Creativity amid piracy: an agent-based model of the incentives to create in a piratic digital world

Jason Rhinelander*
Queen's University

Abstract

Digital piracy of copyrighted works has had a significant, negative effect on the incentive to create new, easily-copied works. This paper develops an agent-based model of digital creativity to explore the effects of digital piracy on the incentives of heterogeneous creators and explores potential remedies to the piracy problem. The paper proposes and makes use of a multivariate Bayesian foundation for representing and updating agent beliefs, demonstrating how this belief structure can be used to obtain sophisticated results in this type of model. It also adds an extension to this technique to model forgetting that can be used to allow beliefs to more rapidly adjust to changing circumstances in a model. The model is able to qualitatively capture several observed and theorized trends in the piratic digital world: among others, it finds a long-run decrease in the utility of agents due to the reduced quantity and quality of new works. The explored policy responses are able to restore some of the of the welfare loss due to piracy, with a policy of limited detection and fines performing better in most ways than proposed public compensation policies.

1 Introduction

High-speed Internet has revolutionized many aspects of modern life, particularly when it comes to the consumption of creative works—whether written, recorded, or filmed. Long before firms adapted to these new opportunities, Internet users reaped the benefit in a big way by engaging in digital "piracy:" unauthorized, unpaid downloads of copyrighted material at a massive scale through online file-sharing facilitating websites such as Napster and The Pirate Bay. To creators, this widespread piracy has been considerably negative: while the Internet opened up new, cheaper distribution channels for copies of creative works, it also enabled digital piracy to take away much of a

*Correspondence: Department of Economics, Dunning Hall, Queen's University, Kingston, ON, Canada, K7L 3N6. Tel.: (613) 572-7442. E-mail: rhinelaj@econ.queensu.ca. Web: www.imaginary.ca

creator's market power by acting as a competitor with identical copies offered at a negligible price.[1] As a result, profits from creative work sales have considerably declined as piracy has eroded the demand for (legitimate) copies. The long-run effects of this are worrying: with decreased profitability of creation, the number of new creations and the quality of those works must suffer.

This paper explores the issue of piracy through the use of an agent-based model that captures the salient details of digital creative work creation and consumption. The model begins in an environment without piracy: copies of created works can only be purchased from creators, at the creator-determined price. Potential creators choose to create (or not) in this environment based on their own ability and their prediction of the likely returns to creation.

Next the model introduces a form of "piracy." This is modelled as a network between agents with fixed connections established at random. The advent of piracy allows agents to obtain copies of created works both from network-adjacent neighbours at a low cost, and from the original creator of the work at whatever price the creator sets.

Finally multiple policy experiments are conducted where government collects a lump-sum tax from agents and uses this to enact one of three policies: public funding proportional to the number of downloads; public funding based on consumer votes; and a mechanism of detection and fines for users engaged in piracy.

This paper applies an agent-based model (ABM) to the problem for several reasons. First, the ABM developed here allows analysis and comparison of both the short-run and long-run effects of piracy and the attempted policies. Second, using an ABM allows considerable heterogeneity of both consumer preferences and creator abilities while maintaining tractability. Third, this approach allows analysis of both the intensive margin—the quality of created works—and extensive margin—the number of works created over time. Finally, it allows addition and comparison of considerably different policy responses to piracy.

With an ABM approach agents must learn and forecast their way in the model environment based on past observations in the simulated environment. This paper addresses the problem by developing a learning process using a Bayesian econometric foundation where potential creators hold "beliefs" about the distribution of the parameters of multivariate, linear equations. Using these beliefs agents make predictions about the profitability of creation and, based on these predictions, decide whether to create and how to price copies of previously created works. As new works are created and sold, agents update their belief distributions in response to model observations, which in turn affects future creation choices.

This paper's contribution is threefold: first it develops a model for analyzing the effects of digital piracy and exploring multiple potential policy remedies to the piracy problem. Second, it proposes a multivariate Bayesian approach to modelling beliefs in an agent-based model, providing a belief structure that allows learning, prediction,

[1] Typically the cost of piracy is not directly monetary, but in terms of the opportunity cost of finding the pirated material, typically including viewing online advertising as part of the process.

and forgetting. Finally, a major component of the software developed for this paper is an open source software library aimed at developing sophisticated agent-based models such as the one developed here.

The effects on welfare in the developed model exhibits two considerably different aspects. In the short run, piracy reduces profits and the quality of new works, but this effect is overwhelmed by the significant utility gains resulting from past works suddenly becoming available at low cost through piracy. In the long run, however, piracy's effects are drastically different: Lower sales drive a significant decline in profits, which in turn reduces the number of new works created. Over the long term, this reduction in new material leads to a significant reduction in welfare; although piracy provides low-cost access to creative works, there are far fewer creative works to consume.

Having built a model of creativity and piracy that reflects observed trends in digital creative industries, the paper then explores three different policy responses to the piracy problem. The first potential remedy explored essentially embraces piracy, but introduces a lump-sum tax that is redistributed to creators in proportion to downloads of their works. This policy fails by most metrics: by rewarding downloads it fails to provide sufficient incentive to create high quality works. As a result the market is awash in low-quality works, but the significant quality deficit results in low profits, low utility, and an overall social welfare that is even lower than the situation under piracy.

The second remedy tweaks the first policy by having agents vote for their favourite works, attempting to correct the insufficient quality incentive by redistributing the collected tax in proportion to the number of votes received rather than number of downloads. This largely restores the incentive to create high-quality: quality of created works exceeds that even of the pre-piracy environment. It fails, however, to entirely restore welfare: despite correcting the quality deficit, the policy is only able to partially restore the number of works created.

The final policy explored is a more traditional response to piracy of having government collect a tax to fund piracy detection with fines assigned to individuals caught pirating. Even after accounting for the cost of funding the policy, this is able to restore social welfare to its pre-piracy level by allowing some of the benefits of piracy without the debilitating effects of widespread, uninhibited piracy.

The main findings of this paper are: First, that piracy is debilitating to the continued flow of creative output by reducing the profits creators receive; creators respond by reducing their effort or dropping out of the creative market, effecting a long-term reduction in creative output. The second result is that certain policy responses—such as simply paying creators in proportion to the downloads they attract—can restore the number of created works, but can do so by flooding the market with low-quality, low-value works. Correcting this by introducing a voting mechanism by consumers improves the situation. Third, a mechanism of imperfect copyright enforcement that only dissuades heavy piracy but effectively permits light piracy is able to effectively restore much of piracy's negative effects. Finally, this paper shows how a multivariate Bayesian belief mechanism can be effectively used to model sophisticated agent behaviour, learning, and adaptation in an agent-based model.

1.1 Creativity and copyright

Creative works are, in general, characterized by several properties. First, they have a very high fixed cost of creation: the vast majority of the cost incurred by a creator comes in creating the original, master copy. Second, the cost of copying and distribution is low: once the initial work is created, creating copies of that work for people to consume is comparatively negligible. Third, the demand to any particular individual is binary: consumers only want 0 or 1 copies of a work; partial copies and additional copies have no value to a consumer. Fourth, more than most economic goods, there is significant heterogeneity among consumers: consumers have strongly varied preferences for the genre and characteristics of created works they consume. Finally, the creative talent underlying creative works varies enormously across individuals: some creators are endowed with considerable natural talent, while others may require considerably more effort to create a work of comparable quality.

The low cost of copying, in particular, leads to a simple free-rider problem: once created, someone other than the author could copy the work, and thus the author would be unable to earn back sufficient profit to have made creation worthwhile in the first place. The traditional approach to inducing creativity, as thoroughly discussed in Landes and Posner (1989), has been through the use of copyright, dating back 1710 with Britain's Statute of Anne.

The theoretical argument behind copyright, however, rests on the premise that copying can be effectively prevented. Technological progress in digital content distribution has certainly challenged this assumption: copying between computers and computerized devices has become considerably easier, and of higher—and often *perfect*—quality. On its own, however, this is not a particularly monumental change: such copying has existed for many years. Though it has become easier, cheaper, higher-quality and faster, it is unlikely that localized sharing alone *without* the Internet would have introduced fundamental market changes. As in the Bakos, Brynjolfsson, and Lichtman (1999) and Bergstrom and Bergstrom (2004) models, such localized sharing can, under certain circumstances, actually increase profits.

The more confounding change has been the massive increase in the breadth of sharing due to high-speed Internet and technologies for the distribution of digital files such as the decentralized BitTorrent protocol. While copying was still available in the pre-high-speed Internet era, it could occur only at a very small scale: any large-scale distribution was relative easy to detect and stop under copyright laws; the penalties of being caught combined with the high probability of being caught eliminated widespread copyright violation. The relative anonymity and enormous scale of the Internet, however, has made enforcement of copyrights exceedingly difficult. High-speed Internet coupled with digital copying has elevated digital piracy's effect on copyright from minor to debilitating.

This suggests a sixth characterization for the list above that applies to *digital* creative works in a high-speed Internet world: perfect, low-cost copying *by consumers*. Anyone who has previously obtained a digital copy of a work (whether legitimately or not) can, in the digital, high-speed Internet era, provide unlimited copies of that work for others at negligible cost.

Responses to the piracy problem have been varied: early attempts at encryption of digital content to prevent its copying have failed.[2] Legal threats and lawsuits for detected cases of infringement have been another attempt at mitigating the use of piracy,[3] but seem to have had little effect at stemming the tide of illicit file sharing. Other industry response has been an attempt to shut down online file-sharing websites with inter-governmental assistance, and though one notable example—Napster—was shut down in 2001, other sites such as The Pirate Bay have proven remarkably resilient to being (permanently) shut down by law enforcement agencies. Finally, the proliferation of relatively low-cost streaming services such as Netflix for video and Spotify for music (to name just a few of many competing services) suggest an erosion of market power: pricing power is constrained by piracy's illicit competition.

Various papers have looked at this decline in music sales in particular. Barker and Maloney (2012) identifies significant lost sales in the music industry due to the existence of peer-to-peer file-sharing as an alternative source of obtaining music. The Recording Industry Association of America—admittedly a self-interested party in the discussion—reports in the prominent "Piracy" section of its website a halving of music industry sales revenue from $14.6 billion in 1999 to $7.0 billion in 2013 (RIAA (2015)). Connolly and Krueger (2006) finds evidence of a significant shift of music artists seeking revenue through concert performances rather than album sales, hypothesizing that this shift is a result of declining album revenue due to piracy.

Oberholzer-Gee and Strumpf (2010), backed up by Handke (2006, 2012), find some evidence that the production of music, books, and movies has substantially increased, rather than decreased, since the availability of file sharing. While there are many feasible explanations for this increase, one hypothesis proposed by Oberholzer-Gee and Strumpf is that the quality of the increased number of produced works has declined: creators are creating more works, but those works are of lower quality (and thus have a lower fixed cost of initial creation). Waldfogel (2012) attempts to investigate this trend in quality by using music critic assessment: he finds some evidence for declining quality over time, but suggests that the decline has mostly levelled off since about 2000, though quality remained low through the end of his 2010 analysis period.

In response to the piracy crisis, Varian (2005) suggests a bevy of possibilities in a world without (effective) copyright: making legitimate copies cheaper than pirated copies; focussing on the selling of complements; selling subscriptions; embedding ad-

[2]For example, the Content Scramble System (CSS) used by DVDs was broken in 1999, about 3 years after the introduction of DVDs. The encryption used for the following generation of Blu-Ray discs—though significantly more complex than that found on DVDs—lasted only a few months before being circumvented.

[3]The most notable examples being the United States cases *Capitol Records, Inc. v. Thomas-Rasset*, which, through various appeals, awarded damages of between $54,000 and $1.92 million, with a final award of $222,000, for copyright infringement of 24 songs; and *Sony BMG Music Entertainment, et al. v. Joel Tenenbaum*, which awarded damages of $675,000 for 30 infringed songs.

vertising; monitoring with collection and restribution of fees; pure public provisioning; prizes and commissions; and various others. As he readily admits, however, “all of these business models have their problems, of course, and none is likely to yield any sort of social optimum.”

1.2 Models of copying and piracy

Several theoretical papers have attempt to model the impact of piracy on the incentive for creation. Besen and Kirby (1989) proposes a static model with a single monopolist selling copies of an already-created work. The model features original copies sold by the monopolist and imperfect substitutes in the form of reduced-quality copies made from the originals. In the model, the producer is made worse off (as long as the marginal cost of copying is non-increasing), while consumers of both originals and copies gain.

Bae and Choi (2006) develop a model of software piracy that begins with a short-term effect of piracy on a single firm with an existing software product selling to users with heterogeneous individual demand for the software. Bae and Choi add to this a long-term improvement decision: the firm may improve its product quality, but does so differently in response to whether or not it faces piracy. As in Besen and Kirby (1989) they find a short-run benefit to consumers of piracy, but find a long-run reduction in quality. Thus the effect is ambiguous and likely negative in the long run: piracy provides cheaper access for both legitimate and illegitimate buyers, but the response of the monopolist to reduce effort in turn reduces the long-run value of the software.

Harbaugh and Khemka (2010) look at piracy among heterogeneous consumers with different types of copyright enforcement targetting. The model, like the ones discussed above, assume that copying results in inferior goods. They find that under such heterogeneity, it is considerably more efficient to concentrate enforcement efforts at high-valuation customers than to apply enforcement more broadly. This concentrated enforcement causes the firm to set a supra-monopoly price and sell only to high-valuation customers, while lower-valuation customers obtain pirated copies. This is able to partially restore both the firm's profits and consumer surplus relative to a broader enforcement effort modelled as increased piracy costs.

The above papers, and other related papers in the literature, typically model a single monopolist in a static setup with a pre-existing product. To some extent, this addresses only one part of copyright, but is a less directly important social issue: the socially optimal copyright for a pre-existing work is to have no copyright at all. It is rather copyright's incentive for the creation of new works that gives it social value.[4]. The model developed in this paper uses an agent-based model that aims to qualitatively capture many of the effects of piracy, both empirical and theoretical, discussed above. This is done by building a dynamic model of multiple creators where the social value comes not only from the welfare gains of a single work, but the welfare gains of a regular flow of new creative works.

[4]Bae and Choi is notable in its long-run product enhancement feature, but still does not explicitly model the initial creation decision.

1.3 Agent-based modelling, learning, and beliefs

Agent-based modelling and, specifically, its application to economics problems, is, at its core, a philosophy that many problems of interest to economists can be addressed by computational simulation of behaviour rather than mathematical solutions to equilibria. Within the discipline of Economics, computational modelling is an alternative to mathematical equilibrium-based solutions. Tesfatsion (2006) provides an excellent overview of the strengths and weaknesses of ABMs at addressing economic problems. In contrast to mathematical approaches, the technical emphasis of an ABM is on structure rather than equilibrium. This implies that model specifications become far less limited by equilibrium tractability requirements.

One of the most contributive areas for agent-based models, rather than confirming or contradicting theoretical results, is to address problems outside the realm of traditional economic theory due to no known feasible approach to calculating an equilibrium, much less to characterizing the full set of equilibria of a model. Miller and Page (2007) offer a very broad introduction to the computational modelling of social models where ABMs have been used. Arthur (2006) surveys various agent-based modelling approaches specifically looking at out-of-equilibrium behaviour, where equilibrium may not exist, may not exhibit a steady-state, or may not be reachable by agents in a model without endowing agents with highly-sophisticated optimization abilities.

Just as ABMs have, in their various forms, been applied in a variety of ways to many areas of interest to economists, the behavioural foundation of ABMs varies substantially. Some models, such as those in Conway (1970)'s "Game of Life" follow very simple, but arbitrarily-selected, rules, observing outcomes under those rules. Some models simply have agents repeatedly choosing a best action conditional on the previous-period actions of other agents and look for equilibrium. For example, Plümper and Martin (2008) offer a political party positioning and voting model where parties poll all possible political positions in any period and jump to the best one; they find that, in such a model, parties tend to revolve around central positions without actually occupying the centre. Another general modelling approach, such as that in Keen and R. Standish (2010, 2015) and the so-called "KMP" models pioneered by Kollman, Miller, and Page (1992), has agents exploring the space near their current choice to make incremental improvements over time.

An alternative to these primitive behavioural rules is the increasingly important use in ABMs of incorporating learning into agents' behaviours. Within the broad category of agent learning are various fundamentally different approaches to modelling learning. Selten (1991) divides these approaches into three main types: rote learning, imitation, and belief learning. Brenner (2006) addresses various types of learning being actively used in agent-based models; most pertinent to this paper is his (brief) discussion of Bayesian learning—though he discusses only simple Bayesian updating of a single parameter—and ordinary least-squares learning.

The basic Bayesian learning approach is to combine Bayesian learning about probabilities with expected utility. Under such a model, utility under a state of the world is known, but the probability distribution over states of the world must be learned.

The learning, following Brenner's notation, specifies a belief model where agents have an estimation of a probability of a hypothesis h at time t, denoted $p(h,t)$, where $h \in \mathcal{H}$ defines a discrete set of hypotheses. After observing a model event $e(t)$ in time t, the agent calculates the probability of the event $P(e(t) \mid h)$ under each hypothesis, and updates beliefs for all hypotheses $h \in \mathcal{H}$ according to Bayes' Rule:

$$p(h, t+1) = \frac{P(e(t) \mid h)p(h,t)}{\sum_{\widetilde{h} \in \mathcal{H}} P(e(t) \mid \widetilde{h})p(\widetilde{h}, t)}$$

subject to the condition $\sum_{h \in \mathcal{H}} p(h,t) = 1$. Given the set of probabilities over events, agents can then choose an action to maximize expected utility conditional on these probabilities. That is, the agent chooses the action a from all possible actions that maximizes expected utility:

$$\sum_{h \in \mathcal{H}} p(h,t) \sum_{e_t \in \mathcal{E}_t} u(e,a)P(e \mid h)$$

where $\mathcal{E}_t$ is the set of possible events in period t. This decision process, combined with updating the probability distribution as new events occur, determine the decisions of agents in the model over time.

Another learning model applied to agent-based models is that of using least-squares estimation of the relationships between variables that are allowed by a linear model, such as that of Marcet and Sargent (1989) and Bullard and Duffy (1994). The basic structure of such a model is that agents use a (researcher-selected) linear specification between parameters of interest, and use events in the simulation as new observations in the agent's belief model, who then use the parameter estimates of the linear equation to choose future actions. One significant drawback of the approach, however, is that it rests on implicit assumptions that the data-generating process underlying the data is constant—that is, that there is a unknown but true value of $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ that is to be discovered. In an ABM, however, this is often not true: each period's actions is based on various observations that were themselves influenced by the estimations of the previous action.

One of the contributions of this paper is to model belief learning (detailed in Section 3) by combining both the increased sophistication of the multidimensional, linear model structure of least-squares learning with the rational incorporation of new information implicit in Bayesian updating. This is done through the use of Bayesian econometric modelling of beliefs. This allows both the advantages of the Bayesian degree-of-belief interpretation, yields a model that directly incorporates uncertainty about belief parameters, and allows multivariate relationships between the parameters of beliefs. It also provides a simple way to model forgetting which aids in the development of a dynamic, evolving model.

This paper is divided into several sections. Section 2 describes the specifics of the model, while section 3 describes the role and functioning of beliefs, learning, and prediction in the model. Section 4 then describes the specific model parameters used to perform model simulations, and Section 5 discusses the results of these simulations. Section 6 concludes the paper and suggests some potential future extensions to the model and modelling technique of this paper.

2 The Model

The model developed in this paper explicitly incorporates the main characteristics described in the previous section. The baseline model is a digital, no-piracy world. To this baseline piracy is introduced, allowed to run its course, and finally one of three policy responses is enacted.

This paper's model has one category of good, "books." A large but finite number of agents in the model function as "readers" who derive utility from reading books, and can also choose to become "authors" by writing and selling copies of a book. For ease of discussion, the terminology used for this model is that of a book market ("books," "readers," "authors," etc.), but the model proposed is intended to represent an abstract digital, creative work, where the initial fixed cost to create the work is high, and the cost of making new copies is low. Thus where this paper refers to "books," "authors," and "readers," the model could instead be thought of as applying, for example, to "songs," "artists," and "listeners."

The model proceeds over several hundred periods with reading and authorship decisions taken in each period. Authors create books based on forecasted profitability; readers in turn purchase copies of books based on the utility derived from reading previously unread books.

2.1 Books

The fundamental goods in this model are identical copies of books. A book, denoted b, is defined by:

$$b = (r_b, q_b) \tag{1}$$
$$r_b \in [-\phi, \phi]^n \tag{2}$$
$$q_b \in \mathbb{R}_+ \tag{3}$$

where r_b is a multidimensional set of book characteristics, and q_b represents the quality of the book. Two r_b values close together, i.e. two books with similar characteristics, can be thought of as belonging to the same genre. As will be detailed below, readers have a subjective view of book characteristics, but an objective view of book quality.

In order to simplify the creative decision process, the characteristic space of r_b is assumed to wrap at its boundaries in each dimension; the distance between any two points is considered to be the length of the shortest path between those points, which may cross one or more boundaries. This wrapping is done to eliminate "corners" in the characteristic space so that, *ex ante*, all characteristic locations are indistinguishable from one another. This is in contrast to some models where explicit edges and corners are a desirable model feature representing extreme positions.

For 1-dimensional space, characteristics can be interpreted as points on the circumference of a circle, as in the "circular city" model of Salop (1979), with distances between points measured as the shortest arc length between those points. In two dimensions, the space is a torus—topologically equivalent to the surface of a doughnut with a hole in it; or as a square where moving beyond an edge of the square moves into the space on the opposite size of the square. In higher dimensions the space constitutes a hypertorus. For any dimensionality, points near $-\phi$ and points near ϕ on a given dimension are near each other, with exactly $-\phi$ and ϕ representing the same coordinate. Figure 1 depicts one-dimensional book characteristics as points on a circle, while Figure 2 depicts a few shortest paths in a two-dimensional characteristic space.

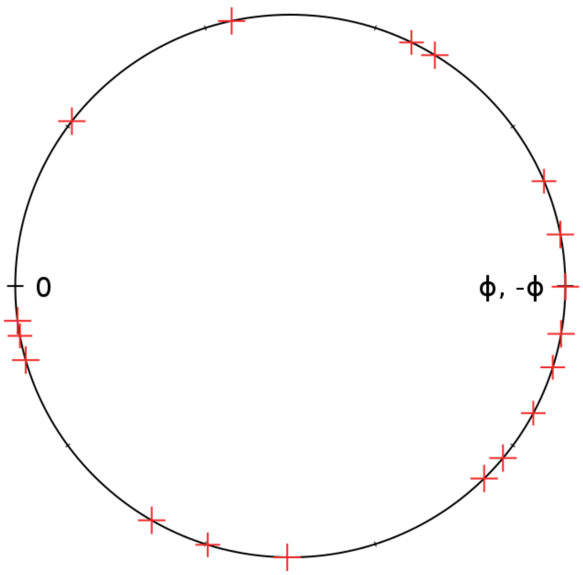


Figure 1: One-dimensional characteristic depiction. The points $-\phi$ and $+\phi$ represent the same point.

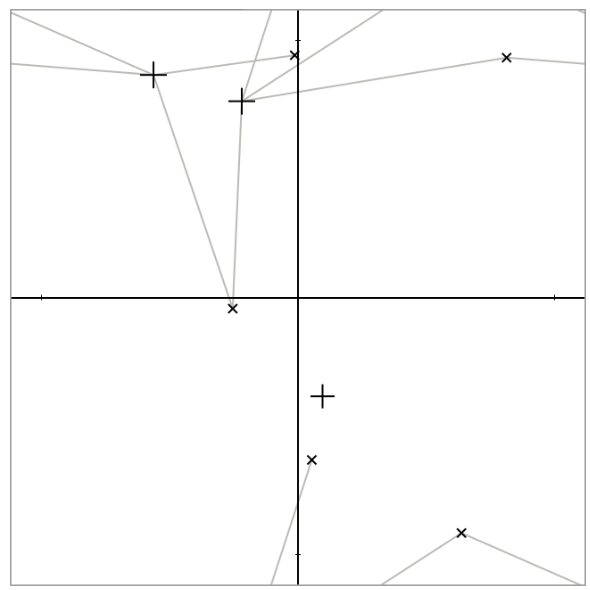


Figure 2: Two-dimensional characteristics and shortest path depictions. The centre of the graph is at coordinates $(0, 0)$, while the edges are at $\pm\phi$ in each dimension.

2.2 Agents

An agent i in this model is defined by $(r_{i,t}, L_{i,t-1}, \alpha_i)$ at time t. $r_{i,t}$ is a location in the book characteristic space representing his ideal book characteristics; $L_{i,t-1}$ is the agent's library of previously-obtained books; and α_i is a parameter reflecting the agent's authorship ability.

Agents play two roles: they are readers, consuming previously unread books written by other authors; and they are authors who create books and sell copies of those books to other readers.

2.2.1 Reading decision

A agent i in the model derives utility from reading, and from an outside numéraire good. In each period he chooses a set of books $B_t \subseteq \boldsymbol{B}_t$ and m_t to solve:

$$\max_{m_t, B_t} u(m_t, B_t; r_{i,t}) \equiv m_t + f(B_t; r_{i,t}, L_{i,t-1}) - z(\#B_t)$$
$$\text{s.t. } m_t + \sum_{b \in B_t} p_{b,t} \leq y + R_{t-1} - C_t$$

where:

- m_t is the outside numéraire good.
- B_t is the set of books the reader chooses to buy.
- $\boldsymbol{B}_t$ is the set of all books available for sale at time t.
- $r_{i,t} \in [0, \phi]^n$ is the reader's ideal book characteristics, which moves randomly over time, reflecting readers' changing tastes.[5]
- $f(B; r_{i,t}, L_{i,t-1})$ is the utility of reading the set of books B given the reader's current ideal book characteristics, $r_{i,t}$, and the reader's "library": the set of all books that have been previous read by the reader, $L_{i,t-1}$. Readers only receive utility from *previously unread* books: that is, adding a book to B that is already in $L_{i,t-1}$ does not increase the value of $f(\cdot)$.
- $z(\#B)$ is an opportunity cost of reading $\#B$ books in the current period, and is increasing in the number of books. This can be thought of either as reduced enjoyment from reading too many books at once, or in a labour market interpretation, as foregone income from increased leisure (i.e. reading).
- $p_{b,t}$ is the current price of book b.
- y is a fixed, exogenous, per-period income, identical across readers. This also determines the baseline utility, i.e. the utility of a hypothetical reader who does not participate in the book market at all.
- R_{t-1} and C_t are the net revenue and costs of authorship, discussed below.

To simplify the optimization problem readers are assumed myopic in their book selection: they do no forecasting of future availability of books or the future prices of books, but rather make a myopic, utility-maximizing selection based on the current selection of available books, prices, and their available income. This myopia, combined with the lack of utility for re-reading, imply that reader utility is dependent on a constant supply of new works: in the absence of new books in the world, a reader's utility will soon decline to his exogenous income level.

[5]This randomness can be interpreted as readers' changing tastes over time. It is also a simplifying assumption to help justify that the characteristic space can be treated as essentially uniform: any reader "hotspots" do not persist.

2.2.2 Authorship decision

Any agent in the model can choose to create a book at the beginning of period t. For convenience, the term "author" will be used to refer to agents who choose to create a new work.[6] An author choses to create a new book when doing so is profitable. Book creation, however, is expensive with both a high fixed cost, c_{create}, and an additional effort component, $\ell \geq 0$, chosen by the author. Author effort directly determines the quality of the book according to:

$$q(\ell) = \alpha_i \ln(\ell + 1) \tag{4}$$

where α_i here is a heterogeneous, reader-specific innate authorship ability of reader i.

Book authorship not only incurs a cost, but also takes an expected $T_{create} \geq 0$ periods to complete.[7] During this time, the author is unable to start an additional book (though still participates as a reader in the model). In the special case $T_{create} = 0$, creation occurs instantly—that is, the book becomes available on the market in the same period in which creation was undertaken. The creation costs, $C_{create} + \ell$, are incurred immediately when book writing is initiated.

The book characteristics, r_b, are determined by the author's current ideal book characteristics, $r_{i,t}$, at the time of undertaking the creation.[8]

2.3 Book market

Upon completion of a book, the book is immediately added into the author's library, $L_{i,t}$, and is available for the author to sell on the market, to attempt to earn a profit. The seller side of the book market is controlled by the authors of each book. For each completed book, an author chooses a price $p_{b,t}$ for period t and sells copies of it, taking into account:

- A low marginal cost, c_{unit}, of creating copies of a book. Copies can be created from the author's copy instantly on demand.
- A fixed maintenance cost, c_{maint}, of keeping a book available for sale for a period. If desired, authors can keep multiple books on the market at once, incurring c_{maint} for each book.
- Once an author believes a book to be no longer profitable, a book can be removed from the market. As a model simplification, it is assumed that removal is permanent.

[6]This is, however, only a notational convenience: all agents are potentially authors.

[7]Specifically, creation takes one of $\{T_{create} - 1, T_{create}, T_{create} + 1\}$ periods to create, with equal probability of each; the value is realized only after the creation decision is made. This randomness helps to spread out creation across simulation periods, avoiding cycles resulting in a significant number of authors creating exactly every T_{create} periods.

[8]This assumption can be interpreted as authors choosing to write in the genre they know best. Since random agent movement over time makes all locations characteristic space *ex ante* indistinguishable, the author has no reason to prefer any particular characteristic point over any other.

The choice of books to keep on the market and the decision of whether to undertake a new creation together determine C_t in the author's budget constraint, (4):

$$C_t = \mathbb{1}_{create}(c_{create} + \ell) + c_{maint} \cdot (\#M_t)$$

where $\mathbb{1}_{create}(\cdot)$ is an indicator function equal to its argument if the author chooses to create, and otherwise equal to 0. $\#M_t$ is the number of the author's (finished) books that the author has decided to keep on the market in period t.

On the buyer side of the market, readers observe prices, book characteristics, and book quality, and choose whether or not to buy books by maximizing their utility, (4). The aggregate decisions of readers then determine each author's net revenue, received as income in period $t+1$:

$$R_t = \sum_{b \in M_t} (p_{b,t} - c_{unit}) s_{b,t}$$

where M_t is the set of the author's books on the market in period t and $s_{b,t}$ is the number of sales of book b in period t.

2.4 Timing

The timing of agent actions in period t are as follows:

1. Agents incorporate new information from the previous period's observations into their beliefs.

2. Agents receive exogenous income, y, plus any net revenue, R_{t-1}, from sales of their books in the previous period.

3. If an agent is not already busy writing a new book, that agent can choose whether or not to undertake authorship in this period, and if so, the effort level ℓ to invest in the creation. The effort and fixed cost are subtracted from this period's available income.

4. If an agent has any books that have either just been completed, or are still on the market, he chooses whether or not to keep the book on the market, and if so, the price at which to sell the book. The market maintenance cost of all books made available on the market is subtracted from this period's available income.

5. Readers choose the set of books to read that maximizes their utility, and incur the cost of obtaining those books. Any leftover income is spent on the numéraire good.

2.5 Author knowledge and beliefs

At the beginning of period t, potential authors (i.e. all readers) have an information set I_t consisting of observations of each book's:

- Characteristics, r_b
- Quality, q_b
- Price history, $p_{b,\tau}$
- Market status history (i.e. still on the market or removed), $b \in M_\tau$
- Sales history, $s_{b,\tau}$

for all periods $\tau < t$. This information is used to decide when to author and how to price authored works.

Authors maintain two separate but related beliefs using this information about book authorship that evolve over time. The first is a belief, Π, about lifetime profit, $\pi_t(b)$, of a book whose creation is started in period t. The second is a belief, S, about the single-period demand, $s_t(b)$, for a book on the market in period t.

In principle, these two beliefs are manifestations of the same thing—a belief about reader demand—but in practice the complexity of the model environment makes it intractable to derive profitability over several periods from the individual demand. Actions taken by other agents between the time the book is undertaken and released to the market may change the demand between the time a book is undertaken and the time it becomes available on the market. Moreover demand may itself be only loosely predictable within a single period, but forecasts of profitability may be considerably more accurate when considered as an aggregate over the multiple periods of a book's lifetime.

Lifetime profit and single-period demand are random variables, Π and S, respectively, that are linear functions of variables in I_t with parameters $\boldsymbol{\beta}$. $\pi_t(b)$ and $s_t(b)$, are observations of these random variables generated within the model. Beliefs are given by the joint distribution over the parameters $\boldsymbol{\beta}$ and error terms in the linear equations for Π and S. This distribution in turn results in distributions of Π and S.

Each belief is used at a different point in the book's life cycle. The distribution of Π is used to determine whether to begin writing,[9] according to whether or not the effort level ℓ^* that maximizes:

$$\begin{aligned} \mathrm{E}\left\{\Pi \mid q_b = q(\ell)\right\} - (c_{create} + \ell) \\ \text{s.t. } c_{create} + \ell \leq y + R_{t-1} \end{aligned} \tag{5}$$

results in a positive expectation. If so, the author begins creation with expected-profit-maximizing effort ℓ^*.

The second belief, that is, the distribution of S, is used once authors have completed a book so as to decide how to price copies of that book for period t. It is also used to price previous books that have not yet left the market, and determines when the author

[9] Note that, as previously mentioned, authors can only undertake one creation at a time, and so this decision is only available when authors do not have a current book in progress.

decides to remove the book from the market. That is, the author chooses a price for $p_{b,t}$ for each book $b \in M_t$ to maximize expected current period profit from that book:

$$\mathrm{E}\left\{S \mid p_{b,t}, b\right\}\left(p_{b,t} - c_{unit}\right) - c_{maint} \tag{6}$$

removing the book from the market if this maximized expectation is negative.

The specific form of both beliefs are linear equation Bayesian belief models; discussion of how such a model is constructed, how it incorporates new information, and how it is used for prediction is deferred until Section 3.

2.6 Modelling piracy

The model described thus far acts as a benchmark stage against which the effects of piracy can be measured. This is done by introducing piracy as a change to the model and observing its effects.

Introducing piracy requires the creation of a network between readers. This network is an undirected graph, with a fixed proportion of the *potential* links between pairs of readers established at random and fixed over time.[10] Piracy operates over this network by allowing readers to obtain copies of books from network-adjacent readers at a small cost c_{piracy}.

Piracy changes readers' optimization problem, (4), by introducing an additional choice of pirated books to obtain from network-adjacent readers. The updated utility function is as follows where additions relative to (4) have been emboldened for emphasis:

$$\begin{aligned} \max_{m_t, B_t, \boldsymbol{P_t}} \; & u(m, B_t, \boldsymbol{P_t}; r_{i,t}) \equiv m + f(B_t \cup \boldsymbol{P_t}, r_{i,t}) - z(\#(B_t \cup \boldsymbol{P_t})) \\ & \text{s.t. } m + \sum_{b \in B_t} p_{b,t} \boldsymbol{+ c_{piracy}(\# P_t)} \leq y + R_{t-1} - C_t \end{aligned} \tag{7}$$

The reader chooses B_t from the set of on-market books, as before piracy, but now also chooses $P_t \subseteq \boldsymbol{L_t}$, where $\boldsymbol{L_t} = L_{1,t-1} \cup L_{2,t-1} \cup \ldots \cup L_{n,t-1}$, that is, the set of all books available to a reader from the n readers adjacent to the reader in the piracy network. It should be noted that authors do not distribute copies of their own, on-market books via the network.

An assumption implicit in this structure is that pirated material does not instantaneously spread through the network. Instead, the books available to a reader at time t is the set of neighbour's books that were obtained—whether through market purchases or piracy—as a result of the neighbour's book choices in period $t-1$. This allows authors some potential to still earn profits: the author has one period of sales during which he only competes for sales with other authors and the pirated copies of other books, but not with pirated copies of his own book.

[10] A network of N nodes has precisely $\frac{N(N-1)}{2}$ potential distinct links between nodes, and a fixed proportion of these links are established by selecting the desired number of potential edges randomly.

A second characteristic of the piracy structure is that it has no opportunities for gains from piracy beyond the direct gain to the reader of obtaining a copy of a book. That is, the cost of piracy, c_{piracy}, is an incurred cost—the opportunity cost of searching for and downloading a pirated copy—but not a payment to the individual providing the pirated copy.

The introduction of piracy has the effect of enlarging a reader's potential set of available books. Previously readers could choose only from books kept on the market by their authors. Under piracy, however, on-market books plus any books possessed by adjacent readers, even if those books are no longer on the market, are available to readers. Moreover, these books are available to a reader at the lower cost c_{piracy}, while delivering identical utility.[11] A book copy obtained via piracy and a copy of the same book obtained purchasing from its author are identical goods: reader utility depends on $B \cup P$, which is obviously identical whether a book is included in P or B.

Figure 3 depicts a potential piracy network: Reader A can obtain pirated copies in period t of any books obtained in previous periods by readers B, C, or D. A copy possessed only by E prior to period t would not become available via piracy to A until $t+1$—and then only if first obtained during period t by B, C, or D. Copies of that book would, however, be available in period t to E's neighbours, B, D, and F. When choosing a set of books at time t, each reader's immediate neighbours make available copies of all books they possess.

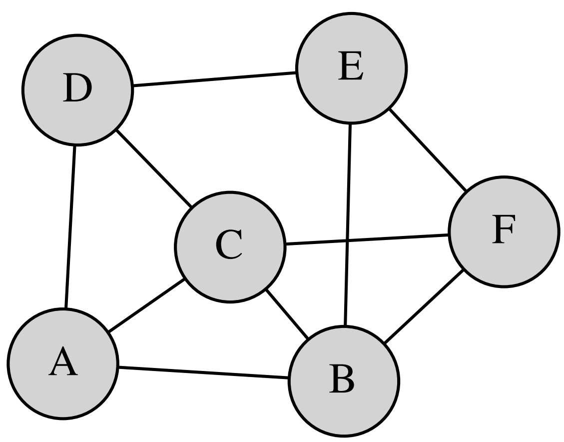


Figure 3: Small piracy network depiction.

The authorship decision procedure under piracy is unchanged: potential authors still undertake creation when the expected profit of creation is positive, and price (or withdraw) existing books according to their belief about per-period demand. Any changes that occur in the model are thus driven by changes in the distributions of

[11]In theory, nothing requires that c_{piracy} be less than $p_{b,t}$, but as long as c_{piracy} is close to (or below) c_{unit}, it will typically be the case: an author would remove a book from the market rather than price the book at or below his marginal cost of copies, c_{unit}.

those beliefs, *not* by any change in the fundamental behavioural structure underlying authorship.

To model that knowledge of the invention of piracy is likely to make creators less certain in their beliefs formed in the pre-piracy environment, beliefs receive a one-time "weakening"—discussed in detail in Section 3.3—that allows them to update beliefs more quickly in response to observations from the newly-piratic environment.

2.7 Policy responses

After many periods of the model under piracy, once a long-run equilibrium has been reached, a potential social planner remedy to the piracy problem is introduced. Three policy mechanisms are explored: two based on public provisioning, and one based on making pirating more expensive through a detection-and-fine mechanism.

The first public provisioning mechanism is based on two of the responses to piracy as proposed by Varian (2005): first a notion of public provisioning of creative works; and second a monitored fee and transfer system. This is modelled here through a tax and transfer mechanism.

The public provisioning system acts as a provider of all off-market books, available to all readers at marginal cost $c_{public} \equiv \min\{c_{unit}, c_{piracy}\}$.[12] A fixed, per-period lump-sum tax, τ, is collected from each agent at the beginning of each period, and is redistributed to all authors at the end of each period in proportion to the number of public downloads of each author's books. It is also assumed that authors cannot download copies of their own works, so as to eliminate the possibility of an author, for example, downloading many copies of his own book to increase his share of public revenue.

As a simple illustrating example, suppose a lump-sum tax of 10 is collected from each of 100 agents in period t, who choose, in aggregate, to download the works of three authors: 10 copies of a book written by author A, 25 and 35 copies of two books written by B, and 30 copies of a book written by C are downloaded. Under the first policy, A would receive 100, B would receive 600, and C would receive 300.

The public provisioning mechanism enlarges readers' choice sets once again: readers may now choose to buy on-market books, to obtain pirated copies from network-adjacent readers, or to obtain off-market books via the public provisioning system. That is, (7) is further amended to:

$$\begin{aligned} \max_{m,B,P,\boldsymbol{G}} \; & u(m,B,P,\boldsymbol{G};r_{i,t}) \equiv m + f(B \cup P \cup \boldsymbol{G}, r_{i,t}) - z(\#(B \cup P \cup \boldsymbol{G})) \\ \text{s.t. } & m + \sum_{b \in B} p_{b,t} + c_{piracy}(\#P) + \boldsymbol{c_{public}(\#G)} \leq y - \boldsymbol{\tau} + R_{t-1} - C_t \end{aligned} \tag{8}$$

where $Z \subseteq \boldsymbol{B}_t^{\complement}$, i.e. all books not on the market in time t. Note also that R_{t-1} is implicitly changed here: it includes both profit from private market sales, *plus* any earned profit from the public provisioning transfer.

[12] In other words, the public provider can use either the author's distribution technology, or the piracy distribution technology, and chooses whichever is cheaper.

As under piracy, the authorship decision *structure* is unchanged: any changes to model actions come though changes in the distribution of beliefs in response to new information in the policy environment.

The potential life cycle of a book is affected by public provisioning. Under the market (and piracy) environments, authors only earn profits from sales in the private market; now authors earn from those sales but *also* earn from public downloads after the book leaves the private market. Thus where it would be irrational to write a book without an intention to bring it to the market at all,[13] that is a valid strategy here: authors can write to obtain *only* public earnings for the book. Also note that piracy, while having a much reduced role, still exists. It is, however, only relevant for private-market books; off-market books are available (legally) from the public provider at a cost at least as low as the cost of obtaining through piracy.

The second policy is an adaptation of the first that assigns a fixed number of votes to each agent. Agents allocate their votes in proportion to the utility realized from each book obtained from the public market in the current period. Multiple votes may be cast for the same book, if a reader so desires,[14] but votes may only be cast for copies a reader (and not the author) obtained from the public market in the current period. The collected tax is then redistributed to authors in proportion to votes received rather than the number of downloads.

To amend the earlier illustrating example, suppose that under the second policy readers each cast three votes in proportion to the utility they received from the downloaded books. Suppose that, in aggregate, readers cast: 20 votes cast for A's book, 75 votes for each of B's two books, and 10 votes for C's book.[15] Then A with 11.1% of the votes would receive a transfer of 111, B with 83.3% of the votes would receive 833, and C with the remainder of the votes would receive 56.

The reading and authorship decisions are unchanged from the public sharing policy. Any differences in the model comes about from changes in author beliefs in response to the different public payment structure.

The third policy option explored is a policy of costly, imperfect detection mechanism with fines for readers found to have pirated books in the current period. This amends

[13] This outcome *can* happen in the simulation, particularly when T_{create} is large and the number of belief observations is relatively low: per-period demand beliefs could change between the authorship decision time and the authorship completion time such that the author no longer predicts an ability to earn back the fixed market cost, c_{maint}.

[14] i.e. a reader who received twice as much utility from reading book a as from reading book b would cast twice as many votes for book a as he would for book b.

[15] Note that the total number of votes need not add up to 300: a reader who downloads no public books in a period does not vote in that period. Thus in this example, only 60 of the 100 readers cast votes.

the reader's utility maximization from (7) to the expected utility maximization:

$$\max_{m,B,P} \mathbf{E}u(m,B,P;r_{i,t}) = m - \mathbf{E}\,\boldsymbol{\Psi}(\#\boldsymbol{P},\boldsymbol{\tau}) + f(B\cup P, r_{i,t}) - z(\#(B\cup P))$$
$$\text{s.t. } m + \sum_{b\in B} p_{b,t} + c_{piracy}(\#P) \leq y - \boldsymbol{\tau} + R_{t-1} - C_t \qquad (9)$$

where $\Psi(p,\tau)$ is a random variable that equals $\psi(p)$—the fine if caught pirating p books in a period—with probably $\gamma(p,\tau)$—the probability of being caught when pirating p books—and 0 with probability $1-\gamma(p)$. Both $\psi(p)$ and $\gamma(p,\tau)$ are assumed to be increasing in the number of pirated books, p, and $\psi(p)$ is furthermore assumed (weakly) convex. τ—the per-agent lump-sum tax—is now used to fund the detection mechanism, and hence it assumed that $\gamma(p,\tau)$ is also increasing in τ: that is, that the probability of detection increases with increased funding. Any collected fines under this policy are redistributed from the infringing reader to the authors of infringed works in equal proportion.

In contrast to the earlier problem, readers now maximize expected utility taking into account the probability of being caught pirating and the associated fine if caught. To simplify the problem, it is not assumed that readers must save for a potential fine: rather the fine (if caught) is removed directly from the reader's utility.

It is assumed in this paper that $\gamma(p,\tau)$ is the cumulative density function of a normal distribution, $\gamma(p,\tau) = \Phi\left(\frac{p-\mu(\tau)}{\sigma(\tau)}\right)$, i.e. the c.d.f. of a normal distribution with mean $\mu(\tau)$ and standard deviation $\sigma(\tau)$. For example, with $\mu(\tau)=7, \sigma(\tau)=3$, detection probabilities would be as shown in Figure 4. Figure 5 shows the expected value of the fine for this detection probability combined with the linear fine $\psi(p)=50p$.

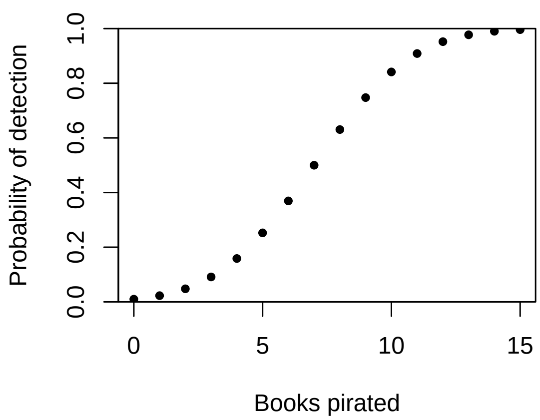


Figure 4: Detection probability

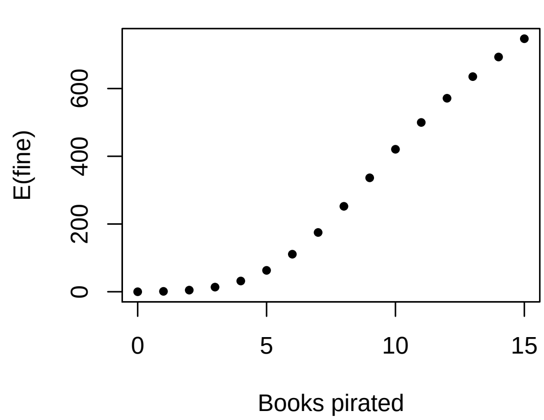


Figure 5: Expected fines

Such a detection mechanism is imperfect in that it rarely detects low levels of piracy while frequently detecting high levels. This has the effect on readers that readers can, at a very low expected fine, engage in limited piracy, but it puts a substantial damper on readers attempting to engage in high levels of piracy. Thus readers ought to be more likely to purchase legitimate copies from the market due to the reduced dissemination of pirated material through the network. Combined, these two effects should increase the size of the potential market for a given book, thus allowing authors to earn higher profits.

3 Agent beliefs

The actions of authors in this paper's model are driven by "beliefs," as has been discussed generally in section 2.5. This section gives a more concrete definition of beliefs, both in terms of general equations and the specific functional forms used in this paper's model.

A "belief" in this model is defined as a joint distribution over the of values of the coefficients, $\boldsymbol{\beta}$, and error term variance, σ^2, of a linear equation of general form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \varepsilon_t \sim \text{iid}(0, \sigma^2) \tag{10}$$

where, in the usual econometric notation, $\boldsymbol{y}$ is an n-vector of dependent variables, $\boldsymbol{X}$ is a $n \times k$ matrix of independent variables, $\boldsymbol{\beta}$ is a k-vector of coefficients, and $\boldsymbol{\varepsilon}$ is a vector of unobserved error terms following an independent distribution with mean 0 and variance σ^2.

As a concrete example for the model in this paper, an author is assumed to use the following linear equation to relate the profitability of a book to the book's quality and the number of books on the market in recent periods:

$$\Pi = \beta_1 + \beta_2 q_b + \beta_3 q_b^2 + \beta_4 M_{t-1} + u, \qquad u \sim \text{NID}(0, \sigma_u^2) \tag{11}$$

The author's belief is a joint distribution over the four $\boldsymbol{\beta}$ coefficients and the error term variance, σ_u^2, in (11). As authors observe the lifetime profits of existing books in the model environment, they incorporate this new data by updating the distribution over $\boldsymbol{\beta}$ and σ_u^2, with uncertainty over and correlation between the parameters being captured in the parameters of the distribution.

In practice, this distribution over $\boldsymbol{\beta}$ and ε are assumed to follow a joint, multivariate normal-gamma distribution. This assumption is important for two reasons: first, this is a well-defined distribution with only a few parameters from which draws can be taken (as detailed in the next section). Second, when combining a prior following this distribution with new information, the posterior remains in the same family, thus yielding a conjugate prior. This is particularly important when updating new information, discussed below, as the prior and posterior are then represented by different parameters of the same joint distribution family.

3.1 Prediction from beliefs

Given a particular distribution over $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$—i.e. a "belief" in (10)—an agent can make predictions about the world by averaging the variable of interest—e.g. profit—for a vector of values of $\boldsymbol{X}$ over many draws from the joint distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$.

For example, to evaluate the profitability of creating a book at time t, an author knows the value of M_{t-1}, but is free to choose q, the book's quality.[16] The author's predicted profit in (5) for any given q is thus:

$$E(\Pi \mid q, M_{t-1}) = \frac{1}{S} \sum_{s=1}^{S} \beta_1^s + \beta_2^s q + \beta_3^s q^2 + \beta_4^s M_{t-1} + u^s$$

where each $\boldsymbol{\beta}^s, u^s$ is an independent draw from the joint distribution over $\boldsymbol{\beta}$ and u. By reusing this set of S draws, the author can evaluate multiple q values to determine the value of q that maximizes the predicted profit.[17] Although not needed with the quasilinear utility of agents in this model, this technique could easily be adapted to agents with non-linear utility by taking an average over a utility function (rather than simply the profit value, as above).

The precise calculations needed to draw from the normal-gamma distribution used here are described in Appendix C.

3.2 Updating beliefs

The discussion of beliefs so far has simply assumed the existence of a belief distribution, but an equally important part of the discussion is how that belief distribution is dynamically updated over time in response to new observations in the model environment. The basic principle here is standard: the belief in time t becomes the prior in period $t+1$; this prior plus new observations form the posterior for $t+1$, and it is this posterior distribution that is used for predictions for actions in period $t+1$.

Beliefs thus evolve recursively over time in a natural way:

$$\begin{aligned} Belief_t &+ Data_t \rightarrow Belief_{t+1} \\ Belief_{t+1} &+ Data_{t+1} \rightarrow Belief_{t+2} \\ Belief_{t+2} &+ Data_{t+2} \rightarrow Belief_{t+3} \\ &\qquad \ldots \end{aligned} \tag{12}$$

From the recursive formulation, it is apparent that the distribution of the posterior distribution in each period must be computationally feasible to draw from, as it will become the prior in the subsequent period. Hence the importance, discussed above, that the distribution over parameters is assumed to following the normal-gamma distribution: both priors and posteriors will follow this same distribution family.

The precise calculations for combining a prior with data to generate a posterior follow Bayesian econometric calculations, which essentially derive from a continuous,

[16] Of course the author actually chooses an effort level ℓ, but any choice of ℓ uniquely determines the resulting book's quality through (4).

[17] Reusing the same set of draws is computationally important: if different draws were used, each new set of draws would imply a slightly different maximizing q, and thus a maximizing algorithm could not converge to a maximum.

multivariate application of Bayes' Theorem, and are detailed in Appendix C. The main result, however, is that the mean and covariance of the posterior β are determined by weighted sums of the information in the prior with the information in the new observations.

One obvious problem with the recursive nature of (12) is that it requires an initial distribution at $t = 0$. The approach used in this paper is to begin from a completely non-informative prior. This prior contributes nothing whatsoever to the determination of the posterior parameters; it also cannot actually be used for prediction. It has the property, however, that once enough observations have been accumulated the combination of non-informative prior and observations yields a usable prior with parameters equal to the maximum likelihood estimates of the linear model. This (informative) posterior then becomes the prior for subsequent updates, as previously described.

This approach, however, still requires some method of generating initial observations: this is accomplished in this paper by specifying initial random actions that are sufficient to generate sufficient data from the model so as to "bootstrap" the model's agents' beliefs.[18]

3.3 Belief weakening

This paper extends the above multivariate Bayesian econometric framework by adding a concept that will be referred to as belief "weakening," whereby the influence of past observations is decayed over time. Weakening is specifically modelled as a manipulation applied to a posterior before it is used as the subsequent period's prior. Specifically, the covariance matrix of the $\boldsymbol{\beta}$ parameters, $\overline{V}$, is multiplied by the square of a weakening factor $w > 1$, as follows, when transitioning from prior to posterior:

$$\underline{V}_{t+1} = w^2 \overline{V}_t \tag{13}$$

This is, in other words, a mean-preserving spread of the posterior when used as the subsequent period's prior. In effect, as can be seen from the equations in Appendix C, this weakens the influence of past observations, giving relatively more explanatory power to new observations. This also amends the recursive overview of belief updating in (12): the first term on each line is now a *weakened* version of the belief that was held in the previous period. Because of the recursive nature of beliefs, this decay is exponential: the influence of older observations is continually deweighted relative to new information.

This weakening concept is further extended to allow w to have different values in different periods; in most periods, w should be a relatively small value ($w = 1.02$ is used in this paper's main results), but in periods of turbulence—such as the initial simulation periods, or immediately after the introduction of a significant model environment change such as piracy—larger w values are used to allow beliefs to quickly adapt to changes in the model environment.

[18]It must also be verified that the initial parameters do not affect the model results.

There are several justifications for the use of prior weakening in a model such as that developed in this paper. First, because actions influence beliefs which influence future actions, *etc.* the "true" coefficients of (10) need not be constant. Instead a model could simply approach a fixed-point equilibrium over many periods where the coefficients only eventually become constant.[19] If every observation is given equal weight, old observations—when the model was in a very different aggregate state—are treated as equally important to predicting the current model state as recent observations; predictions would thus be biased towards the old state of the world, even when more relevant, more recent observations should yield more information about the current state of the world.

A second reason to have beliefs decay over time it allows for agents to react quickly to fundamental shifts in the model, even if the effects of that fundamental change are unknown: by applying a large decay immediately after fundamental model changes, beliefs can adjust quickly, rather than by being held back by observations from before the fundamental change. For example, when piracy is introduced into this paper's model, a large weakening factor $w = 2$ is applied to beliefs.

A third reason to employ weakening is to deal with the initial conditions problem mentioned in the previous section: by combining the suggested initial actions approaches with sustained, relatively large weakening in initial simulation periods, the effect of initial, exogenously determined actions or model-specified initial priors can be decayed away. This helps ensure that a model is driven not by initial conditions, but rather by the complex interactions of the agents within a model.

A fourth justification for weakening comes from the fact that agents in an economic model are, epistemologically, meant to model human beings, who indeed do not have perfect recall. Cognitive psychology, going back to the work of Tversky and Kahneman (1973), suggests that human perceptions of probability are governed by an availability bias: that more weight is given to more readily available memories. And indeed more readily available memories are those that are more recent or more significant. The weakening approach used here is able to capture this: the belief distributions are weighted towards more recent observations, but more significant observations remain more significant than other observations observed in the same time frame, even as their significance decays through compounding prior weakening.

3.4 Authorship beliefs

The model of this paper employs two specific belief distributions, as previously discussed in Section 2.5: one about the parameters of the long-term profitability of creation, and one about the parameters of per-period demand for a book.

For predicting lifetime book profits, an author in this paper's model relies on the

[19]Though even this is not guaranteed: a model could also contain endogenous cycles where coefficients forever fluctuate.

linear equation introduced in section 3.1, and repeated here for further discussion:

$$\Pi = \beta_1 + \beta_2 q_b + \beta_3 q_b^2 + \beta_4 M_{t-1} + u, \qquad u \sim \text{NID}(0, \sigma_u^2) \tag{11}$$

Π is the random variable of a book's lifetime profit, not including initial creation costs. q_b is the book's quality, as chosen by its author. M_{t-1} here is the average number of books on the market over the $T_{create}+1$ periods before the book was created, allowing for the possibility of reduced profits in a more crowded book market. The use of this average, rather than the number of books in the previous period, is designed to eliminate possible cyclical behaviour in the profit prediction resulting from short-term market size fluctuations. Note also that there are no terms related to book characteristic locations, owing to the assumptions described in Section 2 regarding an *ex ante* identical characteristic space.

For predicting demand for a book in a given period, authors use the linear model:

$$\begin{aligned} S = \gamma_1 + \gamma_2 p_{b,t} + \gamma_3 q_b + \gamma_4 S_{b,t-1} + \gamma_5 T_{b,t-1} + v, \qquad v \sim \text{NID}(0, \sigma_v^2) \\ \text{s.t. } \gamma_2 \leq -.05, \gamma_3 \geq 0, \gamma_5 \leq -1 \end{aligned} \tag{14}$$

S is the random variable of the number of sales of a book will have in period t; $p_{b,t}$ is the price of the book in period t; and q_b is the book's quality. $S_{b,t-1}$ is the number of cumulative sales of the book up to and including period $t-1$, included to reflect the fact that each previous sale reduces the number of readers who will consider the book (since readers never buy books they already own). $T_{b,t-1}$ is the number of periods a book has gone without sales so as to allow authors to reflect an observation that a lack of sales in previous periods is negatively correlated with sales in future periods.

An author's predictions about sales are determined by his belief—that is, his distribution over the $\boldsymbol{\gamma}$ coefficients and error term variance, σ_v^2, in (14). This belief is updated in response to observations of the sales of books in the model using observations of sales, $s_{b,t}$, of all market books in the model.

The three linear restrictions impose notional limits on this joint distribution: $\beta_2 \leq -0.05$ is a restriction that demand curves are always at least slightly downward sloping, i.e. at a higher price, *ceteris paribus*, sales will be lower. $\beta_3 \geq 0$ is stating that, all else equal, any increase in quality should translate into an increase in quantity demanded. This essentially follows from the utility structure of agents: agents receive higher utility from higher quality books, and thus there should be (weakly) more people willing to buy, *ceteris paribus*, at a higher quality. The restriction $\beta_5 \leq -1$ serves as a recognition that a book that has had several periods without sales in recent periods is increasingly unlikely to see future sales, even when other parameters suggest that there should be positive demand for the book.

The restrictions are seldom strongly binding in equilibrium in practice, but often help constrain the belief models in the early, often chaotic, periods of the model. Even if some observations appear to violate the constraints, readers are, by assumption, incapable of believing, no matter what observations imply, that sales will be higher at a higher price. In computational implementation, this means that reader prediction is modelled by taking draws from a truncated multivariate distribution.

4 Model implementation

Hitherto the model discussion has been kept general in terms of specific functional forms for reader and author actions. Actually simulating the agent-based model, however, requires specific values for the various parameters of the model. This section covers the specific choices used for simulation of the model, then discusses the results of analyzing many runs of the agent-based model.

Each simulation consists of 620 periods: the first 20 periods are a burn-in period with relative large belief weakening between periods. The next 200 periods are run without piracy. In $T_{piracy} = 221$ piracy is introduced and allowed to run for another 200 periods; finally, in $T_{public} = 421$, public provisioning is added and the simulation runs for a final 200 periods. For analysis of results, averages of variables of interest are taken over the last 25 periods of each stage. In practice, 200 periods is more than enough for the simulation to reach a stable state, i.e. where fluctuations are minimal and there is no obvious trend in the model status over time.

4.1 Functional specifications

The subutility functions $f(B, r_{i,t})$ and $z(\#B)$ in (4), (7), and (8), which translate a set of books into a utility value through:

$$u(m, B; r_{i,t}) \equiv m + f(B, r_{i,t}) - z(\#B) \tag{15}$$

are specified as:

$$f(B, r_{i,t}) \equiv \sum_{b \in (B \setminus L_{t-1})} \left(q_b - 3 \left\| r_b - r_{i,t} \right\|^2\right) \tag{16}$$

$$z(\#B) \equiv (\#B)^2 \tag{17}$$

The utility of an unread book is linear in the book's quality, and in the square of the distance between the book's characteristics, r_b, and the reader's most preferred characteristics, $r_{i,t}$. The purpose of $B \setminus L_{t-1}$ is to ensure that the utility of an already-owned book is 0.[20] The coefficients in the equations (1, -3, and 1) were selected by trial and error so as to condition the model into one where characteristic differences, quality, and the opportunity cost each play a meaningful role in the model equilibria.

The author creation ability function, (4), is chosen to have parameter $\beta = 0$, which simplifies the function (through L'Hôpital's Rule), to $q(\ell) = \alpha_i \ln(\ell + 1)$. The distribution of α_i, individual author ability, across readers is uniform, with end-points randomly chosen as discussed in the following section.

Weakening of prior beliefs is applied at three different rates. In the first 20 periods, priors are weakened by a factor of $w = 1.5$. This means that, for beliefs after $t = 20$,

[20] In actual implementation, it is easier to simply ensure that previously-read books are not added to the choice set B in the first place; the two approaches are functionally identical.

observations from $t = 1$ have 0.045% of the weight of observations from $t = 20$ in determining beliefs. After the initial period, weakening is reduced to $w = 1.02$, under which it takes 35 periods for an observation to decay to half of the weight in determining belief distributions relative to new observations. When the model transitions from one environment to another—i.e. when piracy is introduced in T_{piracy} or public sharing introduced in T_{public}, beliefs are weakened by a one-time factor of $w = 2$, which means the weight of all pre-transition observations in determining beliefs is instantly halved relative to new observations, after which weakening returns to its gradual $w = 1.02$ level.

Predictions for profit and demand are made by averaging the values over 100 draws from the belief distributions.

4.2 Simulation parameters

The agent-based model developed in this paper has a large number of parameters; of interest is how these parameters affect the simulation results across the three main stages of the simulation. Parameters are drawn at random for each simulation, as described below, and then many runs of the simulation are examined to investigate the statistical relationship between parameters and simulation results.

The parameters selected for each simulation are:

- The number of readers in a simulation is drawn from $\mathcal{U}\{100, 200\}$, where $\mathcal{U}\{a, b\}$ denotes the discrete uniform distribution, that is, a discrete distribution with an equal probability of each integer z satisfying $a \leq z \leq b$.
- The number of dimensions of book characteristics is fixed at 2.
- The density of the simulation (in terms of average number of readers per unit squared) is drawn from $\mathcal{U}(0.25, 4)$, where $\mathcal{U}(a, b)$ denotes the continuous uniform distribution. The density, combined with the number of readers, implicitly defines ϕ, the book characteristic boundary in (2).
- Readers' ideal book characteristics follow a random walk in a uniformly-distributed random direction in each period, with the distance drawn from a χ^2_1 distribution rescaled to have mean s, where s is drawn for each simulation from $\mathcal{U}(0.1, 0.9)$.
- Per-period, per-agent exogenous income, y, is fixed at 1000, which is sufficiently large to seldom bind in practice.
- The fixed creation cost is drawn from $\mathcal{U}(50, 250)$.
- The creation time is drawn from $\mathcal{U}\{0, 5\}$.
- The relative authorship ability values, α_i in (4), are drawn for each reader in a simulation from $\mathcal{U}(0, \overline{\alpha})$. $\overline{\alpha}$ itself is drawn from $\mathcal{U}(5, 15)$ for each simulation.
- Market maintenance cost, c_{maint}, is drawn from $\mathcal{U}(0, 50)$.
- Per-copy unit costs of books is drawn from $\mathcal{U}(0, 10)$.
- Per-copy piracy cost is drawn from $\mathcal{U}(0, 10)$.
- The proportion of random links to establish in the network used for piracy is drawn from $\mathcal{U}(0.05, 0.25)$. For a 150-reader simulation, this means there will be between 559 and 2794 network edges.

- The policy is selected, with equal probability, from the three policies
- The policy lump-sum tax, τ, is drawn from $\mathcal{U}(1, 100)$. This same distribution is used whichever policy is in effect.
- For public sharing with voting, the number of votes per reader is drawn from $\mathcal{U}\{1, 5\}$.
- For the catch-and-fine policy, the probability of being caught is equals the c.d.f. of a normal distribution with mean $10 - 0.1\tau$ and standard deviation 3.
- For the catch-and-fine policy, the fine equals F times the number of copies of books pirated in the caught period, where F for each simulation is drawn from $\mathcal{U}(0, 100)$.
- Initial behaviour—before enough observations have been accumulated to give readers informed beliefs—is according to the following simple rules. In theory these values should not (because of the considerable early period weakening) affect results; varying them enables statistical verification that this is indeed the case.
  - A reader initiates book creation in a period with fixed probability drawn for each simulation from $\mathcal{U}(0.1, 0.4)$.
  - An author who has randomly chosen to write then selects an effort level $\ell \sim \mathcal{U}(\underline{\ell}, \overline{\ell})$, where $\underline{\ell}$ and $\overline{\ell}$ are drawn for each simulation according to $\underline{\ell} \sim \mathcal{U}(0, 50)$ and $\overline{\ell} \sim \mathcal{U}(\underline{\ell} + 10, \underline{\ell} + 100)$.
  - When deciding on the market price, the author draws a random price from $\mathcal{U}(c_{unit} + \underline{p}, c_{unit} + \overline{p})$, where $\underline{p}$ and $\overline{p}$ are drawn for each simulation according to $\underline{p} \sim \mathcal{U}(0, 10)$ and $\overline{p} \sim \mathcal{U}(\underline{p} + 10, \underline{p} + 40)$.
  - In subsequent periods, the author keeps the book on the market with a probability drawn for each simulation from $\mathcal{U}(0.25, 0.75)$.
  - For a book remaining on the market for an additional period, the author cuts the book's markup over marginal cost to a fraction a of the previous period's markup, i.e. $p_{b,t} = c + a(p_{b,t-1} - c)$. The fraction a is drawn for each simulation from $\mathcal{U}(0.25, 0.75)$.

  Once data on profitability and per-period demand are sufficient to inform beliefs, the above rules are no longer used: authors create according to their profit belief, and price according to their market demand belief.

4.3 Software implementation

The software developed for this paper fits, generally, into two separate components, both written in C++, and both open source projects. The first is a general purpose agent-base model library called Eris, used here but designed to be applicable to agent-based models in general. Eris deals with much of the mechanics of the simulation: managing agent interactions, optimization orderings, some basic agent functionality, and provides the implementation underlying the general Bayesian belief structure described here. An overview of the basic structure of a model written in Eris is provided in Appendix A.

The second part of the implementation is the specific creativity model. This contains all of the specifics of creativity-model specific components: for example, agent optimization over books, the piracy network, the specific variables of the belief structures, and providing agents that collect and (for the public sharing policies) redistribute collected taxes.

Links to the code and documentation of Eris and the creativity model developed in this paper are given in Appendix A. Both are freely available open source projects.

5 Results

The agent-based model described here was simulated 10230times, with parameters as described in the previous section; approximately one-third use each of the described policy responses.[21] For each simulation, many different variables of interest are calculated and averaged over the last 25 periods of each simulation stage, and first 25 periods of the piracy and policy stages. Some quantities are further averaged over agents or books, where appropriate: for example, net utility is calculated by averaging utility values over both periods and agents.[22] Book-specific values are similarly averaged over all books seen in the period: for example, book profit is calculated as total book profit minus total book costs observed over the period, divided by the number of books with profits or costs incurred over the sample period. Variables such as the number of books written, number of pirated copies, and reader spending are similarly averaged over the analyzed period.

This section discusses the results of the model by first looking at how different parameters affect the general categorization of simulations, and then by looking at the outcomes of the model among those simulations that remain viable even under piracy and public sharing.

5.1 Data categorization

Simulations are first categorized according to whether or not they have writing activity. Any simulation with fewer than 5 new books written per hundred readers over the entire 25 analysis periods was considered to have "no writing" in that stage. This categorization provides a crude look into whether simulation parameters are working through the belief and behaviour of agents in a reasonable way.

Table 1 shows the breakdown of simulations into different categories of writing activity. The second column shows simulations that were not able to sustain writing activity even before the introduction of piracy, while the third through sixth columns show the number of simulations with at least five books per hundred readers written over the analyzed periods.[23]

[21]Specifically 3399use the public sharing policy, 3377use the public sharing with voting policy, and 3454use the catch-and-fine policy.

[22]i.e. the average of $25a$ observations, where a is the number of agents in the simulation.

[23]That is, five books in total over the entire analysis period, or an average of at least 0.2 books per

Policy	total sims.	✗pre-pir.	✓pre-pir. ✗piracy ✗policy	✓pre-pir. ✗piracy ✓policy	✓pre-pir. ✓piracy ✗policy	✓pre-pir. ✓piracy ✓policy
Public sharing	3399	626	85	1747	10	931
Public voting	3377	621	581	1216	61	898
Catch and fine	3454	642	470	1408	41	893

Table 1: Pre-piracy, piracy and policy categorizations. ✓indicates that there was writing during the associated simulation stage; ✗indicates stages without writing.

From the second column, it can be observed that there are some simulations that simply cannot sustain writing activity even in a no-piracy world. These simulations tend to be characterized by simulation parameters that make writing relatively expensive (i.e. higher fixed costs) and a lower upper-bound of authorship abilities.

From this an apparent difference between the policies can be plainly seen: public sharing appears substantially more effective at restoring at least some creative activity. Among simulations whose randomized set of parameters led to a complete abandonment of writing under piracy, 95.4% of simulations see a recovery under the public policy, while the sharing-with-voting and catch-and-fine policies have recovery rates of only 67.7% and 75.0%, respectively.

Looking at the parameters that lead to the above results, some interesting relationships emerge which are worth discussing. The full set of parameter distributions for each data category are detailed in Appendix D.

The first thing to note is that the distribution of the initial behavioural rule draws do not differ in any meaningful way across simulation categories. This is a highly desirable result for an agent-based model: initial parameters do not have a statistically significant effect on model outcomes.

The density of the characteristic space—that is, the average number of readers per square unit—plays a big role in determining whether writing occurs in the benchmark pre-piracy stage. In more dense environments, writing is more likely. This result is not particularly surprising given the model setup: more dense environments means that there are more readers, on average, with ideal characteristics within a fixed distance of any given book's characteristics. Thus books should face higher demand, *ceteris paribus*. Density seems to play a much less significant role, however, in determining how simulations are affected post-piracy, *conditional* on the simulation having had pre-piracy writing.

T_{create}, the mean number of periods required to create a book, seems to have a big effect on whether or not a simulation has consistent writing through all stages. The mean creation time is noticeably lower in simulations without either piracy or public writing, and noticeably higher in simulations with writing in all stages. This is due to the competitive effects of the variable: when books take longer to create, there are fewer

hundred readers per period.

books available in general, leading to less competition and thus higher profits. This holds even under piracy: not only are there fewer legal books, but also fewer pirated books competing with a new book release.

The range of the authorship ability distribution, $\alpha_i \in \mathcal{U}(0, \overline{\alpha})$, is also considerably correlated with a simulation's category. 75% of simulations with writing in all stages have $\overline{\alpha} > 10.19$, while 75% of simulations without pre-piracy writing have $\overline{\alpha} < 8.50$. Simulations with writing that stops under piracy (whether it resumes under public writing or not) have a distribution of $\overline{\alpha}$ that is fairly broad, but somewhat more concentrated in the middle of the distribution than the unconditional, uniform distribution from which the parameter is drawn, suggesting that simulations with a higher upper bound to author ability are more likely to continue to have some creative activity under piracy, while those with low upper bounds on ability are less likely to have been able to sustain creativity even in the non-piratic model.

5.2 Analysis of simulations with writing

In order to perform a more sophisticated analysis of results, a data set consisting of the observations with writing in all stages was analyzed for each of the three policies. While simulations without writing under piracy are still interesting, it becomes impossible to measure some of the effect since values of interest such as "mean revenue per book" are simply not defined when no books are created. Though additional analysis could be done by using a truncated regression model with different model specifications, the current analysis simply calculates coefficients conditional on a simulation having produced writing activity throughout all simulation stages. Thus the results discussed in this section are conditional on having a set of simulation parameters that does not have a total abandonment of writing in any of the analyzed periods.

This simulations are analyzed both graphically—by plotting graphs showing the distribution across simulations of measured simulation variables—and statistically, by looking at how the variables of interest move across simulations. This section includes a few pertinent graphs of key simulation variables and summary tables of the simulation statistics. A more comprehensive set of graphs is presented in Appendix E.

The statistical analysis of simulations uses an average treatment effect model with two treatments, "piracy" and "policy," which are each divided into short-run and long-run effects. Short-run effects here means the average variable value over the first 25 periods of the piracy or policy stage; long-run means the average value over the last 25 periods of the relevant stage. Each simulation thus provides five observations: observed values from the end of the pre-piracy stage; values from the beginning of and the end of the piracy stage; and values from the beginning and end of the public sharing stage. Thus the model used for a single variable of interest is:

$$\begin{aligned} y = \beta_1 &+ \beta_2 \mathrm{I}(\mathit{piracy})\mathrm{I}(\mathit{short\text{-}run}) + \beta_3 \mathrm{I}(\mathit{piracy})\mathrm{I}(\mathit{long\text{-}run}) \\ &+ \beta_4 \mathrm{I}(\mathit{policy})\mathrm{I}(\mathit{short\text{-}run}) + \beta_5 \mathrm{I}(\mathit{policy})\mathrm{I}(\mathit{long\text{-}run}) + u, \quad u \sim \mathrm{iid}(0, \sigma^2) \end{aligned}$$

where y is the variable of interest (e.g. book quality), and $\text{I}(\cdot)$ are indicator functions equal to 1 when the value of the variable is from a piracy, policy, short-run (i.e. first 25) or long-run (last 25) analysis period, and 0 otherwise. The estimated coefficients of this model thus provide the mean of y under pre-piracy (β_1), with the remaining estimators capturing the difference of the variable relative to β_1, the long-run, pre-piracy value.

5.2.1 Average values and effects

The following variables of interest are estimated as described above, using a seemingly-unrelated regression model to estimate all coefficients simultaneously.

- *Mean net utility per reader*: The utility level compared to a model without any writing or reading activity, in which every reader would simply spend his entire income on the numéraire good. Note that this value is influenced positively only by utility received from reading: book creation costs, maintenance costs, and unit costs (and, under piracy, the cost of pirating) impact the value negatively; revenue earned by authors doesn't affect net utility at all, since such revenue is simply a transfer from one agent to another.
- *5th percentile, median, and 95th percentile of readers' mean net utilities*: This is included to enable consideration of how the distribution of net utility changes in addition to the mean. Note that the quantiles are calculated *after* averaging utility across analyzed simulation periods but before averaging across readers.
- *Mean books written per period per hundred agents*: The measures the rate of new creation within a simulation.
- *Mean book quality, q_b*: Provides a measure of book quality. The value is the average over all books written during the analyzed period.
- *5th percentile, median, and 95th percentile of book quality*: Included to allow analysis of the distribution of quality of created books.
- *Mean debut price*: The average initial price of just-finished books. Books that are not released at all or that are released directly to the public market (under the public sharing policies) are not included in this calculation.
- *Average book revenue*: Calculated as total book revenue over the analysis period divided by the total number of books created during the sample period. This includes public sharing tax revenue paid out to an author of a publically-downloaded book.
- *Book profit*: Book revenue (including public sharing revenue, if applicable) minus creation, maintenance, and marginal costs, averaged over books written during the analysis period.
- *Mean author creation scale*: That is the α_i in (4) of the author of books created during the analysis period. Note that this is averaged over books, not over authors.
- *5th percentile, median, and 95th percentile of author creation scale.*
- *Mean book author effort*: The average ℓ in (4) that authors chose to spend on a book created during the analysis period. Note that, like the creation scale, this value is averaged over books, not over authors.

- *5th percentile, median, and 95th percentile of books' author effort.*

The results are summarized in the Tables 2–4; more detailed regression results are found in Appendix F. The constant terms provide some insight into the basic pre-piracy model, but mainly serve as a reference point for analysis of the other parameters.

	Constant	Δ Pir.(SR)	Δ Piracy	Δ Pol.(SR)	Δ Policy
net utility	224.5	154.3	-84.14	-120.8	-119.8
net utility$_{5\%}$	143.6	144.0	-32.58	-76.80	-93.35
net utility$_{50\%}$	164.3	211.7	-32.25	-74.07	-87.26
net utility$_{95\%}$	541.3	-55.69	-334.2	-340.3	-262.7
books written	7.317	0.5120	-3.082	-0.1448	3.564
book quality	64.19	-23.81	-6.665	-17.68	-26.40
book quality$_{5\%}$	49.08	-36.51	-11.73	-27.45	-34.87
book quality$_{50\%}$	64.24	-31.48	-5.427	-19.19	-35.79
book quality$_{95\%}$	79.26	-2.817	-5.537	-5.775	-2.060
book $p_{t=0}$	26.61	10.05	18.24	9.394	1.499
book revenue	1986.	-1621.	-1112.	-1545.	-1299.
book profit	742.5	-948.7	-613.2	-612.4	-468.7
book author α	9.876	-1.694	0.1640	-0.9992	-1.954
book author $\alpha_{5\%}$	8.108	-4.352	-0.6070	-3.058	-4.438
book author $\alpha_{50\%}$	9.889	-1.363	0.4285	-0.6383	-1.738
book author $\alpha_{95\%}$	11.61	-0.04941	0.06245	0.01627	-0.09161
book author ℓ	695.1	-326.0	-265.5	-383.2	-286.8
book author $\ell_{5\%}$	458.1	-392.4	-237.8	-364.7	-375.2
book author $\ell_{50\%}$	686.5	-408.3	-265.8	-418.3	-405.1
book author $\ell_{95\%}$	955.1	-94.12	-284.5	-301.8	123.2

Table 2: Average effects—public sharing policy

5.2.2 Piracy effects

Piracy's effects, as shown in Table 2, are drastically different when comparing the short-run or long-run effects. In the short run, welfare increases considerably for almost everyone: only a few at the upper end of the utility scale—a few highly successful authors earning large profits—see a decrease in utility. Average book revenue sees a considerable drop, and average book profit becomes negative as, in the short run, authors are still learning about the new environment, and response by reducing book quality and increasing book prices. Neither is entirely successful: revenues decline greatly, and average book profit becomes negative.

In the long-run, piracy's effects look much worse: welfare is down across the agent distribution (and particularly so at the high-end) and the rate at which new books are written declines by nearly half. Book quality recovers somewhat from the short-run piracy level, but is still lower than its pre-piracy level. The initial price of new books on the market increases considerably, reflecting the decreased competition of having fewer books being created.[24] The higher price is not, however, met by an increase in revenue or profit: both are significantly lower: while books earn higher profits in the initial period, piracy effects a significant reduction in subsequent periods, once readers start making books available to other readers through piracy via the network.

The distribution of books' authors' ability shifts slightly under piracy, becoming slightly broader. The effort actually spent on creation, however, drops noticeably: piracy makes it no longer worthwhile to spend as much effort on creation because the effect of additional effort—a higher quality book—is not worth as much under piracy. Before piracy high quality books, which deliver larger consumer surplus, allow higher prices and profits. In the piratic environment, however, much of the surplus is retained by readers by avoiding the sales price through the use of piracy.

In summary, the piracy mechanism of this model leads to a considerable reduction in the profitability of writing, which in turn leads to a sharp reduction in the number of works created. The authors who still produce reduce their effort level and this reduces the quality of created books. The end result is a considerable reduction in social welfare: there are fewer books to read, and those books are of lower quality thanks to piracy eroding the incentive to create at higher quality levels.

5.2.3 Policy 1: public sharing

Introducing public sharing as an attempted fix to piracy by rewarding authors for downloads is, in fact, able to more than restore the number of books being written, but does so by inducing mostly low-quality works. It fails to restore utility, which is in fact worse than piracy for the most part; only the very highest agents in the utility distribution see a utility increase; these are predominantly the authors receiving payment via the public transfer mechanism.

The reason for the lack of recovery of utility, despite the huge boost in created works, becomes apparent when looking at the properties of books being created. Average book quality drops considerably, as do most of the quality quantiles: only the very high end of the quality distribution remains close to its pre-piracy level. Median book quality, however, drops to less than half of the pre-piracy level, while the relative drop at the bottom of the quality distribution is even larger. The public sharing mechanism encourages much more writing, but does so by introducing a mechanism that induces a large increase of predominantly low quality books.

[24]Though intuitively one might think that piracy should reduce the price, the initial price of the book doesn't actual face piracy, since no one yet owns the book. Analysis of individual simulation does reveal, however, that the *second*-period price drops considerably—if it exists at all. Often the piracy effect is so large that books only remain on the market for one period, while in pre-piracy they tend to remain on the market longer. The myopia assumption of reader optimization likely also plays a role here: readers do not forecast that a book will become available via piracy at some point in the future, which is why some profitable writing still exists.

Despite the drop in quality, however, average prices of *marketed* books return to about the same level they had before piracy was introduced. This is slightly misleading, however, as a great many books—particularly low quality ones—are simply released directly to the public market; only a few books have sufficient quality to be worth selling privately at all. Average book revenue drops precipitously, even more than under piracy, though average book profit ends up in between the pre-piracy and piracy levels.

The quality effect can also be seen in the author ability and effort levels put into creation by authors: there is a large increase in the number of low-quality, low-effort works created, as can be seen in Figures 6 and 7. This is coupled with large increase of low-ability authors participating in the market. Meanwhile, the upper end of the effort distribution shows an interesting uptick: public sharing induces a few, high-ability authors to keep producing high quality works, which is able to earn a profit amid all the low quality works flooding the market. High quality books still sell privately at reasonably high prices, then are quickly relegated to the public market, where the high quality (at low price) attracts more readers and thus rewards the author with an additional share of the public funding.

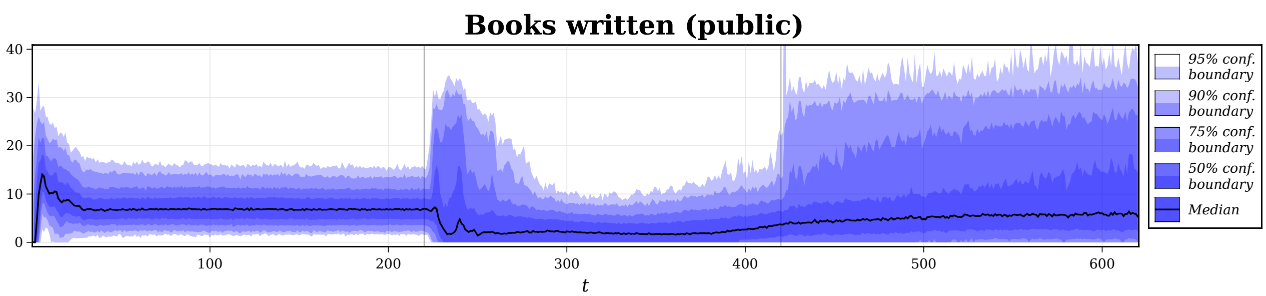


Figure 6: Simulation distribution—books written (per 100 agents) under public sharing

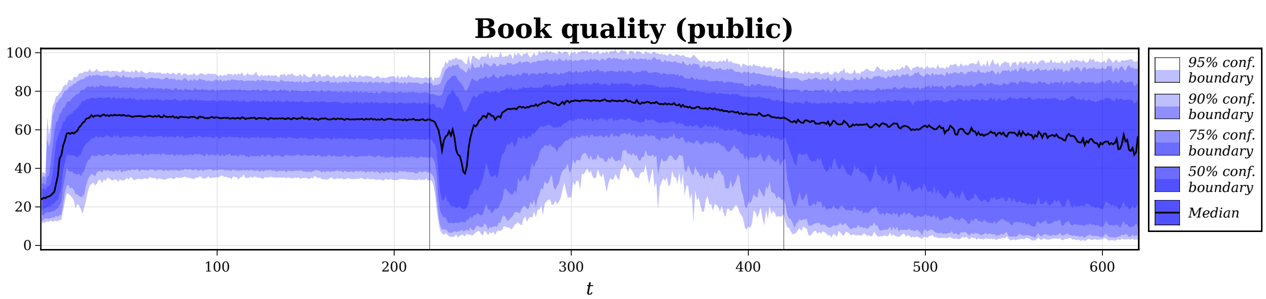


Figure 7: Simulation distribution—average book quality under public sharing policy

The reason the public sharing remedy tends to induce such an increase in works of low quality is that it significantly weakens the link between quality and prices. Authors' beliefs (and thus strategies) respond to the environment in such a way as to produce works that are good enough to attract public downloads—and thus a piece of the public payoff—but there is limited incentive to pursue higher quality because there is limited opportunity to earn a reward on that higher quality. In the pre-piracy world, higher quality meant an ability to charge higher prices, to reach most customers, or a combination of the two: any reader already willing to buy a copy of a book of quality q_b at price p would be willing to buy a copy of quality $q_b + 1$ at price $p + 1$, and there will typically be more readers willing to buy a book of quality $q_b + 1$ at price p. Under public sharing this link becomes weaker: profits are increasingly generated from the public provisioning mechanism, which only rewards downloads. Thus increasing the quality from q to $q + 1$ may reach more individuals, but doesn't earn any more from those individuals. The other problem is that the increase in the number of books written, and the rapid availability of those books through the public sharing system, means that even books kept on the private market face increased competition, further driving down profitability.

5.2.4 Policy 2: public sharing with voting

To address the failure of the public sharing mechanism, an alternative policy was used where consumers of copies of public works cast a limited number of votes for their favourite works in approximate proportion to the utility the reader received from reading the book. Table 3 shows the results of simulations with this policy response.

	Constant	Δ Pir.(SR)	Δ Piracy	Δ Pol.(SR)	Δ Policy
net utility	232.3	159.4	-88.82	-115.8	-66.49
net utility$_{5\%}$	149.7	150.8	-36.03	-68.32	-42.92
net utility$_{50\%}$	170.8	219.2	-36.02	-67.31	-42.03
net utility$_{95\%}$	548.6	-54.91	-335.9	-339.7	-176.1
books written	7.595	0.7403	-3.276	-2.526	-1.967
book quality	64.96	-25.58	-6.411	-7.691	1.464
book quality$_{5\%}$	49.49	-38.97	-11.25	-16.15	2.628
book quality$_{50\%}$	64.94	-34.47	-5.248	-4.944	2.003
book quality$_{95\%}$	80.48	-2.994	-5.496	-5.640	-1.592
book $p_{t=0}$	26.42	10.64	19.84	14.92	7.776
book revenue	2013.	-1672.	-1109.	-1493.	-864.3
book profit	758.3	-980.1	-610.3	-580.6	-217.4
book author α	9.972	-1.824	0.1887	0.06084	0.5729
book author $\alpha_{5\%}$	8.167	-4.685	-0.5893	-1.141	0.9324
book author $\alpha_{50\%}$	9.981	-1.458	0.4839	0.4524	0.6211
book author $\alpha_{95\%}$	11.76	-0.05685	0.05724	0.06034	0.06467
book author ℓ	700.3	-346.9	-253.7	-291.4	-53.79
book author $\ell_{5\%}$	455.4	-412.3	-220.0	-284.5	-56.16
book author $\ell_{50\%}$	690.2	-441.9	-255.4	-289.2	-52.30
book author $\ell_{95\%}$	970.8	-98.87	-280.5	-294.6	-56.24

Table 3: Average effects—public sharing with voting policy

The voting policy is able to do considerably better than the basic public sharing policy. Average welfare ends up improved from its level under piracy, with most of the increase at the upper-end of the distribution, i.e. among authors.

Unlike the basic version of the policy, it does this by much more directly encouraging quality; in fact, the average quality of books ends up slightly higher than the pre-piracy quality. Unfortunately, piracy still takes its toll: the number of books written is still reduced compared to pre-piracy. The distribution of authors here shows a definite shift towards higher-ability authors; the shift towards higher ability combined with the reduced number of works created results in moderate reduction in author effort—but because the authors who find writing profitable are, on average, of higher ability, book quality is restored to its pre-piracy level.

The contrast in policies is even more apparent in a comparison between Figures 8 and 9 and the equivalent Figures 6 and 7 for the basic public sharing policy.

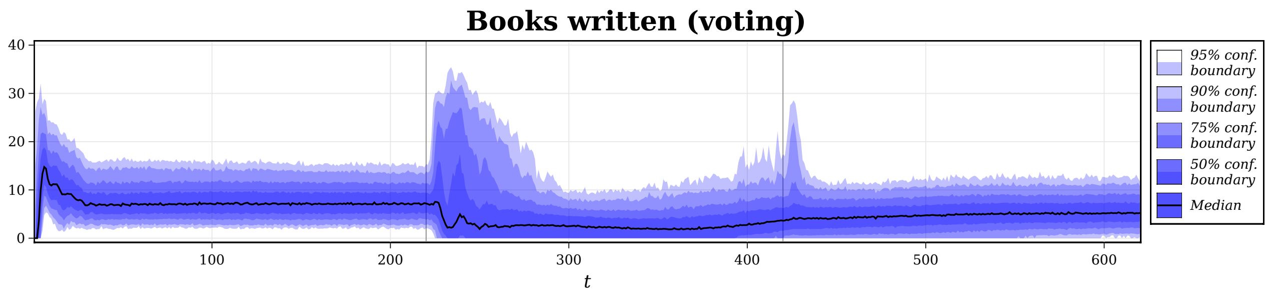


Figure 8: Simulation distribution—books written (per 100 agents) under public sharing

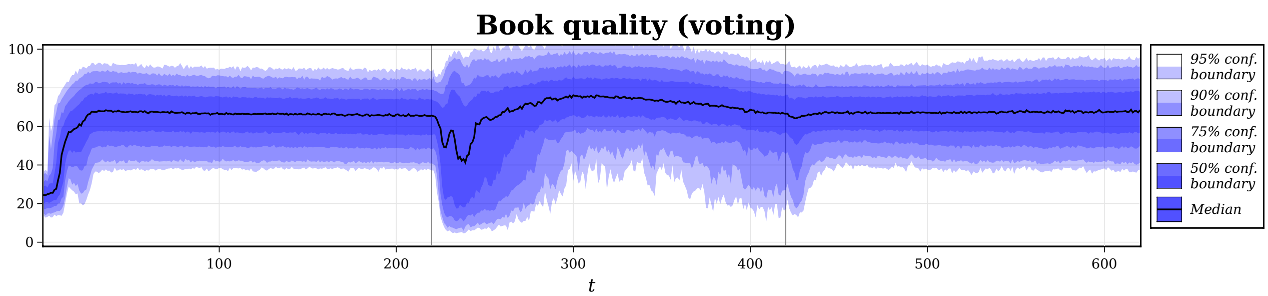


Figure 9: Simulation distribution—average book quality under public sharing with voting

5.2.5 Policy 3: detection and fines

The effects of the third policy explored—detection and fining of those engaged in piracy—is shown in Table 4.

This policy is able to restore average utility to approximately its pre-piracy level, despite the costs (on average 50 per reader per period) of the program. Looking at the books being written, there is a slight decrease in quality relative to pre-piracy levels, and a slight increase in the debut price of market books. Revenue and profits both decrease, but both decreases are relatively small compared to the effect under piracy.

	Constant	Δ Pir.(SR)	Δ Piracy	Δ Pol.(SR)	Δ Policy
net utility	238.8	155.4	-84.17	-98.34	2.199
$\textit{net utility}_{5\%}$	154.0	146.6	-30.03	-88.60	-19.89
$\textit{net utility}_{50\%}$	174.7	217.1	-29.23	-52.94	38.83
$\textit{net utility}_{95\%}$	571.7	-70.89	-343.7	-290.9	-133.4
books written	7.583	1.036	-3.110	0.6058	0.04669
book quality	65.84	-26.20	-4.931	-12.09	-1.401
$\textit{book quality}_{5\%}$	49.98	-39.22	-9.852	-25.88	-0.8676
$\textit{book quality}_{50\%}$	65.85	-35.06	-3.025	-8.729	-1.035
$\textit{book quality}_{95\%}$	81.80	-2.965	-5.100	-5.080	-3.499
book $p_{t=0}$	26.69	9.915	20.92	11.98	5.082
book revenue	2065.	-1701.	-1133.	-859.6	-484.1
book profit	794.0	-1006.	-630.2	-374.6	-225.5
book author α	10.09	-1.893	0.3365	-0.5377	0.2193
book author $\alpha_{5\%}$	8.220	-4.783	-0.3536	-2.590	0.2403
book author $\alpha_{50\%}$	10.10	-1.554	0.6374	-0.09424	0.2501
book author $\alpha_{95\%}$	11.93	-0.05281	0.08727	0.01461	0.03485
book author ℓ	712.2	-353.1	-257.4	-318.3	-132.6
book author $\ell_{5\%}$	462.3	-416.8	-222.5	-348.4	-79.01
book author $\ell_{50\%}$	702.9	-446.5	-250.4	-329.2	-128.3
book author $\ell_{95\%}$	987.9	-101.8	-305.2	-225.8	-200.2

Table 4: Average effects—catch-and-fine policy

Similarly to the voting mechanism, there is an apparently shift in the distribution of authors towards higher end authors, who put in less effort, but result in only a slight decrease to book quality.

The reason why this policy works is that it is able to reduce the amount of piracy that occurs in the model, without eliminating it completely, due to the imperfect detection mechanism. By essentially permitting low levels of piracy (by having a low probability of detection), this is able to balance some of the gains of piracy—providing accessing to some works below market prices—with the increased demand resulting from it becoming costly to pirate a great deal. This shift in piracy and purchasing patterns is best depicted in Figures 10 and 11, which show the distribution of average numbers of books bought and pirated across the three stages of the catch-and-fine policy simulations.

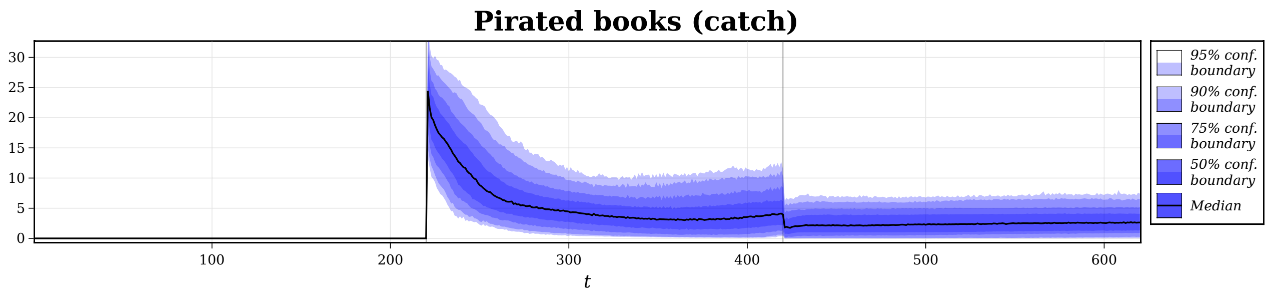


Figure 10: Simulation distribution—average number of pirated books per reader under catch-and-fine policy

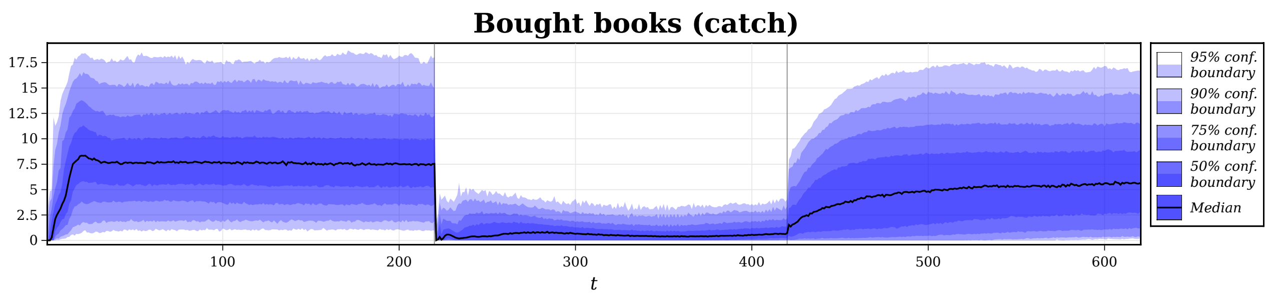


Figure 11: Simulation distribution—average number of book purchased per reader under catch-and-fine policy

This reduction allows authors to still earn substantial profits, and thus still write, while still allowing some of the gains in welfare resulting from piracy (most prominently displayed in the short run piracy effects). Thus some of the positive effects of piracy—making books available much longer and at lower costs—are allowed to persist, while the debilitating effects of piracy are reduced by keeping the level of piracy at a lower level.

6 Conclusion

The issue of piracy is one of immense importance to the long-term future of creative expression. Traditional approaches—based on copyright's government decree and enforcement—have largely been circumvented in a piratic, digital world. As this model captures, piracy significantly reduces profits, eroding the incentives of creators, leading to a loss in creative output and a reduction in the effort put into creative works. Some potential solutions to the piracy problem, as explored in this model, have the potential to partially ameliorate the problem, but introduce their own distortions that don't necessarily result in a strong social welfare improvement.

What seems clear is that responses to piracy—such as a movement in the real world towards subscription services, which are in many ways similar to the public sharing remedy explored in this model—is unlikely to fully recapture the incentives of creation. Ironically, the technological change that has lead to piracy may somewhat ameliorate the problem, at least in those industries—most notably the music industry—where technological change has significantly reduced creation costs. Nevertheless it is imperative that we continue to investigate potential solutions to piracy as the alternative—the significant decline to creative expression—implies a long-run diminishing of quality output.

The effects of piracy on creativity—which has traditionally been a relatively difficult problem to model in economics—is well captured by the model developed in this paper. More importantly, this paper is able to offer a useful benchmark against which we can compare the long-run effects of various remedies to the piracy problem. Of the three remedies explored here, one was, by most metrics, a failure to restore creative incentives, but a useful failure nonetheless: it failed because it induced more "junk" to be created by tying the incentive too closely to the quantity of creation without providing sufficient incentive for the quality of those creations. The second corrected this by associating public payoffs with reader enjoyment, which restored the quality incentive, but only partially restored welfare. The third policy, to catch and fine for piracy, was a considerable success: it largely restored the incentive to create while still providing some of the benefits—low cost access to a limited number of works—to exist. Although under this policy authors end up slightly worse off than before piracy, the social welfare gains of the limited piracy allowed to exist more than balanced out the losses to authors.

The model's novel use of Bayesian econometric models of beliefs seems to work well: it gives a natural way to incorporate new data, forget old data, and offers—through the prior and posterior distributional approach—a good fit to the evolution of beliefs over time. Moreover it opens up various possibilities—such as the linear restrictions used in this paper—that provide a rich set of tools for modelling agent beliefs and actions in agent-based models.

Finally, there is much work left to be done with respect to creativity and piracy. It is no longer sufficient to assume a working copyright system: piracy has effected significant changes both through its direct effects—lost sales due to piracy—and its indirect effects—reduced market power of creators. By ignoring the problem at the social planning level—i.e. by forcing creators to simply adapt—is not an effective solution: it will likely lead to a long-term loss to consumer welfare through its effected reduction in high-quality, high-value creative works.

6.1 Future work

There are a multitude of ways to extend the models presented in this paper. One of the unique advantages of an agent-based model is that it is easily open to extensions. The following are a few suggestions.

Belief models in this simulation are fixed. An interesting approach would be to specify several alternatives for each model, and use Bayesian model comparison techniques to have agents choose which model to use in each period.

The model could be enhanced by attempting multiple belief specifications so as to check the robustness of the results to the belief model specifications.

The simulation model structure here is relatively simple, and open to almost infinite extensions, which in many cases could serve to make the model more realistic. The following lists a few of the more interesting ideas that have come to mind or been suggested:

- Agent lifespans. When agents die, they would be replaced by new agents with non-informative beliefs; this would allow the model to keep moving, even when beliefs of (living) agents don't support creating. It would also open up a reason for books to continue to exist on the market by opening up a small but constant stream of new agents with empty libraries.
- Learning about ability. Currently authors know exactly how they convert effort into quality, but this seems unrealistic: in reality, creators learn their abilities through creative attempts.
- Practice makes perfect. Introducing improvements in an author's quality/effort function as the author creates would match the idea that authors (or musicians, or …) in the world improve over time.
- In contrast, authors might be considered to have a limited lifetime capacity, with a reduction in skill as they create, reflecting that authors might run out of new ideas as they create more.
- Quality information via the network. Currently quality is observed directly, but it would be interesting to explore a more opaque version of quality where quality information cannot be directly observed but can be transmitted through the network. This signal could additionally be noisy: readers could, for example, transmit their utility gain (which includes a distance penalty) instead of just the quality.

Finally, it would be interesting to look at the marginal effects of various simulation parameters on various model parameters of interest. This would map out the relationship between simulation parameters and model results in greater detail, thus allowing more precise determination of how each parameter affects each model outcome variable.

References

Arthur, W. Brian (2006). "Out-of-equilibrium economics and agent-based modeling". In: *Handbook of computational economics* 2, pp. 1551–1564.

Bae, Sang Hoo and Jay Pil Choi (2006). "A model of piracy". In: *Information Economics and Policy* 18.3, pp. 303–320.

Bakos, Yannis, Erik Brynjolfsson, and Douglas Lichtman (1999). "Shared Information Goods". In: *The Journal of Law and Economics* 42.1, pp. 117–156.

Barker, George Robert and Tim John Maloney (2012). "The impact of free music downloads on the purchase of music CDs in Canada". In: *ANU College of Law Research Paper* 4.

Besen, Stanley M. and Sheila Nataraj Kirby (1989). "Private Copying, Appropriability, and Optimal Copying Royalties". In: *The Journal of Law & Economics* 32.2, pp. 255–280.

Brenner, Thomas (2006). "Agent learning representation: advice on modelling economic learning". In: *Handbook of computational economics* 2, pp. 895–947.

Bullard, James and John Duffy (1994). *Learning in a large square economy.* Federal Reserve Bank of St. Louis.

Connolly, Marie and Alan B. Krueger (2006). "Rockonomics: The economics of popular music". In: *Handbook of the Economics of Art and Culture* 1, pp. 667–719.

Conway, John (1970). "The Game of Life". In: *Scientific American* 223.4, p. 4.

Geweke, John (1991). "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities". In: *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Seattle*, pp. 571–578.

Handke, Christian W. (2006). "Plain destruction or creative destruction? Copyright erosion and the evolution of the record industry". In: *Review of Economic Research on Copyright Issues* 3.2, pp. 29–51.

— (2012). "Digital copying and the supply of sound recordings". In: *Information Economics and Policy* 24.1, pp. 15–29.

Harbaugh, Rick and Rahul Khemka (2010). "Does copyright enforcement encourage piracy?" In: *The Journal of Industrial Economics* 58.2, pp. 306–323.

Keen, Steve and Russell Standish (2010). "Debunking the theory of the firm—a chronology". In: *real-world economics review* 53, pp. 56–94.

Kollman, Ken, John H. Miller, and Scott E. Page (1992). "Adaptive Parties in Spatial Elections". In: *The American Political Science Review* 86.4, pp. 929–937.

Koop, Gary (2003). *Bayesian Econometrics.* John Wiley & Sons Ltd.

Landes, William M. and Richard A. Posner (1989). "An economic analysis of copyright law". In: *The Journal of Legal Studies*, pp. 325–363.

Marcet, Albert and Thomas J. Sargent (1989). "Convergence of least squares learning mechanisms in self-referential linear stochastic models". In: *Journal of Economic theory* 48.2, pp. 337–368.

Miller, John H. and Scott E. Page (2007). *Complex Adaptive Systems*. Princeton University Press.

Oberholzer-Gee, Felix and Koleman Strumpf (2010). "File sharing and copyright". In: *Innovation Policy and the Economy, Volume 10*. University of Chicago Press, pp. 19–55.

Plümper, T. and C.W. Martin (2008). "Multi-party competition: A computational model with abstention and memory". In: *Electoral Studies* 27.3, pp. 424–441.

Rhinelander, Jason G. (2016a). *Creativity - Agent-based model of creativity and piracy*. URL: `https://git.imaginary.ca/eris/creativity`.

— (2016b). *Eris - agent-based economic modelling library*. URL: `https://git.imaginary.ca/eris/eris`.

Rodriguez-Yam, Gabriel, Richard A. Davis, and Louis L. Scharf (2004). "Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression". In: *Unpublished PhD thesis*.

Salop, Steven C. (1979). "Monopolistic competition with outside goods". In: *The Bell Journal of Economics*, pp. 141–156.

RIAA (2015). *Scope of the Problem*. URL: `http://www.riaa.com/physicalpiracy.php?content_selector=piracy-online-scope-of-the-problem` (visited on June 12, 2015).

Selten, Reinhard (1991). "Evolution, learning, and economic behavior". In: *Games and Economic Behavior* 3.1, pp. 3–24.

Standish, Russell K. and Steve Keen (2015). "Rationality in the Theory of the Firm". In: *World Economic Review* 2015.5, pp. 101–106.

Tesfatsion, Leigh (2006). "Agent-based computational economics: A constructive approach to economic theory". In: *Handbook of computational economics* 2, pp. 831–880.

Tversky, Amos and Daniel Kahneman (1973). "Availability: A heuristic for judging frequency and probability". In: *Cognitive Psychology* 5.2, pp. 207–232.

Varian, Hal R. (2005). "Copying and copyright". In: *Journal of Economic Perspectives*, pp. 121–138.

Waldfogel, Joel (2012). "Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster". In: *Journal of Law and Economics* 55.4, pp. 715–740.

Appendices

A Eris: agent-based economic modelling library

This project is based on the open source Eris library, an object-oriented software library written in the C++ programming language[25] by the author of this paper to serve as a framework for agent-based modelling in general, with a focus on useful tools for agent-based computational economics in particular.

Eris itself is designed for problems that can be divided into sequential iterations supporting multiple types of interaction between stages. The simulation itself consists of programming "objects" where each object represents an agent, a market, or a good, or simply an object without any model that still contains necessary computational tasks.

An iteration of the simulation is divided into several inter-period stages and intra-period stages. Any simulation object can register itself as an "optimizer" for one or more stages by implementing the relevant optimization class, which essentially just lets the main simulation loop know what methods to call.

Each iteration of the simulation consists of, in order:

- The simulation time period, t, is incremented.
- Inter-period optimization phase:
  - "Begin" phase
  - "Optimize" phase
  - "Apply" phase
  - "Advance" phase
- Intra-period optimization phase:
  - "Initialize" phase
  - "Reset" phase
  - "Optimize" phase
  - "Reoptimize" phase, which lets any member restart the entire optimization process for all members at the "Reset" phase.
  - "Apply" phase, invoked if no reoptimizing agent resets the optimization process during the "Reoptimize" phase.
  - "Finish" phase.

Thus inter-period optimization is where actions are decided for the upcoming period. In this paper's model, this is the stage of the model where beliefs are updated; where authors receive the proceeds of book sales from the previous period; where authors decided whether or not to keep their authored books on the market, and if so at what price; and where agents decided whether or not to undertake the authoring of a new book.

Intra-period optimization is designed to capture actions that are taken within a period. In this paper's model, this consists of readers deciding which books to purchase

[25] Specifically the C++11 specification of the language.

or pirate.

The notional intention of the separation into individual phases is to allow agents to decide their actions in the "Optimize" phases, but not actually implement those actions until the "Apply" phase. This is explicitly meant to allow models where agents take simultaneous actions using the current state of the world (in the "Optimize" phases) before the state of the world actually changes as a result of the decided upon actions (in the "Apply" phase). Thus all readers see, for example, the same number of sales for a given book; there is no difference resulting from the purely computational reason that reader i happened to have its optimization code before reader j. This essentially allows models to implement simultaneous decision processes among agents, even though the underlying computational implementation is not simultaneous.

The "inter-Begin," "inter-Advance," "intra-Initialize," and "intra-Finish" phases are simply to allow extra timing steps in a simulation. In this paper's simulation, they are used only for computational cleanup: updating cached variables (for example, the list of on-market books) used to increase the simulation performance.

The "Reoptimize" phase exists to allow simulations the ability to have a sort of auctioning process, where the intra-period optimization needs to be recursive: for example, allowing a virtual Walrasian auctioneer that sets a market-clearing price by manipulating a price until the quantity available just clears. This phase is not used in this paper.

Eris provides various helper classes to facilitate agent-base modelling such as a Bundle class to store bundles of goods; basic agents which possess a bundle of assets; Market classes that simplify market transactions; code to handle positions, including the wrapped characteristic space used in this paper's creativity model; and the basic Bayesian linear implementation used in this paper. As a library, it is not designed to provide an agent-based model itself, but rather to provide a framework that takes care of the underlying pieces and structure needed to build a sophisticated agent-based model.

The Eris library is free software, available under the GPL license (v3 or later), from `https://git.imaginary.ca/eris/eris`. Documentation of its API can be generated from the source code itself; a pre-built version of the documentation is maintained at `https://imaginary.ca/eris/api/annotated.html`. Eris consists of over 13,000 lines of code, including complete API documentation.

The creativity model developed in this paper is likewise free software, also available under the GPLv3+, from `https://git.imaginary.ca/eris/creativity`. Its API documentation is likewise available from the source code itself, and at `https://imaginary.ca/eris/creativity/annotated.html`. The creativity code consists of over 18,000 lines of code (including documentation).

The Bayesian belief implementation is contained within the eris project's `eris/learning` directory; the BayesianLinear and BayesianLinearRestricted classes handle the basic algorithms. Corresponding subclasses in the creativity project's `creativity/belief` provide the specific functional forms for the beliefs used in the creativity application.

Instructions for compiling the code are in each project's README.md file, also available directly from the project pages linked above.

Both projects are © Jason Rhinelander 2016, but are freely available for use and modification under the terms of the GNU General Public License version 3 (or later).

B Running the creativity model

There are two main methods of invoking a creativity simulation: through a graphical interface (`creativity-gui`, built using the GTK+ library), which provides a graphical representation of a single simulation; and through a command-line interface (`creativity-cli`). Simulation results can either be stored in memory (only for the GUI interface) or written to disk (for both interfaces), and this stored file can later be replayed in the GUI to observe the simulation in detail.

The project additionally includes several utility programs to aid in running and analyzing simulations:

- `creativity-info` produces summary information for a stored simulation.
- `creativity-data` generates data rows for analysis from one or more stored simulation files.
- `creativity-results` performs the statistic analysis described in Section 5.
- `creativity-random` that replaces specially-formatted arguments with random draws, then invokes the simulation command-line interface.
- `creativity-series` generates time-series outputs over the simulation periods from a set of simulation files.
- `creativity-series-quantiles` converts time-series output files into quantiles.
- `creativity-series-graphs` converts time-series output files (or quantile files) into time-series quantile plots.
- `run-sim.sh` which invokes creativity-random to run a single simulation with the arguments drawn from the distributions described in Section 4.2.

Each command (including `creativity-gui` and `creativity-cli`) accepts a `--help` argument that details the usage of the command.

C Bayesian belief computational details

Given a set of n new observations, $\boldsymbol{y}$ and $\boldsymbol{X}$, the following Bayesian econometric equations govern the combination of prior and data to calculate the posterior distribution parameters. $\underline{x}$ and $\overline{x}$ represent prior and posterior variables, respectively:

$$\overline{V}^{-1} = \underline{V}^{-1} + X^\top X \tag{18}$$

$$\overline{V}^{-1}\overline{\boldsymbol{\beta}} = \underline{V}^{-1}\underline{\boldsymbol{\beta}} + X^\top y \tag{19}$$

$$\overline{\nu} = \underline{\nu} + n \tag{20}$$

$$\overline{\nu s}^2 = \left(y - X\overline{\boldsymbol{\beta}}\right)^\top \left(y - X\overline{\boldsymbol{\beta}}\right) + \underline{\nu s}^2 + \left(\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}\right)^\top \underline{V}^{-1}\left(\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}\right) \tag{21}$$

Thus the posterior can be seen as being determined by a weighted sum of the information in the prior and the information in the new data. Note that these formulae differ in presentation from that typically used for Bayesian econometric (for example, in Koop (2003)), but are numerically identical and a more useful representation when Bayesian econometric models are being employed as a continually updated model of beliefs.

Equation (18) is more commonly represented as

$$\overline{V} = (\underline{V}^{-1} + X^\top X)^{-1}$$

The former is preferred in the context of this model for two reasons. The first is computational: computing and storing a matrix inverse is rarely desirable as it usually involves some loss of numerical precision; if the belief made use of $\overline{V}$ rather than $\overline{V}^{-1}$, the subsequent period would thus require an inverse of an already-inverted matrix, and so the imprecision introduced by a matrix inversion would be unnecessarily compounded over time: it is computationally simpler to thus simply store the $\overline{V}^{-1}$ value. Moreover, the $\overline{\boldsymbol{\beta}}$ solution to (19) can be solved more precisely and efficiently using computational linear equation solution methods that do not rely on the explicit computation of an inverse. Despite the fact that it is thus actually the $\overline{V}^{-1}$ matrix used, not $\overline{V}$, the notation is expressed as an inverse to be consistent with other Bayesian econometric literature.

The second preference for the representation as an inverse is due to the belief initialization problem discussed at the end of Section 3.2. Because the belief begins with a non-informative prior, the right-hand side of expression (18) may not be invertible at all: that is, $\overline{V}$ may not exist. Indeed it is the invertibility of this expression that defines when enough data has been accumulated to make a belief usable.

In empirical use of Bayesian econometric models, these issues do not arise: the combination of prior and data (and very often each on its own) is usually enough to make the existence of the inverse a non-issue, and the use of a single inverse, rather than an recursive set of inversions over repeated updating, is not a concern.

The calculation of $\overline{\boldsymbol{\beta}}$ in (19) is more typically expressed as:

$$\overline{\boldsymbol{\beta}} = \overline{V}(\underline{V}^{-1}\underline{\boldsymbol{\beta}} + X^\top X\widehat{\boldsymbol{\beta}})$$

where $\widehat{\boldsymbol{\beta}}$ is the usual OLS vector $\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top y$. The first modification is to use $\overline{V}^{-1}$ rather than its inverse, as described above. The second modification is required because the latter equation is not feasible for the updating needed in this model: it is simply not calculable when $X^\top X$ is not full rank. Note, however, that since $X^\top X\widehat{\boldsymbol{\beta}} = X^\top X(X^\top X)^{-1}X^\top y = X^\top y$, equation (19) is equivalent whenever $X^\top X$ *is* full-rank; moreover it is exactly the correct value to use even when this is not the case.

Equation (21), repeated here, differs substantially from Koop's version of the formula:

$$\overline{\nu s}^2 = \left(y - X\overline{\boldsymbol{\beta}}\right)^\top \left(y - X\overline{\boldsymbol{\beta}}\right) + \underline{\nu s}^2 + \left(\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}\right)^\top \underline{V}^{-1}\left(\overline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}\right) \tag{21}$$

$$\overline{\nu s}^2 = \underline{\nu s}^2 + ns^2 + (\widehat{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})^\top \left[\underline{V} + (X^\top X)^{-1}\right]^{-1} (\widehat{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}) \tag{Koop}$$

where s^2 and $\widehat{\boldsymbol{\beta}}$ are the usual OLS quantities. As above, this formula requires both an invertible $X^\top X$ matrix, and relies upon multiple matrix inversions. Under these conditions it can be shown that these two expressions are equivalent by noting that both equations are simply the sum of squared residuals of both the prior and new data. The three terms in (21) are the new data SSR at $\overline{\boldsymbol{\beta}}$, the prior SSR at $\underline{\boldsymbol{\beta}}$, and the change in the prior's SSR resulting from the change from $\underline{\boldsymbol{\beta}}$ to $\overline{\boldsymbol{\beta}}$.

Koop's version of the formula calculates the same quantity, but expressed as the prior's SSR, the new data's SSR at what β would equal without the prior, plus a more complicated term that captures the change in both SSRs resulting from the shift of the prior from $\underline{\boldsymbol{\beta}}$ to $\overline{\boldsymbol{\beta}}$ and the shift in the intermediate OLS estimate from $\widehat{\boldsymbol{\beta}}$ to $\overline{\boldsymbol{\beta}}$. The two approaches are numerically equivalent.

Under weakening with factor $w > 1$, the prior for $t+1$ is calculated as $\underline{V}_{t+1} = w^2 \overline{V}_t$; the other distribution hyperparameters are used without adjustment $((\underline{\boldsymbol{\beta}}, \underline{\nu}, \underline{s}^2)_{t+1} = (\overline{\boldsymbol{\beta}}, \overline{\nu}, \overline{s}^2)_t)$. Effectively this means that the first value in the matrix sum in (18) is scaled down by the value $w^{-2} < 1$, which gives relatively more weight to the new information $X^\top X$. This in turn effects shifts in the other parameters through equations (19) and (21), similarly re-weighting results in favour of the new information. [26]

As for (21) under weakening, the value is no longer the change in the SSR of the prior data, but now the change in the SSR of the prior data with the weakening factor w^{-2} applied to the data. In effect, this is the change in the sum of squared residuals of the prior data, with the size of this change weakened by the amount of weakening that has been applied. This is desirable: large changes in $\overline{\boldsymbol{\beta}}$ allowed by weakening should not imply large increases in the estimation of the error term because these represents shifts in the environment, not larger errors in the environment.

C.1 Drawing from Bayesian econometric models

Given T rows of data, $\boldsymbol{X}^*$, an agent can make predictions about the associated T values of a function the dependent variable of interest in the model environment, $\boldsymbol{y}^*$, by taking draws from the distribution, according to the relationship:

$$\boldsymbol{y}^{*(i)} = \boldsymbol{X}^* \boldsymbol{\beta}^{(i)} + \boldsymbol{\varepsilon}^{*(i)} \tag{22}$$

where $\boldsymbol{\beta}^{(i)}$ and $\varepsilon^{(i)}$ represent the ith draw from the underlying distribution of these parameters:

$$\begin{aligned} \boldsymbol{\beta}^{(i)} | h &\sim \mathcal{N}(\overline{\boldsymbol{\beta}}, h^{-1}\overline{V}) \\ \varepsilon^{(i)} | h &\sim \mathcal{N}(\mathbf{0}_T, h^{-1} I_T) \\ h &\sim \Gamma(\overline{\nu}/2, 2/(\overline{s}^2 \overline{\nu})) \end{aligned} \tag{23}$$

[26] This can be seem immediately from recursive expansion of $\underline{V}_{t+1}$ using (18):

$$\underline{V}^{-1}{}_{t+1} = X_{(t)}{}^\top X_{(t)} + w^2 X_{(t-1)}{}^\top X_{(t-1)} + w^4 X_{(t-2)}{}^\top X_{(t-2)} + w^6 X_{(t-3)}{}^\top X_{(t-3)} + \dots$$

where the hyper-parameters $\overline{\boldsymbol{\beta}}$, $\overline{V}$ ($k \times k$, symmetric, positive definite), $\overline{\nu}$, and $\overline{s}^2$ are calculated as detailed from the prior and data as described above.

Each draw i from this joint distribution yields a T-vector of predictions, $\boldsymbol{y}^{*(i)}$, with S draws performed in total.[27] Letting $f(y)$ denote a function of interest to the belief-holding agent (e.g. utility or profit), this yields predictions of:

$$\mathrm{E}(f(y_t)|X_t) \approx \sum_{i=1}^{S} f(y_t^{*(i)}), \quad t = 1, \ldots, T$$

In the simple case of $f(y) = y$, this simply yields the mean predicted y_t for each X_t. If $f(y)$ is a non-linear function—for example, the utility function of a risk-averse agent—this procedure obtains a prediction of expected utility that incorporates the likelihood of extreme values and the correlation between parameters as captured in the belief distribution over the parameters. It also, by including a draw from the error distribution, incorporates the uncertainty due to unobserved factors.

To draw from a model with linear restrictions, such as the demand model in section 3.4, the drawing procedure becomes more complicated: if the restriction doesn't bind too often, one can simply use the above procedure, discarding draws that don't satisfy the linear constraints. When restrictions impose a very small (probabilistically) region for β, another procedure such as Gibbs sampling is needed; Geweke (1991) provides such an implementation; Rodriguez-Yam, Davis, and Scharf (2004) proposes another for a multivariate normal with considerably better mixing properties; it is this latter technique, with an adaptation to draw from the required multivariate-t distribution, that this paper uses.

[27] S here refers to the number of draws actually used for inference. Note, however, that one common technique for obtaining such draws involves Gibbs sampling, where each draw is conditional on the previous value of the draw. Gibbs samplers often employ burn-in—throwing away of a set of initial draws—and thinning—performing and discarding intermediate draws. Under either of these approaches, the total number of draws actually performed computationally will exceed S, possibly by a significant amount.

D Simulation categories parameter distributions

D.1 Public sharing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	148.9	29.54	100	104	122	147	174	196	200
density	1.063	0.8236	0.25	0.2848	0.4482	0.7676	1.391	2.806	3.987
reader__step__mean	0.4808	0.2353	0.1003	0.1333	0.2759	0.4613	0.6762	0.8669	0.8987
ability__max	7.345	2.059	5	5.182	5.728	6.681	8.458	11.65	14.82
creation__fixed	157.8	57.56	50.61	61.12	109.9	163.4	204.8	241.7	249.8
creation__time	2.431	1.717	0	0	1	2	4	5	5
cost__market	28.06	13.63	0.2175	4.151	17.28	29.67	39.78	47.72	49.99
cost__unit	5.497	2.919	0.05289	0.6001	3.073	5.764	8.006	9.636	9.964

Table D1: Parameter values for 626 simulations without pre-piracy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	157.1	30.37	101	106.2	128	166	184	196.8	200
density	2.167	1.044	0.2816	0.4817	1.331	2.026	3.177	3.729	3.969
reader__step__mean	0.4959	0.2332	0.1012	0.1673	0.2815	0.5313	0.6937	0.8658	0.8914
ability__max	9.425	2.797	5.099	5.64	6.841	9.397	11.44	13.87	14.75
creation__fixed	154.4	59.05	52.61	62.79	103.2	160.8	201.4	239.2	246.5
creation__time	3.212	1.103	0	2	2	3	4	5	5
cost__market	22.25	13.7	1.098	1.935	11.5	20.74	32.92	45.59	49.42
cost__unit	4.226	2.659	0.06129	0.4456	2.493	3.864	5.814	9.192	9.971
cost__piracy	4.625	3.184	0.01068	0.2074	1.54	4.432	7.546	9.337	9.925
piracy__links	0.1452	0.05256	0.05069	0.06148	0.1068	0.1376	0.1902	0.2356	0.2447
policy__tax	45.69	26.98	2.129	6.292	24.07	42.74	65.13	90.89	99.93

Table D2: Parameter values for 85 simulations without piracy or policy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	152.6	29	100	106	127	155	179	195.7	200
density	2.174	0.9871	0.2548	0.6341	1.351	2.146	2.958	3.777	3.985
reader__step__mean	0.5178	0.2349	0.1002	0.1437	0.3163	0.5243	0.7289	0.8703	0.8999
ability__max	10	2.69	5.014	5.753	7.879	9.85	12.26	14.48	15
creation__fixed	149.4	56.87	50	61.62	98.97	147.5	197.3	238.8	250
creation__time	2.212	1.741	0	0	1	2	4	5	5
cost__market	23.38	14.89	0.02473	1.827	10.09	22.3	36.41	47.28	49.98
cost__unit	5.134	2.852	0.0202	0.6067	2.758	5.185	7.6	9.504	9.996
cost__piracy	4.942	2.905	0.003088	0.4723	2.377	4.952	7.414	9.446	9.992
piracy__links	0.1521	0.05737	0.05	0.06033	0.1045	0.153	0.2025	0.2396	0.25
policy__tax	50.76	28.88	1.015	5.946	25.44	50.46	75.57	95.6	99.98

Table D3: Parameter values for 1747 simulations with no piracy writing, but recovery under the policy

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	154.4	30.94	110	111.4	125.5	163	178	189.6	190
density	2.824	0.8374	1.675	1.753	2.103	2.805	3.523	3.916	3.985
reader_step_mean	0.4302	0.2827	0.1035	0.1039	0.1468	0.4873	0.6335	0.7798	0.8058
ability_max	9.966	2.842	5.9	6.262	7.181	10.91	12.02	13.33	13.48
creation_fixed	154.7	68.7	52.56	69.11	98.55	145.4	223.8	231.6	233.4
creation_time	3.1	1.197	1	1.45	2.25	3	4	4.55	5
cost_market	15.25	9.048	5.617	5.667	9.155	11.55	22.41	29.02	30.16
cost_unit	3.596	2.172	0.1671	0.3358	2.231	3.904	5.39	6.03	6.096
cost_piracy	4.543	3.009	0.38	0.7201	2.085	4.735	7.138	8.38	8.937
piracy_links	0.1756	0.06344	0.06233	0.06846	0.1574	0.1882	0.2235	0.2418	0.2441
policy_tax	55.33	29.23	6.621	7.758	39.01	66.72	72.99	84.79	92.11

Table D4: Parameter values for 10 simulations with piracy writing but not policy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	144.1	29.24	100	104	119	140	169	194	200
density	2.727	0.9049	0.2905	0.9925	2.079	2.897	3.46	3.887	3.991
reader_step_mean	0.4705	0.2324	0.1002	0.1248	0.2725	0.4554	0.6723	0.8419	0.9
ability_max	11.76	2.369	5.077	7.016	10.2	12.12	13.71	14.78	15
creation_fixed	150.4	57.85	50.04	61.8	100.2	148.7	201.1	241.3	249.8
creation_time	3.026	1.59	0	0	2	3	4	5	5
cost_market	25.6	14.07	0.04068	2.684	13.6	26.53	38.33	46.99	49.71
cost_unit	4.688	2.756	0.0242	0.4852	2.327	4.617	6.983	9.197	9.951
cost_piracy	5.203	2.938	0.01504	0.5373	2.666	5.253	7.75	9.578	9.972
piracy_links	0.1441	0.05773	0.05063	0.05728	0.09699	0.1387	0.1935	0.2393	0.2496
policy_tax	50.97	28.08	1.049	5.722	28.64	50.28	73.5	95.62	99.96

Table D5: Parameter values for 931 simulations with writing in all stages

D.2 Public sharing with voting

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	149.5	29.08	100	105	124	149	174	195	200
density	1.12	0.8397	0.2504	0.2948	0.4807	0.8536	1.465	3.083	3.986
reader_step_mean	0.4887	0.2251	0.1003	0.144	0.2971	0.4865	0.6816	0.851	0.8975
ability_max	7.31	2.134	5.001	5.119	5.678	6.619	8.317	11.78	14.82
creation_fixed	157.5	57.36	50.05	61.71	112	159.7	207.8	243.2	249.9
creation_time	2.293	1.739	0	0	1	2	4	5	5
cost_market	28.08	13.78	0.761	4.991	17.06	30.1	39.6	47.85	49.81
cost_unit	5.7	2.699	0.0262	1.01	3.664	6.03	7.99	9.621	9.997

Table D6: Parameter values for 621 simulations without pre-piracy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	151.5	29.06	100	104	127	154	176	196	200
density	2.216	1.029	0.2646	0.508	1.407	2.209	3.108	3.78	3.981
reader_step_mean	0.5261	0.2223	0.1059	0.1658	0.334	0.538	0.7168	0.8616	0.8958
ability_max	8.128	2.232	5.012	5.405	6.497	7.597	9.345	12.75	14.89
creation_fixed	156.7	58.14	50.44	60.4	109.3	160.8	206.5	244	249.6
creation_time	2.26	1.667	0	0	1	2	4	5	5
cost_market	24.17	14.61	0.06161	1.757	11.29	24	37.08	47.21	49.9
cost_unit	5.292	2.868	0.02981	0.6045	2.918	5.422	7.824	9.493	9.976
cost_piracy	4.698	2.853	0.01519	0.3651	2.179	4.692	7.068	9.292	9.983
piracy_links	0.1637	0.05695	0.05029	0.06287	0.1181	0.1706	0.2136	0.2426	0.25
policy_tax	47.81	27.81	1.101	7.117	22.83	46.63	70.43	93.65	99.73
policy_votes	2.997	1.45	1	1	2	3	4	5	5

Table D7: Parameter values for 581 simulations without piracy or policy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	152.9	28.2	100	107	129	155	176	196.2	200
density	2.148	0.9604	0.3067	0.7038	1.351	2.092	2.912	3.735	3.999
reader_step_mean	0.511	0.2298	0.1007	0.14	0.3172	0.5138	0.7147	0.8618	0.8996
ability_max	10.6	2.408	5.042	6.566	8.764	10.63	12.54	14.4	15
creation_fixed	146.9	58.78	50.02	57.6	94.11	145.6	197.4	240.1	249.8
creation_time	2.237	1.714	0	0	1	2	4	5	5
cost_market	23.8	14.58	0.02713	2.117	10.86	23.69	36.55	47	49.91
cost_unit	4.807	2.905	0.0179	0.445	2.228	4.654	7.39	9.371	9.996
cost_piracy	5.075	2.825	0.01963	0.5219	2.779	5.187	7.43	9.48	9.991
piracy_links	0.1498	0.05807	0.05017	0.05961	0.09804	0.1514	0.1992	0.2405	0.2497
policy_tax	50.25	28.63	1.087	6.812	24.15	49.95	74.85	95.79	99.94
policy_votes	2.991	1.441	1	1	2	3	4	5	5

Table D8: Parameter values for 1216 simulations with no piracy writing, but recovery under the policy

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	150.7	31.73	100	106	121	151	183	196	198
density	2.573	0.997	0.4452	0.7992	1.881	2.7	3.407	3.865	3.996
reader_step_mean	0.4764	0.2345	0.1055	0.1186	0.2634	0.4897	0.6481	0.818	0.8902
ability_max	8.813	2.832	5.069	5.335	6.478	8.433	10.78	14.37	14.91
creation_fixed	158.9	57.48	50.38	68.56	105.3	165.7	207.8	237.2	246.2
creation_time	3.049	1.596	0	0	2	3	4	5	5
cost_market	25	13.89	0.1254	2.507	16.01	26.07	36.14	48.86	49.67
cost_unit	4.401	3.105	0.06049	0.5059	2.032	3.535	7.216	9.454	9.896
cost_piracy	4.54	2.477	0.5301	1.415	2.524	4.086	6.832	9.506	9.995
piracy_links	0.1626	0.05662	0.06154	0.07666	0.1084	0.1787	0.2112	0.2392	0.2485
policy_tax	51.57	29.92	1.235	9.5	24.49	50.86	79.65	93.48	97.79
policy_votes	3.148	1.412	1	1	2	3	4	5	5

Table D9: Parameter values for 61 simulations with piracy writing but not policy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	143.2	29.57	100	103	116	139	167.8	191.1	200
density	2.752	0.8726	0.2622	1.106	2.163	2.879	3.483	3.916	3.997
reader_step_mean	0.4729	0.2354	0.1004	0.1342	0.2579	0.4514	0.6789	0.8503	0.8987
ability_max	11.93	2.164	5.247	7.751	10.56	12.26	13.72	14.8	15
creation_fixed	146.5	58.26	50.23	58.99	96.77	143.1	198	239.8	249.4
creation_time	3.134	1.502	0	0	2	3	4	5	5
cost_market	24.97	13.93	0.06388	2.527	13.51	25.13	36.28	47.28	49.99
cost_unit	4.703	2.864	0.003403	0.3693	2.272	4.445	7.098	9.42	9.999
cost_piracy	5.126	2.818	0.01063	0.6022	2.723	5.271	7.556	9.379	9.997
piracy_links	0.1438	0.05853	0.05005	0.05751	0.09274	0.1427	0.1932	0.2384	0.2498
policy_tax	51.26	28.69	1.036	6.344	27.89	50.12	76.41	94.83	99.92
policy_votes	2.993	1.412	1	1	2	3	4	5	5

Table D10: Parameter values for 898 simulations with writing in all stages

D.3 Catch-and-fine policy

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	150.1	28.74	100	106	126	148.5	177	195	200
density	1.052	0.7824	0.2504	0.2833	0.4656	0.8108	1.374	2.724	3.969
reader_step_mean	0.4802	0.2302	0.1004	0.1334	0.2919	0.4706	0.675	0.86	0.8987
ability_max	7.372	2.132	5.001	5.118	5.74	6.718	8.459	11.83	14.53
creation_fixed	161	57.05	50.57	62.81	117.8	164.3	211.8	243.2	249.9
creation_time	2.364	1.712	0	0	1	2	4	5	5
cost_market	28.1	13.8	0.07822	4.654	17.17	28.22	40.14	48.28	50
cost_unit	5.486	2.764	0.006885	0.9225	3.155	5.713	7.894	9.644	10

Table D11: Parameter values for 642 simulations without pre-piracy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	150.5	28.88	100	103	126.2	151.5	174	196	200
density	2.129	1.018	0.2543	0.5178	1.287	2.159	2.933	3.785	3.999
reader_step_mean	0.5234	0.227	0.1028	0.1575	0.3234	0.5252	0.7269	0.8719	0.8971
ability_max	8.224	2.266	5.002	5.356	6.501	7.82	9.417	12.61	14.97
creation_fixed	158.5	56.94	50.71	63.63	110.5	160.3	207.2	243.1	249.4
creation_time	2.464	1.633	0	0	1	2	4	5	5
cost_market	23.47	14.58	0.009585	2.318	10.78	21.88	35.58	47.75	49.97
cost_unit	5.095	2.897	0.02626	0.4424	2.708	5.168	7.498	9.494	9.975
cost_piracy	4.763	2.874	0.0108	0.4003	2.374	4.574	7.04	9.314	9.985
piracy_links	0.1586	0.05663	0.05068	0.06518	0.1105	0.1594	0.2059	0.2422	0.2496
policy_tax	41.32	26.56	1.526	5.054	19.13	37.27	61.94	89.1	99.91
policy_fine	43.42	29.38	0.1466	3.779	17.14	37.96	68.4	92.89	99.47

Table D12: Parameter values for 470 simulations without piracy or policy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	153.9	28.59	100	106	130	155	178	196.6	200
density	2.175	0.9945	0.2646	0.6353	1.35	2.168	3.003	3.763	3.999
reader_step_mean	0.5294	0.2265	0.1021	0.1562	0.3342	0.5457	0.7306	0.8633	0.9
ability_max	10.49	2.585	5.039	6.196	8.476	10.52	12.68	14.4	15
creation_fixed	144	57.86	50.22	56.84	93.69	142.2	191.8	236.2	250
creation_time	2.132	1.69	0	0	1	2	4	5	5
cost_market	24.41	14.46	0.09387	2.662	11.52	23.44	36.71	47.6	49.99
cost_unit	4.917	2.839	0.001187	0.5737	2.512	4.907	7.299	9.52	9.993
cost_piracy	4.903	2.867	0.0102	0.4319	2.498	4.754	7.453	9.494	9.998
piracy_links	0.1495	0.05712	0.05003	0.0626	0.09977	0.1492	0.1978	0.2412	0.25
policy_tax	53.42	28.43	1.074	7.077	30.06	53.83	78.53	95.82	100
policy_fine	50.46	28.4	0.0141	5.253	26.2	49.91	74.17	95.58	99.91

Table D13: Parameter values for 1408 simulations with no piracy writing, but recovery under the policy

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	148.2	25.29	102	115	125	149	167	189	195
density	2.437	0.9874	0.3014	0.5954	1.834	2.356	3.358	3.687	3.831
reader_step_mean	0.5102	0.2594	0.1201	0.1477	0.247	0.5261	0.7701	0.8599	0.8731
ability_max	8.566	2.537	5.061	5.346	6.84	7.845	9.658	13.24	14.73
creation_fixed	154.1	58.96	57.39	68.42	115.8	153.1	207.9	245.1	249.6
creation_time	2.341	1.51	0	0	1	2	3	5	5
cost_market	24.75	14	0.39	2.692	14.32	22.47	38.42	46.06	46.98
cost_unit	5.118	2.773	0.6294	0.9869	2.656	4.947	7.961	8.619	9.687
cost_piracy	5.454	3.058	0.253	0.9178	2.821	5.487	7.762	9.838	9.85
piracy_links	0.1594	0.05741	0.05212	0.05904	0.1309	0.1646	0.1992	0.2369	0.2441
policy_tax	40.21	26.93	3.025	8.123	21.23	34.29	54.01	94.85	99.9
policy_fine	36.79	28.52	1.077	3.864	12.48	27.16	59	85.64	97.36

Table D14: Parameter values for 41 simulations with piracy writing but not policy writing

Parameter	Mean	s.e.	Min	5th %	25th %	Median	75th %	95th %	Max
readers	143.6	28.47	100	103	119	141	167	192	200
density	2.739	0.8939	0.2532	1.078	2.131	2.925	3.465	3.9	3.999
reader_step_mean	0.4666	0.2322	0.1002	0.1301	0.2721	0.4454	0.6605	0.865	0.8996
ability_max	11.99	2.182	5.532	7.677	10.57	12.43	13.78	14.72	14.99
creation_fixed	146.5	57.43	50.02	59.94	95.63	144.9	195.7	238.4	249.6
creation_time	3.084	1.523	0	0	2	3	4	5	5
cost_market	25.71	14.19	0.003049	2.599	14.23	25.31	38.27	47.3	49.86
cost_unit	4.719	2.844	0.004141	0.4423	2.277	4.551	7.143	9.425	9.99
cost_piracy	5.237	2.936	0.0002158	0.4337	2.683	5.327	7.742	9.577	9.998
piracy_links	0.1453	0.05826	0.05009	0.06002	0.09386	0.1394	0.1978	0.2378	0.2496
policy_tax	49.82	29.28	1.078	4.681	24.6	50.29	75.66	95.32	99.97
policy_fine	52.78	28.79	0.04753	6.067	28.2	54.09	77.7	95.25	99.99

Table D15: Parameter values for 893 simulations with writing in all stages

E Simulation time series

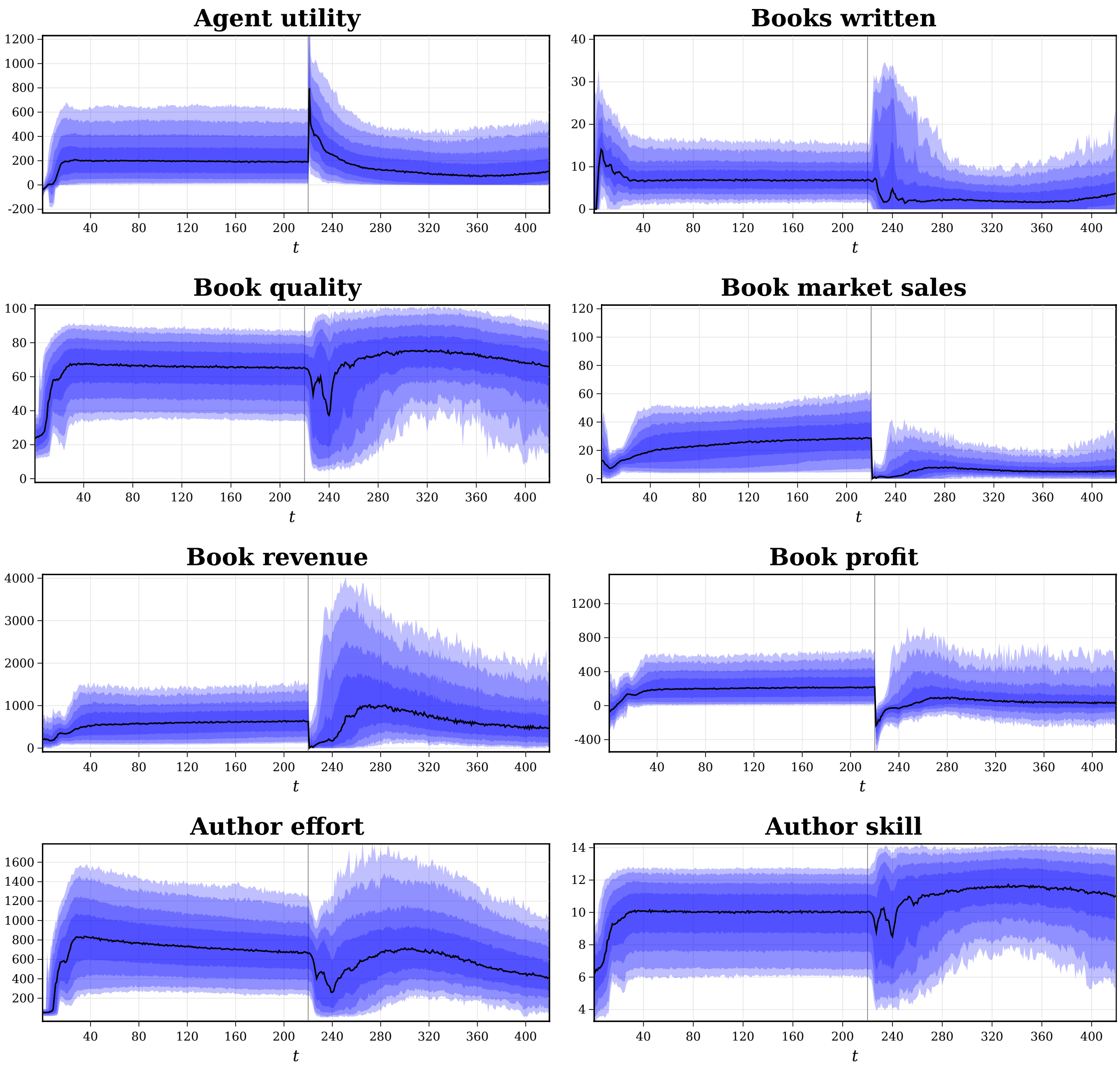
E.1 Pre-piracy and piracy stages

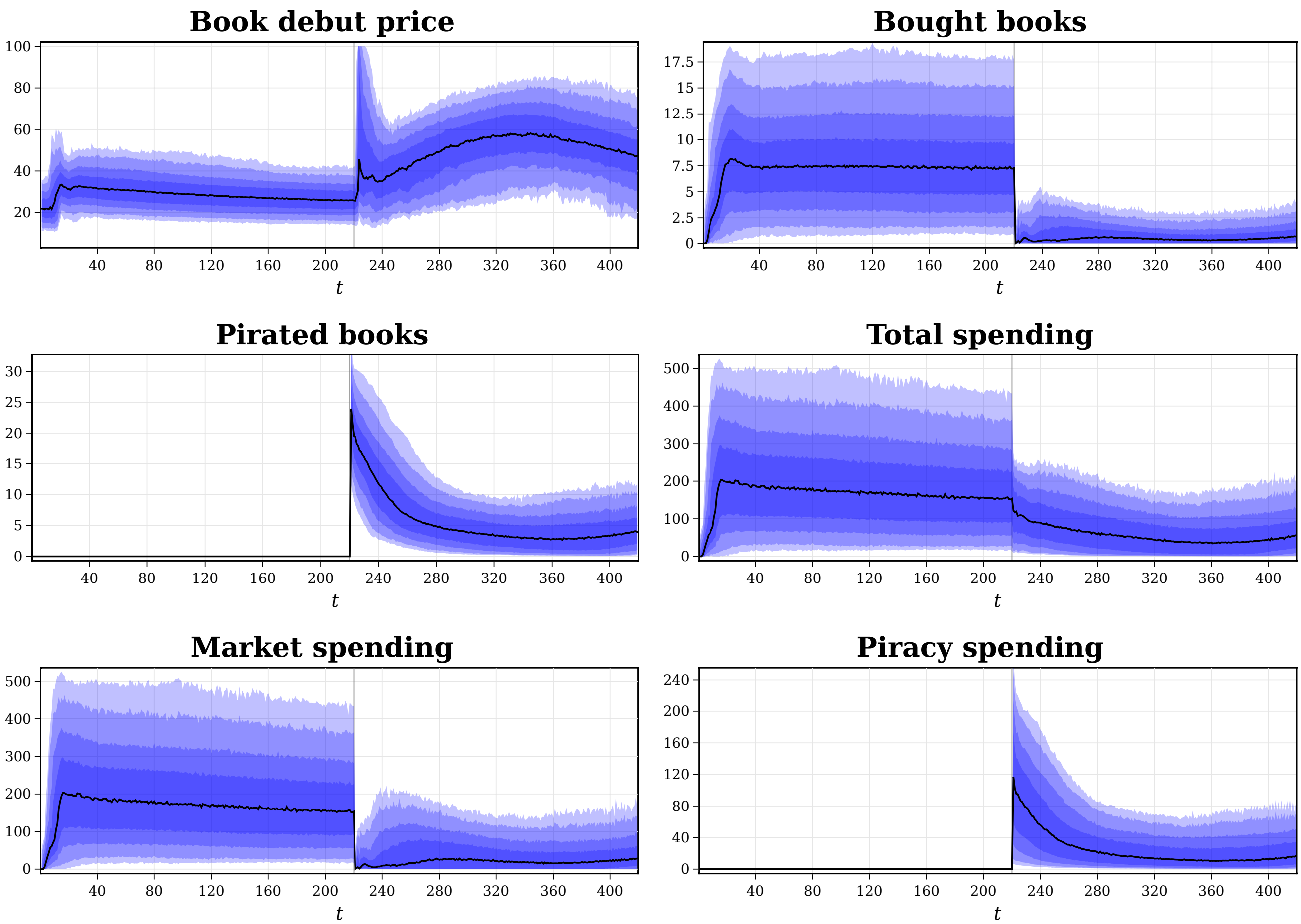
The following graphs show the distribution of the means of various variables through the pre-piracy and piracy stages across the analyzed simulations. Each graph shows the median as the black line and bands indicating 50%, 75%, 90%, and 95% of simulations. Only simulations with writing in each stage are included. The vertical line at $t = 220$

indicates the introduction of piracy.

The values for each simulation are constructed by averaging across all agents or books within each period of a simulation. Book-specific values (such as quality, revenue, and author effort) are per-book averages. Agent utility, bought books, and spending values are averaged across simulation agents. The "Books written" variable is the number of books written per 100 agents in the simulation.

The values depicted are generated from the "public sharing" policy data; it should be noted, however, that there is very little discernable difference between the pre-policy graphs for any of the three policies.

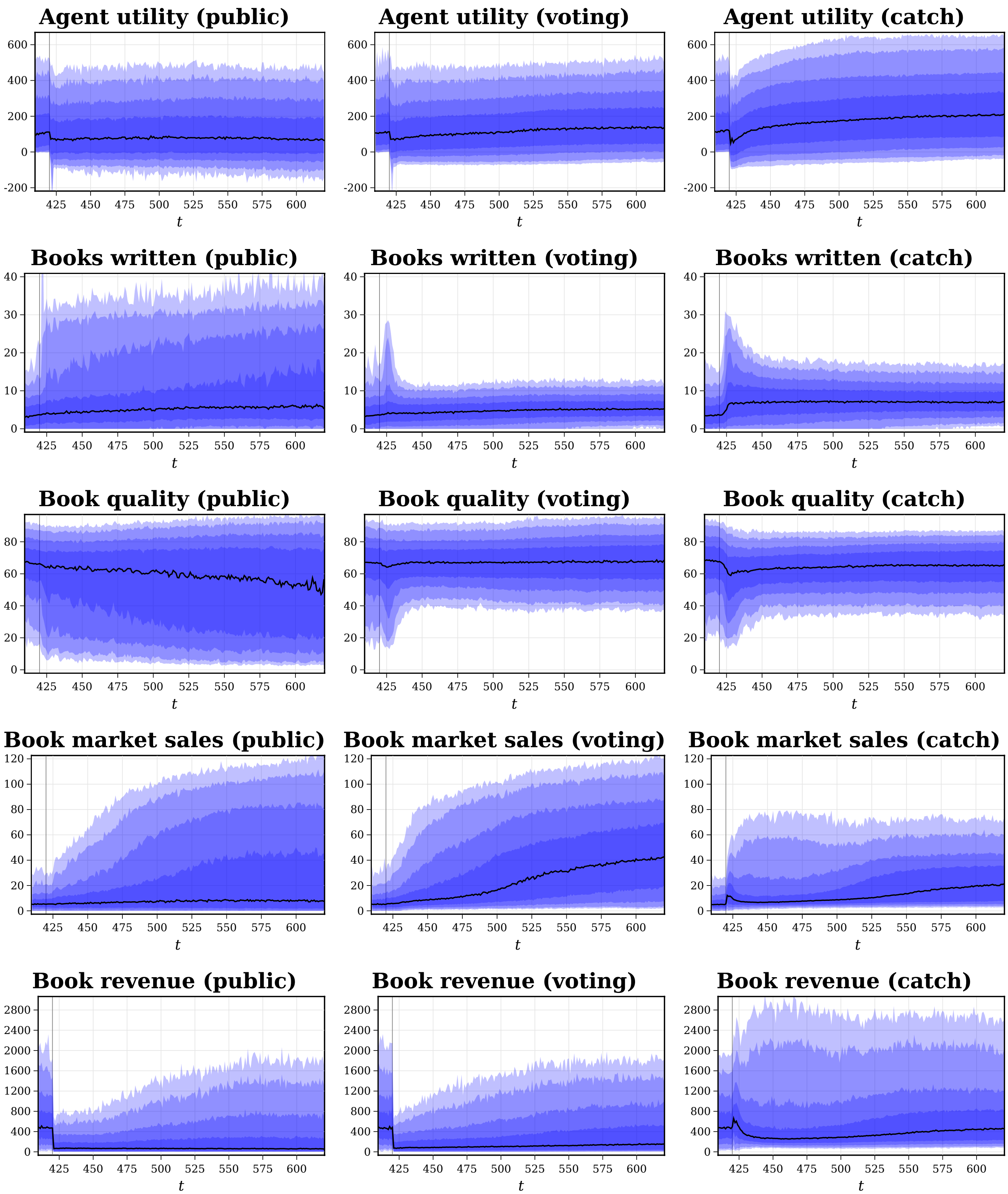


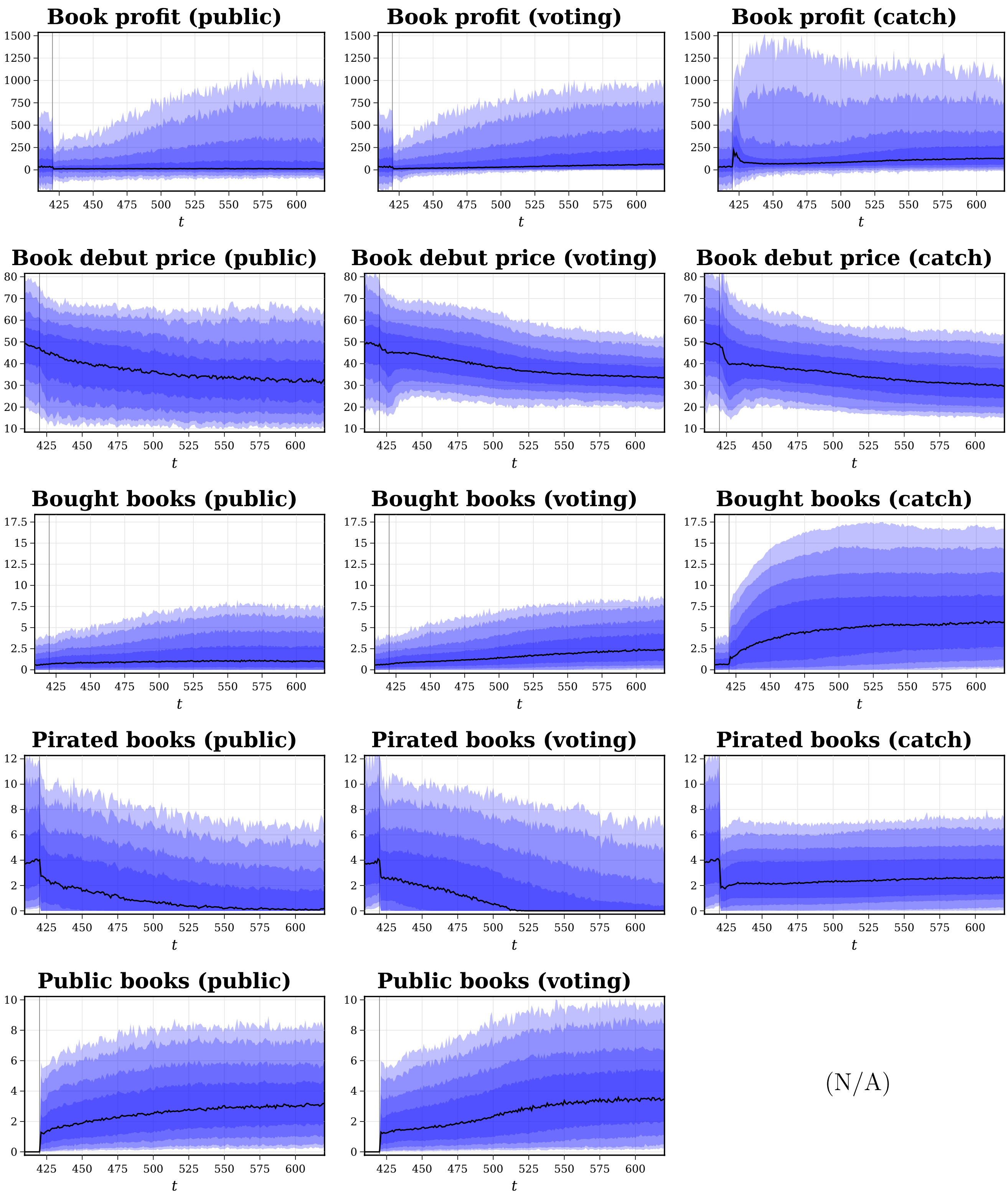


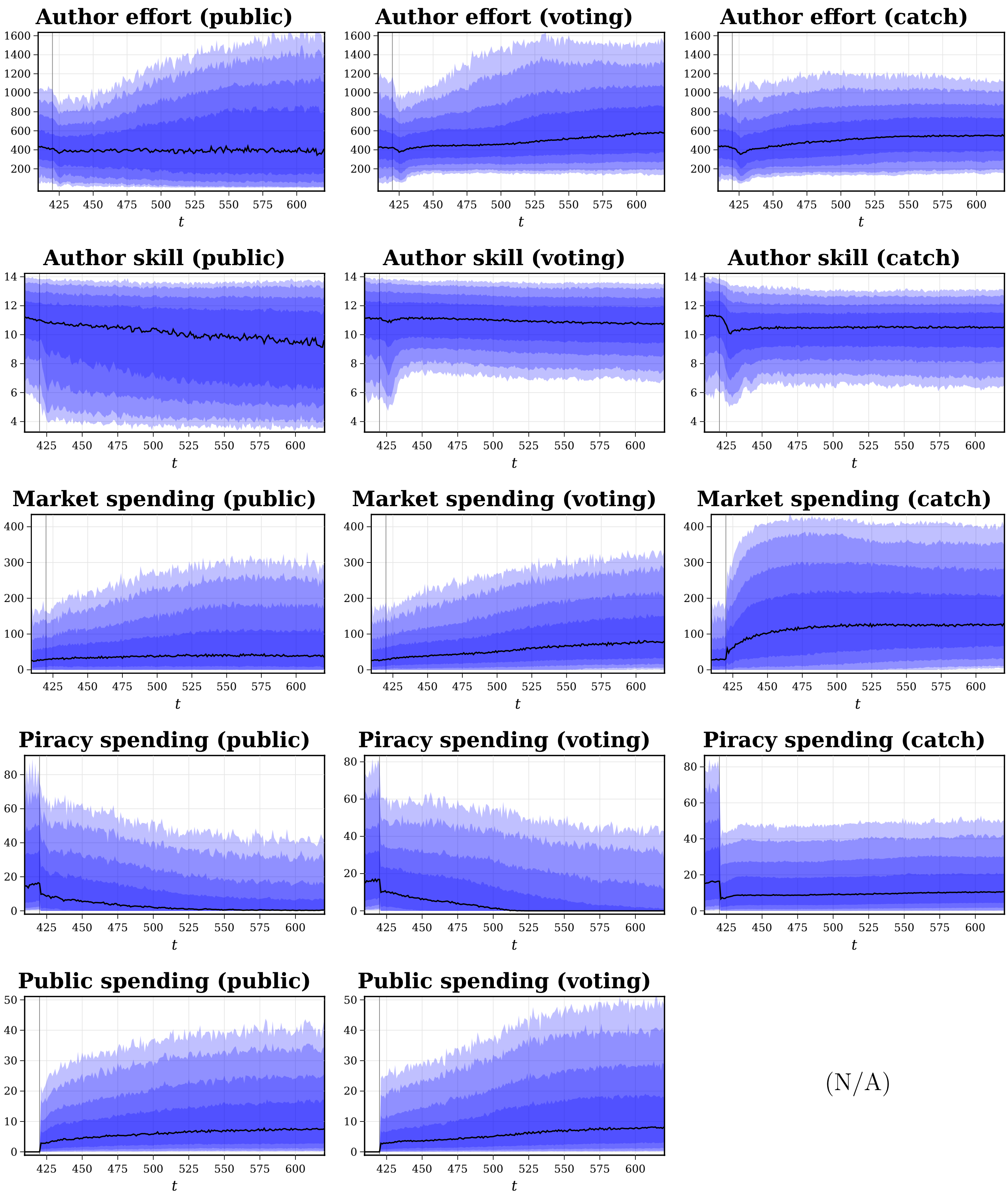
E.2 Policy stages

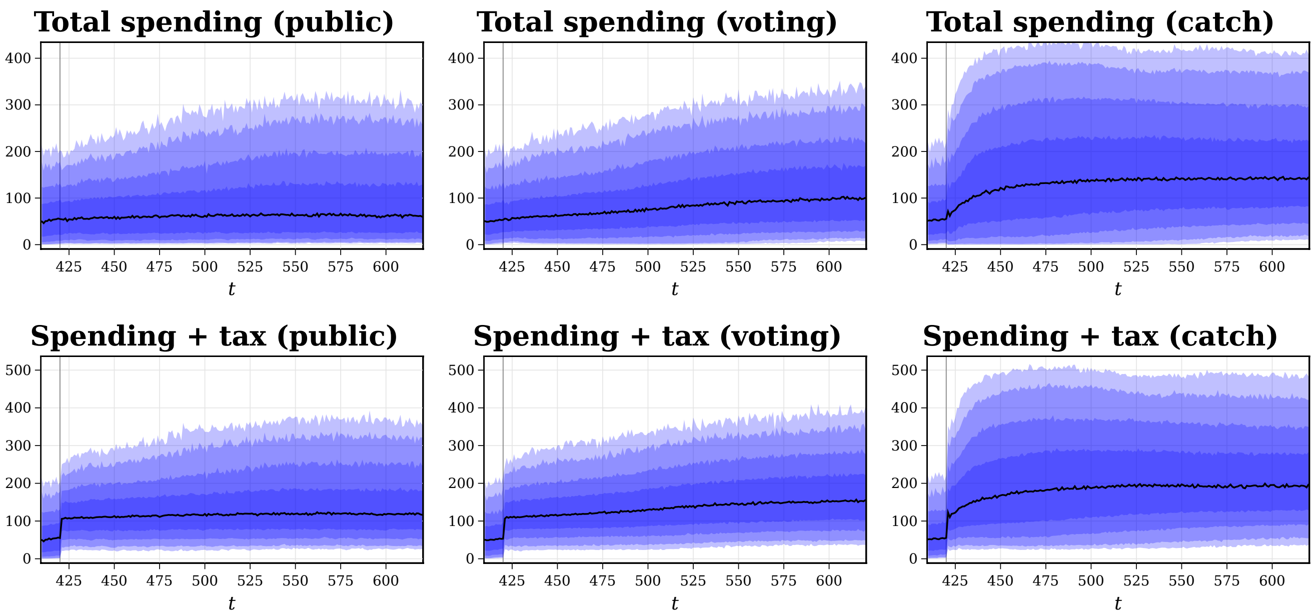
The following show the responses of the same variables depicted above to the introduction of the three policies explored in this paper. The introduction of the policy is indicates by the vertical line at $t = 420$. Piracy data (from $t = 410$ to $t = 419$) is included for comparison.

"Public" indicates the public sharing policy; "voting" indicates the public sharing with voting policy; and "catch" indicates the catch-and-fine policy. The calculation and interpretation of the different variables is identical to that in the pre-policy graphs depicted above.


Agent utility (public)
Agent utility (voting)
Agent utility (catch)
Books written (public)
Books written (voting)
Books written (catch)
Book quality (public)
Book quality (voting)
Book quality (catch)
Book market sales (public)
Book market sales (voting)
Book market sales (catch)
Book revenue (public)
Book revenue (voting)
Book revenue (catch)
t


Book profit (public)
Book profit (voting)
Book profit (catch)
Book debut price (public)
Book debut price (voting)
Book debut price (catch)
Bought books (public)
Bought books (voting)
Bought books (catch)
Pirated books (public)
Pirated books (voting)
Pirated books (catch)
Public books (public)
Public books (voting)
(N/A)
t


Author effort (public)
Author effort (voting)
Author effort (catch)
Author skill (public)
Author skill (voting)
Author skill (catch)
Market spending (public)
Market spending (voting)
Market spending (catch)
Piracy spending (public)
Piracy spending (voting)
Piracy spending (catch)
Public spending (public)
Public spending (voting)
(N/A)
t



F Average effects of piracy and policy

F.1 Public sharing

(N.B.: *** indicates significance at .001 level, ** at .01, and * at .05.)

Equation 1: net_u ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	224.5	5.197	43.19	0.000	***
piracy×SR	154.3	7.350	20.99	1.883×10^{-93}	***
piracy×LR	-84.14	7.350	-11.45	6.156×10^{-30}	***
policy×SR	-120.8	7.350	-16.44	5.176×10^{-59}	***
policy×LR	-119.8	7.350	-16.30	4.137×10^{-58}	***

Equation 2: net_u_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	143.6	4.201	34.18	5.548×10^{-228}	***
piracy×SR	144.0	5.941	24.24	3.516×10^{-122}	***
piracy×LR	-32.58	5.941	-5.484	4.374×10^{-8}	***
policy×SR	-76.80	5.941	-12.93	1.453×10^{-37}	***
policy×LR	-93.35	5.941	-15.71	3.194×10^{-54}	***

Equation 3: net_u_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	164.3	4.765	34.48	1.396×10^{-231}	***
piracy×SR	211.7	6.738	31.41	3.971×10^{-196}	***
piracy×LR	-32.25	6.738	-4.787	1.748×10^{-6}	***
policy×SR	-74.07	6.738	-10.99	9.343×10^{-28}	***
policy×LR	-87.26	6.738	-12.95	1.094×10^{-37}	***

Equation 4: net_u_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	541.3	9.233	58.62	0.000	***
piracy×SR	-55.69	13.06	-4.265	2.037×10^{-5}	***
piracy×LR	-334.2	13.06	-25.59	4.820×10^{-135}	***
policy×SR	-340.3	13.06	-26.06	1.266×10^{-139}	***
policy×LR	-262.7	13.06	-20.12	2.795×10^{-86}	***

Equation 5: books_written_pc ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	7.317	0.1393	52.53	0.000	***
piracy×SR	0.5120	0.1970	2.599	0.009380	**
piracy×LR	-3.082	0.1970	-15.65	8.669×10^{-54}	***
policy×SR	-0.1448	0.1970	-0.7351	0.4623	
policy×LR	3.564	0.1970	18.09	1.060×10^{-70}	***

Equation 6: book_quality ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	64.19	0.6719	95.54	0.000	***
piracy×SR	-23.81	0.9502	-25.05	7.896×10^{-130}	***
piracy×LR	-6.665	0.9502	-7.014	2.665×10^{-12}	***
policy×SR	-17.68	0.9502	-18.61	1.629×10^{-74}	***
policy×LR	-26.40	0.9502	-27.78	5.506×10^{-157}	***

Equation 7: book_quality_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	49.08	0.7068	69.44	0.000	***
piracy×SR	-36.51	0.9995	-36.53	2.684×10^{-256}	***
piracy×LR	-11.73	0.9995	-11.74	2.352×10^{-31}	***
policy×SR	-27.45	0.9995	-27.46	1.113×10^{-153}	***
policy×LR	-34.87	0.9995	-34.88	2.225×10^{-236}	***

Equation 8: book_quality_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	64.24	0.8923	71.99	0.000	***
piracy×SR	-31.48	1.262	-24.94	8.908×10^{-129}	***
piracy×LR	-5.427	1.262	-4.301	1.738×10^{-5}	***
policy×SR	-19.19	1.262	-15.21	5.781×10^{-51}	***
policy×LR	-35.79	1.262	-28.36	5.230×10^{-163}	***

Equation 9: book_quality_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	79.26	0.6222	127.4	0.000	***
piracy×SR	-2.817	0.8800	-3.201	0.001379	**
piracy×LR	-5.537	0.8800	-6.292	3.425×10^{-10}	***
policy×SR	-5.775	0.8800	-6.562	5.889×10^{-11}	***
policy×LR	-2.060	0.8800	-2.341	0.01926	*

Equation 10: book_p0 ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	26.61	0.4084	65.16	0.000	***
piracy×SR	10.05	0.5775	17.40	1.016×10^{-65}	***
piracy×LR	18.24	0.5775	31.58	5.814×10^{-198}	***
policy×SR	9.394	0.5775	16.27	7.348×10^{-58}	***
policy×LR	1.499	0.5775	2.595	0.009494	**

Equation 11: book_revenue ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	1986.	22.53	88.18	0.000	***
piracy×SR	-1621.	31.86	-50.90	0.000	***
piracy×LR	-1112.	31.86	-34.92	9.470×10^{-237}	***
policy×SR	-1545.	31.86	-48.49	0.000	***
policy×LR	-1299.	31.86	-40.76	1.035×10^{-309}	***

Equation 12: book_profit ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	742.5	12.25	60.60	0.000	***
piracy×SR	-948.7	17.33	-54.76	0.000	***
piracy×LR	-613.2	17.33	-35.39	1.713×10^{-242}	***
policy×SR	-612.4	17.33	-35.34	6.983×10^{-242}	***
policy×LR	-468.7	17.33	-27.05	1.596×10^{-149}	***

Equation 13: book_author_scale ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	9.876	0.07732	127.7	0.000	***
piracy×SR	-1.694	0.1093	-15.49	8.943×10^{-53}	***
piracy×LR	0.1640	0.1093	1.500	0.1336	
policy×SR	-0.9992	0.1093	-9.138	9.400×10^{-20}	***
policy×LR	-1.954	0.1093	-17.87	4.589×10^{-69}	***

Equation 14: book_author_scale_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	8.108	0.1088	74.55	0.000	***
piracy×SR	-4.352	0.1538	-28.30	2.331×10^{-162}	***
piracy×LR	-0.6070	0.1538	-3.946	8.052×10^{-5}	***
policy×SR	-3.058	0.1538	-19.88	2.011×10^{-84}	***
policy×LR	-4.438	0.1538	-28.85	3.352×10^{-168}	***

Equation 15: book_author_scale_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	9.889	0.07969	124.1	0.000	***
piracy×SR	-1.363	0.1127	-12.09	3.697×10^{-33}	***
piracy×LR	0.4285	0.1127	3.802	0.0001454	***
policy×SR	-0.6383	0.1127	-5.663	1.578×10^{-8}	***
policy×LR	-1.738	0.1127	-15.42	2.514×10^{-52}	***

Equation 16: book_author_scale_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	11.61	0.07451	155.8	0.000	***
piracy×SR	-0.04941	0.1054	-0.4689	0.6392	
piracy×LR	0.06245	0.1054	0.5927	0.5534	
policy×SR	0.01627	0.1054	0.1544	0.8773	
policy×LR	-0.09161	0.1054	-0.8693	0.3847	

Equation 17: book_author_effort ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	695.1	9.865	70.46	0.000	***
piracy×SR	-326.0	13.95	-23.37	4.177×10^{-114}	***
piracy×LR	-265.5	13.95	-19.03	9.446×10^{-78}	***
policy×SR	-383.2	13.95	-27.46	9.806×10^{-154}	***
policy×LR	-286.8	13.95	-20.56	7.202×10^{-90}	***

Equation 18: book_author_effort_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	458.1	7.321	62.57	0.000	***
piracy×SR	-392.4	10.35	-37.90	2.519×10^{-273}	***
piracy×LR	-237.8	10.35	-22.97	1.553×10^{-110}	***
policy×SR	-364.7	10.35	-35.23	1.658×10^{-240}	***
policy×LR	-375.2	10.35	-36.24	9.918×10^{-253}	***

Equation 19: book_author_effort_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	686.5	11.04	62.16	0.000	***
piracy×SR	-408.3	15.62	-26.14	1.877×10^{-140}	***
piracy×LR	-265.8	15.62	-17.02	5.433×10^{-63}	***
policy×SR	-418.3	15.62	-26.78	7.987×10^{-147}	***
policy×LR	-405.1	15.62	-25.94	1.877×10^{-138}	***

Equation 20: book_author_effort_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	955.1	14.30	66.77	0.000	***
piracy×SR	-94.12	20.23	-4.653	3.360×10^{-6}	***
piracy×LR	-284.5	20.23	-14.06	5.362×10^{-44}	***
policy×SR	-301.8	20.23	-14.92	3.532×10^{-49}	***
policy×LR	123.2	20.23	6.090	1.220×10^{-9}	***

F.2 Public sharing with voting

(N.B.: *** indicates significance at .001 level, ** at .01, and * at .05.)

Equation 1: net_u ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	224.5	5.197	43.19	0.000	***
piracy×SR	154.3	7.350	20.99	1.883×10^{-93}	***
piracy×LR	-84.14	7.350	-11.45	6.156×10^{-30}	***
policy×SR	-120.8	7.350	-16.44	5.176×10^{-59}	***
policy×LR	-119.8	7.350	-16.30	4.137×10^{-58}	***

Equation 2: net_u_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	143.6	4.201	34.18	5.548×10^{-228}	***
piracy×SR	144.0	5.941	24.24	3.516×10^{-122}	***
piracy×LR	-32.58	5.941	-5.484	4.374×10^{-8}	***
policy×SR	-76.80	5.941	-12.93	1.453×10^{-37}	***
policy×LR	-93.35	5.941	-15.71	3.194×10^{-54}	***

Equation 3: net_u_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	164.3	4.765	34.48	1.396×10^{-231}	***
piracy×SR	211.7	6.738	31.41	3.971×10^{-196}	***
piracy×LR	-32.25	6.738	-4.787	1.748×10^{-6}	***
policy×SR	-74.07	6.738	-10.99	9.343×10^{-28}	***
policy×LR	-87.26	6.738	-12.95	1.094×10^{-37}	***

Equation 4: net_u_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	541.3	9.233	58.62	0.000	***
piracy×SR	-55.69	13.06	-4.265	2.037×10^{-5}	***
piracy×LR	-334.2	13.06	-25.59	4.820×10^{-135}	***
policy×SR	-340.3	13.06	-26.06	1.266×10^{-139}	***
policy×LR	-262.7	13.06	-20.12	2.795×10^{-86}	***

Equation 5: books_written_pc ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	7.317	0.1393	52.53	0.000	***
piracy×SR	0.5120	0.1970	2.599	0.009380	**
piracy×LR	-3.082	0.1970	-15.65	8.669×10^{-54}	***
policy×SR	-0.1448	0.1970	-0.7351	0.4623	
policy×LR	3.564	0.1970	18.09	1.060×10^{-70}	***

Equation 6: book_quality ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	64.19	0.6719	95.54	0.000	***
piracy×SR	-23.81	0.9502	-25.05	7.896×10^{-130}	***
piracy×LR	-6.665	0.9502	-7.014	2.665×10^{-12}	***
policy×SR	-17.68	0.9502	-18.61	1.629×10^{-74}	***
policy×LR	-26.40	0.9502	-27.78	5.506×10^{-157}	***

Equation 7: book_quality_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	49.08	0.7068	69.44	0.000	***
piracy×SR	-36.51	0.9995	-36.53	2.684×10^{-256}	***
piracy×LR	-11.73	0.9995	-11.74	2.352×10^{-31}	***
policy×SR	-27.45	0.9995	-27.46	1.113×10^{-153}	***
policy×LR	-34.87	0.9995	-34.88	2.225×10^{-236}	***

Equation 8: book_quality_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	64.24	0.8923	71.99	0.000	***
piracy×SR	-31.48	1.262	-24.94	8.908×10^{-129}	***
piracy×LR	-5.427	1.262	-4.301	1.738×10^{-5}	***
policy×SR	-19.19	1.262	-15.21	5.781×10^{-51}	***
policy×LR	-35.79	1.262	-28.36	5.230×10^{-163}	***

Equation 9: book_quality_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	79.26	0.6222	127.4	0.000	***
piracy×SR	-2.817	0.8800	-3.201	0.001379	**
piracy×LR	-5.537	0.8800	-6.292	3.425×10^{-10}	***
policy×SR	-5.775	0.8800	-6.562	5.889×10^{-11}	***
policy×LR	-2.060	0.8800	-2.341	0.01926	*

Equation 10: book_p0 ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	26.61	0.4084	65.16	0.000	***
piracy×SR	10.05	0.5775	17.40	1.016×10^{-65}	***
piracy×LR	18.24	0.5775	31.58	5.814×10^{-198}	***
policy×SR	9.394	0.5775	16.27	7.348×10^{-58}	***
policy×LR	1.499	0.5775	2.595	0.009494	**

Equation 11: book_revenue ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	1986.	22.53	88.18	0.000	***
piracy×SR	-1621.	31.86	-50.90	0.000	***
piracy×LR	-1112.	31.86	-34.92	9.470×10^{-237}	***
policy×SR	-1545.	31.86	-48.49	0.000	***
policy×LR	-1299.	31.86	-40.76	1.035×10^{-309}	***

Equation 12: book_profit ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	742.5	12.25	60.60	0.000	***
piracy×SR	-948.7	17.33	-54.76	0.000	***
piracy×LR	-613.2	17.33	-35.39	1.713×10^{-242}	***
policy×SR	-612.4	17.33	-35.34	6.983×10^{-242}	***
policy×LR	-468.7	17.33	-27.05	1.596×10^{-149}	***

Equation 13: book__author__scale ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	9.876	0.07732	127.7	0.000	***
piracy×SR	-1.694	0.1093	-15.49	8.943×10^{-53}	***
piracy×LR	0.1640	0.1093	1.500	0.1336	
policy×SR	-0.9992	0.1093	-9.138	9.400×10^{-20}	***
policy×LR	-1.954	0.1093	-17.87	4.589×10^{-69}	***

Equation 14: book__author__scale__5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	8.108	0.1088	74.55	0.000	***
piracy×SR	-4.352	0.1538	-28.30	2.331×10^{-162}	***
piracy×LR	-0.6070	0.1538	-3.946	8.052×10^{-5}	***
policy×SR	-3.058	0.1538	-19.88	2.011×10^{-84}	***
policy×LR	-4.438	0.1538	-28.85	3.352×10^{-168}	***

Equation 15: book__author__scale__median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	9.889	0.07969	124.1	0.000	***
piracy×SR	-1.363	0.1127	-12.09	3.697×10^{-33}	***
piracy×LR	0.4285	0.1127	3.802	0.0001454	***
policy×SR	-0.6383	0.1127	-5.663	1.578×10^{-8}	***
policy×LR	-1.738	0.1127	-15.42	2.514×10^{-52}	***

Equation 16: book__author__scale__95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	11.61	0.07451	155.8	0.000	***
piracy×SR	-0.04941	0.1054	-0.4689	0.6392	
piracy×LR	0.06245	0.1054	0.5927	0.5534	
policy×SR	0.01627	0.1054	0.1544	0.8773	
policy×LR	-0.09161	0.1054	-0.8693	0.3847	

Equation 17: book__author__effort ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	695.1	9.865	70.46	0.000	***
piracy×SR	-326.0	13.95	-23.37	4.177×10^{-114}	***
piracy×LR	-265.5	13.95	-19.03	9.446×10^{-78}	***
policy×SR	-383.2	13.95	-27.46	9.806×10^{-154}	***
policy×LR	-286.8	13.95	-20.56	7.202×10^{-90}	***

Equation 18: book__author__effort__5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	458.1	7.321	62.57	0.000	***
piracy×SR	-392.4	10.35	-37.90	2.519×10^{-273}	***
piracy×LR	-237.8	10.35	-22.97	1.553×10^{-110}	***
policy×SR	-364.7	10.35	-35.23	1.658×10^{-240}	***
policy×LR	-375.2	10.35	-36.24	9.918×10^{-253}	***

Equation 19: book__author__effort__median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	686.5	11.04	62.16	0.000	***
piracy×SR	-408.3	15.62	-26.14	1.877×10^{-140}	***
piracy×LR	-265.8	15.62	-17.02	5.433×10^{-63}	***
policy×SR	-418.3	15.62	-26.78	7.987×10^{-147}	***
policy×LR	-405.1	15.62	-25.94	1.877×10^{-138}	***

Equation 20: book__author__effort__95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	955.1	14.30	66.77	0.000	***
piracy×SR	-94.12	20.23	-4.653	3.360×10^{-6}	***
piracy×LR	-284.5	20.23	-14.06	5.362×10^{-44}	***
policy×SR	-301.8	20.23	-14.92	3.532×10^{-49}	***
policy×LR	123.2	20.23	6.090	1.220×10^{-9}	***

F.3 Catch and fine

(N.B.: *** indicates significance at .001 level, ** at .01, and * at .05.)

Equation 1: net__u ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	224.5	5.197	43.19	0.000	***
piracy×SR	154.3	7.350	20.99	1.883×10^{-93}	***
piracy×LR	-84.14	7.350	-11.45	6.156×10^{-30}	***
policy×SR	-120.8	7.350	-16.44	5.176×10^{-59}	***
policy×LR	-119.8	7.350	-16.30	4.137×10^{-58}	***

Equation 2: net__u__5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	143.6	4.201	34.18	5.548×10^{-228}	***
piracy×SR	144.0	5.941	24.24	3.516×10^{-122}	***
piracy×LR	-32.58	5.941	-5.484	4.374×10^{-8}	***
policy×SR	-76.80	5.941	-12.93	1.453×10^{-37}	***
policy×LR	-93.35	5.941	-15.71	3.194×10^{-54}	***

Equation 3: net__u__median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	164.3	4.765	34.48	1.396×10^{-231}	***
piracy×SR	211.7	6.738	31.41	3.971×10^{-196}	***
piracy×LR	-32.25	6.738	-4.787	1.748×10^{-6}	***
policy×SR	-74.07	6.738	-10.99	9.343×10^{-28}	***
policy×LR	-87.26	6.738	-12.95	1.094×10^{-37}	***

Equation 4: net__u__95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	541.3	9.233	58.62	0.000	***
piracy×SR	-55.69	13.06	-4.265	2.037×10^{-5}	***
piracy×LR	-334.2	13.06	-25.59	4.820×10^{-135}	***
policy×SR	-340.3	13.06	-26.06	1.266×10^{-139}	***
policy×LR	-262.7	13.06	-20.12	2.795×10^{-86}	***

Equation 5: books_written_pc ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	7.317	0.1393	52.53	0.000	***
piracy×SR	0.5120	0.1970	2.599	0.009380	**
piracy×LR	-3.082	0.1970	-15.65	8.669×10^{-54}	***
policy×SR	-0.1448	0.1970	-0.7351	0.4623	
policy×LR	3.564	0.1970	18.09	1.060×10^{-70}	***

Equation 6: book_quality ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	64.19	0.6719	95.54	0.000	***
piracy×SR	-23.81	0.9502	-25.05	7.896×10^{-130}	***
piracy×LR	-6.665	0.9502	-7.014	2.665×10^{-12}	***
policy×SR	-17.68	0.9502	-18.61	1.629×10^{-74}	***
policy×LR	-26.40	0.9502	-27.78	5.506×10^{-157}	***

Equation 7: book_quality_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	49.08	0.7068	69.44	0.000	***
piracy×SR	-36.51	0.9995	-36.53	2.684×10^{-256}	***
piracy×LR	-11.73	0.9995	-11.74	2.352×10^{-31}	***
policy×SR	-27.45	0.9995	-27.46	1.113×10^{-153}	***
policy×LR	-34.87	0.9995	-34.88	2.225×10^{-236}	***

Equation 8: book_quality_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	64.24	0.8923	71.99	0.000	***
piracy×SR	-31.48	1.262	-24.94	8.908×10^{-129}	***
piracy×LR	-5.427	1.262	-4.301	1.738×10^{-5}	***
policy×SR	-19.19	1.262	-15.21	5.781×10^{-51}	***
policy×LR	-35.79	1.262	-28.36	5.230×10^{-163}	***

Equation 9: book_quality_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	79.26	0.6222	127.4	0.000	***
piracy×SR	-2.817	0.8800	-3.201	0.001379	**
piracy×LR	-5.537	0.8800	-6.292	3.425×10^{-10}	***
policy×SR	-5.775	0.8800	-6.562	5.889×10^{-11}	***
policy×LR	-2.060	0.8800	-2.341	0.01926	*

Equation 10: book_p0 ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	26.61	0.4084	65.16	0.000	***
piracy×SR	10.05	0.5775	17.40	1.016×10^{-65}	***
piracy×LR	18.24	0.5775	31.58	5.814×10^{-198}	***
policy×SR	9.394	0.5775	16.27	7.348×10^{-58}	***
policy×LR	1.499	0.5775	2.595	0.009494	**

Equation 11: book_revenue ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	1986.	22.53	88.18	0.000	***
piracy×SR	-1621.	31.86	-50.90	0.000	***
piracy×LR	-1112.	31.86	-34.92	9.470×10^{-237}	***
policy×SR	-1545.	31.86	-48.49	0.000	***
policy×LR	-1299.	31.86	-40.76	1.035×10^{-309}	***

Equation 12: book_profit ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	742.5	12.25	60.60	0.000	***
piracy×SR	-948.7	17.33	-54.76	0.000	***
piracy×LR	-613.2	17.33	-35.39	1.713×10^{-242}	***
policy×SR	-612.4	17.33	-35.34	6.983×10^{-242}	***
policy×LR	-468.7	17.33	-27.05	1.596×10^{-149}	***

Equation 13: book_author_scale ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	9.876	0.07732	127.7	0.000	***
piracy×SR	-1.694	0.1093	-15.49	8.943×10^{-53}	***
piracy×LR	0.1640	0.1093	1.500	0.1336	
policy×SR	-0.9992	0.1093	-9.138	9.400×10^{-20}	***
policy×LR	-1.954	0.1093	-17.87	4.589×10^{-69}	***

Equation 14: book_author_scale_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	8.108	0.1088	74.55	0.000	***
piracy×SR	-4.352	0.1538	-28.30	2.331×10^{-162}	***
piracy×LR	-0.6070	0.1538	-3.946	8.052×10^{-5}	***
policy×SR	-3.058	0.1538	-19.88	2.011×10^{-84}	***
policy×LR	-4.438	0.1538	-28.85	3.352×10^{-168}	***

Equation 15: book_author_scale_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	9.889	0.07969	124.1	0.000	***
piracy×SR	-1.363	0.1127	-12.09	3.697×10^{-33}	***
piracy×LR	0.4285	0.1127	3.802	0.0001454	***
policy×SR	-0.6383	0.1127	-5.663	1.578×10^{-8}	***
policy×LR	-1.738	0.1127	-15.42	2.514×10^{-52}	***

Equation 16: book_author_scale_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	11.61	0.07451	155.8	0.000	***
piracy×SR	-0.04941	0.1054	-0.4689	0.6392	
piracy×LR	0.06245	0.1054	0.5927	0.5534	
policy×SR	0.01627	0.1054	0.1544	0.8773	
policy×LR	-0.09161	0.1054	-0.8693	0.3847	

Equation 17: book_author_effort ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	695.1	9.865	70.46	0.000	***
piracy×SR	-326.0	13.95	-23.37	4.177×10^{-114}	***
piracy×LR	-265.5	13.95	-19.03	9.446×10^{-78}	***
policy×SR	-383.2	13.95	-27.46	9.806×10^{-154}	***
policy×LR	-286.8	13.95	-20.56	7.202×10^{-90}	***

Equation 18: book_author_effort_5th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	458.1	7.321	62.57	0.000	***
piracy×SR	-392.4	10.35	-37.90	2.519×10^{-273}	***
piracy×LR	-237.8	10.35	-22.97	1.553×10^{-110}	***
policy×SR	-364.7	10.35	-35.23	1.658×10^{-240}	***
policy×LR	-375.2	10.35	-36.24	9.918×10^{-253}	***

Equation 19: book_author_effort_median ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	686.5	11.04	62.16	0.000	***
piracy×SR	-408.3	15.62	-26.14	1.877×10^{-140}	***
piracy×LR	-265.8	15.62	-17.02	5.433×10^{-63}	***
policy×SR	-418.3	15.62	-26.78	7.987×10^{-147}	***
policy×LR	-405.1	15.62	-25.94	1.877×10^{-138}	***

Equation 20: book_author_effort_95th ~ const + [piracy×SR] + [piracy×LR] + [policy×SR] + [policy×LR]

	Coeff.	std.err.	t-stat	p-value	
const	955.1	14.30	66.77	0.000	***
piracy×SR	-94.12	20.23	-4.653	3.360×10^{-6}	***
piracy×LR	-284.5	20.23	-14.06	5.362×10^{-44}	***
policy×SR	-301.8	20.23	-14.92	3.532×10^{-49}	***
policy×LR	123.2	20.23	6.090	1.220×10^{-9}	***