number\ of\ titles\ in\ P1_i$ control for authors' intrinsic productivity levels.

Table A5: # of P2 Titles written (Robust T in brackets)

	Fiction	Nonfiction
P1 Revenue from P0 and P1 titles per year	0.528	1.52
	[1.950]	[1.698]
P1 Revenue from P0 and P1 titles per year ^ 2	-0.0121	-0.119
	[-0.923]	[-1.104]
P1 Titles per year	4.223	3.454
	[16.93]	[6.444]
P0 Titles per year	-0.1069	0.2887
	[-0.651]	[0.945]
Company Author	1.641	9.29

	[0.791]	[3.791]
Major Publisher	0.2461	-0.2469
	[1.352]	[-1.678]
Self-Publish	0.3689	4.001
	[1.119]	[1.747]
Constant	1.227	0.6581
	[6.843]	[2.166]
Genres (Fixed Effects)	55	1,347
Observations	6,130	14,582
R-squared	0.519	0.376

The results in Table A5, which mimic Tables 2 in the main paper, are consistent with the results in the main paper. Table A6 shows the percentage of revenues in P1 due to prior revenues, and these percentages are substantially smaller than those in Table 3 in the main paper.[2]

Table A6: Impact of Period 1 Revenues on Period 2 Production		
	Fiction	Nonfiction
Generated # of New Titles	444	2,499
Share of all P2 Titles	1.44%	5.26%
Generated Revenue from New Titles	$439,000,000	$1,210,000,000
Share of P2 Revenue	13.72%	20.38%

[2] We also examined a sample of authors who published their first book in P1 and also published at least one book in P2, to get a feel for how payments affect production along the intensive margin, with no entry or exit in P2. The results for this sample are similar to the results for long-term authors.

Appendix 4: Non-Parametric Estimates

To allow the nonlinearities of the backward bending labor supply curve discussed in the labor economics literature, in the main text we assumed that the number of books written in period 2 is affected by both revenues in period 1 and the square of revenues in period 1. An alternative is to be agnostic about the functional form and estimate a non-parametric regression that does not assume such a quadratic relationship.

One problem with non-parametric estimation is that it is much less efficient than parametric estimation, and the number of observations needed increases with the number of covariates (i.e., curse of dimensionality). While we have many observations for authors with relatively low levels of revenues in periods 1, there are very few authors with high levels of revenues as required to non-parametrically estimate the effect of revenues in period 1 on the number of titles written in period 2. Therefore, we cannot include all the covariates we included in the main text when running non-parametric regressions. We only include the amount of revenues in period 1 and the number of titles written in period 1 as covariates. Even limiting our regressions to include these two covariates the marginal effect is not estimable for high levels of revenues in P1 where the number of observations is limited.[3] Table A7 presents non-parametric regression estimates of the marginal effects computed at various levels of revenues in period 1.[4]

Table A7: Non-Parametric Estimates

Predicted Mean Number of Titles in P2 based on Revenues in P1

Revenues in P1 in US$	Fiction	Non-Fiction
10,000	0.6362955	0.3516525
100,000	1.420567	0.5544971
500,000	4.356943	1.373541
1,000,000	6.898817	2.227732
2,000,000	4.307746	3.282402
5,000,000	1.021679	1.583435

[3] It is unsurprising that the marginal effects are not estimable for high levels of revenues in P1 since there are few authors in the upper tail of the distribution of authors (e.g., 9 fiction (36 non-fiction) authors have revenues higher than US$5,000,000 in period 1).

[4] We computed these regressions using the Stata 15 command "npregress kernel" and the mean number of titles in P2 was computed using the command "margins" at the default mean level of the number of titles in P1.

10,000,000	not estimable	0.3457805
20,000,000	not estimable	not estimable

The number of titles in P1 is included as a control variable.

The results in Table A7 show that higher revenues in period 1 predict a higher number of titles written in period 2 for lower levels of revenues in period 1 and a backward bend for high levels of revenues in period 1. While the backward bend is consistent with the backward bend we showed in the main text, the backward bend occurs for considerably lower levels of revenues in period 1 compare to the results in the main text.

Since our main goal in running the nonparametric regressions was to confirm whether our choice of a quadratic functional form makes sense, we believe that the results indicate that it does. With that conclusion, we believe it is best to use the more efficient parametric estimates found in the paper.

Appendix 5: Impact on summary statistics from truncating the fiction data to those observations with less than 65 units sold

The main text describes the summary statistics for works of fiction when observations with fewer than 65 units sold are excluded (Tables 1a and 1b). This was done to put fiction and nonfiction works on an equal footing and thus make them comparable to one another. The resulting summary statistics in the text, therefore, do not reflect the actual values for the entire fiction sample. The following tables present the summary statistics including fiction authors who sold less than 65 units.

Tables A8a and A8b - Summary Statistics - Fiction

Table A8a: List-Price Sales Revenues for Authors					
variable	N	Mean	Median	min	max
Yearly P1 rev from P1 books	64,129	$9,373	$88	$1	$24,900,000
Yearly P2 rev from P2 books	64,129	$3,764	$0	$0	$9,767,030
Yearly P2 rev from P2 books (only authors with new P2 books)	13,371	$18,054	$186	$1	$9,767,030

Table A8b: New Titles					
variable	N	mean	p50	min	max
Titles per year published P1	64,129	0.59	0.50	0.17	14.33
Titles per year published P2	64,129	0.16	0.00	0.00	17.86

Titles per year published P2 (excludes zero)	13,371	0.76	0.50	0.14	17.86

The first item to note is that the number of observations is much larger in tables A8a and A8b than in tables 1a and 1b in the main text, with roughly four times as many observations. As expected, the mean values of revenues are much lower in these tables, roughly between a third and a fourth of the values in the main text. The median revenue values are lower by a much larger percentage, except for the variable "Yearly P2 rev from P2 books," which is zero in both the tables in the main text and here. The number of titles is also smaller in tables A5a and A5b than in tables 1a and 1b in the main text, but much less so than for the revenue variable.

Appendix 6: The Possible Bias from Using P1 Revenue/Title as a Proxy for P2 Revenue/Title

Table A9 provides more detail on the predicted P2 titles and revenues due to P1 revenues, for fiction and nonfiction works. All the rows that represent calculated values have a formula showing the calculation where the term "R" refers to the row with the data, usually from this table but sometimes referencing Table 3 in the main text. The non-calculated values were generated from our data set as explained in the paper.

Rows 1 and 2 provide the number of authors who publish only in P1 or who publish in both P1 and P2. Row 3 calculates the share of authors who publish only in P1.

Rows 4 and 5 provide the number of predicted P2 titles induced by P1 revenues for authors who publish only in P1 or who publish in both P1 and P2, respectively. Row 6 uses that information to determine the share of new titles due to authors who only publish in P1 (and not in P2).

The text mentions the 80% of authors who publish in P1 only, found on the third line of this table, and the 28% and 54% of new titles induced by P1 revenue for P1-only authors found on line 6.

Rows 7 and 8 provide details on the predicted generated revenues for authors publishing in both periods or P1-only authors, respectively. These predicted revenues, as shown in equation (2) in the text, are the product of the proxy for revenue per title and the number of predicted generated titles, summed over all authors.

Rows 9 and 10 calculate average revenue per title for the predicted titles in P2 based on P1 revenues for two-period authors and P1-only authors, respectively. The P1-only authors have their predicted revenues calculated using the P1 revenue/title as a proxy. The two-period authors have their predicted revenues calculated using P2 revenue/title as a proxy. Thus the generated revenue/title results in rows 9 and 10 are merely the (total generated revenue)/(total generated titles) for fiction and nonfiction works.

Table A9: Examining Data to Inform Use of P1 Rev/Title as Proxy		
	Fiction	Nonfiction
1. Number of P1-only authors	50,172	84,940
2. Number of two-period authors	12,461	20,841
3. P1-only share of authors R1/(R1+R2)	80.10%	80.30%
4. Generated # of New Titles from P1-only authors (using equation 1)	321	1,680
5. Generated # of New Titles from two-period authors (using equation 1)	844	1,413
6. P1-only share of generated new titles R4/(R1 Table 3)	27.59%	54.32%
7. Generated revenue from two-period authors (using equation 2)	$281,000,000	$833,000,000
8.Generated revenue from P1-only authors (using equation 2)	$171,000,000	$1,817,000,000
9. Generated rev per title in P2 for two-period authors R7/R5	$333,093	$589,718
10. Generated rev per title in P2 for P1-only authors R8/R4	$532,084	$1,081,741
11. P2 generated rev/title compared to P1 generated rev/tit R9/R10	62.60%	54.52%
12. Adjusted generated revenue from P1-only authors R8*R11	$107,048,574	$990,549,123
13. Adjusted Induced P2 Revenue as share of actual P2 rev (R12+R7)/(R4 Table 3)	30.80%	45.36%

Row 11 compares the predicted generated revenue/title for P1-only authors using the P1 value as proxy with the predicted revenue/title for two-period authors where the P2 revenue/title value was used as a proxy. The results show that the P1 revenue/title proxy is a larger number than the P2 proxy. If the higher P1 values was thought to be due to a bias in the analysis (as we explain in footnote 41 in the text and in the following paragraph), then the portion of predicted generated revenue coming from P1-only authors would be biased upward.

It could be argued that such a bias exists because authors in P1 who experience a positive random shock in sales will be predicted to produce more P2 titles than would be expected given their actual quality (assuming they are not one of the few individuals on the backward bending portion of the curve) and their proxy for quality (their P1 revenue/title value) will also be too high. The product of this overly high number of predicted P2 titles and the overly high revenue/title in P1 would lead to overly high predicted P2 revenues from these authors in the P-1 only group. Similarly, authors with a negative shock in P1 would be predicted to produce

too few titles in P2, and the P1 proxy for quality for those authors in the P1-only group would be too low. Thus, authors with positive P1 shocks would be over-represented in the back of the envelope calculations for P1-only authors, and authors with negative shocks will be underrepresented, leading to an overstatement of the importance of P2 generated titles induced by P1 revenues.

We don't know how important this potential bias is. Instead, we assume that such a bias is likely to be responsible for less than the entire difference in the higher P1 proxy value relative to the P2 proxy value. If that were the case, we can make a rough adjustment in order to determine the maximum change in the back of the envelope calculations if the bias were removed. That is done in rows 12 and 13.

Since the higher value of P1 revenue per title only affects the estimates for P1-only authors, we take the P1-only values in row 8 and adjust them downward by the values in row 11 to put the predicted revenues based on P1 proxy values on the same footing as the revenues generated by two-period authors using P2 values as a proxy for revenue per title. This is found in row 12.

In row 13 we use the lowered values from row 12 in a recalculation of the last row in Table 3. These are the values that appear in the paper when discussing the impact of the different proxy values.

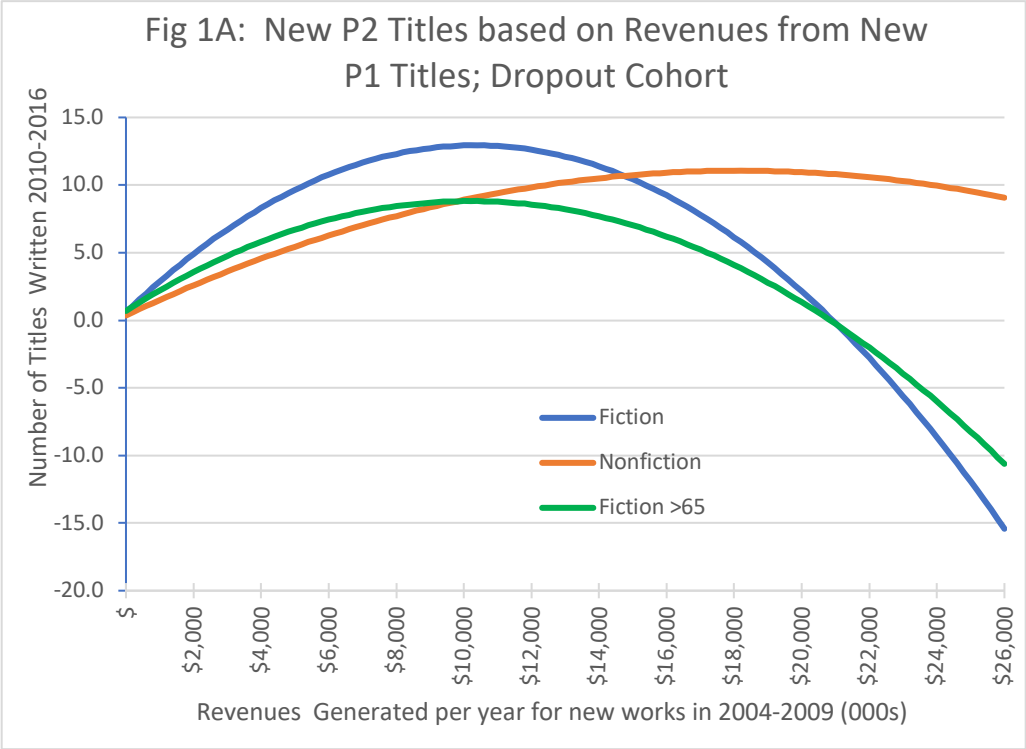
The most extreme (and unrealistic) supposition would be to assume that the proxy value for P1 revenue/title is zero. In that case P1-only authors and their predicted P2 titles disappear from the analysis, rows 8 and 12 are zeroed out, and the values in row 13 would be 22% and 21%.

Appendix 7: Fiction Regressions with truncated data

The regressions in the main text indicate that authors of fiction and nonfiction works respond somewhat differently to revenues. The two samples differ in that the nonfiction sample is truncated to observations where more than 65 books are sold in a year. One question that naturally arises is whether the difference between fiction and nonfiction works is due to the truncating of the fiction data.

To answer that question, we reran the fiction regressions, limiting the included works to those which sold more than 65 units per year. This should put the data from fiction authors on an equal footing with nonfiction data. When we do that we find that the truncated fiction data provides results similar to the full fiction data, and not similar to those from the nonfiction data.

This can be most easily seen in Figure 1A, which contains the fiction and nonfiction curves from Figure 1 in the text, but also includes the fiction results when observations selling fewer than 65 units are removed. The truncated fiction curve has similar curvature to the regular fiction curve and has a backward bend at a similar level of revenues, but reaches a lower maximum point and is not quite as curved.



The implications of this curve, with regard to our back of the envelope calculations of the importance of revenues in new titles found in Table A10, is similar to the values for the complete fiction data set. The share of P2 revenue from P2 titles induced by P1 revenues increases from 36% to 45%, which is closer to the nonfiction value, but still considerably smaller. Thus, some of the difference between the fiction and nonfiction results in the text may be due to the truncation of the nonfiction data, but it would seem to be a minor part of the difference.

Table A10: Impact of Period 1 Revenues on Period 2 Production		
	Fiction	Fiction>65
Generated # of New Titles	1,197	788
Share of all P2 Titles	3.04%	3.83%

Generated Revenue from New Titles	$479,000,000	$522,000,000
Share of P2 Revenue	35.98%	45.30%

The regression results that lead to the previous chart and table are found in Table A11.

Table A11: # of Fiction P2 Titles written (Robust T in brackets)		
	All observations	Obs with > 65 books
P1 Revenue per year	2.34	1.59
	5.42	[4.58]
P1 Revenue per year ^ 2	-0.110	-0.077
	-4.56	[-4.43]
P1 Titles per year	1.63E+00	2.881935
	13.94	[11.92]
Company Author	-3.14E-01	-0.4511012
	-7.98	[-3.13]
Major Publisher	1.10E+00	0.5697134
	16.49	[6.96]
Self Publish	-1.16E-01	-0.1295318
	-6.38	[-1.2]
Constant	-4.03E-01	-0.8199098
	-6.02	[-5.65]
Observations	62,633	13,892
R-squared	0.18	0.25