Demand Distribution Dynamics in Creative Industries: the Market for Books in Italy*

Edoardo Gaffeo,[a] Antonello E. Scorcu,[b] Laura Vici,[b]

[a] *Department of Economics and CEEL, University of Trento, Italy*
[b] *Department of Economics, University of Bologna, Italy*

Abstract

We study the distribution dynamics of the demand for books in Italy. We find that for each of three broad sub-markets in which the book publishing industry can be classified – Italian novels, foreign novels and essays – sales over a three-year sample can be adequately fitted by a power law distribution. Our results can be plausibly interpreted in terms of a model of interactions among buyers exchanging information on the books they buy.

Keywords: Book publishing industry; Information transmission; Power law distribution.
JEL classification: L82, Z1.

---

1. Introduction

The decision to buy a book represents an activity in which social influences play a capital role. The choice of a title from the shelves of a bookshop is often the outcome of a decision process where personal attitudes are combined with a bundle of information from acquaintances' experiences through word-of-mouth on the one hand, and public sources like reviews on newspapers or magazines on the other one. Social interaction and information transmission, therefore, are likely powerful forces shaping the economics of the book publishing industry. Surveys conducted on a regular basis among consumers seem to support this assumption. In Italy (see e.g. AIE, 2005), for instance, 55% of the people that buy books affirm to possess all the information they need about the title they want to buy before entering a bookshop. For frequent readers (i.e., more than 12 books read per year), such a share rises to 69%. More than 50% of Italian readers state their book selection process is a by-product of word-of-mouth effects, while 16% and 12% of them are mostly influenced by booksellers and book reviews, respectively.

As in all creative industries characterized by strong social influences and sequential discovery of quality (Caves, 2000) - like the movie industry or performing arts - the demand for new books can never be known a priori (the famous *nobody knows* principle), but it evolves dynamically over time according to decentralized information processes. The *New York Times*' columnist Martin Arnold has admirably summarized this feature of the publishing business, by noting that "[…] *those who say there are no absolutes in the gamble that is book publishing are almost right. Actually, there is one. Ultimately the success of a novel depends on that mystical force called word of mouth*" (quoted in Beck, 2007). In turn, the supply has to adjust accordingly to meet a continuously evolving demand. The evolution of the system gives rise to a market characterized by complex dynamics, in which a new-entry bestseller selling 100,000 copies in a month cohabits with a poetry title selling a few copies per year, and all of them are sold at roughly the same final price.

In addition to the social dimension affecting its consumption, a book is a typical experience good, the value of which can be fully recognized only after it is consumed. Furthermore, the cost structure of a creative process leading to a published result exhibits initial high fixed costs, while the marginal cost of production, that is the cost of producing copies from the original and of distributing them, is generally very low.

Unlike other cultural goods (i.e., radio broadcasting or historic monuments), however, books are rival and excludable, so that the probability to observe market

failures should be ruled out almost by definition. As a result, from an industrial organization perspective the book publishing industry is customarily seen as a monopolistic competitive market which works properly, through a traditional supply chain composed of production, wholesale, distribution and retail (Allen and Curwen, 1991). Far from suggesting that problems of signal extraction on the unknown quality of a title could add to the complexity of the book business, the consensus view explicitly maintains that "[…] *author reputation, book reviews, book clubs and word-of-mouth create a fair amount of transparency.* […] *The characteristics of book markets are not that different from many other markets*" (Canoy *et al.,* 2006).

Moving from these premises, in this paper we aim at shedding some light on the role played by peer-to-peer influences among heterogeneous buyers on book sales, by empirically exploring some of the leading features of the demand for books in Italy. Instead of recurring to structural econometric estimates from a reduced form equation, we prefer to focus directly on the distributional characteristics of the data. The key argument of the research methodology we employ is that alternative assumptions on the decision-making process adopted by prospective buyers should lead to distinct invariant (i.e., long run) distributions of sales among all the titles available on the market. Conversely, the emergence of certain regularities on the way the data are distributed should allow us to discriminate among alternative models of consumption behavior, or on the plausibility of their underlying assumptions.

Summing up our main results, we find that the Power Law (or Pareto) distribution with infinite variance represents a reasonable statistical model for fitting the right tail of the book sales' distribution for each of three broad categories in which the book publishing industry can be classified, i.e. Italian novels, foreign novels and essays. Such a result, which may be naturally interpreted in terms of the invariant distributional outcome of an information contagion model, suggests that books are extremely risky products and that word-of-mouth is likely to add to this uncertainty, rather than to help reducing it. Furthermore, we find that the scaling exponents characterizing the sales distributions evolve in time in a cyclical fashion, signaling the possibility of interesting interactions between the openings of new books, the transmission of peer-to-peer information and seasonal effects.

The paper is organized as follows. Section 2 provides an overview of the book publishing industry, with a special focus to Italy. Section 3 introduces some theoretical and empirical issues regarding the book demand distribution, and discusses a simple choice-theoretic model aimed at capturing its main empirical characteristics. Section 4

reports our empirical results, obtained exploiting a bi-monthly data-set for books sold in Italy over the period 1994-96. Finally, Section 5 concludes.

2. The book publishing industry in Italy

While in the economic literature a description of the key features of the books market have been offered for several countries, see e.g. U.S. (Szenberg M. and E. Youngkoo Lee, 1994; Greco, 2000), Denmark (Hjorth-Andersen (2000) and U.K. (Allen and Curwen, 1991), far less attention has been devoted so far to Italy. Since our empirical analysis is based on Italian data, it seems worthwhile to motivate our analysis and to briefly summarize the economics of the book publishing industry by reporting some descriptive statistics for Italy. Further details on the size and characteristics of the market for books in Italy can be found e.g. in AIE (2005).

In spite of being extremely popular and its core business being well established, the book publishing industry across industrialized countries, and therefore in Italy as well, is not very large in absolute terms. In 2003, the value of total shipments (books, CD-Rom, collection editions and exports) by Italian producers was around €3.5 billions. In that same year, about 4,000 editors issued 55,000 titles, of which around 70% were new editions. The total number of printed copies was 254 millions, to which we must add *circa* 77 millions copies (referring to 1659 titles) sold attached to newspapers or magazines.

The number of people (adults aged 14-year and older) who declared to have bought at least one book during the previous year was 17.6 millions, while the number of people (with more than 6 years) who declared to have read at least one book (in addition to school-books) was 23 million (41.4% of the Italian population), of whom 60% were women. The number of frequent readers (at least one book per month) is much lower, however, being equal to 2.8 millions. The reading-ratio for metropolitan areas is 47%, while it decreases to 40% for cities with less than 50,000 inhabitants, suggesting that people living in more concentrated place have a higher probability to be a reader. On average, individuals spend € 44 in books per year (including school and university textbooks). However, given that a mere 13% of Italian households count for 51% of books sold, central tendencies could be misleading.

Most publishers segment their market by making books available in hardcover and paperback versions at different prices. In terms of titles, 23% of the volumes available at bookstores and other shops (like supermarkets) in 2003 were priced less than €7.75, while an additional 17% cost less than €15.50. In terms of printed copies, 67% of the

total has an end-sales price lower than €15.50. The average pondered price charged for new books was €18.52.

The publishing industry is notoriously highly concentrated, in terms of the market shares of both publishing firms and titles. Just 220 publishers issued more than 50 titles in 2003; four large publishing groups (Mondadori, Rizzoli, De Agostini, Messaggerie Italiane), plus 50 middle size publishers covered around 90% of the whole market if measured by total book sales. As far as the market shares of titles is concerned, it suffices to note that in 2003 the average edition for books has been equal to 4,800 copies, slightly lower than the mean value registered in the 1990s (6,000). During the same period, however, *The DaVinci Code* by Dan Brown and *Harry Potter and Order of the Phoenix* by Joan Rowlings sell more than 500,000 copies each. Just to fix ideas, if we assume that the first 10 bestsellers sell 100,000 copies each at the average pondered price (€18.52), it turns out that the C10 concentration index for products is equal to a staggering 65% of the total market for books. This back-of-the-envelope calculation returns figures in line with the international experience. Data reported for the U.S. market in Becker (1991), for instance, reveal that for a representative publisher the coefficient of variation in total sales from its two-year old catalogue exceeds 129% for hardcover fiction and 177% for nonfiction books, respectively.

The book publishing industry, therefore, is dominated by extreme events and the *nobody knows* principle, a fate shared with other industries for which social influences on individual demands matter a lot and the *ex-ante* uncertainty on the success or failure of new products is so high that it cannot be reliably estimated, as the movie or the pharmaceuticals industries. New titles are inherently risky products: what makes a book successful is in general the number of readers of the first two or three months, which in turn depends on the composite interaction of advertising and fortuitous events triggering chain reactions which leads to selling stars or to flops, generally following hump-shaped sales patterns (Beck, 2007). Bandwagon effects and/or informational cascades origin from a complex stochastic process that can go anywhere, implying a time-increasing variance of the process. For consumers, the importance of previous information about the quality of a book is even more important than, say, knowledge about the opinions of earlier movie-goers, as books are not only expensive but also and above all time consuming. Moreover, differently from movies (if we exclude VHSs and DVDs) books are durable goods.

These facts are at the heart of several contractual arrangements aimed at minimizing risk. New authors are usually asked by a publisher to sign a contract which includes a "first refusal" clause that guarantees to the latter the right to decide whether to edit the

manuscript or not, by submitting it to a refereeing process. After the author has published successfully his first manuscript, in turn, she is generally offered a long-term exclusive contract, implying that the author commits herself with the same editor for the manuscripts to follow under the implicit assumption that reputation effects contribute to sales. As the demand cannot effectively be forecasted for any particular piece of craftwork, however, royalties are established on the basis of a fixed-end price. Fixed-end pricing allows also for cross-subsidization of a large number of less successful projects by few, but unknown *ex-ante*, more successful ones due to positive bandwagon effects. In addition, van der Ploeg (2004) suggests that fix-price agreements cause more variety in titles published even in the absence of cross-subsidization, although at the cost of lower sales of each title due to monopolistic pricing. These contractual practices underline the "*winner-take-all*" nature of the book industry: many titles cause losses for publishers, but few of them allow substantial gains able to compensate the relevant losses produced by unsuccessful editions.

In summary, revenues in the publishing industry are dominated by few bestsellers in the thick upper tail of the sales distribution, and the profits of the sector come only from these few industry productions. This is the reason why any publishers' attempt at reducing the fundamental uncertainty surrounding the sector (by hiring highly reputed writers, publishing sequels, investing in adverting, and so on) is likely to fail. Information cascades drive the consumers' selection process implying a variance much higher (potentially infinite) than the one we would expect if the business were Gaussian, with almost unpredictable word-of-mouth selection processes and an expected value of sales without any attraction power.

A crucial issue, therefore, appears to be that of empirically modelling the dynamics of demand, in order to assess whether some useful regularities may be detected which may help in modelling the forces at work in this market.

3. Empirical and theoretical issues

The property of non-predictability of final demand for cultural goods implies that ex-ante knowledge of buyers' preferences and the intrinsic merit of artistic works may not suffice to predict the final market configuration. The complex interaction of sequential discovery of quality through word-of-mouth, reputation effects, advertising and publishers' distributional strategies entail that the selection among many alternative outcomes and the speed of convergence toward them are driven by the accumulation of many random historical events. Among them, the spreading of information on

quality by early buyers is predominant. Therefore, social interactions must be taken into account if we aim at properly understanding the book publishing industry.

3.1 Distributional analysis

The empirical detection of social interactions in markets is well-known to be affected by several identification issues emerging from the so-called *reflection problem* (Manski, 1993). In a nutshell, whenever the propensity of an agent to behave in some way varies positively with the mean behavior in a reference group, observations of market outcomes do not allow the researcher to assess whether individual behaviors conform to the mean behavior, or group behavior simply reflects the aggregate behavior of isolated individuals. While several identification procedures have been successfully proposed (Brock and Durlauf, 2000; Manski, 1997), all of them relies on the availability of micro data or spatially organized observations for the same phenomenon (Glaeser and Scheinkman, 2001). Alas, given the data at our disposal none of these strategies are available.

In this paper we follow an alternative route in trying to empirically detect social interactions, i.e. by treating book sales as stochastic dynamic processes driven by information transmission, and studying the distributional features of the statistical attractors to which these dynamic processes converge (DeVany and Walls, 1996). We start by recognizing that the occurrence of some bestsellers among a large population of books selling a very limited amount of copies should imply a book sales distribution characterized by a positive support and a very long and heavy right tail. In principle, the exact profile of this tail should convey information on the stochastic process generating it, as different models lead to different testable predictions on how *large* observations relate to the bulk of the distribution.

In order to assess the legitimacy of our working hypothesis, the first step consists in performing a preliminary analysis based on a representational technique traditionally employed in *hard sciences* like physics, the so-called Zipf plot, consisting in plotting the log of the rank versus the log of the variable being analyzed (Stanley *et al.*, 1995). Briefly, suppose that we observe M realizations $(s_1, \ldots, s_M)$ of a non-negative random variable S, whose cumulative distribution function is $F(s)$. If we rank the data from the largest to the smallest, with the index i being the rank of s_i, we obtain:

$$\ln i = \ln\left[1 - F\left(s_i\right)\right] + \ln M \ . \tag{1}$$

7

The rank-size relation is extremely useful as it accentuates the tails of the distribution. In particular, while the typical shape of many well-known fat-tail distributions, like the lognormal or the Weibull ones, is typically downward bending if plotted in a log-log plot, the appearance of a linear rank-size relationship in double-logarithmic paper can be immediately read as the sign of a power law distribution. The power law distribution corresponds to the asymptotic limit of the Pareto distribution:

$$\Pr[S \geq s] = \left(\frac{s}{s_0}\right)^{-\alpha} \tag{2}$$

where s_0 is lower threshold ($S \geq s_0$) and $\alpha > 0$. Provided that $[1 - F(s)] \sim s^{-\alpha}$, where $f(x) \sim g(x)$ represents that the limit of their ratio goes to 1 as x grows large, the double logarithmic plot of sales versus rank appears as a straight line with slope $-\alpha$. It follows that in a power law distribution the tails fall to the power α, which leads to much heavier tails than other models commonly used in describing how demand distributes in monopolistically competitive markets, such as Gaussian or exponential distributions. Roughly speaking, if the data on book sales are power law distributed, the average value of total revenues and profits are dominated by few bestsellers. Furthermore, as discussed below, Paretian behaviors are generally due to stochastic processes in which all possible outcomes are equally likely ex-ante, i.e. the statistical translation of the *nobody knows* principle. As a result, the book publishing business is an extremely risky one, much riskier than predicted by the now conventional wisdom among the profession.

Figure 1 reports a representative sample of Zipf plots of book sales distributions in Italy, randomly selected from a data-set reporting bi-monthly sales in bookstores during the period 1994-96. Each point represents (the log of) a rank-size pair, where size refers to the total number of copies sold by a title during a time window elapsing two months. More details on the data will be provided in Section 4. Notice that in all cases of Figure 1 a linear relationship appears. More importantly, as we will show with more details in Section 4 the same feature emerges regularly in our data-set, suggesting that a power law behavior in the demand for books in Italy seems to be pervasive. As shown in Section 4, all estimates yields $\alpha < 2$ with a remarkable goodness of fit to a linear relationship.
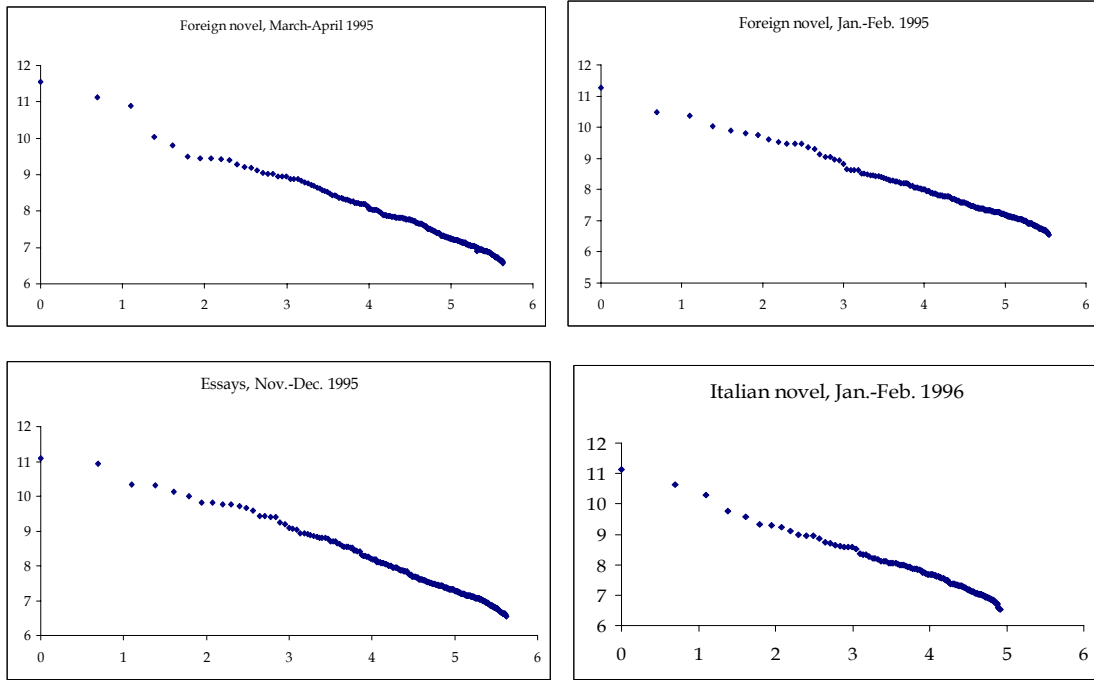
Figure 1. Representative sample of the book sales distribution in Italy.

The relevance of the Pareto distribution (2) can be immediately acknowledged by noting that a Paretian tail behavior with $\alpha < 2$ is a distinguishing attribute of a family of distributions known as Lévy-stable (Lévy, 1925). In particular, the Generalized Central Limit Theorem (Gnedenko and Kolmogorov, 1954) states that the only possible limiting distribution for sums of independent and identically distributed random variables is a Lévy-stable distribution. The Gaussian distribution is just a special case of the above, obtained by imposing that each of the constituent random variables has finite variance.

3.2 Theory

Our theoretical benchmark is given by a simple extension of the *information contagion* model by Arthur and Lane (1993). A large (possibly infinite) number of agents $i = 1, 2, \ldots$, are arranged on a line, the order of which is exogenously fixed and known to all. Agents enter sequentially into the books market, where they can choose one among a

continuously increasing number of alternative titles m = 1, 2, …, according to their internal representations – expressed in terms of probability distributions – of the characteristics or *quality* of books.

The true intrinsic cultural merit of a book, q_m, can be fully assessed only after the book itself has been read (experience good). However, purchasers can exploit all the available prior information to form beliefs on its worthiness. In particular, the quality of the generic book m may be gauged by means of public information on its performance characteristics, which takes the form of a normal prior probability distribution $N_m \sim N(\mu_m, \sigma_m^2)$. One can think of public information as advertising by the publisher as the book enters the market, or publicly available reviews published on newspapers or magazines. Furthermore, an agent can collect information by randomly sampling $n \le Z$ among the people who proceeded him in the row, asking them about the quality of the book they read. The signal i obtains from each of the n_m people who bought the book m, with $\sum_{m=1}^{M} n_m = n$ and M the total number of books already issued as the agent i enters the market, is:

$$X_m = q_m + \varepsilon \tag{4}$$

where ε is a random observational error, with $\varepsilon \sim N(0, \sigma_{ob}^2)$.

As the ith consumer enters the market, he has a probability p to choose a brand new title (M+1), on which no personal experience can be collected, while with probability (1 – p) he can switch to one old titles $m \in (1, M)$. In this last case, the agent i processes all the information he has collected to obtain a posterior distribution of qualities. Such a processing is performed by taking the convolution of the Gaussian integrals corresponding to the prior and to the information collected by other purchasers. The average expected posterior utility is then:

$$\mu_{post,m} = \frac{1}{n_m + \alpha_m} \left[n_m \mu_m^* + \alpha_m n_m \right]. \tag{5}$$

Thus, the posterior mean signal represents an average of different polled opinions, with a weighting factor that is inversely proportional to the variance of the distributions:

$$\alpha_m = \frac{\sigma_{ob}^2}{\sigma_m^2}. \tag{6}$$

If the new book is effectively chosen and the consumer has experienced its quality, the following period it is treated as an old book.

Each consumer is endowed with a constant absolute risk aversion utility function defined on the internal representations associated to the quality of the M issued books:

$$u(\mu_m) = \begin{cases} -\exp(-2\lambda\mu_m) & \text{if} \quad \lambda > 0 \\ \mu_m & \text{if} \quad \lambda = 0 \end{cases} \tag{7}$$

so that the objective function of the ith agent is to maximize a linear function of the mean and the variance of the posterior probability associated to the quality of the book m:[1]

$$u_m = \frac{1}{n_m + \alpha_m}\left(n_m\mu_m^* + \alpha_m n_m - \lambda\sigma_{ob}^2\right) \tag{8}$$

where the constant λ measures the degree of risk aversion: the larger λ is, the more risk averse the agent is. Upon computing u_m for each book in $(1, M)$, consumers choose the book with the highest expected utility.

The long-run dynamical properties of the choice-theoretic model (8) can be characterized in terms of Markovian processes. Let us assume that the sampling procedure is one with replacement, and define

$$T_j^m = \begin{cases} 1 & \text{if the } j^{th} \text{ purchase is } m \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = 1, 2, \dots$$

and

---

[1] See for example Sargent (1987, pp.154-155).

$$s_m(j) = \sum_{k=1}^{j} T_k^m .$$

Thus, making use of Lemma 4.1 in Arthur and Lane (1993), we introduce the conditional probability:

$$P\left[T_{n+1}^m = 1 \middle| T_1^m, ..., T_n^m\right] = f\left(\frac{s_m(n)}{n}\right) \tag{9}$$

where f is a polynomial of degree at most Z, whose coefficients depend on n, λ and μ_m, μ^*_m and σ_m for $m = 1, \ldots, M$. In other terms, the probability of a new purchase of the book m depends only on its current proportion. This implies that the process for a consumer who has switched to buying old books with probability $(1 - p)$ possesses an associated *urn function* (Arthur *et al.*, 1987)

$$f(x_m) = P\left[T_{n+1}^m = 1 \middle| \frac{s_m(n)}{n} = x_m\right] \tag{10}$$

and that the whole choice sequence with a fixed positive probability p to choose a brand new book each time a consumer enters the row is equivalent to an infinite Polya process, that is one in which the number of bins can grow without any bound.

The urn function turns out to be a complicated polynomial, whose coefficients depend on the parameters of the model and the sampling mechanism. Unfortunately, analytical solutions can be obtained explicitly for special combinations of parameters only. Nevertheless, Chung *et al.* (2003) prove that the limiting occupancy-rate distribution for an infinite Polya urn belongs to just one among three broadly defined classes. In particular, as we let the probability a new ball is placed in an existing urn (in our case, a new customer purchases an incumbent book) to be proportional to s_m^γ, with the parameter $\gamma \in \mathbf{R}$, Theorems 3.1, 4.1 and 4.2 in Chung *et al.* (2003) state that:

  i)   if $\gamma > 1$, one bin dominates;
  ii) if $\gamma = 1$, the limit probability distribution function associated to the random vector $(s_1, \ldots, s_M)$ satisfies:

$$P[S_m = s_m] \propto c s_m^{-(1+\alpha)} \qquad\qquad (11)$$

that is a power law distribution with $\alpha = \dfrac{1}{1-p}$, and c is a constant;

iii) if $-\infty < \gamma < 1$, the distribution of bin sizes decreases exponentially under rather mild conditions.

The three conditions correspond to different models of information transmission among peers. While the degenerate sales distribution of point *i*) emerges from a preferential attachment scheme with positive feedback, the presence of power law or exponential tails can be immediately interpreted in terms of linear interaction between buyers or of negative feedback due to congestion effects, respectively. The shape of the invariant market-share distribution, in particular along its tails, returns information useful to make inference on the underlying interaction mechanism.

This model helps to put on firmer choice-theoretic foundations the notion of *returns to information*, which has recently gained momentum in the empirical analysis of cultural markets (Walls, 1997; Maddison, 2005; Giles, 2007). In this literature, increasing returns to information – i.e., evidence of concavity in Zipf's plots due to autocorrelation in growth rates - are generally interpreted as a signature of *informational cascades* (Bikhchandani *et al.*, 1992), that is situations where it is optimal for agents choosing sequentially to ignore their own preferences and imitate the actions of the agents ahead of them. Our model suggests that when we replace the assumption of agents imitating the actions of others with that of agents exchanging information on the quality of the items, increasing returns to information should have far more drastic consequences than the mere appearance of concavity in Zipf's plots: they should lead instead to a degenerate asymptotic distribution, that is one superstar whose market share approaches 1. De Vany and Lee (2001) use agent-based simulations to reach exactly the same result. It turns out that the mechanism through which private information is transmitted from peer to peer – that is, weather it is passed through the observation of the action of others or through word-of-mouth – is crucial in making indirect inference on the degree of returns to information.

4. Data description and estimation results

The empirical analysis contained in this paper is based on proprietary data (Source: Demoskopea) on books' sales, organized as bi-monthly samples of unitary prices and number of volumes sold in Italy between January 1994 through December 1996. For each year we dispose of five bi-monthly samples only, given that during the traditional summer holiday period (July-August) data are not collected.[2] As a result, we can examine 15 book sales distributions (sub-samples) separately, as well as track their evolution in time. The dataset includes only books which succeeded in selling more than 600 copies in each two-month window. This means that only the upper tail of the whole book sales distribution is considered ($s_0 = 600$), and that the total number of titles in the tails is allowed to change period by period. Finally, individual observations refer exclusively to sales by traditional retailers. This is not likely to seriously affect the generality of our results, as non-traditional distributive channels (e.g., virtual bookshops or book clubs) were not very important during the sampled period. Incidentally, this continue to remain true for Italy even in more recent years, as Internet and book clubs sales are estimated to represent 1-1.5 and 4 percent of the whole market, respectively (Canoy *et al.*, 2006).

From an industrial organization perspective, the relevant market for differentiated goods is the one composed of products that are close demand or supply substitutes. In the case of books, close substitutability occur among titles sharing the same artistic canons. In order to preserve a sufficient number of observations, we choose the broadest possible definition of monopolistically competitive markets for books, classifying data into three very general genres: Italian novels, foreign novels and (non-novel) essays. Foreign titles are systematically translated in Italian, as in Italy the demand for foreign books in original language is so negligible to be considered null. As the most successful books and long-sellers are sometimes published in different editions (hardcover and paperback) and issued by several publishers (under multiple-license contracts), we pool multiple entries of the same title.

The final dataset includes a total of 4510 books written by 2294 authors. In particular, 1116 authors have written 1685 essays (with a minimum of 1 to a maximum of 28 books per author), 354 Italian novelists have written 764 novels (with a

---

[2] As pointed out by a referee, the lack of data for a period in which the demand is likely to be particularly high due to more spare time available by customers could bias our analysis, especially as regards the apparent seasonality of the scaling exponent α. This argument clearly reinforce the need to take our conclusions cautiously.

production ranging from 1 to 27 books per author) and 824 foreign novelists have produced 2061 books (with a production of 1 to 46 books per authors). The sample contains one the most successful Italian novel of the last thirty years, *Va dove ti Porta il Cuore* by Susanna Tamaro. Stephen King is the foreign novelist with the highest number of books translated, Isabel Allende is the most sold writer, while Ken Follett is the foreign author with the highest revenues. Unsurprisingly, an analysis of books by authors shows that the most successful authors are also the most prolific ones.

We estimate the power law characteristic parameter α in the linear regression model:

$$\text{Log}(i) = \beta - \alpha \, \text{Log}(s_i) + u \qquad (12)$$

by recurring to three alternative methods: 1) Robust (White heteroskedasticity adjusted) Ordinary Least Squares; 2) Hamilton robust regression (Hamilton 1991); 3) Quantile (median) regression.[3] Table 1 reports the estimated values of the coefficient α for the upper tail of the book sales distribution in terms of number of copies sold for the three segments of the market, for each period considered. All the three different methods seem to provide results that incorporate the same information.

Point estimates suggest that the values of α are always significantly different from 2 and, therefore, the estimated degree of uncertainty in the book publishing market is too high to be compatible with a Gaussian distribution. Recall that the lower is the estimated α, the heavier is the right tail of the empirical distribution, with the mean acting less and less as an attractor of the expected value. Revenues, and therefore profits, are largely dependent from few extreme events which can be hardly anticipated. Even if all segments are characterized by infinite variance, the degree of uncertainty can be nevertheless by ranked in terms of the relative value registered by the characteristic exponent α. Form this viewpoint, the segment of novels (in particular, Italian novels) turns out to be riskier than essays. Such a result holds true independently of the method used in estimations.

---

[3] Other possible methods are the Huber regression that refers to a bounded influence regression where observations are reweighed using criteria suggested by Huber (1964) and Mosteller and Tukey (1977); and the Trimmed least squares, which corresponds to a least squares method applied to the central deciles of the data.

Table 1. Estimates of the scaling exponent α for all three markets. *a*: White's robust OLS estimates; *b*: robust regression estimates (Hamilton); *c*: median regression estimates. All parameters statistically significant at the 5% level. The goodness of fit R^2 is higher than 0.94 in each case.

Sample	Italian novels			Foreign novels			Essays		
	a	*b*	*c*	*a*	*b*	*c*	*a*	*c*	*c*
94.1	1,39	1,12	1,31	1,38	1,34	1,36	1,32	1,25	1,27
94.2	1,33	1,14	1,18	1,34	1,23	1,27	1,33	1,32	1,33
94.3	1,21	1,05	1,09	1,51	1,46	1,46	1,37	1,29	1,32
94.5	1,04	1,15	1,12	1,33	1,42	1,41	1,13	1,26	1,23
94.6	0,95	0,99	0,98	1,11	1,09	1,09	1,01	1,07	1,09
95.1	1,07	1,06	1,06	1,2	1,2	1,21	1,35	1,31	1,34
95.2	1,17	1,18	1,16	1,19	1,16	1,18	1,29	1,26	1,27
95.3	1,18	1,05	1,12	1,28	1,25	1,26	1,39	1,1	1,16
95.5	1,06	1,03	1,06	1,07	1,07	1,06	1,16	1,14	1,16
95.6	1,01	0,93	0,95	0,91	1,02	1,04	1,11	1,06	1,08
96.1	1,13	1,13	1,12	1,12	1,03	1,06	1,4	1,26	1,31
96.2	1,2	1,19	1,18	1,15	1,03	1,05	1,44	1,28	1,33
96.3	1,26	1,14	1,18	1,19	1,14	1,17	1,45	1,53	1,5
96.5	1,15	1,12	1,12	1,2	1,09	1,11	1,3	1,25	1,26
96.6	0,98	0,89	0,91	1,1	1	1,07	1,09	0,97	1,01

It seems worthwhile to note that the estimated αs show a remarkable variation over time, which seems to follow a seasonal pattern: during the Christmas time, sales increase and seem to concentrate upon few titles on every market segment, making the business – if possible - even more risky. These seasonal effects is so strong to threat the existence of a mean value (see e.g. the Italian novel Nov.-Dec. samples for years 1994, 1995 and 1996).

As we aggregate sales according to the number of copies totally sold by each author (i.e., by pooling all his/her titles present in the list of most-sold books during a certain period), several considerations lead us to predict a higher market concentration (i.e., a lower α). First, bestseller writers are also the most prolific ones, so that their books have a higher probability to be in the market at any point in time, *ceteris paribus*. Second, only titles which have proven to be a hit abroad are translated in Italian. Third, as a newly issued title by a star-writer reaches a top position in the ranking for sales,

sales of previous books by that same author tend to increase as well due to bandwagon and reputation effects. Combining the three phenomena, we expect the author effect to be particularly strong for the novel sub-market and relatively weaker for essays, as the specialization of the subjects dealt with in essays makes superstar effects less likely.

The empirical evidence reported in Figure 2, where we compare the estimated αs for writers and books, seems to somehow support our a-priori expectations. It appears that the time series of the estimated αs for authors is systematically lower than that for books, meaning that the author effect is present in all sub-markets. Furthermore the scaling exponents calculated for novelists are in general lower than that for essay-writers. The uncertainty in the life cycle of writers (particularly novelists) is greater than the uncertainty surrounding the life-cycle of their books; focusing a catalogue on novelists, in turn, is more risky than publishing essay-writers.

This effect is confirmed when we compare scaling parameter estimates for books and authors as regards revenues, as reported in Table 2. Furthermore, Table 2 reports also estimates of the scaling parameters calculated on the top 10% of samples, given that theoretical arguments suggest that only for very large observations the true tail behaviour is detected, due to traditional estimates based on least squares regressions being sample-size sensitive (Borak *et al.*, 2005). Results for the whole distribution are always smaller in absolute value than that on the upper 10% tail, suggesting that predictions on the entire distribution (or larger tails) are systematically biased. However, even if sales on the far-right tails (i.e., bestsellers only) seems to be less concentrated, the theoretical variance is still infinite and the book market, characterized by *wild* fluctuations, proves to be extremely risky.

Similar results hold also for other creative industries. De Vany and Walls (2004), for example, report estimates of α in the range 1.3 to 1.7 for motion picture box office revenues in the U.S. and many other countries, but the value of scaling parameter shrinks to 0.4 when revenues of actors in their artistic lifetime are considered. By the same token, Rock&Roll performers are characterized by a more hazardously line of business (artists' concert revenues in the U.S. are power law distributed with $\alpha = 0.45$) than their promoters ($\alpha = 0.55$), who play the same role of publishers in the book industry (Connolly and Krueger, 2006). From this viewpoint, the good news emerging from our paper is that the career of a typical writer seems to be surrounded by lower risk than rock stars or actors, although uncertainty regarding success is still so high to be practically impossible to safely guess about it.
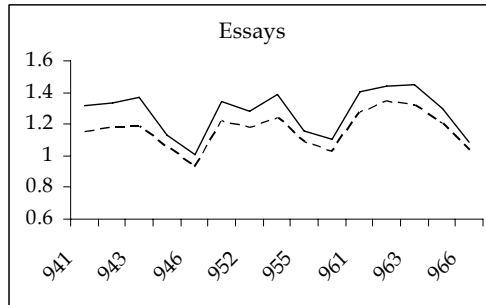
a)

Italian novels

b)

Foreign novels

c)

Essays
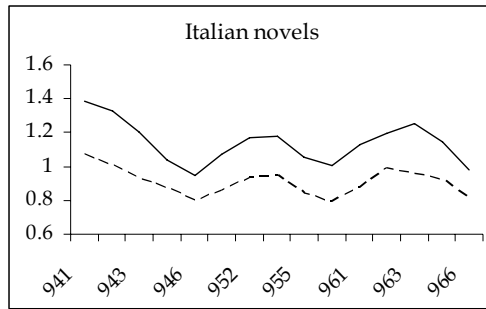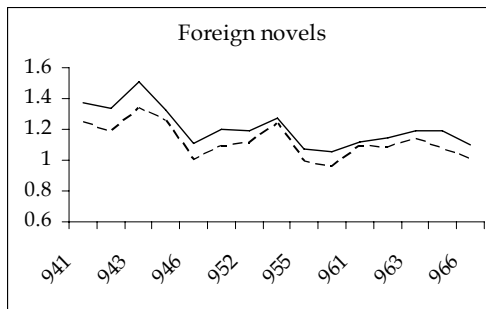
Figure 2. Estimates of α for writers (dashed line) and books (solid lines).

Table 2. Full sample and top 10% estimates of total sales, revenues, and sales by authors. *a*: White's robust OLS estimates; *b*: robust regression estimates (Hamilton); *c*: median regression estimates.

	Italian novels, Quantity, Titles				Foreign novels, Quantity, Titles				Essays, Quantity, Titles		
	a	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*
Total	0,75	0,61	0,64	Total	0,82	0,73	0,73	Total	0,96	0,73	0,84
10%	1,39	1,42	1,21	10%	1,32	1,17	1,08	10%	1,4	1,25	1,21

	Italian novels, Revenues, Titles				Foreign novels, Revenues, Titles				Essays, Revenues, Titles		
	a	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*
Total	0,62	0,54	0,59	Total	0,6	0,62	0,59	Total	0,68	0,67	0,67
10%	1,24	1,27	1,15	10%	1,25	1,14	0,95	10%	1,4	1,33	1,27

	Italian novels, Quantity, Authors				Foreign novels, Quantity, Authors				Essays, Quantity, Authors		
	a	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*
Total	0,65	0,53	0,55	Total	0,67	0,55	0,58	total	0,82	0,72	0,74
10%	1,09	0,97	0,82	10%	1,05	0,86	0,83	10%	1,36	1,36	1,26

	Italian novels, Revenues, Authors				Foreign novels, Revenues, Authors				Essays, Revenues, Authors		
	a	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*
Total	0,59	0,52	0,54	Total	0,58	0,5	0,52	Total	1,37	1,44	1,41
10%	1,07	0,94	0,9	10%	0,99	0,79	0,78	10%	0,71	0,69	0,81

As a final empirical exercise, in Table 3 we compare the average of the αs for quantities estimated through all bimesters, to the one obtained by pooling all data for each submarket over the whole period 1994-1996. It turns out that the pooled estimate is smaller than the average of bi-monthly observations, a result due to the correlation among successes and failures - captured by the pooled estimate, but not by the average over bi-monthly estimates - driven by the dynamics demand selection process. This is precisely what one should expect when facing a Lévy-stable industry. Conditional on having sell a certain number of copies in the past, the expected sales of a book continues to increase with current sales, and this continues as more copies are sell. In

other terms, a success or a failure can protract over time, making a success more successful and a failure more flopping.

Table 3. Average estimates *vs*. pooled estimates

	OLS	Rreg	MedR
Italian novels			
α	0,77	0,56	0,65
Average α	1,142	1,078	1,1027
Foreign novels			
α	0,85	0,73	0,76
Average α	1,2053	1,1687	1,1867
Essays			
α	0,97	0,75	0,85
Average α	1,276	1,2233	1,244

5. Conclusions

The book publishing industry is a suitable place to look at for inspecting the nexus between the way private information is transmitted from peer to peer and final market outcomes. The demand for a published piece of literature develops sequentially as readers discover whether they like it or not, and communicate their experience to other potential readers. The dynamic disclosure of preferences can generate both hits and flops through information bandwagon, although nobody knows ex-ante what makes a hit or a flop, or when it will happen. In spite of the extreme uncertainty surrounding the fate of a single title, some regularities at the market level can be usefully detected and exploited, however. In this paper we offer a simple model of information transmission and Bayesian updating which shows how alternative invariant (i.e., steady-state) market share distributions can be derived from assumptions on the degree of returns to information in markets characterized by word-of-mouth. In particular, if agents are allowed to communicate directly about the quality of an item, increasing returns to information ($\gamma > 1$) implies a long-run degenerate demand distribution, with one superstar approaching a market share of 1. If the flow of information on the quality of the book exhibits constant returns ($\gamma = 1$), in turn, demand distributes itself among items according to a Pareto distribution. Finally, an

exponential distribution emerges in the long-run when the transmission of information is bounded by decreasing returns ($\gamma < 1$).

These results are applied to a dataset reporting sales in a representative sample of Italian bookshops from 1994 to 1996. We invariably find that the demand distribution is Pareto, with a characteristic exponent ranging between 0.9 to 1.5. Since we do not find any significant departure from the Pareto law towards a degenerate distribution or an exponential distribution, we conclude that the information transmission in the Italian market for books is likely to be characterized by constant returns. The dynamics of the book demand distribution supports the *nobody knows*, the *winner takes all* and the *success breeds success* principles, all of them features that the publishing industry shares with other creative industries (DeVany and Walls, 1996, 1999, 2004). Furthermore, the degree of uncertainty surrounding the market for books follows a seasonal patterns, with Christmas being the most risky period of the year.

Two possible extensions of the analysis have been conceived but left for future research. First, alternative estimation methodologies will be employed to check for the robustness of our results. Second, we will make use of data on demand to control for the presence of economies of scale for publishers, along the lines developed in Sornette (2002).

References

Allen, W. and Curwen P. (1991), *Competition and Choice in the Publishing Industry*. London, Institute for Economic Affairs.

Associazione Italiana Editori (2005), *Rapporto sullo Stato dell'Editoria in Italia* - 2005. Milan, AIE.

Arthur, B. and D. Lane (1993), Information contagion, *Structural Change and Economic Dynamics*, **4**:81-104.

Arthur, B., Ermoliev, Y. and Y. Kaniovski (1987), Path-dependent processes and the emergence of macro-structure, *European Journal of Operational Research*, **30**:294-303.

Beck, J. (2007), The sales effect of word of mouth: a model for creative goods and estimates for novels, *Journal of Cultural Economics*, **31**:5-23.

Becker, G. (1991), A note on restaurant pricing and other examples of social influences on prices, *Journal of Political Economy*, **99**:1109-1116.

Bikhchandani, S., Hirshleifer, D. and I. Welch (1992), A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy*, **100**:992-1026.

Borak, S., W. Hardle, and R. Weran (2005), Stable distributions, *SFB 649, Discussion Paper 2005-008*.

Brock, W. and S. Durlauf (2000), Interactions-based models, in Heckman, J. and E. Leamer (eds), *Handbook of Econometrics*, Vol.5. Amsterdam, North-Holland.

Canoy, M., van Ours, J. and F. van der Ploeg (2006), The economics of books, in Ginsburgh, V. and D. Thorsby (eds), *Handbook of the Economics of Art and Culture*. Amsterdam, North-Holland.

Caves, R. (2000), *Creative Industries*. Cambridge, Harvard University Press.

Chung, F., Handjani, S. and D. Jungreis (2003), Generalizations of Polya's urn problem, *Annals of Combinatorics*, **7**:141-153.

Connolly, M. and D. Krueger (2006), Rocknonomics: the economics of popular music, in Ginsburgh, V. and D. Thorsby (eds), *Handbook of the Economics of Art and Culture*. Amsterdam, North-Holland.

De Vany, A. and C. Lee (2001), Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues, *Journal of Economic Dynamics and Control*, **25**:593-614.

De Vany, A. and D. Walls (1996), Bose-Einstein dynamics and adaptive contracting in the motion picture industry, *Economic Journal*, **106**:1493-1514.

DeVany, A. and Walls W.D. (1999), Uncertainty in the movie industry: Does star power reduce the terror of the box office? *Journal of Cultural Economics*, **23**:285-318.

De Vany, A. and D. Walls (2004), Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar, *Journal of Economic Dynamics and Control*, **28**:1035-1057.

Giles, D. (2007), Increasing returns to information in the US popular music industry, *Applied Economics Letters*, **14**:327-331.

Glaeser, E. and J. Scheinkman (2001), Measuring social interactions, in Durlauf S. and P. Young (eds), *Social Dynamics*. Washington, Brookings Institution Press.

Greco, A. (2000), Market concentration levels in the U.S. consumer book industry: 1995-1996, *Journal of Cultural Economics*, **24**:321-336.

Gnedenko, B. and A. Kolmogorov (1954), *Limit Distributions for Sums of Independent Random Variables*. Reading, Addison-Wesley.

Hamilton, L. (1991), How robust is robust regression?, *Stata Technical Bulletin*, **2**:21-26.

Hjort-Andersen, C. (2000), A model of the Danish book market, *Journal of Cultural Economics*, **24**:27-43.

Huber, P. (1964), Robust estimation of a location parameter, *The Annals of Mathematical Statistics*, **35**:73-101.

Ijiri, Y. and H. Simon (1964), Business firm growth and size, *American Economic Review*, **54**:77-89.

Ijiri, Y. and H. Simon (1974), Interpretations of departures from the Pareto curve firm-size distributions, *Journal of Political Economy*, **82**:315-332.

Lévy, P. (1925), *Calcul de Probabilités*. Paris, Gauthier Villars.

Maddison, D. (2005), Increasing returns to information and the survival of Broadway theatre productions, *Applied Economics Letters*, **11**:639-643.

Manski, C. (1993), Identification of endogenous social effects: the reflection problem, *Review of Economic Studies*, **60**:531-542.

Manski, C. (1997), Identification of anonymous endogenous interactions, in Arthur, A., Durlauf, S. and D. Lane (eds), *The Economy as an Evolving Complex System* II. Reading, Addison-Wesley.

Mosteller, F. and J. Turkey (1977), *Data Analysis and Regression*. Boston, Addison-Wesley.

Nolan, J. (2001), Maximum likelihood estimation and diagnostics for stable distributions, in Barndorff-Nielsen, O., Mikosch, T. and S. Resnick (eds), *Lévy Processes*. Boston, Birkhauser.

van der Ploeg, F. (2004), Beyond the dogma of the fixed book price agreement, *Journal of Cultural Economics*, **28**:1-20.

Sargent, T. (1987), *Macroeconomic Theory*. New York, Academic Press.

Sornette, D. (2002), Economy of scales in R&D with block-busters, *Quantitative Finance*, **2**:224-227.

Stanley, M., Buldyrev, S., Havlin, S., Mantegna, R., Salinger, M. and E. Stanley (1995), Zipf plots and the size distribution of firms, *Economics Letters*, **49**:453-457.

Szenberg, M. and E. Youngkoo Lee (1994), The structure of the American book publishing industry, *Journal of Cultural Economics*, **18**:313-322.

Walls, D. (1997), Increasing returns to information: evidence from the Hong Kong movie market, *Applied Economics Letters*, **4**:287-290.