

Digital Disintermediation and the Market for Ideas*

Christian Peukert

University of Zurich

Imke Reimers

Northeastern University

May 23, 2017

Abstract

Electronic products and decreased marginal production and distribution costs have led to an upsurge of new products, increasing the variety available to consumers. In many markets, digitization has further allowed inventors to market their products directly to consumers, without having to find an intermediary and split the revenues. This new distribution channel affects the relationship between inventors and firms in the market for ideas. We study this relationship by examining contracts for over 100,000 book ideas from 2004 to 2016. We utilize quasi-experimental variation across time and genres to show that increased competition from self-publishing platforms leads to licensing contracts that favor the author more. In addition, an increase in new products allows publishers to learn more about an idea's likely success before committing resources. In markets in which a product's appeal is notoriously difficult to predict, such improvements can have large impacts on efficiency and total surplus in the short run and in the long run.

*Preliminary draft. Peukert: christian.peukert@uzh.ch, Reimers: i.reimers@northeastern.edu

1 Introduction

Digital technology and the Internet have brought about changes in media and information markets in recent years. A move towards electronic products has decreased marginal production and distribution costs substantially in many industries. This has led to an upsurge of new products, increasing the variety available to consumers (Waldfogel, 2012; Aguiar and Waldfogel, 2015), and leading to potentially changes in consumer and producer surplus (Brynjolfsson et al., 2003). While the implications on the market for the final product have been studied to some extent, the impact on the market for ideas – the relationship between the creators of the ideas and their distributors – is less understood.

A decrease in the costs of distribution can allow inventors to market their products directly to consumers, without having to find an intermediary and splitting the costs. This can improve the inventor’s bargaining position when a contract is set up. At the same time, digitization is often accompanied by improvements in the information environment, perhaps allowing firms to better predict the commercial success of a product before entering into a contract. In markets in which an idea’s success is notoriously difficult to predict, such an improvement can have large impacts on efficiency and surplus. This paper examines and quantifies the impacts of these two forces on the contracts between inventors (authors) and firms (publishers) in the book publishing industry.

Media industries have adopted and embraced digital technology through different mechanisms and at different points in time. While news, music and movie industries began to see effects of digitization around the turn of the millennium, the book industry did not move in that direction until several years later, when Amazon introduced the *Kindle* in late 2007. Moreover, while movie and music industries have experienced digitization on the demand side (for example, through online piracy), the book industry has evolved through changes on the supply side first. This makes book publishing a particularly salient industry for studying the impact of digitization on the relationship between creators and distributors in the market for ideas.

Traditionally, an author has had to find a publisher who was willing to publish the book, often with the help of an agent. Since publishers have little incentive to publish books which will not sell, they are quite selective in choosing which book to adopt, and many authors cannot publish their books. If a match is found, a contract is set up in which the author licenses the book’s rights

to the publisher. The contract includes an advance payment to the author – to be paid out before any copies are sold – as well as royalties for each copy that is sold beyond the advance payment. While the royalty payments tend to be quite constant across books and over time at about 15% of the revenue, advances vary significantly across books, authors, and publishers, from a few thousand dollars to over a million, depending on the predicted success of the book in the product market.

This process of discovery and distribution can be highly inefficient if publishers do not accurately predict the success of an idea. While many rejected book ideas were likely inconsequential, it is quite possible that some high-quality ideas were “falsely” rejected in this market and never reached consumers despite a large market potential. At the same time, most books which were published did not sell well as well as expected or hoped for, losing the publishers large amounts of money. In addition to the obvious direct effects of some ideas never reaching the consumers, such inefficiencies can lead to under-investment in types of ideas which can create large consumer surplus in the long run.

More recently, the introduction of the Kindle and the sudden rise in the popularity of e-books have provided a new channel with which authors can reach consumers. Instead of relying on traditional publishing houses to recognize (bet on) a book’s appeal, authors can now publish their works directly via self-publishing platforms, thus circumventing the gatekeeper. These self-publishing platforms often require a small fee for making the book available, and they take a share of the revenue from each book sold. Both distribution channels have advantages and disadvantages. While self-publishing platforms do not offer the author an advance, authors receive a much larger share of the revenues if they self-publish.¹ At the same time, traditional publishers are better able to market the book, and since self-publishing is often limited to electronic books (which require an e-reader), traditional publishers can reach more consumers because physical books do not require any additional devices.²

Regardless of the relative sizes of these benefits and costs, the option to self-publish directly improves the author’s outside option and increases the level of competition among publishers. This

¹The most popular platforms include Lulu, Smashwords, and Amazon’s Kindle Direct Publishing, and all offer similar deals for authors. Amazon, for example, offers authors a platform to publish their work with a royalty rate of up to 70 percent of revenue, but no advance payments. See <https://kdp.amazon.com/help?topicId=A30F3VI2TH1FR8>.

²In 2015, 45% of American adults owned a tablet computer and 19% of American adults owned a designated e-book reader. See http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/pi_2015-10-29_device-ownership_0-01/.

would strengthen the author’s bargaining position, causing an increase in the publisher’s payments to the author in some cases, even if the quality of the idea remains unchanged.³ But if publishers have to pay more for the same products, they will reject more book ideas – the bottleneck becomes even narrower.

The ability to reach consumers directly also leads to more products entering the market at all, and there is information in the demand for these titles. Self-publishing – the digital disintermediation in book publishing – facilitates the exploration of “good” books, allowing publishers to make more informed decisions about other book ideas. This indirect effect of improved information could make both publishers and consumers better off: publishers are less likely to lose money on book deals if they accurately predict the book’s success, and more “good” books enter the market, with traditional publishers allocating the most resources to those ideas which are worth the most.

This paper introduces a theoretical model of competition and information in the market for ideas, in which we allow for uncertainty about an idea’s appeal, and learning about this appeal through related products. We make several predictions about the transfers between firms and inventors as well as about the ability to predict an idea’s success *ex ante*, and we test these results empirically.

Empirical research on the relationship between creators and distributors in the market for ideas is scarce for two reasons. First, the researcher has to observe data on these ideas, which are usually hard to come by. Second, causal inference further demands plausibly exogenous variation in entry costs. Our empirical setting allows us to deal with both issues. We take advantage of the fact that e-books and self-publishing arrived fairly suddenly (with the large-scale diffusion of Amazon’s Kindle) and that this arrival affected different genres differently. While books of most genres are still predominantly published through traditional publishers, authors and readers of romance novels have largely embraced self-publishing platforms. By 2012, about 40% of all romance novels in the USA Today Top 150 bestseller lists were originally self-published, compared to only about 5% of the books of other genres.⁴

We examine contracts of over 100,000 book and rights deals from 2004 to 2016, and we utilize variation in the genres’ propensity for self-publishing to estimate the effect of this disintermediation

³Whether the author gets a better deal depends on the relative demands for traditionally and self-published books, as well as the relative cost structures.

⁴See section 3.2 for this evidence as well as for reasons for this variation.

on book deals in a difference-in-differences model. We find that romance novels experience much larger increases after the arrival of self-publishing than books of other genres, and we provide evidence that the increases in advance payments are due to added competition on the publisher level, and not supply- or demand-driven.

We then analyze the publishers' ability to predict an idea's success by examining the correlation between advance payments and commercial success before and after digitization, again taking advantage of variation in the propensity for self-publishing across genres. We find that publishers have become better at predicting the commercial success of romance novels, compared to other genres and compared to the pre-digitization era. Self-publishing can indeed improve the efficiency of the market by allowing firms to invest more money in the best ideas, while avoiding investment in ideas which will not be successful. In addition, even if self-publishing facilitates the discovery of high-quality ideas, most of these ideas are eventually published through traditional channels regardless of their original path. Hegde and Luo (2014) show that publication of patents through a credible, centralized institution mitigates information costs for buyers and sellers. Similarly, we argue that publication through traditional publishing institutions facilitates transaction in the products market.

In examining how inventors and firms interact in the market for ideas, our paper follows a long string of literature analyzing the optimal commercialization strategy of new products (see Gans and Stern, 2003). Many of these papers consider the role of intellectual property rights and appropriability in an inventor's commercialization decision (Katz and Shapiro, 1986; Gans et al., 2008). While we keep the question of appropriability in mind, we are more interested in how an increasingly digital world and a change in the level of competition at the firm level affect the relationships and licensing agreements between firms and investors in the market for ideas.

Our results provide insights for many other settings in which powerful gatekeepers select those ideas which eventually end up on the market as new products. Obvious other examples are music and movies (Luo, 2014), but our findings have implications for new ideas and products beyond media industries. They extend to any markets in which innovators (or inventors) license their products to downstream firms (Arora and Fosfuri, 2003), for instance in chemicals (Grindley and Nickerson, 1996) and in the semiconductor industry (Grindley and Teece, 1997). More recently, traditional investors, who can be compared to traditional publishers, increasingly face competition

from crowdfunding platforms concerning the financing of ideas and inventions (Agrawal et al., 2013, 2015). Determining the driving forces behind changes in the contracts between authors and publishers provides insights into innovators' incentives to create new products and their optimal commercialization strategies beyond book publishing.

Our paper highlights the importance of information when eventual success of a product or idea are unknown ex ante. We find that the digital distribution channel allows traditional publishers to invest more in higher-quality ideas, thus providing support for Nagaraj (2015), who finds that information from NASA's Landsat satellite mapping almost doubles the rate of significant gold discoveries.

The remainder of the paper proceeds as follows. In section 2 we explain the mechanisms which lead to changes in the market for ideas by introducing a model of incomplete information and competition in the market for ideas. Section 3 describes the data in more detail and provides preliminary evidence of the competition and information mechanisms. Section 4 shows results from a difference-in-differences analysis of deal advances, and section 5 examines implications regarding the quality of products and the efficiency of the market. Finally, we conclude in section 6.

2 Theory

The arrival of the e-book and the rise of self-publishing affect the market for ideas and the relationship between authors and publishers in two ways, both of which lead to a change in the advances that publishers pay to authors. First, self-publishing directly raises the level of competition on the firm side, which leads to firms (publishers) increasing their offers to make themselves and their services more attractive on the margin. Second, the digital environment allows publishers to learn more about the product and its appeal among consumers before acquiring the license to sell a book. This learning may happen as a previously unknown author gains fame through Internet blogs, self-published titles, or YouTube channels.⁵ As more products reach consumers (for example, through self-publishing), traditional publishers learn more about a particular genre's and topic's appeal among consumers, which further allows them to make more informed decisions when evaluating future ideas. We now introduce a simple (preliminary) model of incomplete information and com-

⁵For example, Andy Weir's *The Martian* was originally published as a series of blog posts before being acquired by Crown Publishing and turned into a major motion picture.

petition in the market for book ideas. The model predicts several testable results as a low-cost option for inventors (writers) to circumvent traditional institutions is introduced.

Suppose there are two traditional publishers, $j \in \{1, 2\}$, who are presented a book idea. Both publishers predict the market appeal of the idea in stage 0. Given the predicted appeal and standardized royalty rates r , each firm decides whether or not to make an offer, and if it makes an offer, it chooses a lump sum transfer A_j to the author in the ideas market in stage 1. Finally, the firm chooses its price p_j and associated quantity q_j in the product market to maximize profit and profits are realized in stage 2.

The true market (consumer) demand for the book is linear and given as $p = a - q$.⁶ The two publishers may not have full information about the book's market appeal and guess the demand function as

$$p_j = a + e_j - q, \quad (1)$$

where e_j is the error with which firm j estimates the demand intercept for the book idea. This error term is a random variable with mean μ_j and variance σ .⁷

Given its predicted demand function, firm j 's expected profit before the lump sum payment is

$$\pi_j(q_j) = q_j(a + e_j - q_j)(1 - r) - cq_j - F, \quad (2)$$

where r is the fraction of the book's price which is given to the author (the royalty), c is the marginal cost associated with distributing one more unit to consumers, and F is a fixed cost of setting up production, copy editing, and marketing of a title, which we assume to be common across publishers.⁸ Firm j chooses q_j^* to maximize equation 2 in the second stage.

Given the profit maximizing q_j^* , firm j makes an offer to the author if $\pi_j(q_j^*) > 0$ in stage 1. The offered advance A_j is bounded above by this $\pi_j(q_j^*)$. Suppose without loss of generality that $\mu_1 > \mu_2$. If both firms expect positive profit, then firm 1 will win the book, and it will pay an advance A_1 such that $\pi_1^* > A_1 > \pi_2^*$. Of course, if neither firm expects a positive profit from

⁶We are concerned about the market for ideas, and the products market is merely there to determine the firm's variable profits, rather than identifying prices and quantities. Thus, the functional form of the market demand is of secondary importance.

⁷We assume that the variance of the error does not depend on the publisher, whereas its mean does. Thus, the publishers' relative decisions about the advances only depend on μ_j . Under risk neutrality, the σ term becomes irrelevant altogether. For now, we ignore the role of σ .

⁸This fixed cost F can depend on e_j . We plan on adding this in future iterations of the paper.

publishing the title, neither firm will make an offer and the title will not be produced. If both firms made this decision based on a negative draw of e_j , it is possible that a book which “should” be produced will not reach consumers at all.

Now we add a third option to the market for ideas: a self-publishing platform with lower marginal costs and a higher royalty rate, r^s , but also with a lower demand curve.⁹ The book’s true self-publishing demand curve is given as $p_s = \lambda a - q$, with $0 < \lambda < 1$. If the author self-publishes the book, her expected profit is

$$\pi_w^s = q_w(\lambda a + e_w - q_w)r^s - c_w q_w - \beta F, \quad (3)$$

where e_w is a random variable with mean μ_w and variance σ ; $c_w < c$, and β is positive, allowing for fixed costs of setting up and marketing the book to vary by channel as well.

With this third option present, we can distinguish between three cases. First, under perfect information, the errors e_j and e_w have mean and variance 0. Traditional publishers and self-publishing remain in the market, and who obtains the rights to a book idea depends on the relative costs (c_w , c , and β), and the relative sizes of the demand curve intercepts, as given by λ . Books which have low costs in self-publishing compared to their demand in the product market are more likely to be self-published. This result holds even as we allow for uncertainty in the shape of nonzero error terms e_j and e_w .

Second, we acknowledge that prediction technologies are not perfect, and we introduce uncertainty and asymmetric expectations about demand. This uncertainty implies that traditional publishers can make positive profits, while both publishers and self-publishing continue to exist. Importantly, this implies that some “good” books will be self-published, while some “bad” books are published traditionally. Now, which publisher obtains the rights to the book idea depends in the relative costs and market demand, as well as the draws of e_j and e_w .

Finally, we allow authors and publishers to make better predictions about an idea’s appeal as more similar products enter the market. Self-publishing allows more products to reach consumers, even if some of these products only do so in limited fashion. If the absolute values of μ_j and μ_w are

⁹This structure fits self-publishing in the book industry well, but it also matches other industries in which an inventor is able to retain larger portions of the profit but may not have the established infrastructure to reach all consumers.

decreasing functions of N (the number of similar books in the product market), then publishers make fewer mistakes in choosing book ideas and offering advances, and authors make fewer mistakes in choosing outlets. In the limit, the market moves toward one with perfect information, and traditional and self-publishing still coexist as described above.

The model thus makes three main predictions which we can test empirically with the right data. First, books with relatively low costs of and high market demand for self-publishing will disproportionately be self-published. Second, advances for traditionally published works will increase on average, as the second-best option increases for some authors. Third, with more titles entering the market via self-publishing, traditional publishers will be better able to predict demand for new ideas, increasing the correlation between advance payments and market success.

3 Data and Preliminary Evidence

3.1 Data

Our data base consists of two datasets, which are connected with one another on the author level. First, we collect information on licensing contracts between traditional publishers and authors through publishersmarketplace.com. Second, we obtain a measure of market success by collecting the weekly Top 150 book bestseller lists from USA Today.

We have access to data on 102,310 advance deals – involving 56,773 different authors – closed between January 2004 and December 2016 in the United States. Publishers Marketplace is a premium community for authors, agents and publishers in which agents post book deals and the involved entities.¹⁰ Using natural language processing, we are able to extract information about writers, types and genres of books, types of deals (digital, print, or audio book, or movie, TV, and international rights), and publishers, editors and agents. Most importantly, we quantify the size of advance payments for a subset of about 25% of these deals, observing whether an advance payment was *nice* (less than \$50,000), *very nice* (\$50k to \$99k), *good* (\$100k to \$250k), *significant* (\$251 to \$499k), or *major* (more than \$500k).

Figure 1 shows the distribution of advance categories for all deals, listing the midpoints of the deal categories on the x-axis, and the relative frequency of each category on the y-axis. While most

¹⁰Agents have an incentive to post deals because these deals make them attractive to potential clients (authors).

deals involve small advances (62 percent of deals are “nice”), a non-negligible number of deals (11 percent) is in the “major” category. Advances for book deals are significantly larger than those for rights deals, although the heterogeneity of rights deals (they could pertain to international rights as well as TV or movie rights) makes a direct comparison difficult.

We supplement these data with information on the commercial success of different titles and authors through the USA Today weekly top 150 bestseller lists from October 1993 to October 2016.¹¹ We observe titles by 8244 authors, for whom we know the first appearance on the list, their highest ranking, the genre, and the publisher of the book. In addition to providing a measure of market success, we can determine whether a title entered the bestseller lists as a self-published product or via traditional channels.

3.2 Preliminary Evidence and Identification

Recall from section 2 that we aim to test three hypotheses. First, books which can reach large markets with low costs via self-publishing will be self-published more often. Second, self-publishing leads to larger advances in the traditional market for book ideas. And third, self-publishing leads to better predictability of success in the products market. Here, we provide evidence for the first prediction, and we later use this evidence to identify the effect of self-publishing on the transactions in the ideas market conditional on a deal being completed.

We argue that some genres are inherently better suited for self-publishing than others. More precisely, romance novels (such as E. L. James’ *Fifty Shades of Grey*, but also Nicholas Sparks’ *The Notebook*) are relatively easy to write as less research is needed than for (e.g.) historical nonfiction books. Advertising and marketing costs are lower than in other genres as well: romance novel readers are a close-knit group which often communicates via online communities, allowing readers to learn about new books via word of mouth, rather than through costly advertising campaigns often employed by traditional publishers. Finally, the nature of many romance novels might make readers reluctant to read them in public. Self-publishing platforms circumvent this problem by predominantly publishing e-books, a format for which the book’s cover is not seen by others. For these reasons, one would expect romance novels to be more susceptible to self-publishing than other genres.

¹¹These data are extended from Waldfogel and Reimers (2015).

Self-publishing as a Viable Option

Figure 2 suggests that self-publishing provides a serious and lucrative alternative for romance novels. It shows that romance novels are more likely to be self-published than other genres, conditional on being a USA Today bestseller. The difference becomes particularly evident after 2010, as self-publishing became a successful “mainstream” distribution channel. In 2011, as many as 20% of bestsellers in the romance category had a self-publishing background, and this number increased to about 50% by 2013. During that period, less than 5% of the bestsellers in other categories were ever self-published. Note that this does not imply that romance novels have become more popular overall. Rather, those romance novels which are successful are more likely to be self-published, compared to their non-romance counterparts.

Figure 2 shows that the probability of success depends much less on the publishing platform for romance novels than for other genres (suggesting a relatively large λ for this genre), but it remains to be determined whether authors of romance novels respond to these relative success probabilities. To this end, we collect information on all books available through the popular self-publishing platform Smashwords, finding that 23% of authors and 34% of books fall into the categories of romance and erotica novels – a much larger percentage than that of the deals in Publishers Marketplace. We interpret this as evidence that books with low costs and relatively large chances of success in self-publishing are indeed disproportionately more likely to take this route.

Romance Novels as the Treated Group

To test whether (and in what ways) the option to self-publish has changed the competitive environment for traditional publishers, we make use of this variation in the propensity to self-publish. If self-publishing affects the traditional publishers’ willingness to offer large advance payments, we will see larger increases in book advances for genres which are more likely to be self-published, compared to other genres. We thus use romance and erotica novels as our treated group in our main analysis.

Of course, we need to ensure that no other variables shifted differently for romance novels than for other genres in order for us to identify a treatment effect of self-publishing. For example, 50

Shades of Grey by E.L. James – the best-selling book of 2012¹² – was first published on May 25, 2011, and it may have led to a shift in demand for romance novels more generally. If that were the case, an increase in advance payments for that category would simply reflect relative increases in demand, rather than a response to self-publishing.

While booksellers and literary agents confirmed in conversations with us that romance novels have always been a popular genre, and that books like *Fifty Shades of Grey* did not seem to change any priors among publishers regarding the profitability of romance novels, a more data-driven analysis is instructive. To that end, we create a measure of pseudo-sales for each entry in the USA Today Top 150 bestseller lists, assuming a power distribution of sales within these rankings.¹³ That is, we assign sales as the inverse of the rank in a given week: $sales_j = \frac{1}{rank_j}$. We then calculate the share of romance novel sales among all bestseller sales for each week, and we follow these shares over time.

Figure 3 displays the fraction of pseudo sales over time, focusing only on those books which were never self-published. It becomes clear that traditionally published romance novels did not become more successful over time, providing support for the identifying assumption that romance novels did not change but for the introduction of self-publishing, and it is unlikely that changes in the market for book ideas are solely demand-driven. In addition, the fact that traditionally published romance novels do not become more popular overall leads us to believe that the mean quality of ideas which are presented to publishers does not increase either, implying that any changes in the market are direct results of the new self-publishing platform.

Evidence of Changes in Advances

We first ask whether advances changed as a result of the new distribution channel. Figure 4 shows the annual averages of the advance deal sizes for romance novels and for all other genres.¹⁴ The categories follow similar trends before 2008. After 2008 (the year after the introduction of the Kindle), there is a small increase in advances for romance novels, compared to a slight drop in advances for other genres. More pronounced, there is a sharp increase in deal sizes for romance

¹²See <http://www.usatoday.com/story/life/books/2013/01/16/100-best-selling-books-of-2012/1839803/>.

¹³There has been an active tradition of translating sales ranks into quantities utilizing the common wisdom that sales-rank relationships tend to obey power laws (see Chevalier and Goolsbee (2003); Brynjolfsson et al. (2003); Reimers and Waldfogel (2017)). We follow this tradition here.

¹⁴We use the midpoints of each deal size category to calculate these averages.

novels after 2010 (the introduction of the iPad), whereas the deal sizes for all other categories remain almost constant. We examine these differences in more detail in section 4.

Evidence of Predictability

Next, we ask whether publishers become better able to predict a book idea’s success as self-publishing becomes more prevalent. Figure 5 shows the fraction of authors on Publishers Marketplace who first appear on the USA Today Top 150 bestseller list within 24 months of the deal (“future stars”). The fact that more deals were made with future stars among the romance genre throughout the time period of our study suggests that publishers have been able to predict the future success of romance novels better than the success of non-romance books. After 2010 – concurrent with the large rise in self-publishing among romance bestsellers – the ability to predict bestsellers among romance novels increased further, with an increase in the share of future bestsellers among romance deals from about 2% to 5%. We interpret this as evidence that self-publishing can be used as a discovery tool, and we examine this mechanism further in section 5.

4 Empirical Strategy and Results

The above figures suggest that the relationship between authors and traditional publishers changes as the option to reach consumers directly becomes more viable. Examining the mechanisms behind these changes can help provide insights into who benefits from and who is hurt by such disintermediation, and it allows us to shed light on the inefficiencies in the market for ideas when a product’s success cannot easily be predicted. We explore these mechanisms by asking three questions. First, did book deals between traditional publishers and authors change as a result of the emerging self-publishing channel? That is, did advances change differently for romance novels than for authors of other genres? Second, is there evidence that self-publishing leads to changes in the information environment? Third, if changes in the information environment affect the relationship between authors and publishers, what are the implications for the quality of works which will be licensed? We begin by examining the sizes of the advances.

4.1 Baseline results

Our baseline model is a difference-in-differences set-up in which romance novels are the treated group. We estimate

$$\text{LogSize}_{i,t} = \alpha + \beta R_j + \gamma \text{After}_t + \delta \text{After}_t \times R_j + \nu_e + \varepsilon_{i,t}, \quad (4)$$

where $\text{LogSize}_{i,t}$ is the log of the advance payment (based on the midpoints of the observed size categories) for book i in year t , R_j is a dummy that is 1 if the author writes in the romance genre, and After_t is turned on after the year 2008. Because some editors have more resources than others, we include editor fixed effects ν_e , and we cluster standard errors at the editor level for the same reason.

We then estimate a specification with a more flexible structure, allowing the effect to vary over time, as

$$\text{LogSize}_{i,t} = \alpha + \beta R_j + \sum_{t=2004}^{2014} (\gamma^t \mu_t + \delta^t \mu_t \times R_j) + \nu_e + \varepsilon_{i,t}, \quad (5)$$

where μ_t are year dummies, and the remaining variables are as described above. The excluded year is 2008, to facilitate a comparison of pre-Kindle and post-Kindle years. This specification allows us to test the common trends assumption, and it gives an indication of how the effect evolves over time.

Estimation results for book deals are reported in table 1. Average advance payments in the romance genre do not significantly differ in the years before e-reading devices were introduced to the market, but they increase significantly thereafter. Interestingly, the point estimates increase substantially after 2010. This closely follows the introduction of Apple’s iPad in April 2010, and it coincides with the publication of *Fifty Shades of Grey* (May 2011) – another phenomenon of self-publishing, which we cannot identify separately. The large and significant increases in advances for book deals imply a large effect of self-publishing on average book advance payments, providing empirical support for our second hypothesis.

4.2 Placebo Test: Rights Deals

While figure 3 indicates that the romance genre did not become more popular overall, there might still be other factors which we did not account for, and which affected the size of advances. Here, we provide a placebo test by examining the size of payments for rights deals, rather than book deals.

We would not expect to see an effect of self-publishing on the size of rights deals for two reasons. First, self-publishing does not provide a viable outside option for rights deals. Traditional publishers are much better equipped to distribute books internationally (for example, by employing professional translators), or to set up a movie or TV show. This implies that traditional publishers do not face significant competition by the new platform in any genre. Second, rights deals are only negotiated for successful books, as has been the case long before self-publishing. Conditional on having previously published, self-publishing platforms do not provide any additional information with regards to acquiring rights to books. Consequentially, one would not expect any relative changes in the size of advances for rights deals.

We repeat the analysis from equations 4 and 5. Columns 1 and 2 of Table 2 look at all rights deals. They show that the effect of self-publishing disappears in this market, providing confirmation that our identifying assumption holds. If we had found a significant effect, this would suggest an increased popularity of romance novels overall. Since we do not see such an effect, it is unlikely that advances for romance novels rose because more people wanted to read them.

We take this robustness check one step further by specifically examining deals for international rights, taking into account the possibility that there might be a spike in demand for romance *books* (because they can now be read in secret), whereas the demand for erotic *movies* may not have changed. In that case, one would not expect an effect for movie rights, but an effect for international rights would still be likely. Columns 3 and 4 of Table 2 show that this is not the case, suggesting that the increases in advances are neither demand driven, nor due to changes in the supply and quality of ideas.

4.3 Beyond average effects

Table 1 suggests that, on average, the option to self-publish makes authors better off. However, this does not necessarily hold true for all authors, independently of the book idea’s inherent market potential. To test which authors benefit the most, we investigate how the number of deals changes across size categories.

Results of a linear probability model using indicator variables for the respective size categories are reported in Table 3. Across all columns of the table, the year-romance interactions in the pre-Kindle era (2004 to 2007) have coefficients that are not significantly different from zero (with one exception for nice deals – the smallest observed deal size – in 2004). The pre-trends are similar across the treatment and control groups, across the distribution of deal sizes.

Another potential pitfall is due to the fact that the size of the advance is reported for just about 25% of the deal observations. For the other deals, we only observe that a book idea has been licensed to a traditional publisher. While non-reporting might be correlated with the deal size or change over time, our identification strategy would only fail if there was a systematic difference in the trends of reporting between romance book ideas and those in other genres. Looking at the year-romance interactions in column 1 of Table 3, we find no evidence of this, as all point estimates are close to zero and statistically insignificant.

With the identifying assumptions satisfied, we can move to the interpretation of the year-romance coefficients after 2008. The table shows that the number of nice, very nice, good and significant deals does not change much, whereas the number of major romance deals (column 6) increases significantly. With self-publishing, traditional publishers are more willing to offer larger advances to those books with the highest market potential, and this difference is strong enough to drive the average effect.

This could be evidence that increases in advances are driven by changes in the information environment, as publishers would not offer advances over \$500,000 unless they are sure of the commercial success of the idea. Similarly, when examining the variance of advance sizes over time and across genres, we find that this variance increases for romance novels, compared to other genres, which also indicates that publishers know more about the likely success (or lack thereof) of romance novels after 2008 than they do for other genres.

5 Quality of Works

The above analysis shows that the option to self-publish increases advances paid to the author. The model attributes these changes in advances to changes in competition for ideas, as the author’s outside option improves, given certain cost and demand structures. The model next predicts that as more books enter the market, publishers become better able to predict the commercial success of individual ideas. They can thus take advantage of self-publishing as a tool for discovering “talented” authors.

If this is the case, then most authors benefit from this exploration mechanism, either because they get higher advances for same-quality books, or because they have a chance to reach the market where traditional publishers would previously have acted as gatekeepers. However, authors who turn out to be less successful than previously expected (lemons) will be hurt if they are no longer able to secure a book deal. For publishers, the effect is less straightforward. Without improvements in the information environment, the rise of self-publishing would unambiguously be bad for traditional publishers. But an improved ability to predict an idea’s success would mean that it is not all bad news.

5.1 Estimation Strategy and Results

If publishers become better at predicting success as a result of the self-publishing platforms, one would expect that the size of the advance becomes a better predictor of success for romance authors after 2008: publishers will be willing to pay large advances if they are fairly certain of success, and they would offer small advances for less promising works. We test for this by estimating the probability that an author becomes a bestselling author as a function of her book’s genre, the size of the advance, and whether the deal was made after 2008, in a triple-differences analysis:

$$I(usa)_{j,t+1} = \beta_0 + \beta_1 R_j + \beta_2 DealSize_{i,t} + \beta_4 R_j \times DealSize_{i,t} + \beta_5 R_j After_t + \beta_6 DealSize_{i,t} After_t + \delta After_t \times R_j \times DealSize_{i,t} + \mu_t + \nu_e + \varepsilon_{i,t},$$

where $I(usa)_{j,t+1}$ is 1 if author j appears on the USA Today bestseller lists in the future, $DealSize$ is the size of the advance of deal i (in 5 categories), and μ_t and ν_e are time and editor fixed effects,

respectively.¹⁵ In this regression, we limit our analysis to book deals for authors who have not yet appeared on the bestseller lists because the success of previously published authors is comparatively “easy” to predict regardless of the structure of the ideas market. We record deal sizes linearly, and as dummies for each of the five size categories.

This regression shows that the size of the advance is a good predictor of an author’s future success regardless of time period or genre. The (unreported) coefficients on the size of the deal are positive and highly statistically significant throughout, with a one-category increase in the advance payment leading to an increase of the probability of becoming a bestseller in the future of up to 3.15 percentage points on average. With a mean of 4.7% of authors becoming bestsellers after a deal, the increase is quite sizable.

Table 4 shows the coefficients on the interaction terms of romance and the deal size after 2008, thus picking up the differential effect of a larger deal category for romance novels in the digital age, compared to non-romance novels and compared to the pre-digital age. The coefficients suggest that advances for such deals (which are most affected by self-publishing) become even better predictors of success. Importantly, though, the improvement in prediction precision is focused around major deals becoming more successful, rather than the smallest deals becoming less successful. This result holds whether we look at the Top 150 or the Top 50, and whether we follow these authors for one year or longer. Given the low mean probability of success, the coefficients on major romance deals after 2008 are not only statistically significant, but also economically highly significant.

5.2 Efficiency

The ability of traditional publishers to predict success has direct implications for the consumer surplus generated by this market for ideas. While self-publishing platforms allow any and all authors to reach consumers, thus making more works available to readers, these platforms do not typically reach *all* consumers. Self-publishing platforms usually specialize in making books available electronically, so that consumers who do not own an e-reading device may not be able to read self-published books.

Ideally, the “best” books are published traditionally, so they can reach all consumers. But this means that publishers and authors have to come to an agreement. If the “best” authors get the

¹⁵We estimate this as a linear probability model for ease of interpretation.

largest deals, these authors become more likely to remain in the traditional market – available to all consumers – despite an appealing outside option of self-publishing and receiving up to 70 percent of the revenue from sales.¹⁶ Table 3 suggests that the increase in advance payments to romance authors is largely driven by a surge in major deals for that category, and table 4 shows that these deals go to the “right” authors. The arrival of self-publishing seems to make the book publishing market more efficient in the sense that authors of “good” books receive larger incentives to publish traditionally and reach all consumers.

Of course, publishers benefit from correctly predicting the quality of a book because it would allow them to offer an advance that is high enough for an author to accept the offer, but not higher than the revenue they would receive from selling the book. If self-publishing improves such an information environment, one would expect that the variation in deal sizes increases as self-publishing becomes more prevalent. Indeed, another (unreported) difference-in-differences analysis shows that the variance in deal sizes for romance authors has increased after 2008, compared to the change in the variance in deal sizes for non-romance authors. Self-publishing also seems to make the market for ideas itself more efficient by helping publishers and authors invest most resources into those ideas with the highest ex-post appeal to consumers.

6 Conclusion

The role of intermediaries has changed significantly as digital technology has gained importance in many markets. With both the production and distribution of products becoming cheaper, it has become easier for the creators of products – the inventors – to become entrepreneurs and bring their products to consumers without the help of intermediaries. Intuitively, this leads to changes in the terms of new contracts between the inventors and firms because the inventor’s outside option changes. How these contracts change, and more importantly how the incentives to innovate and the distribution of products are affected has been difficult to assess in the past, although welfare implications are large.

We create a novel dataset of book advance deals and ex-post success of products, and we take advantage of variation in the appeal of self-publishing across genres, to measure the effect

¹⁶Of course, publishers could give these authors higher advances, thus increasing the effective royalty rate.

of self-publishing platforms on deals between traditional publishers and authors. We find that self-publishing leads to larger advances, and we rule out other explanations for the comparatively large increase in advances. The relationship is likely driven by two different effects: increased competition for publishers, and an improved information environment for talent discovery. Our results are indicative of large short-term welfare effects in settings where an innovator can choose to become an entrepreneur or to seek an employment relationship with a firm.

Just as importantly, a reallocation of resources toward those ideas with the largest ex-post appeal among consumers has large long-term implications on the market for ideas. Traditional institutions which have a comparative advantage in marketing products and reaching a lot of consumers can continue to exist alongside new platforms which allow inventors to reach consumers directly. Our model shows that both institutions have merit, and they can even complement each other in removing inefficiencies in the market.

References

- Agrawal, A., Catalini, C., and Goldfarb, A. (2015). “Crowdfunding: Geography, social networks, and the timing of investment decisions.” *Journal of Economics & Management Strategy*, 24(2), 253–274.
- Agrawal, A. K., Catalini, C., and Goldfarb, A. (2013). “Some simple economics of crowdfunding.” Tech. rep., National Bureau of Economic Research.
- Aguiar, L., and Waldfogel, J. (2015). “Revenue, new products, and the evolution of music quality since napster.” Tech. rep., Institute of Prospective Technological Studies, Joint Research Centre.
- Arora, A., and Fosfuri, A. (2003). “Licensing the market for technology.” *Journal of Economic Behavior & Organization*, 52(2), 277–295.
- Brynjolfsson, E., Hu, Y., and Smith, M. D. (2003). “Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers.” *Management Science*, 49(11), 1580–1596.
- Chevalier, J., and Goolsbee, A. (2003). “Measuring prices and price competition online: Amazon.com and barnesandnoble.com.” *Quantitative marketing and Economics*, 1(2), 203–222.
- Gans, J. S., Hsu, D. H., and Stern, S. (2008). “The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays.” *Management Science*, 54(5), 982–997.
- Gans, J. S., and Stern, S. (2003). “The product market and the market for ideas: commercialization strategies for technology entrepreneurs.” *Research policy*, 32(2), 333–350.
- Grindley, P. C., and Nickerson, J. (1996). “Licensing and business strategy in the chemicals industry.” *Technology Licensing: Corporate Strategies for Maximizing Value*. John Wiley and Sons, New York, 97–120.
- Grindley, P. C., and Teece, D. J. (1997). “Managing intellectual capital: licensing and cross-licensing in semiconductors and electronics.” *California Management Review*, 39(2), 8–41.

- Hegde, D., and Luo, H. (2014). “Patent publication and the market for ideas.” *Management Science*.
- Katz, M. L., and Shapiro, C. (1986). “How to license intangible property.” *The Quarterly Journal of Economics*, 567–589.
- Luo, H. (2014). “When to sell your idea: Theory and evidence from the movie industry.” *Management Science*, 60(12), 3067–3086.
- Nagaraj, A. (2015). “The private impact of public mapslandsat satellite imagery and gold exploration.”
- Reimers, I., and Waldfogel, J. (2017). “Throwing the books at them: Amazon’s puzzling long run pricing strategy.” *Southern Economic Journal*.
- Waldfogel, J. (2012). “Copyright protection, technological change, and the quality of new products: Evidence from recorded music since napster.” *Journal of Law and Economics*, 55(4), 715–740.
- Waldfogel, J., and Reimers, I. (2015). “Storming the gatekeepers: Digital disintermediation in the market for books.” *Information Economics and Policy*, 31, 47–58.

Table 1: Diff-in-diff regression: book deals

	(1)	(2)
After	0.245*** (0.028)	
Romance	-0.103* (0.053)	-0.075 (0.075)
Romance \times After	0.198*** (0.061)	
Romance \times 2004		0.008 (0.121)
Romance \times 2005		-0.031 (0.100)
Romance \times 2006		-0.056 (0.092)
Romance \times 2007		-0.073 (0.085)
Romance \times 2009		0.103 (0.088)
Romance \times 2010		0.151* (0.085)
Romance \times 2011		0.092 (0.096)
Romance \times 2012		0.154* (0.092)
Romance \times 2013		0.317*** (0.100)
Romance \times 2014		0.216** (0.105)
Romance \times 2015		0.231** (0.127)
Observations	23123	23123
$\overline{R^2}$	0.537	0.544

Dependent variable: Log dealsize. Editor and year fixed effects are included. Standard errors clustered on the editor-level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Diff-in-diff regression: rights deals

	(1)	(2)	(3)	(4)
	All rights deals		International rights	
after	0.114** (0.0529)		0.0368 (0.0499)	
romance	0.145 (0.0989)	0.148 (0.136)	0.115 (0.0941)	0.0930 (0.135)
after \times romance	-0.0115 (0.100)		-0.0456 (0.0912)	
Romance \times 2004		-0.336 (0.462)		-0.379 (0.348)
Romance \times 2005		-0.164 (0.211)		-0.190 (0.191)
Romance \times 2006		0.110 (0.246)		0.202 (0.240)
Romance \times 2007		-0.00894 (0.157)		0.0192 (0.158)
Romance \times 2009		-0.0713 (0.206)		-0.0318 (0.201)
2010.year \times 2010		0.163 (0.198)		0.144 (0.194)
2011.year \times 2011		-0.117 (0.156)		-0.133 (0.125)
2012.year \times 2012		-0.0859 (0.190)		-0.153 (0.163)
2013.year \times 2013		-0.0218 (0.182)		-0.00265 (0.185)
2014.year \times 2014		0.241 (0.241)		0.192 (0.248)
2015.year \times 2015		-0.183 (0.146)		-0.121 (0.134)
N	7979	7979	6814	6814
adj. R^2	0.609	0.612	0.532	0.534

Dependent variable: Log dealsize. Editor and year fixed effects are included. Standard errors clustered on the editor-level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Diff-in-diff regression by deal size category

	(1) no size	(2) nice	(3) very nice	(4) good	(5) significant	(6) major
Romance	0.016 (0.026)	0.039 (0.034)	0.012 (0.033)	-0.005 (0.028)	0.010 (0.018)	-0.056** (0.023)
2004 × Romance	0.004 (0.043)	-0.110* (0.060)	0.063 (0.040)	0.040 (0.041)	-0.010 (0.030)	0.017 (0.035)
2005 × Romance	0.003 (0.039)	-0.004 (0.049)	-0.024 (0.044)	-0.003 (0.041)	-0.008 (0.026)	0.038 (0.030)
2006 × Romance	-0.051 (0.038)	0.031 (0.045)	-0.038 (0.039)	-0.008 (0.037)	-0.024 (0.023)	0.039 (0.029)
2007 × Romance	-0.029 (0.034)	0.030 (0.043)	-0.022 (0.038)	-0.008 (0.041)	-0.013 (0.025)	0.012 (0.027)
2009 × Romance	-0.028 (0.027)	-0.038 (0.044)	-0.018 (0.038)	0.004 (0.034)	-0.004 (0.021)	0.056** (0.026)
2010 × Romance	-0.022 (0.032)	-0.030 (0.038)	-0.034 (0.035)	0.022 (0.035)	-0.001 (0.023)	0.042* (0.025)
2011 × Romance	0.031 (0.036)	-0.043 (0.043)	0.039 (0.043)	-0.081** (0.036)	0.011 (0.033)	0.074** (0.030)
2012 × Romance	-0.013 (0.034)	-0.035 (0.041)	0.003 (0.040)	-0.068* (0.035)	-0.002 (0.023)	0.102*** (0.035)
2013 × Romance	-0.043 (0.031)	-0.093** (0.041)	-0.022 (0.039)	-0.012 (0.036)	0.001 (0.025)	0.126*** (0.037)
2014 × Romance	0.008 (0.036)	-0.063 (0.050)	-0.033 (0.039)	-0.016 (0.044)	-0.008 (0.031)	0.120*** (0.039)
Year FEs	✓	✓	✓	✓	✓	✓
Observations	81440	21218	21218	21218	21218	21218
$\overline{R^2}$	0.192	0.532	0.062	0.113	0.018	0.204

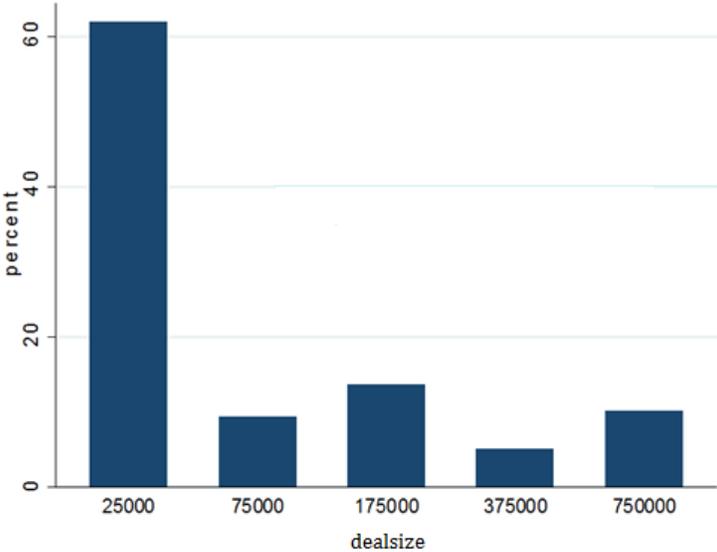
Dependent variable: Indicator for each category (nice: <50k; very nice: 50k-100k; good: 100k-250k; significant: 250k-500k; major: >500k). Editor and year fixed effects. Standard errors clustered on the editor-level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Triple-diff regression: predicting success

	1 year		2 years	
	Top 150	Top 50	Top 150	Top 50
Dealsize = 1	0.0310 (0.0381)	-0.00186 (0.0141)	0.0301 (0.0496)	0.0219 (0.0203)
Dealsize = 2	0.0520 (0.0477)	-0.0167 (0.0151)	0.0319 (0.0623)	-0.00146 (0.0259)
Dealsize = 4	0.0297 (0.0882)	0.000804 (0.0374)	0.0190 (0.110)	-0.0320 (0.0713)
Dealsize = 5	0.155* (0.0797)	0.111* (0.0625)	0.225** (0.109)	0.185** (0.0793)
N	13706	13706	13706	13706

Dependent variable: $I(usa)_{j,t+1}$, as a Top 150 bestseller (columns 1 and 3), or as a Top 50 bestseller (columns 2 and 4), in the following 12 months (columns 1 and 2), or in the following 24 months (columns 3 and 4). Robust standard errors are clustered on the editor-level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Editor fixed effects and month dummies are included, as are indicator variables and simple interactions. Reported coefficients are the triple interactions of “after 2008” \times “romance” \times “dealsize,” where “dealsize” is a categorical variable of the respective deal sizes (1=25k, 2=75k, 3=175k, 4=375k, 5=750k). Dealsize = 3 is the omitted category.

Figure 1: Distribution of deal size categories



Shares as percentages of all deals in the dataset. 25,000 is less than \$50,000; 75,000 is between \$50,000 and \$100,000; 175,000 is between \$100,000 and \$250,000; 375,000 is between \$250,000 and \$500,000; and 750,000 is above \$500,000.

Figure 2: Share of USA Today bestsellers which were originally self-published

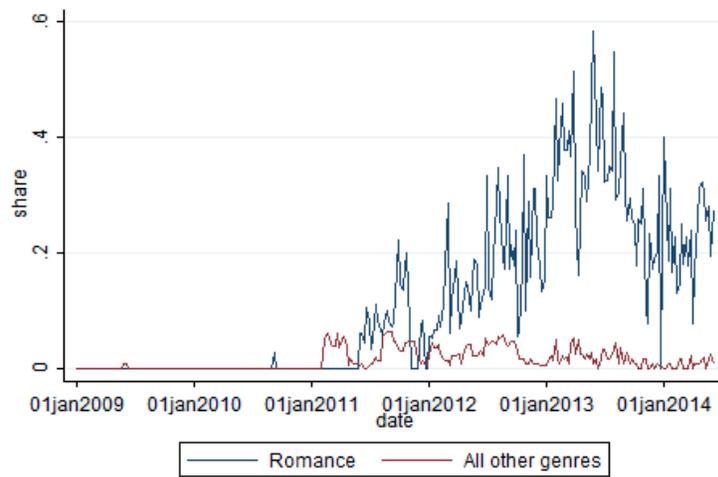
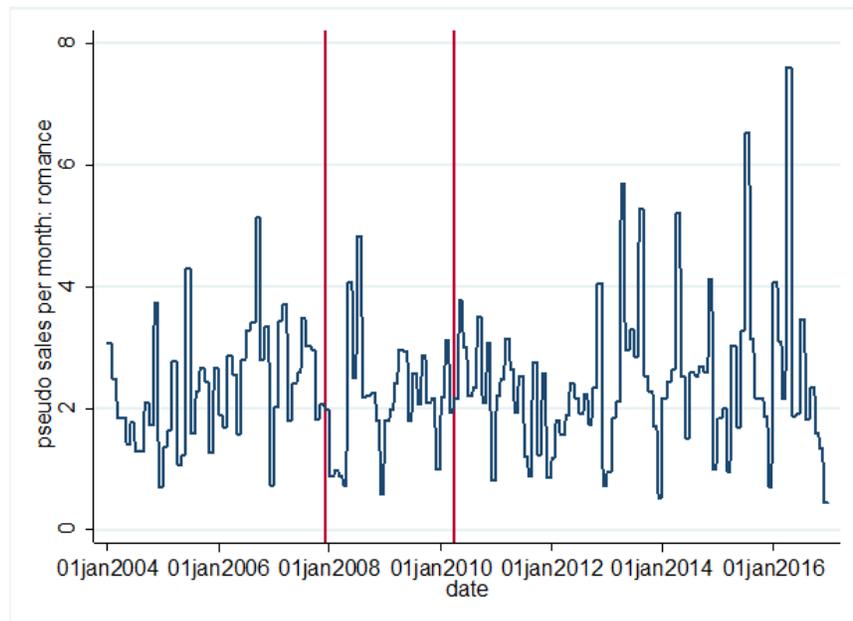


Figure 3: Share of romance sales among bestsellers – traditionally published



Pseudo sales are calculated as $sales = \frac{1}{rank}$. The red lines indicate the introduction of the Amazon Kindle and the iPad, respectively.

Figure 4: Average deal size over time

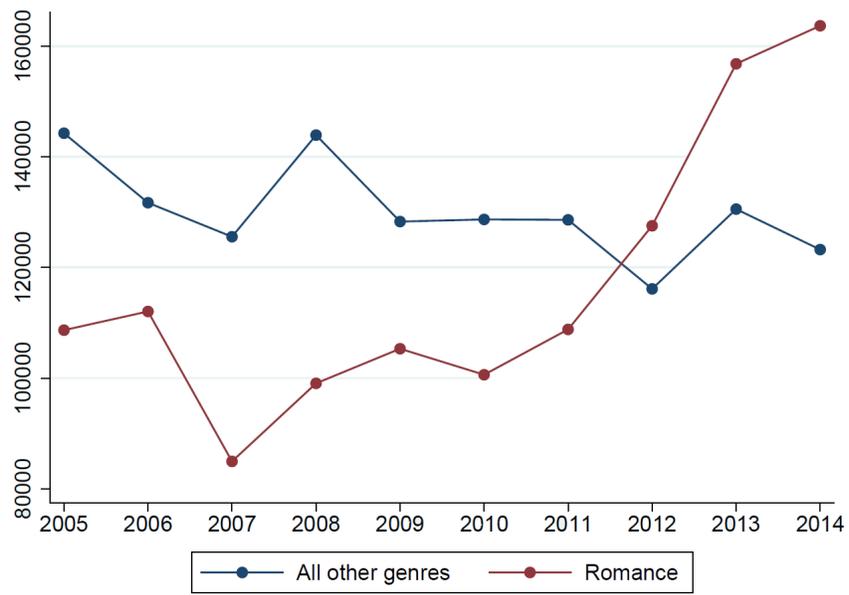


Figure 5: Share of deals with future bestseller authors

