WZB

WISSENSCHAFTSZENTRUM BERLIN
FÜR SOZIALFORSCHUNG

SOCIAL SCIENCE RESEARCH
CENTER BERLIN

Jonathan Beck *

Diderot's Rule

* Ludwig-Maximilians-Universität München

SP II 2008 – 13

April 2008

ISSN Nr. 0722 – 6748

Research Area
Markets and Politics

Research Unit
Competition and Innovation

Schwerpunkt II
Märkte und Politik

Abteilung
Wettbewerb und innovation

Zitierweise/Citation:

Jonathan Beck, **Diderot's Rule**, Discussion Paper SP II 2008 – 13, Wissenschaftszentrum Berlin, 2008.

Wissenschaftszentrum Berlin für Sozialforschung gGmbH,
Reichpietschufer 50, 10785 Berlin, Germany, Tel. (030) 2 54 91 – 0
Internet: www.wzb.eu

ABSTRACT

Diderot's Rule*

by Jonathan Beck

Like many new products, newly released creative goods such as books, music records and movies are sometimes 'surprise' hits but often flops. Experimental and empirical research suggests that it is hard to predict the demand for a new creative good, and therefore its success, even for industry experts. Rules of thumb on the quantitative properties of demand uncertainty exist for various creative industries – including a rule by Denis Diderot (1763) according to which one out of ten published books is a commercial success. Yet, representative evidence on any industry's new-product success rate is scarce. This paper studies new-product success in a random sample of novels. Its empirical strategy to identify success – a simple characterization of author-publisher bargaining combined with a parsimonious model of new-product diffusion – is based on the common observation that word-of-mouth is a crucial success factor in creative industries. Parametric and semi-parametric estimation results corroborate Diderot's rule: between 10 and 15% of novels enjoy significantly positive effects of word-of-mouth.

Keywords: *New-product success rate, demand uncertainty, word-of-mouth, creative industries*

JEL Classification: *D83, L82, Z1*

* E-mail: beck@wl.lmu.de. Mail: Institute for Communication Economics, Munich School of Management, Schackstr. 4/III, 80539 Munich. I thank Joe Clougherty, Michal Grajek, Ulrich Kamecke, Lars-Hendrik Röller and seminar participants at the Royal Economic Society Conference in Warwick, DIW, the London Business School and the WZB for helpful comments. In addition, this paper benefitted from various discussions I had with colleagues and other professionals during an editorial internship at W.W.Norton publishers in New York City. Of course, all remaining errors are mine. I am grateful to *Media Control GfK International* and Uwe Maisch for data provision, to the *Börsenverein des Deutschen Buchhandels* for permission to use it, and to the German Federal Ministry of Education and Research for financial support (project 01AK702A). A part of this paper was written during a visit to the London Business School, which I thank for its hospitality.

ZUSAMMENFASSUNG

Diderot’s Daumenregel

Neu veröffentlichte Kreativgüter wie Bücher, Musikalben und Filme sind, ähnlich anderen neuen Produkten, zwar manchmal "Überraschungserfolge", meistens jedoch Flops. Laut experimentellem und empirischem Forschungsstand ist es selbst für Branchenexperten schwierig, die Nachfrage nach einem neuen Kreativgut, und damit seinen kommerziellen Erfolg, vorherzusagen. Daumenregeln zu den quantitativen Eigenschaften dieser Nachfrageunsicherheit existieren in einigen kreativen Branchen -- unter anderem eine Regel von Denis Diderot (1763), wonach eines von zehn veröffentlichten Büchern ein kommerzieller Erfolg ist. Allerdings mangelt es an repräsentativer Evidenz zu der Erfolgsrate neuer Produkte, gleich in welcher Branche. Dieses Papier untersucht den Erfolg neuer Produkte in einer zufälligen Stichprobe von Romanen. Die verwendete empirische Strategie zur Identifikation von Erfolg -- eine einfache Charakterisierung der Verhandlungen zwischen Autor und Verlag, kombiniert mit einem überschaubaren Modell der Diffusion neuer Produkte -- basiert auf der verbreiteten Beobachtung, dass Mundpropaganda ein entscheidender Erfolgsfaktor in kreativen Branchen ist. Parametrische und semiparametrische Schätzergebnisse bestätigen Diderot's Daumenregel: zwischen 10 und 15% der Romane profitieren von einem signifikant positiven Einfluß von Mundpropaganda.

"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."

(Arnold, 2002)

1 Demand uncertainty and new-product success

A salient characteristic of markets for creative goods like theatrical movies, music records or novels is that demand – and therefore success – is extremely difficult to foresee. According to screenwriter William Goldman (1983), not even industry experts are able to predict a particular movie's box office performance. As a result, box office 'flops' are an empirical regularity: De Vany and Walls (2004) find that only 6.3% of all movies earn 80% of all movie industry profits. Similarly, the market for music records is said to be dominated by a "stiff ratio" – the share of loss-making recordings – of around 90% (Caves, 2000, p. 79; Denisoff and Schurk, 1986, p. 4).

The probably oldest quantification of demand uncertainty in a creative industry has been proposed by Denis Diderot in a 1763 article.[1] He estimates that at most one out of ten published books is a success while four recover costs in the long run and five end up with losses:[2] "*Ajoutez que, de compte fait, sur dix entreprises, il y en a une, et c'est beaucoup, qui réussit, quatre dont on recouvre ses frais à la longue, et cinq où l'on reste en perte*." Diderot's statement is often cited as a rule of thumb in the book trade.[3]

Caves (2000) claims that this high degree of demand uncertainty – what he calls 'the *nobody knows* property' – is peculiar to creative industries. Success of new industrial products has been said to follow the more general '80-20 rule' associated with the Pareto distribution, however, Crawford (1977, p. 51) finds "surprisingly little documentation for the frequent claim that 80% of all new products fail" and surveys a few studies that report failure rates between 30 and 90%. Naturally, the management literature focuses more on the determinants of new product success in an industry than on the success rate itself

[1]Denis Diderot (1713-1784) was a French philosopher, writer and editor-in-chief of the *Encyclopédie*, one of the first encyclopedias.

[2]The article is published in Diderot's collected works (Diderot, 1876); the above quote is from a more recent reprint (Diderot, 2003, p. 61). See Turnovsky (2003) for a review of the general reception of Diderot's article.

[3]For example by Escarpit (1969, p. 123), Tietzel (1995, p. 38) and von Lucius (2005, p. 66).

(see Goldenberg, Lehmann and Mazursky, 2001, and the references therein). Moreover, systematic evidence on the new-product success rate in any industry is scarce for two reasons: first, it is difficult to obtain a representative sample of products, and second, it is difficult to measure success.

This paper studies new-product success in a random and hence representative sample of novels. The release of a new novel as hardcover marks a clear moment of new-product introduction and ensures comparability of performance across titles. The first contribution of this paper is thus a quantification of demand uncertainty a creative industry. As the subject of Diderot's statement, the book industry is interesting in itself, but is is also an archetype for any creative industry. Moreover, since demand uncertainty serves as a building block for many theories that justify the implementation of vertical restraints, results are relevant to policy makers as well as managers. For example, producers of creative goods often allow retailers to return unsold items, but the profitability of this policy depends on the degree of demand uncertainty (see Cachon and Lariviere, 2005, and the references therein). In the model by Deneckere, Marvel and Peck (1997), resale price maintenance is always preferred to flexible prices by a monopolistic producer, but it can affect consumer surplus and thus social welfare positively or negatively depending on the extent of demand uncertainty.

The paper's second contribution is a methodological point concerning the measurement of success in industries in which consumption has a social component. Since cost data are often unavailable or unreliable, standard measures of new-product success rely on absolute sales or bestseller status of a product.[4] As I show below in more detail, this approach misrepresents the perspective of producers, because potent suppliers (ie. 'star' actors or authors) often demand a large share of the revenues that can be expected from the presence of predetermined success factors such as the star herself. For example, Elberse (2007) finds that the involvement of star actors increases expected revenue of producers (film studios), however, it does not increase their company valuation (expected profits).[5] In other words, the fraction of new products that were profitable to

[4]See Elberse, Eliashberg and Leenders (2006) for a review of the multidisciplinary literature that looks at the effect of academy awards, movie ratings, critic reviews or star participation on movie success.

[5]A stream of literature initiated by Rosen (1981) discusses whether extreme superstar earnings can be attributed to superior talent. Giles (2006) and the references therein provide empirical evidence on superstars. Giles (2006) also notes some of the issues in measuring success.

the producer does not equal the fraction of its new products whose sales exceeded some threshold.

Therefore, the empirical strategy to identify new-product success in this paper does not rely on total sales, but uses the common observation that word-of-mouth is a crucial success factor: "when a new creative good appears, social contacts transmit consumers' appraisals at a very low perceived cost to them, giving 'word-of-mouth' its importance for a creative good's ultimate success." (Caves, 2000, p. 173.). In the context of this paper, 'word-of-mouth' is a catch-all phrase for the diffusion of consumer awareness for a product and information about its quality. This includes person-to-person communication, but also (online) product reviews (Dellarocas, 2003) and less direct forms of communication such as blogs, bestseller lists and *Oprah's Book Club*.[6]

Evidence from a recent experiment by Salganik, Dodds and Watts (2006) shows that new-product success is basically unpredictable when consumption is subject to word-of-mouth (or what the authors call "social influence"). Participants in their study were offered to sample and then download previously unknown music. In some experiment groups, participants also received information about the number of a title's downloads by other participants. In a significant number of cases, titles that were sampled but rarely downloaded in a group where participants did not receive this information became 'bestsellers' in groups where participants did receive it; and *vice versa*. Salganik, Dodds and Watts (2006) conclude that "experts fail to predict successes not because they are incompetent [...] but because when individual decisions are subject to social influence, markets do not simply aggregate pre-existing individual preferences."

In what follows, I first consider a parsimonious model of new-product diffusion and word-of-mouth in the context of supplier-producer (author-publisher) contracting (section 2). In particular, I show that, if cost data are not available, measures of word-of-mouth approximate producer success better than nominal sales because the corresponding effects are hard to appropriate by suppliers in *ex ante* bargaining. The model directly leads to a parametric approach to identify the occurrence and effect of word-of-mouth in week-to-week variation in unit sales.[7] However, it also illustrates identification problems

[6]Chevalier and Mayzlin (2006) find that consumer reviews at Amazon.com significantly affect sales – measured by sales rank – relative to Barnesandnoble.com. Sorensen and Rasmussen (2004) study weekly scanner data on book sales to find that even a negative review in the *New York Times* increases the sales of the reviewed title, but not by as much as a positive one.

[7]Moul (2007), who quantifies the average effect of word-of-mouth in motion picture revenues, uses a similar strategy to identify word-of-mouth through intertemporal dynamics of weekly unit sales. However,

associated with more general models of word-of-mouth; therefore, I propose a semiparametric identification method to accommodate theoretical indeterminacies.

In section 3, I then present the data and discuss estimation results for both the parametric and the semiparametric approach. Results indicate that between 10 and 15% of titles enjoy positive word-of-mouth and thus seem to corroborate Diderot's rule. On average, titles that are estimated to enjoy positive word-of-mouth perform better in terms of total sales, however, some of these titles have relatively low sales whereas some of the sample's best sellers are not estimated to enjoy positive word-of-mouth. It is therefore possible that a success measure based on word-of-mouth leads to a different result than a standard measure based on nominal sales.

2 Identifying word-of-mouth and publisher success

As conceptual framework, I use a parsimonious model of new-product diffusion that features two essential ingredients to describe sales-effective word-of-mouth: heterogeneous buyers and intertemporal dynamics. In contrast, most studies in the extensive theoretical and empirical literature on new-product diffusion follow Bass (1969) in assuming that buyers are homogeneous regarding their propensity to buy and differ only in the timing of their purchase (see Van den Bulte and Joshi, 2007, for a review). Yet, word-of-mouth among homogeneous buyers can only affect the intertemporal dynamics of sales, not their overall level. To have an effect on overall sales, word-of-mouth needs to take place between buyers that are heterogeneous in their propensity to buy.

I consider the simplest case of heterogeneity: a two-segment structure, where the population of M potential buyers of a newly released creative good (henceforth "title") consists of two types: N_b buffs and N_c casuals (title subscripts omitted). I thus closely follow Caves' (2000, p. 173) observation that the "distribution of consumers between 'buffs' and 'casuals' strongly influences the organization of an art realm". Buffs buy the title in any case. Casuals only buy if they are exposed to positive word-of-mouth. If there is no positive word-of-mouth about a certain title, its long-run sales are restricted to N_b. The case of negative (sales-destructive) word-of-mouth is discussed in section 2.3.

the specific demand model underlying his analysis (nested logit) is very different from the new-product diffusion model used here, variants of which are widely applied in the marketing literature.

Heterogeneity-driven word-of-mouth has implications for the *ex ante* as well as the *ex post* view on title performance. First, it affects *ex ante* bargaining over contract terms between supplier (henceforth "author") and producer (henceforth "publisher"), and consequently title success from the publisher perspective. Second, it implies that *ex post* studies of aggregate (product-level) sales have an identification problem: the extent of word-of-mouth needs to be inferred by decomposing observed sales into unobserved sales to buffs and casuals. In section 2.2, I therefore consider a specification for intertemporal sales dynamics that can be used to parametrically identify the effect of word-of-mouth. In section 2.3, I discuss semiparametric identification in the context of more general models.

2.1 *Ex ante* bargaining and Diderot's rule

The standard contract between author and publisher grants the publisher the exclusive right to market the author's title. In its typical form, this publishing contract consists of a royalty scheme through which publisher and author share revenues from sold copies. Revenue-sharing is a standard contractual response to demand uncertainty: for example, Dana and Spier (2001) show that, when demand uncertainty realizes only after inventory decisions have been made, revenue sharing is valuable also in manufacturer-retailer contracts (which are outside the scope of this paper).[8]

The advance. An additional – and in our context more important – twist in author-publisher contracts comes in the form of a nonrefundable advance payment. The advance is often interpreted as a device to increase the publisher's *ex post* incentives to market a title, which relates to one of the problems associated with the infeasibility of profit-sharing (Caves, 2003), but the advance may serve other means as well. Hansmann and Kraakman (1992) consider the context of an early contracting stage, before the author has written the book, and study a 'hands-tying' contract where the advance helps publishers to commit to producing the title without detailed knowledge of its contents. Here, I abstract from this stage and interpret the advance generally as the lump-sum fee in a two-part tariff that is contracted upon under common knowledge of some forecast for a title's sales.

[8] Horvitz (1966) and a subsequent literature on economic issues in academic (textbook) publishing notes a number of implications of the royalty scheme for author and publisher marketing incentives. He also discusses in more detail why the seemingly more natural alternative of profit-sharing is rarely observed in practice.

To be more precise, consider the following formal example, denoting by r the author's revenue share ($0<r\leq 1$) and by A the advance on this share. Assuming fixed production costs of C and normalizing the title's wholesale price to one, the publisher's expected gross profits at the time of bargaining are

$$E[\pi] = E[Q] - A - rE[\max\{0; Q - A/r\}] - C, \tag{1}$$

where $E[Q]$ are expected sales. Since A is nonrefundable, it implicitly defines a threshold value for sales (A/r) below which the author's factual revenue share exceeds r. For any value of r, A may be used to appropriate the remaining expected publisher profits, depending on the author's bargaining power. Indeed, in practice royalty rates vary little across different contracts whereas advance payments tend to vary strongly, even in relation to the number of copies finally sold (Caves, 2000, pp. 56ff). Book authors are frequently represented by literary agents who usually earn between 10 and 20% of their client's remuneration and who thus have a direct incentive to achieve a high advance. In fact, agents often attempt to maximize the author's share of expected profits by auctioning publishing rights.[9]

The combination of royalty and advance implies that publisher profits are not monotone increasing in sales Q, but rather in their relation to expectations $E[Q]$.[10] In particular, this is true for titles whose authors have a strong bargaining position and are thus able to pocket much of the expected profits – presumably titles with high $E[Q]$. In the auction case with sufficiently many competing publishers, A will be close to $E[Q]$-C, such that the winning publisher's profits are close to zero in expected terms and positive *ex post* only if sales exceed their expectation.

Forecasting sales. From the publisher perspective, title success therefore depends on the accuracy of sales predictions, which, in turn, is related to word-of-mouth. Suppose a title's potential market has the two-segment buyer structure discussed above. Since buffs buy the title in any case but casuals buy only if there is positive word-of-mouth, expected

[9] An auction is the optimal selling format from the author viewpoint (Bulow and Klemperer, 1996) and has a long tradition in the book industry: see Moldovanu and Tietzel (1998) for an early example and Hansmann and Kraakman (1992) and De Vany and Walls (2004) for further anecdotal evidence.

[10] For authors, in contrast, total sales remain important *ex post* because they are associated with auxiliary revenues, for example from live performances or movie deals.

sales consist of

$$E[Q] = E[N_b] + Pr(word)E[N_c|word], \tag{2}$$

where $Pr(word)$ is the *ex ante* probability that a title receives positive word-of-mouth and $E[N_c|word]$ are expected sales to casuals in that case. The key question is whether it is possible, at the time of bargaining, to predict title-specific values for all of these components. Some predetermined observable characteristics – such as sales of previous titles by the author or the size of the author's fan club – are certainly informative regarding expected sales to buffs (N_b). Author-publisher bargaining is thus likely to operate under common knowledge of $E[N_b]$. In contrast, the findings by Salganik, Dodds and Watts (2006) indicate that predetermined characteristics are unlikely to contain information on the title-specific propensity to receive word-of-mouth ($Pr(word)$) and the corresponding additional sales ($E[N_c|word]$). In that case, parties can at best work with market-level statistics or general principles such as Diderot's rule. For example, suppose the average probability for the occurrence of word-of-mouth ($Pr(word)$) is δ and the average value for the resulting additional sales $E[N_c|word]$ is k times N_b. The maximum advance a publisher is willing to pay is then $\bar{A} = E[N_b](1 + \delta k) - C$.

It follows immediately that, for titles whose authors have strong bargaining power (hence $A \rightarrow \bar{A}$), publishers end up making profits only in case of word-of-mouth, that is, only with probability δ. For titles whose authors have weak bargaining power – presumably titles with low $E[N_b]$ – publishers may bargain down the advance payment. For these titles, however, production costs are relatively more important, which can also lead to negative *ex post* profits in case there is no word-of-mouth.[11]

Empirical implication. In either case, given that contract and cost data are typically not available or unreliable, differences in *ex ante* expectations and advance payments across titles cannot be accounted for empirically. In consequence, observed total sales are not an appropriate success measure from the publisher perspective. Empirical analysis may, however, utilize the fact that the more appropriate success measure, the difference between *ex ante* expectations and *ex post* sales, is particularly affected by the *ex post* extent of word-of-mouth. An approach to assess Diderot's rule empirically is thus to estimate the distribution of positive word-of-mouth across titles. The corresponding

[11] If the negotiated advance payment is $\rho\bar{A}$, where $0<\rho\leq 1$, publisher profits are negative *ex post* if there is no word-of-mouth and $C > E[N_b](1 - \frac{\rho}{1-\rho}\delta k)$.

interpretation of Diderot's rule holds that $\delta = \frac{1}{10}$: one out of ten titles enjoys positive word-of-mouth.[12] An implementation of this empirical approach requires (*i*) a representative sample of titles and (*ii*) a method to identify the presence of positive word-of-mouth in sales.

The following model of new-product diffusion illustrates that, with reasonable assumptions on how sales to buffs and eventually casuals distribute over time, it is possible to not only identify the existence of positive word-of-mouth but also to quantify its effect on a title's overall sales (that is, to estimate N_c). In section 2.3, I discuss more general models and semiparametric identification.

2.2 Intertemporal sales dynamics

As a matter of notation, it is more convenient to consider the total number of potential buyers of a title M and its share of buffs θ ($\frac{N_b}{M}$), instead of N_b and N_c. Time-invariant predetermined variables that may affect N_b and thus M, such as a title's characteristics and price, can be omitted in this section.[13] Operating within a continuous-time framework, denote by $F_b(t)$ the c.d.f. of a title's sales to buffs, that is, cumulative sales to this group at time t divided by its population (θM). Similarly, $F_c(t)$ is the c.d.f. of sales to casuals and $f_b(t)$ and $f_c(t)$ are the corresponding densities.[14]

Buffs buy the title in any case, however, not necessarily in its release week. For example, some may want to first finish the book they are currently reading. A standard assumption is that in every period the title is bought by a fraction p of those buffs who have not bought earlier. In continuous terms, this is a constant hazard rate: $p = \frac{f_b(t)}{1-F_b(t)}$. Since $F_b(0)$=0, we can solve for $F_b = 1 - \exp(-pt)$ and the cumulative number of sales to buffs at time t is

$$n_b(t) = \theta M - \theta M \exp(-pt). \tag{3}$$

In other words, period sales to buffs – first differences of $n_b(t)$ – follow the steady decay pattern typically observed, for example, for blockbuster movies. In terms of the model,

---

[12] An assessment of the part of Diderot's rule that distinguishes between loss making titles and titles that just break even, however, is difficult without cost data. Therefore, I focus on identifying successful titles.

[13] In most European countries, book prices are by law subject to resale price maintenance and thus invariant over time. Even in the unregulated U.S. market, intertemporal price variation is virtually inexistent (Clerides, 2002).

[14] The following specification was developed independently by Van den Bulte and Joshi (2007) and myself (Beck, 2007); in the former paper it is a special case of a more general class of models (discussed in more detail below).

aggregate sales exhibit steady decay in two cases: On the one hand, the number of casuals may be zero (θ=1). On the other hand, θ may be below one but there is no word-of-mouth such that overall sales are limited to N_b and sales dynamics are determined by equation 3. The upper left title in figure 1 provides an example for sales dynamics without word-of-mouth. An important implication for empirical work is that in this case, θ and M are not separately identified. In other words, it is impossible to say how much a title would have sold if it had received some word-of-mouth. As I discuss below, a related identification problem is associated with the possibility of negative word-of-mouth.

Casuals. As long as θ<1 the title under consideration has the potential to benefit from word-of-mouth. In particular, an independent buyer may recommend the product to w ($\geq$0) casuals each period following her purchase. Parameter w can be interpreted as a population average; for example, w=.5 means that one out of two buffs recommends the title to a casual each period after her purchase.[15] The contacted casuals then go ahead and buy the recommended title, unless they have not already done so in response to an earlier recommendation. Since the probability that a casual exposed to word-of-mouth at time t has not been contacted and therefore has not bought earlier is $1 - F_c(t)$, the cumulative number of sales-effective recommendations at time t is $1 - F_c(t)$ multiplied by w and $n_b(t)$. Divided by the overall number of casuals, $(1 - \theta)M$, this amounts to the density $f_c(t)$ and rearranging yields the relationship

$$\frac{f_c(t)}{1 - F_c(t)} = qF_b(t), \tag{4}$$

where $F_b(t) = \frac{n_b(t)}{\theta M}$ and $q = w\frac{\theta}{1-\theta}$ is a reparametrization convenient in empirical applications. Using equation 3 and the fact that $F_c(0)$=0, this differential equation solves for

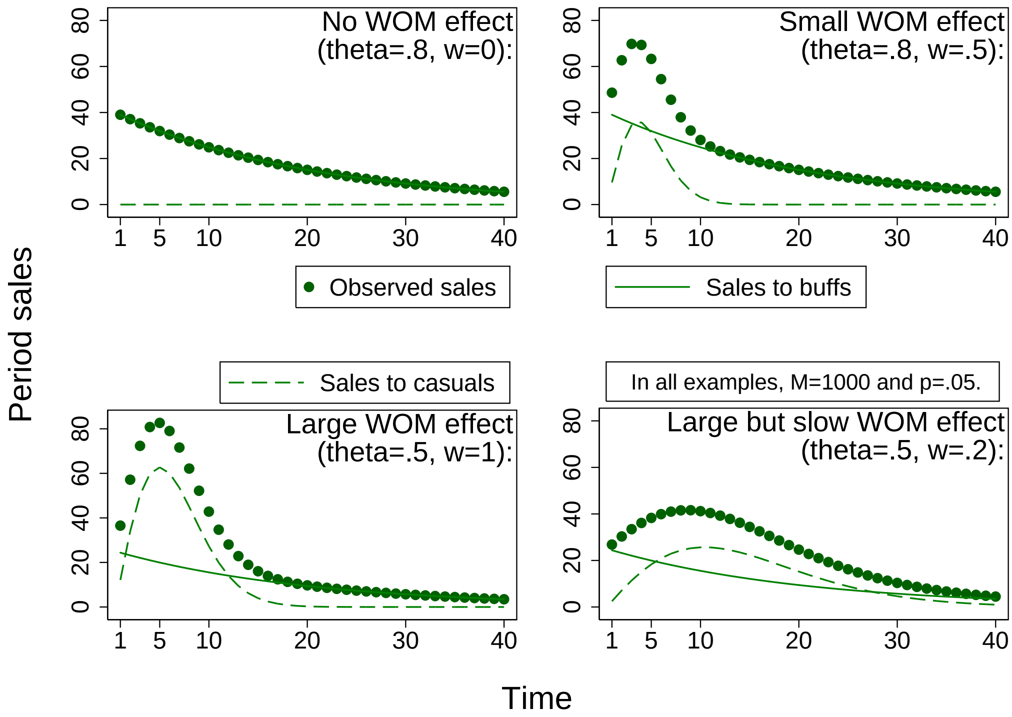
$$F_c(t) = 1 - e^{\frac{q}{p}(1-e^{-pt}-pt)}. \tag{5}$$

Hence, cumulative sales to casuals are $n_c(t) = (1 - \theta)MF_c(t)$, and total cumulative sales at time t are the sum of $n_b(t)$ and $n_c(t)$:

$$N(t) = M(1 - \theta e^{-pt} - (1 - \theta)e^{\frac{q}{p}(1-e^{-pt}-pt)}). \tag{6}$$

[15] A simplifying assumption discussed in more detail below is that casuals do not recommend the title to other people.

Figure 1: Example sales patterns



Period sales within the interval $(t,t-1)$ are then described by $N(t) - N(t-1)$. Figure 1 plots the corresponding sales pattern for four different value combinations of parameters θ and w. Most notably, sales without word-of-mouth follow a convex pattern, whereas with word-of-mouth they are concave (hump-shaped) in early sales weeks: due to an increasing number of buff buyers engaging in word-of-mouth, sales to casuals rise initially. The size and width of the resulting sales hump depends on the intensity of word-of-mouth and on the relative number of casuals.

2.3 Generalizations

Equation 6 provides a functional form that can be used to structurally identify the extent of word-of-mouth in title-specific time series of sales. However, the corresponding results depend on the viability of the model, which can be restrictive in a number of circumstances. First, it only considers positive word-of-mouth, although anecdotal evidence from the movie industry suggests that negative word-of-mouth can also be important. In terms of the model, negative word-of-mouth would not only imply that w=0, but in addition that buffs engage in sales-destructive communication among each other. Negative word-of-mouth aggravates the identification problem already present when w=0; for example, among buff buyers it implies that their overall number (θM) is subject to decay over time: first-week buff buyers dissuade other buffs from their initially planned purchase. As evident from equation 3, it is impossible to distinguish this effect from the hazard rate p. For example, consider the upper left title in figure 1, whose observed sales decline steadily after release. Based on such data, it is impossible to say whether the particular pattern is driven by negative word-of-mouth (decreasing θM over time) or merely by the decay parameter p. Therefore, in order to identify negative word-of-mouth among buffs, more restrictions on parameter p are needed, for example by assuming it to be equal across titles.

Second, in the above model casuals care exclusively for recommendations from buffs. In general, however, positive word-of-mouth within groups may also be sales-effective. Van den Bulte and Joshi (2007) analyze a more general class of models, which allows for positive word-of-mouth within both segments, and derive closed-form solutions. These more general cases have the property that period sales exhibit a 'dip' or are convex for early sales periods, when decreasing sales to buffs are not yet over-compensated by increasing sales due to word-of-mouth. In other words, a hump-shaped sales curve driven by word-of-mouth can have more than one stationary point and only one of these is the global maximum. Van den Bulte and Joshi (2007) present evidence on sales of music CDs that seem to exhibit such a 'dip' in early sales weeks. Since models of new-product diffusion are inherently nonlinear, model complexity increases exponentially for these more general cases. Indeed, estimation based on standard methods can be troublesome, which complicates comparison between model variants. Nevertheless, all cases share the property that word-of-mouth leads to a concave (hump-shaped) pattern around the global maximum of a title's sales curve. As I discuss in more detail below, semiparametric identification of word-of-mouth relies on this property.

Alternative explanations. A related economics literature on the diffusion of new technologies generally explains a hump-shaped pattern in adoptions, which is inherent in an S-shaped cumulative adoption curve, by either decreasing costs or increasing benefits of adoption over time (Hall and Khan, 2003). For a new creative good, however, both price and material product characteristics are usually constant within its short selling period. Therefore, the only possible kind of increase in benefit is one associated with immaterial characteristics: for example, consumers may change their perception of the intellectual value of a title. This view is generally consistent with the occurrence of positive word-of-mouth. Similarly, word-of-mouth can be regarded as a variant of explanations based on information diffusion (Jensen, 1988). Other, 'behavioral' explanations for hump-shaped sales often assume some form of suboptimal behavior by producers or retailers, which is outside the scope of this paper. For example, a publisher may gradually increase marketing efforts for some titles. At least in the book industry, however, traditional instruments such as advertising seem to have modest effects. Instead, industry sources emphasize the effects of public performances by a title's author, which are often out of direct publisher control.[16]

2.4 Parametric identification

The standard approach to estimating parametric models of new-product diffusion with aggregate data is based on period sales, that is, first differences of cumulative sales (Putsis and Srinivasan, 2000) :

$$S(t) = N(t;\phi) - N(t-1;\phi) + \varepsilon_t, \tag{7}$$

where $S(t)$ denotes observed sales of a given title during the period ($t-1$,t), $N(.)$ is the cumulative sales function defined by the particular model, ϕ is the set of model parameters and t=1,...,T. Here, $N(.)$ is defined by equation 6 and ϕ consists of M, θ, p and q. Assuming that the error term ε has the usual least squares properties, parameters may be estimated by nonlinear least squares (NLS). A grid search procedure yields proper initial values for iterative estimation. Through log-transformations, I impose non-negativity for all parameters, and θ, $p \leq 1$.

Neither asymptotic nor small-sample properties of such estimators are known (Boswijk and Franses, 2005), but bias and consistency can be studied by means of a a Monte Carlo simulation. For the present model, NLS estimates are reliable if observations cover a sufficiently large part of a title's life cycle and are not too volatile (Beck, 2007). Furthermore, residual autocorrelation may be present: for example, a television appearance by a title's author may boost sales not only in that but also in the following weeks. The

[16]"Booksellers say author tours, Oprah most effective for marketing books", *Book Publishing Report*, vol. 24, iss. 38; "Suche nach Öffentlichkeit", *Handelsblatt*, iss. 54, 16 March 2006.

procedure to test and account for residual autocorrelation is straightforward (see Beck, 2006b, for more details).

Applied to a title with hump-shaped sales, the model provides estimates for all four parameters and thus identifies both the existence of word-of-mouth (w) and its relative sales effect ($1 - \theta$). Based on a representative sample of titles, this approach therefore yields an estimate of the distribution of positive word-of-mouth in the market – the share of titles with a positive w – and thus allows for an assessment of Diderot's rule.

Yet, such a parametric approach is associated with two main problems. First, equation 3 indicates that the parameters of the model are not identified when sales of a title are steadily decreasing over time. Effectively, estimates for such titles tend to converge at boundary values ($\hat{p}$ or $\hat{\theta}$ equal to zero or one, $\hat{w}$ equal to zero). Results of this kind may indicate that the respective title did not enjoy positive word-of-mouth and hence that w=0, but they may also be driven by data volatility (Beck, 2006b). Second, and more importantly, parametric estimates are meaningful only if the imposed model is a good approximation of the data generating process. If more general forms of word-of-mouth cannot be excluded or if data volatility complicates estimation, semiparametric methods provide useful alternatives to test for the existence of positive word-of-mouth.

2.5 Semiparametric identification

A more general empirical specification for intertemporal sales dynamics is

$$S(t) = g(t) + \varepsilon_t, \tag{8}$$

where $g(t)$ denotes the unknown function according to which period sales distribute over time. Various methods are available to semiparametrically estimate $g(t)$ in order to obtain a smoothed time series $\hat{S}(t)$. In principle, these methods can be viewed as variants of kernel density estimation that differ mainly in the employed kernel and the degree of smoothing. Results are typically invariant to the researcher's choice of kernel but highly sensitive to the chosen bandwidth (Cameron and Trivedi, 2005). Further, a crucial distinction is between global estimators that search for a function that fits best over all available data, and local estimators that smooth over a moving data window

(Ruppert, Wand and Carroll, 2003). Estimators also differ in their treatment of extreme observations ('outliers').

For the purposes of this paper, the locally weighted regression approach (*loess*, following Cleveland, 1979) seems most appropriate. First, as a local estimator it does not exhibit irregularities at the beginning or end of the sample that have been found with global estimators. This is important for the present application because word-of-mouth driven sales humps tend to occur in early sales weeks. Second, the *loess* approach has a high degree of automation, which facilitates application to a large number of titles: only one smoothing parameter has to be chosen and procedures exist to automate even this decision; in particular, I will use the improved Akaike information criterion developed by Hurvich, Simonoff and Tsai (1998). Finally, by iterative reweighting of observations the *loess* estimate is robust to extreme observations, which due to events like television appearances of authors are likely to occur in data of book sales.

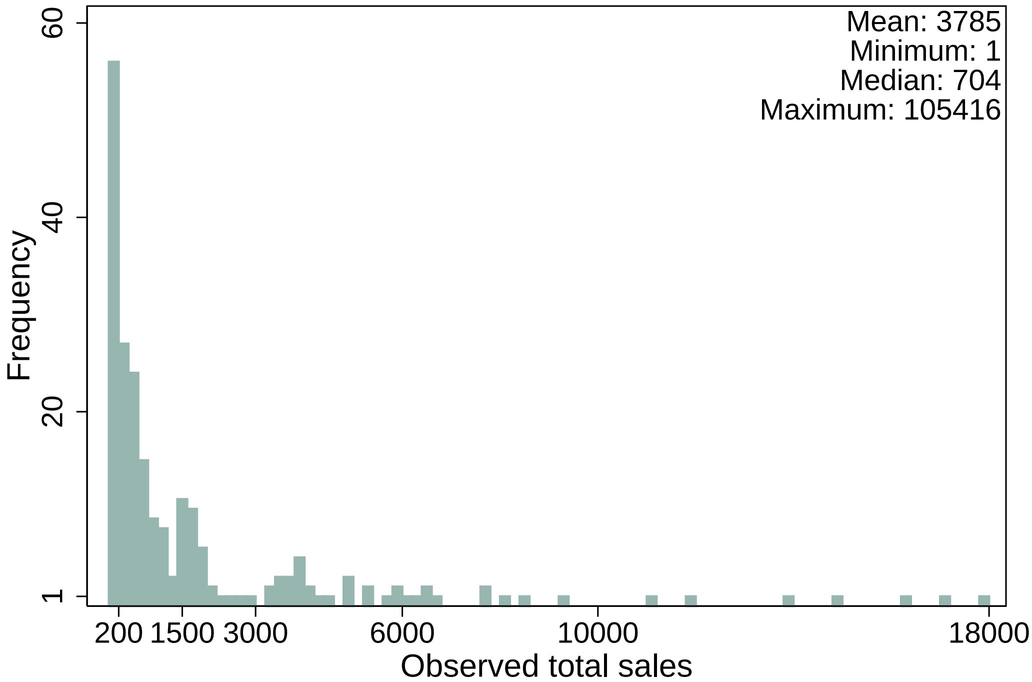
3 Empirical results and Diderot's rule

The data used in this study is a sample of 229 novels randomly drawn from the set of all novels released 2003 as hardcover in Germany. The data come from a marketing research firm that aggregates scanner data from over 750 physical points of sale and all main internet retailers in Germany. In the appendix, I discuss data characteristics and sampling procedure in more detail.

I focus on novels because this segment is most important for the book trade – both in economic and cultural terms – and on hardcover editions because only newly released titles are of interest; in Germany, the paperback edition of a title is delayed, typically by one to two years.[17] As demanded by the data proprietor, I received anonymized data, where all title-, author- and publisher-specific information except for a title and publisher code, sales (by week) and price (constant) had been removed. The sample period ends in summer 2004, providing between 41 and 81 weekly sales observations per title.

[17] This release strategy is a textbook example of intertemporal price discrimination (Clerides, 2002).

Figure 2: Distribution of sales across titles



Note: graph omits 12 titles with observed total sales of over 20000 copies.

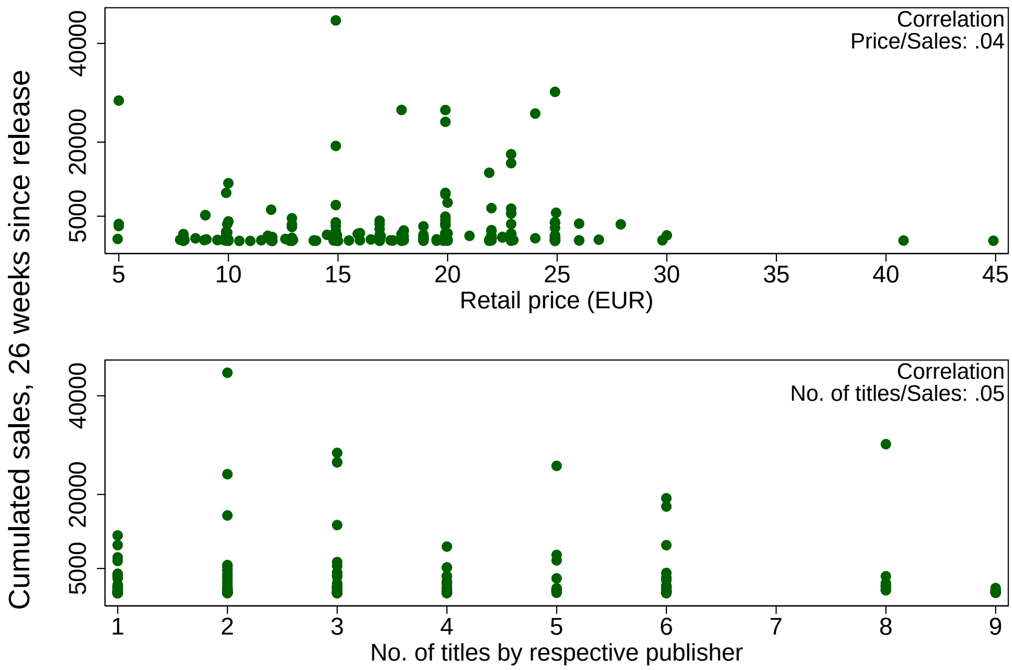
3.1 Sample characteristics

Figure 2 presents a histogram of total sales observed across titles. It does not account for the fact that titles are observed for differently long time periods, however, this turns out to be inessential: most sales take place within the first 26 weeks after release (observed for all titles).[18] The result is a well-known picture: most titles have very low overall sales – about 43% of titles sell less than 500 copies – and only few titles get to five- or six-digit sales figures. In effect the distribution of total sales is skewed: whereas the best sellers drive the sample mean up to 3785 copies, the median title sells 704 copies only. Sorensen (2007) presents a similar graph based on U.S. data.

Bivariate relationships. The data lack detailed information on title characteristics, but one might expect a title's retail price to proxy for characteristics like author reputation or the number of pages (see Beck, 2006a, for evidence from a comprehensive data set of prices and title characteristics). Yet, in aggregate terms there does not seem to be a systematic relationship between a title's sales and its price: figure 3 relates cumulated

[18]Figure 6 in the appendix gives standard kernel density estimates for the distribution of cumulated sales: one including the first 26 sales weeks only and one including all observed weeks for each title. The distribution for sales including all observed weeks is quite similar and only slightly broader than the one including the first 26 weeks only.

Figure 3: No evident relation between sales and price or publisher size



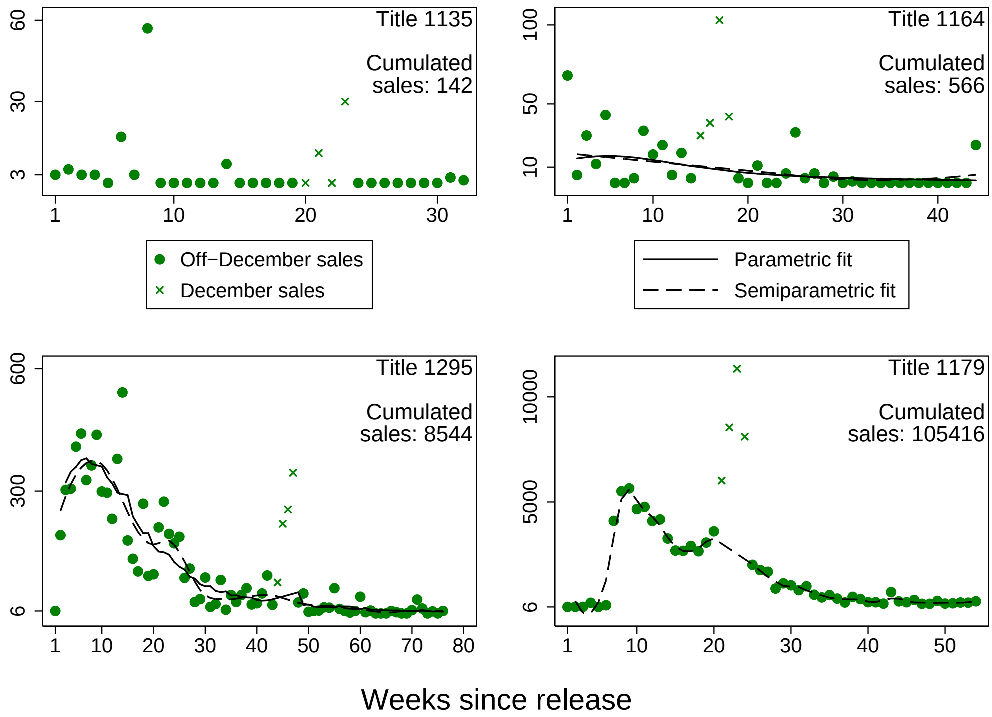
Note: graph omits one title with 26−week cumulated sales of 91939 copies and a price of 19.

sales after 26 weeks to price by title and gives the correlation coefficient, which is close to zero. The second panel in figure 3 relates sales to the number of titles a publisher has in the sample, a measure of firm size constructed from the publisher code. For the median title the measure is 1, but a number of publishers have multiple titles in the sample. There seems to be no correlation between this measure of publisher size and title sales.

Another consequence of title anonymity is that it is difficult to determine exact release dates. Usually, the earliest week with positive sales corresponds to the release week, but for some titles the raw data seem to contain erroneously booked advance orders; therefore, I use a systematic procedure to determine the effective release week (see section A.1 in the appendix for more detail). In any case, first- and second-week sales figures are not comparable because in contrast to theatrical movies, new books do not have a particular weekday for release. One title may be shipped on a Monday and another on a Friday, leaving only one or two sales days for the latter title's first calendar week. Therefore, I omit first-week observations in estimations.

Intertemporal dynamics. The primary information contained in the sample are title-specific dynamics that underly week-to-week variation in sales. To illustrate both the

Figure 4: Example titles



variety of patterns observed in the sample as well as some of the estimation issues, figure 4 presents four example titles. First of all, a significant share of the sampled titles, like title 1135 in the upper left panel of figure 4, has low overall sales and therefore zero sales in many weeks. For such a title, neither econometric method will yield useful results based on week-to-week variation. Altogether, 51 titles (22.2% of the sample) have less than 13 positive off-December sales observations before they reach 95% of cumulative sales (which range from 1 to 796 with a mean of 116). I assume that these titles have not received positive word-of-mouth and I do not attempt to estimate any other parameter econometrically for these titles.

For a number of titles, hump-shaped sales patterns such as those in the lower two panels of figure 4 suggest the existence of word-of-mouth effects. For other titles, such as title 1164 in in the upper right panel, sales variance is relatively high and it is difficult to infer a particular pattern by mere observation. Therefore, figure 4 already includes predicted values from parametric and semiparametric estimations. These predictions turn out similar for titles 1164 (upper right) and 1295 (lower left): the smoothed series of title 1164 decrease quite constantly over time, while both methods indicate an early hump in sales for title 1295. In contrast, sales of title 1179 – which has the highest overall sales in the sample – remain at low levels initially and are hump-shaped only in later weeks. Parametric estimation is troublesome in such a case because, in this particular model of positive word-of-mouth, the sales curve increases right from the start and has at most one stationary point. Therefore, in cases with multiple stationary points such as title 1179, NLS estimation based on equation 6 exhibits converges problems or converges only at boundary estimates for p and θ. Locally weighted regression, instead, is more adaptive to multiple stationary points and can thus identify patterns consistent with more general models of new-product diffusion.

Christmas sales. Another pattern evident in all four examples is that December observations tend to depart quite starkly from whatever trend sales follow before and after December.[19] Obviously, this is driven by the fact that books are popular Christmas presents, an effect that introduces an additional identification problem with December observations. For the main purpose of this paper, it suffices to merely acknowledge that the Christmas effect may lead December observations to deviate (positively) from a title's particular sales pattern before and after December; in practical terms, this amounts to placing zero weight to December observations in estimation. More details on estimation and interpretation of the Christmas effect can be found in section A.2 of the appendix. There, I also present some regression results which suggest, on the one hand, that additional Christmas-driven sales do not have significant second-order effects on post-Christmas sales, and on the other hand, that any potential effect of strategic pre-Christmas release timing by publishers seems to be of minor importance.

3.2 Estimation results

I apply the parametric and semiparametric estimators to all titles in the sample that have at least 13 positive sales observations before they reach 95% of observed cumulative sales (178 titles). Due to the identification problem discussed in section 2.2, parametric estimation based on equation 6 yields degenerate results for a title whose sales pattern is best represented by a monotone decline or does not converge for a title whose sales pattern

[19]No other significant seasonal variation seems to be present. In a panel regression specification following Sorensen (2007), where $S_{i\tau} = (\alpha_i + \alpha_\tau + \beta t_{i\tau})S_{i\tau-1} + \varepsilon_{i\tau}$, τ denotes calendar weeks and $t_{i\tau}$ denotes title i's weeks since release at week τ, all off-December week fixed effects α_τ are insignificant.

Table 1: Distribution of parametric estimates*

	Mean	Minimum	Median	Maximum
Directly estimated:				
$M_i/N_i(T)$	1.1	.901	1.08	1.56
p_i	.103	.0118	.0541	.894
θ_i	.552	.0121	.606	.882
q_i	.914	.0213	.177	22
Indirectly estimated:				
w_i	.529	.0121	.172	10.3

*Summary statistics for 59 title-specific NLS results, based on equation 7 including time dummies for December observations (see section A.2 in the appendix).
For 12 of these titles, estimates base on an adapted version of equation 7 that includes AR(1) errors.

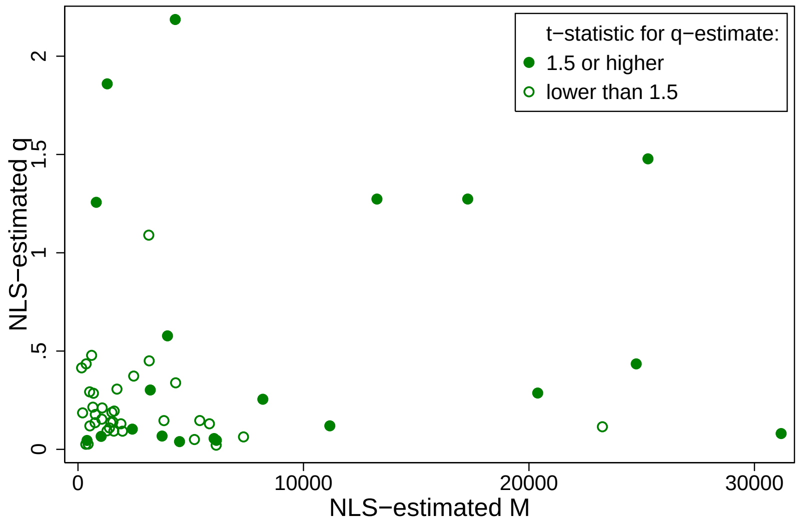
requires a more general diffusion model.[20] Parametric estimation converges and yields nondegenerate results for 59 titles in the sample.

The parametric test for the presence of positive word-of-mouth in sales of title i corresponds to a test for the significance of coefficient estimate $\hat{q}_i$. Within the set of nondegenerate results, $\hat{q}$ is significantly different from zero with 95% confidence for 23 titles. In other words, the parametrically estimated share of titles that received positive word-of-mouth is about 10%, a result that corroborates Diderot's rule. Table 1 summarizes the corresponding coefficient estimates. To facilitate comparison across titles, $\hat{M}_i$ is summarized relative to observed total sales $N_i(T)$.[21] The estimated hazard rate for sales to buffs ($\hat{p}_i$) is .1 on average, however, this average seems to be driven by some titles with high estimates. Mean and median estimates for the share of buff buyers are closer to each other (.55 and .61, respectively) and indicate that buff buyers tend to make up for the majority of sales also for titles with positive word-of-mouth. Figure 5 provides more details on the distribution of q- and M-estimates across titles. Most q-estimates are below .5 and whereas the few larger estimates are almost all significant, a good share of the lower estimates is significant as well. Altogether, across titles that seem to have enjoyed some word-of-mouth there is no evident relationship between its intensity (as measured by q) and overall sales (as measured by M).

[20] I classify converged estimates as degenerate if they fulfill at least one of the following conditions: (i) $\hat{p}$ is smaller than .01 and not significantly different from zero, (ii) $\hat{\theta}$ is smaller than .01 or larger than .99 and not significantly different from zero or one (both with 95% confidence).

[21] Since $\hat{M}_i$ does not include Christmas sales, it can be lower than $N_i(T)$ for titles with a large Christmas share.

Figure 5: Significance of word-of-mouth estimates and correlation with overall sales



Note: graph omits two titles with q−estimates larger than 10 (one insignificant, both M−estimates below 6000) and two titles with M−estimates larger than 32000 (q−estimates significant and below one).

The semiparametric test for the existence of word-of-mouth is directly based on the estimated shape of the sales curve. In particular, it draws on the theoretical result that positive word-of-mouth leads to a concave sales pattern around a title's peak sales. The proper semiparametric test corresponding to a test for q=0 in the parametric case, therefore, is a test for convexity of the sales pattern around peak (maximum) sales. In particular, I reject local convexity if *loess*-predicted peak sales are significantly greater than predicted sales in week 2; that is, if the 95% confidence intervals around these two predictions do not overlap. For example, the *loess* fit for sales of title 1164 in figure 4 is steadily decreasing. Predicted values and confidence intervals for peak sales and sales in week 2 are thus equivalent and convexity cannot be rejected. For titles 1295 and 1179, this test rejects convexity and thus indicates the existence of positive word-of-mouth. Altogether, the semiparametric test rejects convexity for 34 titles and hence suggests that less than 15% of all titles in the sample received positive word-of-mouth, which is also close to Diderot's prediction.

Table 2 compares the distribution of total sales across the identified subsamples of titles with and without indication of word-of-mouth. For both methods of identification, subsample averages of observed total sales are significantly different (t-test, 99% confidence). Titles with signs of positive word-of-mouth indeed perform better in terms of total sales, however, some of these titles have relatively low sales whereas some of the

Table 2: Distribution of total sales across titles with and without word-of-mouth

	No. of titles	Observed total sales Mean	Min.	Median	Max.	Price Mean	PubSize* Mean
Full sample	229	3785.0	1	704	105416	17.40	3.1
Estimation sample	178	4836.2	87	1250.5	105416	17.71	3.4
Parametric test for convex sales (95% confidence):							
not rejected	155	3743.6	87	1022	105416	17.66	3.3
rejected	23	12199.6	354	5911	43200	18.11	3.9
Semiparametric test for (locally) convex sales (95% confidence):							
not rejected	144	3561.6	87	1139.5	43200	17.83	3.3
rejected	34	10234.4	218	1983.5	105416	17.21	3.8

*Number of sampled titles published by the respective publisher.

sample's best sellers are not estimated to enjoy positive word-of-mouth. The two success measures nominal sales and word-of-mouth can therefore lead to contrasting findings. In other words, a success measure based on word of mouth can lead to a different result than a measure based on sales alone, although the two measures are likely to be correlated.

Table 2 also summarizes retail prices and a measure of publisher size (the number of sampled titles published by the respective publisher) for each subsample. For both methods of identification, both average retail prices and publisher size do not differ significantly across titles with and without identified positive word-of-mouth (t-test, 95% confidence). These findings suggest that predetermined title or publisher characteristics are of little help in predicting the occurrence of word-of-mouth.

4 Concluding remarks

Many new products fail commercially, however, this statement holds to different degrees in different industries. Creative industries seem to have especially low new-product success rates, and professionals in these industries share rules of thumb on the extent of demand uncertainty. For example, an old rule going back to Denis Diderot (1763) states that one out of ten published books is a commercial success. Representative evidence on the new-product success rate, however, is scarce for any industry.

In this paper, I study new-product success in a random sample of novels. My empirical strategy to identify a successful title is based on a parsimonious model of new-product diffusion and author-publisher bargaining. I show that indicators of word-of-mouth measure success better than standard measures of total sales and apply both a parametric and a semiparametric estimation method to identify the existence and extent of positive word-of-mouth based on a title's week-to-week variation in sales. Estimation results indicate that between 10 and 15% of titles enjoy positive word-of-mouth and thus corroborate Diderot's rule.

Measures of positive word-of-mouth are imperfectly correlated with a title's total sales and are not correlated with predetermined title characteristics such as a title's retail price. Although based on limited data on title characteristics, these findings support the view that the title-specific extent of word-of-mouth is extremely difficult to foresee and that choice of measure is important in studying new-product success. Based on more comprehensive data such as author-, producer- and eventually consumer-specific information, future research will be able to better understand *how* word-of-mouth affects success of new products whose consumption has a social element.

A Appendix

A.1 Data characteristics

The results in this paper are based on a sample of novels released 2003 as hardcover in Germany. This is a representative sample that I drew myself on location at the data provider (*Media Control GfK International*) from the set of all hardcover novels released in 2003, using a list of computer-generated random numbers. *Media Control GfK International* aggregates scanner data from over 750 points of sale (bookshops, department) as well as all main internet retailers in Germany. Not sampled are direct sales from publishers to consumers, book club sales and mail order sales. Supermarket sales are also not sampled, but they represent a negligible portion of book sales. Altogether the sampled retail channels account for about 66% of total book sales in Germany.

Yet, for the particular segment studied here (novels in hardcover), sales coverage of sampled channels is likely to be much higher. First, publisher direct sales are not very important for popular publications such as novels; direct sales usually concern professional publications. Second, book clubs can be regarded as a secondary market that only becomes important for a title after its diffusion in the primary market (which is studied here). Furthermore, no particular estimation bias arises from this type of sampling. By law, book prices are the same for all retail channels.[22] In theory, buffs may be more inclined to order directly from publishers because they do not need retailer advice; in practice, however, title availability is high and ordering processes are quicker (typically overnight) at stationary bookshops and online retailers. Direct orders from publishers are thus unattractive for non-professional buyers.

The raw data indicate negative sales – books returned by consumers after purchase – for 18 weekly observations. A good share of these take place in January and thus seem to be mis-given Christmas presents. In estimations, I replace sales with value zero in these observations. In addition, data for some titles appear to contain advance orders erroneously booked as sales: sales of 1, 2 or 3 copies followed by a number of zero-sales, long before sales actually take off with two- and three-digit weekly sales.[23] I therefore apply an automatic procedure to identify the most evident cases, namely those in which a

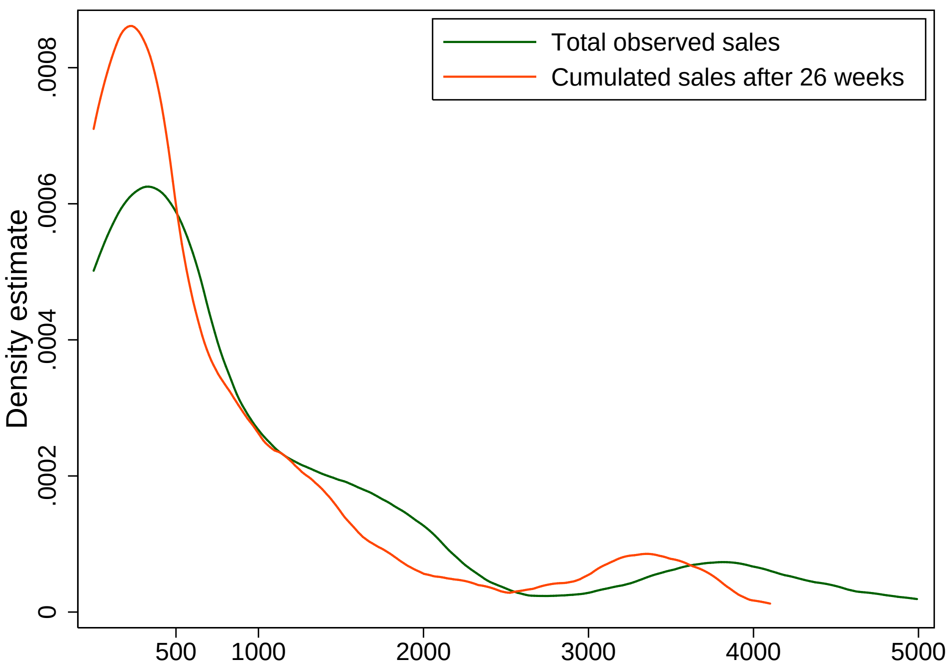
[22] In many European countries including Germany, book prices are by law subject to resale price maintenance; that is, retailers must not offer discounts from the publisher's list price.

[23] In fact, each observation (sales of 1, 2 or 3) may arise from just one pre-ordered copy because some points of sale from which the data were aggregated have a sample weight larger than one.

first sales observation of 1, 2 or 3 copies is below the respective title's average weekly sales (based on weeks with nonzero sales) and is followed by a zero-sale week. I assume that such an observation represents an advance order and add the amount to the following observation of positive sales, which I assume to be the effective release week. Since a few titles seem to exhibit multiple weeks with such advance orders, I repeat the procedure four times. In a similar exercise, I also interpret all first-week sales observations of sales of 1 to 3 copies as advance orders whenever they were below 10% of the title's average sales (based on weeks with nonzero sales). Apart from identifying a title's effective first sales week, these procedures leave results essentially unaffected because first-week sales observations are not used in estimations (see section 3.1).

The initial sample consisted of 307 titles, but many of those were actually not sold before 2004 or very late in 2003. This is not unusual because a title's release year is determined with publication of the publisher's season catalogue, long before the start of the season. For most titles the production process is not yet finalized at that point. Therefore, a late start of observed sales is a sign of delayed release rather than weak demand. In order to ensure a sufficient number of observations per title, I restrict attention to the 229 titles in the sample that began selling before mid-October (week 43 of 2003).

Figure 6: Distribution of cumulated sales across titles



A.2 The Christmas effect

Christmas-driven purchases are usually concentrated on the December weeks (t_D; D=1,...4) and can boost a title's sales significantly, which introduces an additional identification problem. In each of the four December weeks, sales $S_i(t_D)$ consist of three independent parts: the 'usual' model part $f(t)$, an error term ε_{it_D}, and extra Christmas-driven sales $s_i^x(t_D)$. For the moment, the independence of $f(t)$ and $s_i^x(t_D)$ is an assumption, but supportive evidence is presented below.

For parametric identification of $f(t)$, I include time dummies for each December week, whose coefficients (λ_D) are assumed to be non-negative. With respect to estimation of model parameters, this amounts to placing zero weight on December observations only when they are above the title's specific trend. Predicted values from the original function to be fitted are then unresponsive to December spikes, and the λ-coefficients capture all sales in excess of those predicted by the otherwise best-fitting set of parameters. Provided $E[\varepsilon_{it_D}]$=0, the sum of these coefficients may then be interpreted as an estimate for a title's extra Christmas sales: $\tilde{S}_i^x = \sum_D s_i^x(t_D) = \sum_D \hat{\lambda}_D$.[24] On average, parametrically estimated extra Christmas sales $\tilde{S}_i^x$ represent about 10% of the average title's overall sales; for some titles, however, $\tilde{S}_i^x$ represents up to one third of overall sales.

If one is interested primarily in estimating $f(t)$, a simple solution to the Christmas identification problem is to place zero weight on December observations in estimation. I follow this approach for the semiparametric estimation results presented in this paper. The underlying assumption is again that $f(t)$ and extra Christmas sales ($s_i^x(t_D)$) are independent. In order to be able to assess the validity of this assumption, I first use the semiparametric model to obtain a title-specific estimate of extra Christmas sales ($\hat{S}_i^x$). Provided $E[\varepsilon_{it_D}]$=0 and given an estimated smooth function $f(t)$ for observed pre- and post-December sales, I impute December values $f(t_D)$ by interpolation. An estimate for extra Christmas sales in December week t_D is then the difference $S_i(t_D) - f(t_D)$ whenever it is positive, or zero else. Yet, period sales are differently variant across titles (heteroskedastic), which affects this estimate of $s_i^x(t_D)$. As a more robust estimate that enables comparison across titles with different sales variances, I therefore use the upper

[24] Alternatively, one may specify a functional form for $s_i^x(t_D)$ and its relation to $f(t)$ and ε_{it} (see Beck, 2006b, for a parsimonious approach).

limit of the prediction's confidence interval $f^{\bar{CI}}(t_D)$ in calculating extra Christmas sales:

$$\hat{S}_i^x = \sum_{t_D=1}^{4} (S_i(t_D) - f^{\bar{CI}}(t_D)) \, \mathbb{1}(S_i(t_D) > f^{\bar{CI}}(t_D)).$$

Based on this estimate for S_i^x, I assess a potential relationship between a title's extra Christmas sales and its post-Christmas performance by running the following cross-title regression:

$$\frac{\mathbf{N_i^{2004}}}{\mathbf{N_i^{2003}} - \mathbf{\hat{S}_i^x}} = \overset{(1.16)}{\underset{(1.76)}{\mathbf{1.46}}} \; \overset{(-.04)}{\underset{(-.02)}{-\mathbf{.03}}}\, \mathbf{T^{2003}} \; \overset{(-.09)}{\underset{(+.33)}{+.12}} \frac{\mathbf{\hat{S}_i^x}}{\mathbf{N_i^{2003}} - \mathbf{\hat{S}_i^x}} + \varepsilon_i,$$

$$\text{where} \quad i = 1, ..., 176 \; (R^2 = .25).$$

In this regression, N_i^{2004} denotes sold copies observed for title i's in the first half of 2004, N_i^{2003} denotes sold copies observed for title i's since its release in 2003, and T^{2003} denotes the number of weeks title i has been sold in 2003 (52 minus its 2003 release week). In other words, the regression relates a title's level of observed 2004 sales (relative to 2003 non-Christmas sales) to the number of weeks it has been for sale in 2003 as well as to the level of extra Christmas sales (relative to 2003 non-Christmas sales). Bracketed numbers indicate 95% confidence intervals for the estimates (centred).[25] The estimated coefficients indicate, for example for a title released in mid-2003 (hence T^{2003}=26), that on average 2004 sales represent about 68% of 2003 non-Christmas sales. The estimated effect of extra Christmas performance is modestly positive on average but not significantly different from zero (an increase in $\frac{S_i^x}{N_i^{2003} - S_i^x}$ by .25 – about one standard deviation – is associated with an increase in relative 2004 sales by about 3%-points on average.). Hence, additional Christmas-driven December sales do not seem to induce second-order sales effects in the new year.

A second Christmas-related issue that may affect estimation results is strategic release timing: do publishers strategically choose release times for titles that are expected to do well in the Christmas season? A simple approach to this question is to assess whether a title's extra Christmas sales are related to its release date (its distance to Christmas).

[25]Confidence intervals do not account for the fact that $\hat{S}_i^x$ is itself the result of an estimation and are therefore too narrow. Furthermore, the above estimates exclude two titles that due to large sales shocks in 2004 have very large values of N_i^{2004}/N_i^{2003}. In a regression that includes these two titles, all coefficients are insignificant.

Regressing extra Christmas sales, relative to overall non-Christmas sales (N_i), on release week (T^{2003}) gives the following result:

$$\frac{\mathbf{\hat{S}_i^x}}{\mathbf{N_i - \hat{S}_i^x}} = \underset{(.40)}{\overset{(.24)}{\mathbf{.32}}} \quad \underset{(-.004)}{\overset{(-.009)}{\mathbf{-.006\ T^{2003}}}} \quad + \varepsilon_i,$$

$$\text{where} \quad i = 1, ..., 178 \ (R^2 = .12).$$

These estimates indicate that the relationship between release timing and Christmas sales is indeed modestly positive: moving a title's release date one week closer to Christmas is associated with an estimated increase in the relative size of Christmas sales by .006%-points. Yet, it is unclear whether this effect is a strategic one. A confounding effect is that older titles are less attractive as Christmas presents because they have a higher probability that the donee already knows them. Given the relatively modest economic effect and the high variability in the data as indicated by a low R^2, I conclude that strategic release timing with respect to Christmas seems to be of minor importance.

References

Arnold, Martin (2002), Seeking readers with recall, *New York Times*, April 11, 2002.

Bass, Frank M. (1969), A new product growth model for consumer durables, *Management Science*, 15(5): 215–227.

Beck, Jonathan (2006a), Fixed, focal, fair? Book prices with optional resale price maintenance, working paper, September 2006.

— (2006b), The sales effect of word of mouth: A model for creative goods and estimates for novels, WZB Discussion Paper SP II 2006-16.

— (2007), The sales effect of word of mouth: A model for creative goods and estimates for novels, *Journal of Cultural Economics*, 31(1): 5–23.

Boswijk, H. Peter and Philip Hans Franses (2005), On the econometrics of the Bass diffusion model, *Journal of Business & Economic Statistics*, 23(3): 255–268.

Bulow, Jeremy and Paul Klemperer (1996), Auctions versus negotiations, *American Economic Review*, 86(1): 180–94.

Cachon, Gérard P. and Martin A. Lariviere (2005), Supply chain coordination with revenue-sharing contracts: Strengths and limitations, *Management Science*, 51(1): 30–44.

Cameron, A. Colin and Pravin K. Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, New York, NY.

Caves, Richard E. (2000), *Creative Industries*, Harvard University Press, Cambridge, MA.

— (2003), Contracts between art and commerce, *Journal of Economic Perspectives*, 17(2): 73–83.

Chevalier, Judith A. and Dina Mayzlin (2006), The effect of word of mouth on sales: Online book reviews, *Journal of Marketing Research*, 43(3): 345–354.

Clerides, Sofronis K. (2002), Book value: Intertemporal pricing and quality discrimination in the U.S. market for books, *International Journal of Industrial Organization*, 20(10): 1385–1408.

Cleveland, William S. (1979), Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, 74(368): 829–836.

Crawford, Merle C. (1977), Marketing research and the new product failure rate, *Journal of Marketing*, 41(2): 51–61.

Dana, James D. and Kathryn E. Spier (2001), Revenue sharing and vertical control in the video rental industry, *Journal of Industrial Economics*, 49(3): 223–245.

De Vany, Arthur S. and W. David Walls (2004), Motion picture profit, the stable paretian hypothesis, and the curse of the superstar, *Journal of Economic Dynamics and Control*, 28(6): 1035–1057.

Dellarocas, Chrysanthos (2003), The digitization of word of mouth: Promise and challenges of online feedback mechanisms, *Management Science*, 49(10): 1407–1424.

Deneckere, Raymond, Howard P. Marvel and James Peck (1997), Demand uncertainty and price maintenance: Markdowns as destructive competition, *American Economic Review*, 87(4): 619–641.

Denisoff, Serge and William L. Schurk (1986), *Tarnished Gold: The Record Industry Revisited*, Transaction Publishers, New Brunswick, NJ.

Diderot, Denis (1876), Lettre sur le commerce de la librairie, in: *Oeuvres complètes*, vol. 18, Garnier-Frères, Paris.

— (2003), *Lettre sur le commerce de la librairie*, Mille et une nuits, Paris.

Elberse, Anita (2007), The power of stars: Do star actors drive the success of movies?, *Journal of Marketing*, 71(4): 102–120.

Elberse, Anita, Jehoshua Eliashberg and Mark Leenders (2006), The motion picture industry: Critical issues in practice, current research and new research directions, *Marketing Science*, 25(6): 638–661.

Escarpit, Robert (1969), *La révolution du livre*, Unesco/Presses Universitaires de France, Paris, 2nd ed.

Giles, David E. (2006), Superstardom in the US popular music industry revisited, *Economics Letters*, 92: 68–74.

Goldenberg, Jacob, Donald R. Lehmann and David Mazursky (2001), The idea itself and the circumstances of its emergence as predictors of new product success, *Management Science*, 47(1): 69–84.

Goldman, William (1983), *Adventures in the Screen Trade: A Personal View of Hollywood and Screenwriting*, Warner Books.

Hall, Bronwyn H. and Beethika Khan (2003), Adoption of new technology, in: Jones, Derek C. (ed.), *New Economy Handbook*, Academic Press.

Hansmann, Henry and Reinier Kraakman (1992), Hands-tying contracts: Book publishing, venture capital financing, and secured debt, *Journal of Law, Economics and Organization*, 8(3): 628–55.

Horvitz, Paul M. (1966), The pricing of textbooks and the remuneration of authors, *American Economic Review*, 56(1/2): 412–420.

Hurvich, Clifford M., Jeffrey S. Simonoff and Chih-Ling Tsai (1998), Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *Journal of the Royal Statistical Society Series B*, 60(2): 271–293.

Jensen, Richard (1988), Information cost and innovation adoption policies, *Management Science*, 34(2): 230–239.

Mahajan, Vijay, Eitan Muller and Yoram Wind (eds.) (2000), *New-product diffusion models*, Springer, New York, NY.

Moldovanu, Benny and Manfred Tietzel (1998), Goethe's second-price auction, *Journal of Political Economy*, 106(4): 854–859.

Moul, Charles C. (2007), Measuring word of mouth's impact on theatrical movie admissions, *Journal of Economics and Management Strategy*, 16(4): 859–892.

Putsis, William P. Jr. and V. Srinivasan (2000), Estimation techniques for macro diffusion models, in: Mahajan *et al.* (2000), 263–294.

Rosen, Sherwin (1981), The economics of superstars, *American Economic Review*, 71(5): 845–58.

Ruppert, David, Matthew P. Wand and Raymond J. Carroll (2003), *Semiparametric regression*, Cambridge University Press, New York, NY.

Salganik, Matthew J., Peter S. Dodds and Duncan J. Watts (2006), Experimental study of inequality and unpredictability in an artificial cultural market, *Science*, 311(5762): 854–856.

Sorensen, Alan T. (2007), Bestseller lists and product variety, *Journal of Industrial Economics*, 55(4): 715–738.

Sorensen, Alan T. and Scott Rasmussen (2004), Is any publicity good publicity? A note on the impact of book reviews, working paper.

Tietzel, Manfred (1995), *Literaturökonomik*, J.C.B. Mohr (Paul Siebeck), Tübingen.

Turnovsky, Geoffrey (2003), Conceptualising the literary market: Diderot and the lettre sur le commerce de la librairie, *Studies on Voltaire and the Eighteenth Century*, 2003(1):

135–170.

Van den Bulte, Christophe and Yogesh V. Joshi (2007), New product diffusion with influentials and imitators, *Marketing Science*, 26(3): 400–421.

von Lucius, Wulf D. (2005), *Verlagswirtschaft*, UTB, Konstanz.

Bücher des Schwerpunkts Märkte und Politik
Books of the Research Area Markets and Politics

Mark Gradstein, Kai A. Konrad (Eds.)
Institutions and Norms in Economic Development
2007, MIT Press

Johannes Münster
Mobbers, Robbers, and Warriors
2007, Shaker Verlag

Kai A. Konrad, Beate Jochimsen (Eds.)
Der Föderalstaat nach dem Berlin-Urteil
2007, Peter Lang Verlag

Kai A. Konrad, Beate Jochimsen (Eds.)
Finanzkrise im Bundesstaat
2006, Peter Lang Verlag

Robert Nuscheler
On Competition and Regulation in Health Care Systems
2005, Peter Lang Verlag

Pablo Beramendi
Decentralization and Income Inequality
2003, Madrid: Juan March Institute

Thomas R. Cusack
A National Challenge at the Local Level: Citizens, Elites and Institutions in Reunified Germany
2003, Ashgate

Sebastian Kessing
Essays on Employment Protection
2003, Freie Universität Berlin
http://www.diss.fu-berlin.de/2003/202

Daniel Krähmer
On Learning and Information in Markets and Organizations
2003, Shaker Verlag

Tomaso Duso
The Political Economy of the Regulatory Process: An Empirical Approach
Humboldt-University Dissertation, 2002, Berlin,
http://edoc.hu-berlin.de/dissertationen/duso-tomaso-2002-07-17/PDF/Duso.pdf

Bob Hancké
Large Firms and Institutional Change. Industrial Renewal and Economic Restructuring in France
2002, Oxford University Press

Andreas Stephan
Essays on the Contribution of Public Infrastructure to Private: Production and its Political Economy
2002, dissertation.de

Peter A. Hall, David Soskice (Eds.)
Varieties of Capitalism
2001, Oxford University Press

Hans Mewis
Essays on Herd Behavior and Strategic Delegation
2001, Shaker Verlag

Andreas Moerke
Organisationslernen über Netzwerke – Die personellen Verflechtungen von Führungsgremien japanischer Aktiengesellschaften
2001, Deutscher Universitäts-Verlag

Silke Neubauer
Multimarket Contact and Organizational Design
2001, Deutscher Universitäts-Verlag

Lars-Hendrik Röller, Christian Wey (Eds.)
Die Soziale Marktwirtschaft in der neuen Weltwirtschaft, WZB Jahrbuch 2001
2001, edition sigma

Michael Tröge
Competition in Credit Markets: A Theoretic Analysis
2001, Deutscher Universitäts-Verlag

Torben Iversen, Jonas Pontusson, David Soskice (Eds.)
Unions, Employers, and Central Banks
2000, Cambridge University Press

Tobias Miarka
Financial Intermediation and Deregulation: A Critical Analysis of Japanese Bank-Firm-Relationships
2000, Physica-Verlag

Rita Zobel
Beschäftigungsveränderungen und organisationales Lernen in japanischen Industriengesellschaften
2000, Humboldt-Universität zu Berlin
http://dochost.rz.hu-berlin.de/dissertationen/zobel-rita-2000-06-19

Jos Jansen
Essays on Incentives in Regulation and Innovation
2000, Tilburg University

Ralph Siebert
Innovation, Research Joint Ventures, and Multiproduct Competition
2000, Humboldt-Universität zu Berlin
http://dochost.rz.hu-berlin.de/dissertationen/siebert-ralph-2000-03-23/

Damien J. Neven, Lars-Hendrik Röller (Eds.)
The Political Economy of Industrial Policy in Europe and the Member States
2000, edition sigma

Jianping Yang
Bankbeziehungen deutscher Unternehmen: Investitionsverhalten und Risikoanalyse
2000, Deutscher Universitäts-Verlag

DISCUSSION PAPERS 2007

Kai A. Konrad	**Strategy in Contests – an Introduction**	SP II 2007 – 01
Jo Seldeslachts Joseph A. Clougherty Pedro Pita Barros	**Remedy for Now but Prohibit for Tomorrow: The Deterrence Effects of Merger Policy Tools**	SP II 2007 – 02
Stijn Goeminne Benny Geys Carine Smolders	**Political Fragmentation and Projected Tax Revenues: Evidence from Flemish Municipalities**	SP II 2007 – 03
Hilde Coffé Benny Geys	**Measuring the Bridging Nature of Voluntary Organizations: A Note on the Importance of Association Size**	SP II 2007 – 04
Lutz Engelhardt	**Der Entry Standard der Frankfurter Wertpapierbörse als Nachfolger des Neuen Marktes: Zwei Versuche institutioneller Innovation in Deutschland**	SP II 2007 – 05
Beate Jochimsen Robert Nuscheler	**The Political Economy of the German Länder Deficits**	SP II 2007 – 06
Thomas R. Cusack Torben Iversen David Soskice	**Economic Interests and the Origins of Electoral Systems**	SP II 2007 – 07
Lutz Engelhardt	**Arbeits- und Kapitalmarktstruktur als Determinanten von Frühphasen-Wagniskapitalinvestitionen**	SP II 2007 – 08
Benny Geys Jan Vermeir	**Taxation and Presidential Approval: Separate Effects from Tax Burden and Tax Structure Turbulence?**	SP II 2007 – 09
Kristien Werck Bruno Heyndels Benny Geys	**The Impact of 'Central Places' on Spatial Spending Patterns: Evidence from Flemish Local Government Cultural Expenditures**	SP II 2007 – 10
Benny Geys	**How to Make Head or Tail of 'Bridging' and 'Bonding'?: Addressing the Methodological Ambiguity**	SP II 2007 – 11
Tomaso Duso Enrico Pennigs Jo Seldeslachts	**The Dynamics of Research Joint Ventures: A Panel Data Analysis**	SP II 2007 – 12

DISCUSSION PAPERS 2008

Dan Kovenock Brian Roberson	**Inefficient Redistribution and Inefficient Redistributive Politics**	SP II 2008 – 01
Dan Kovenock Brian Roberson	**Coalitional Colonel Blotto Games with Application to the Economics of Alliances**	SP II 2008 – 02
Vito Tanzi	**The Future of Fiscal Federalism**	SP II 2008 – 03
Kai A. Konrad Kjell Erik Lommerud	**Love and Taxes – and Matching Institutions**	SP II 2008 – 04
Benny Geys Jan Vermeir	**Party Cues and Yardstick Voting**	SP II 2008 – 05
Benny Geys Jan Vermeir	**The Political Cost of Taxation: New Evidence from German Popularity Ratings**	SP II 2008 – 06
Kai A. Konrad Dan Kovenock	**The Alliance Formation Puzzle and Capacity Constraints**	SP II 2008 – 07
Johannes Münster	**Repeated Contests with Asymmetric Information**	SP II 2008 – 08
Kai A. Konrad Dan Kovenock	**Competition for FDI with Vintage Investment and Agglomeration Advantages**	SP II 2008 – 09
Kai A. Konrad	**Non-binding Minimum Taxes May Foster Tax Competition**	SP II 2008 – 10
Florian Morath	**Strategic Information Acquisition and the Mitigation of Global Warming**	SP II 2008 – 11
Joseph Clougherty Anming Zhang	**Domestic Rivalry and Export Performance: Theory and Evidence from International Airline Markets**	SP II 2008 – 12
Jonathan Beck	**Diderot's Rule**	SP II 2008 - 13

Bei Ihren Bestellungen von WZB-Papers schicken Sie bitte unbedingt einen an Sie adressierten Aufkleber mit sowie je paper eine Briefmarke im Wert von 0,51 Euro oder einen "Coupon Reponse International " (für Besteller aus dem Ausland)

Please send a self addressed label and postage stamps in the amount of 0.51 Euro or a "Coupon-Reponse International" (if you are ordering from outside Germany) for each WZB-paper requested

Bestellschein *Order Form*

Absender / Return Address:

____

____

____

____

Wissenschaftszentrum Berlin
für Sozialforschung
Presse- und Informationsreferat
Reichpietschufer 50

D-10785 Berlin-Tiergarten

Hiermit bestelle ich folgende(s) Discussion paper(s):

Please send me the following Discussion paper(s):

Bestell-Nr. / Order no.	Autor/in, Kurztitel /Author(s) / Title(s) in brief