Experts, Information, Reviews, and Coordination: Evidence on How Prizes Affect Sales[*]

Nicolas Lagios[†]

Pierre-Guillaume Méon[‡]

Exploiting the award process, we implement a regression discontinuity design to estimate the effect of winning France's main literary prize, the Goncourt. It increases sales, especially for books that sold fewer copies before the announcement, the number of reviews on Amazon, and the probability of them being negative. The effect is partly driven by an increase in word of mouth. Those

[†] Authors' affiliations: Université libre de Bruxelles (ULB), CEBRIG and Dulbéa, Research Fellow F.R.S.-FNRS – Aspirant FNRS, CP-114/03, avenue F.D. Roosevelt 50, 1050 Bruxelles, Belgium. *e-mail: nicolas.lagios@ulb.be*

[‡] Université libre de Bruxelles (ULB), CEBRIG and Dulbéa, CP-114/03, avenue F.D. Roosevelt 50, 1050 Bruxelles, Belgium. *email: p-guillaume.meon@ulb.be*

findings are consistent with a model where the prize provides information on the existence of a book and acts as a quality signal and a coordination device but prompts consumers to read books that are far from their tastes.

I.   INTRODUCTION

If awards, which are meant to reward artistic quality, result in commercial success, they may contribute to reconciling economic and artistic objectives (Canoy *et al.* [2006]). Considering the efforts that producers and publishers devote to ensuring that their productions receive or are shortlisted for an award, they are doubtless convinced of the material consequences of prizes (English [2014]).

By contrast, academic research takes that belief with a grain of salt. The first reason is that causality is difficult to establish. If awards are bestowed on cultural products with characteristics that predestine them for commercial success, then they will correlate with success, because they confirm or predict it, but they will not cause it (Eliashberg and Shugan [1997]; Moser and Nicholas [2013]; Ponzo and Scoppa [2015]). As a result, the effect of awards on sales may be elusive (Ginsburgh [2003]).

The second reason for researchers' scepticism is that the channels of transmission between awards and commercial success, if any, are unclear and have received little empirical attention. Prizes may affect sales through three channels that are not mutually exclusive. First, awards may simply put an artwork under the spotlight, raising awareness of its existence and attracting new consumers, regardless of its quality (Clement *et al.* [2007]; Sorensen and Rasmussen [2004]; Berger *et al.* [2010]). The second channel is that awards may send a signal about the quality of goods that are experience goods, insofar as one does not know beforehand the utility to be gained from watching a movie,

attending a performance, or reading a book. If consumers believe that experts can gauge the quality of artworks and bestow awards on the best of them, then they will interpret those awards as a positive quality signal, and commercial success will ensue (Ashworth *et al.* [2010]; Clement *et al.* [2007]; Ginsburgh and van Ours [2003]; Ginsburgh [2003]; Ponzo and Scoppa [2015]). The third channel is that, if consumers are better off consuming the same artwork as others do, prizes may allow them to coordinate, in line with Adler's [1985] model. Experts and prizes may accordingly work as coordination devices (Loeper *et al.* [2014]). To determine the effect of awards on sales, it is necessary both to address causality and understand how awards operate.

To study whether and how awards affect sales, we use France's most prestigious literary prize, the Goncourt Prize, awarded annually since 1903 to the "best and most imaginative prose work" of the year (Assouline [2013]). Using hand-collected data from the archives of the Académie Goncourt on the confidential votes of the committee, we can address causality by implementing the regression discontinuity design (RDD) approach used by Ponzo and Scoppa [2015]. Specifically, we take advantage of the discontinuity created because the Goncourt is bestowed on the nominated book receiving the highest number of votes in the final round of the selection process.

We improve on that design by adapting it to a dynamic set-up, using a database that reports the weekly sales of each book and that allows us to leverage the time dimension of sales. We can thus control for each book's pre-Goncourt sales trend and avoid any bias in RDD estimates that may appear if the probability of winning the prize is correlated with sales. We can apply that design to the number of sales of each book from its publication date until the 50th week of 2019 and track how winning the prize affects those sales. We essentially focus on the period from 2004 to 2018, over which we

can track the entire sales of nominated books since publication; but we also extend the analysis to identify the effect of the 1954-2018 prize period on sales between 2004 and 2019. We therefore consider between 220 and 854 books. In line with common wisdom, our estimates show that the Goncourt boosts sales by an estimated 350% or an average of 260,000 copies.

By comparing books that marginally won and those that marginally lost, our RD design compares similar books and sifts away the effect of omitted variables that may prompt books to be both more likely to win the prize and more commercially successful, thereby addressing causality concerns. When a perfectly random instrumental variable is not available, like in Ginsburgh and van Ours [2003], the literature has addressed causality thanks to matched samples (Azoulay *et al.* [2014]), the synthetic control method (Chan *et al.* [2014]), or a difference-in-differences approach (Borjas and Doran [2015]). Unlike the matched samples and synthetic control methods, ours does not require to define a comparison group. Instead, it compares books that were all close to winning the prize. Moreover, our RDD estimates are immune to violations of the parallel trends assumption that may plague difference-in-differences estimates. Our estimates can therefore be interpreted as causal.

We go beyond confirming common wisdom and previous literature by testing the channels of transmission between the prize and book sales. Specifically, we report evidence supporting the information channel, whereby the award raises the awareness of potential consumers on the existence of prize-winning books. We do so by conditioning the effect of the prize on sales prior to the winner announcement. We find that the effect of the Goncourt decreases with previous book sales and even becomes statistically indistinguishable from zero for the books that sold the most copies prior to receiving the

prize and were already well known to the public, limiting the scope of the information channel.

We also test an implication of the quality signal channel. To do this, we assess consumer satisfaction by performing a sentiment analysis on consumers' reviews on Amazon.fr, and we find that the Goncourt negatively affects their opinions. Consumers' star ratings show the same negative effect. These findings are consistent with a model where consumers follow the advice of the experts on the prize committee but have tastes that differ from those of the experts. As a result, some do not enjoy the winning book and post negative reviews. This interpretation is backed by the finding that, despite its positive effect on sales, the Goncourt des Lycéens prize has no observable impact on the sentiment of reviews. The Goncourt des Lycéens is based on the same list of books as its namesake but is awarded by a large group of high-school students, whose tastes are arguably closer to the public's than those of experts. The interpretation is also backed by the finding that the effect of the Goncourt on reviewer sentiment is mitigated by the number of reviews posted online. This suggests that peer opinion can compensate for the gap between the committee's tastes and those of consumers, thus preventing them from buying a book they will not enjoy.

Finally, we test the coordinating role of the prize by assessing the role of word of mouth as a mediating factor. We do so by adjusting our regression discontinuity design to a mediation framework. We find that the Goncourt boosts the volume of reviews a book receives on Amazon.fr, which in turn boosts sales, regardless of the tenor of those reviews. This result is in line with the hypothesis that the prize generates a buzz that is informative of the likelihood that consumers will have the opportunity to interact about a book, prompting them to coordinate on reading the winning book because others did so.

This result therefore validates one of the key hypotheses of Adler's [1985] theory of superstars.

This paper contributes to several strands of literature, first and foremost that on awards (Ashworth *et al.* [2010]; Frey and Gallus [2017]; Ginsburgh [2003]; Azoulay *et al.* [2014]; Chan *et al.* [2014]; Ponzo and Scoppa [2015]; Borjas and Doran [2015]). It does so by confirming the causal effect of cultural awards on sales. Furthermore, we provide suggestive evidence of three channels through which literary prizes operate. Specifically, the results suggest that they inform consumers of the existence of the winning book, send an expert signal, and help potential readers to coordinate. Second, the paper contributes to the literature on experts (Clement *et al.* [2007]; Ginsburgh and van Ours [2003]; Hilger *et al.* [2011]; Ginsburgh *et al.* [2019]; Ekelund *et al.* [2020]; Reinstein and Snyder [2005]) by showing that they can be influential when bestowing a prize but may also reduce consumers' satisfaction if the tastes of the former differ from those of the latter. Finally, we add to the understanding of the role of consumer reviews (Babić Rosario *et al.* [2016]; Chen and Wu [2021]; Reimers and Waldfogel [2021]) by showing how consumer reviews are affected by prizes and contribute to the effect on sales.

The rest of the paper is organized as follows. Section II discusses in more detail the three channels through which prizes can affect sales. Section III provides background information on the history and functioning of the Goncourt. Section IV describes the data. Section V discusses the empirical strategy. Section VI presents the baseline findings while Section VII provides evidence on the mechanisms. Section VIII concludes.

II. THEORETICAL FRAMEWORK: PRIZES AS INFORMATION, QUALITY SIGNALS, AND COORDINATION DEVICES

The channels through which a literary prize can affect the sales of a book can be inferred from the conditions necessary for a consumer to decide to buy it. The first necessary condition is tautological: the consumer must be aware of the book's existence. By focusing public attention, a prize raises consumers' awareness of a book and prompts some of them to buy it. That "information effect" is in line with the finding of Sorensen and Rasmussen [2004] and Berger *et al.* [2010] that even negative reviews increase the sales of works by relatively unknown authors. Sorensen [2007] and Cabral and Natividad [2016] also reported evidence that appearing in the list of best-selling books or being the best-selling movie during the opening weekend affects later sales by raising the awareness of consumers.[1]

The second condition for consumers to buy a book is that they must expect the utility of reading it to exceed its total cost, which includes the purchase price and the opportunity cost of reading. A prize does not alter that opportunity cost and is unlikely to affect the price unless publishers react by changing it. The latter reaction is, in any case, impossible in France because of legislation stipulating that the prices of books are determined by publishers, printed on the cover, and cannot be changed during the first two years after publication.[2]

However, a prize can affect the expected utility of reading by reducing quality uncertainty. This is firstly because books are experience goods, whose intrinsic quality cannot be known prior to consumption (Nelson [1970]). Consequently, literary prizes can provide a quality signal. Specifically, if consumers trust committees to select works on the basis of quality, they will expect those books to give them more utility and will buy

prize-winners. In addition to being experience goods, books can also be credence goods insofar as consumers may not be able to fully judge their quality even after reading them (Darby and Kani [1973]; English [2014]). In that case, consumers interested in quality will follow expert opinion if they consider it reliable. The role of experts' views has been reported in the case of books (Clement *et al.* [2007]; Sorensen and Rasmussen [2004]; Berger *et al.* [2010]; Ponzo and Scoppa [2015]), movies (Eliashberg and Shugan [1997]; Reinstein and Snyder [2005]), and wines (Hilger *et al.* [2011]; Dubois and Nauges [2010]; Friberg and Grönqvist [2012]). We refer to this effect as the quality signal effect of prizes.[3]

Another way literary prizes may affect expected utility is by serving as coordination devices. This would happen if consumers received not only intrinsic utility from reading a book but also extrinsic utility from discussing it with other readers. This is the basic premise of Adler's [1985] theory of superstars, whereby consumers have an incentive to coordinate on consuming the same cultural products to maximize the probability of being able to discuss them. Prizes facilitate coordination by providing a focal point. When purchasing an award-winning book, consumers know that many other readers have done the same, so prizes give them a near certainty of being able to talk about it. In the extreme, if reading the book that has received a given prize becomes the norm, then not reading it may result in a social stigma, especially if the award is prestigious. In either case, consumers have an incentive to read. That effect will be channelled by word of mouth if consumers use past sales to infer the probability that they will be able to talk about a book, this will result in a bandwagon effect (Babić Rosario *et al.* [2016]). We refer to that effect as the coordination effect of literary prizes.[4]

III. The Goncourt Prize in a Nutshell

Created at the bequest of Edmond de Goncourt, the Goncourt Prize is the most prestigious French literary prize. It has been awarded yearly since 1903 by a jury of ten experts to the author of "the most imaginative prose work published in the year" (Assouline [2013]). The jury members, chosen by cooptation, are usually prominent figures (writers, essayists, philosophers, screenwriters, etc.) on the French literary scene.[5] Originally, the prize – 5,000 French francs – was intended to allow the winner to live by their pen until their next book. That has since been reduced to the symbolic sum of 10 euros, but the Goncourt provides other advantages such as greater visibility and the acknowledgement of a writer's literary artistry.

The award is bestowed by the Académie Goncourt at the beginning of November after three selection stages that take place between the beginning of September and the end of October, starting from a short-list of around 15 books, then 8 and finally 4.[6] The last stage is divided into rounds, with each jury member casting one vote per round. During the first ten rounds, the prize can only be awarded by an absolute majority; from the 11th to the 13th rounds, a relative majority suffices. If there is a tie, the president's vote counts double in the 14th round. We use that decision mechanism to implement a regression discontinuity design.

IV. Data

Our dataset includes all the nominated books that competed for the Goncourt between 1954 and 2018 and for which data on sales are available. However, since the sales record only goes back to January 1, 2004, we focus on the 2004-2018 awards in our main analyses and use the entire sample in some robustness checks.[7] We thus observe 220

books, including 15 winners.[8] Those 15 constitute the treatment group while the other 205 are the control group. Since we compare winning books with nominated ones, our estimates are a lower bound of the impact of the Goncourt on sales if the mere nomination for the prize already has a commercial effect. For each book, we collected data on its sales, votes, and characteristics.

IV(i). *Data on Sales*

Data on sales were collected from EdiStat, a website for French book industry professionals that publishes figures for weekly sales in mainland France.[9] We observe the number of sales of each nominated work from 2004 to the 50th week of 2019 and can therefore track the total sales of books participating in the 2004 to 2018 editions of the Goncourt.

One important feature of the database is that it reports weekly sales of each book. We can therefore leverage the time dimension of the data to measure the effect of the prize on the flow of sales; we can also control for pre-Goncourt sales trends and avoid any bias caused when the prize may be awarded to books that are already selling well.[10] We thus improve on Ginsburgh [2003] and Ponzo and Scoppa [2015], who use the stock of sales at a given point in time.

Moreover, our database measures actual book sales, as opposed to proxies. By contrast, Ginsburgh [2003] proxies sales by the number of editions of a book, while Ponzo and Scoppa [2015] use the number of aNobii's members who own a given book in their virtual collection.[11] We also improve on Ashworth *et al.* [2010], who use the number of copies reordered by booksellers, because our figures take into account the sales made over the entire life cycle of the book, including those in the shops' initial orders. In addition, we take into account only the copies that have been effectively sold, as opposed

to those ordered by bookshops but not subsequently sold. By contrast, the number of ordered copies would reflect the beliefs of booksellers as to the effect of the prize rather than the effect of the prize itself.

IV(ii). *Data on Votes*

Information on votes comes mainly from the archives of the Académie Goncourt in France. However, as data on the very recent editions of the prize were not available at the time of collection, we supplemented the database by hand-extracting information on votes from news articles and press releases for each missing year.

Regrettably, only the votes of the last round of the final selection are systematically available, be it in the archives or the media. This implies that shortlisted books receiving no votes in the last round have the same zero number of votes as non-shortlisted books. To distinguish between the two groups, we add five votes to all shortlisted books and include them in the votes received in the last round of the final selection.[12] Adding five votes is to some extent arbitrary but can be motivated on the grounds that if a book had received more than that number in a single round, it would have been selected by an absolute majority of the committee and won the prize. This puts an objective cap on the number of votes received by books that did not make it to the final round. Most of all, this vote coding process has no impact on the results, as the identification strategy relies on the comparison of books that nearly received the prize with those that received it with a small victory margin. Accordingly, observations far away from the cut-off have little influence on the estimates. In any case, we show in Online Appendix D4 that the results are not driven by the way that votes are coded.[13]

IV(iii). *Books' Characteristics*

For each book, we also record the publisher and the gender of the author, whether it was adapted for the movies or television, and whether it won any other major prizes.[14]

Table I reports descriptive statistics.[15] It shows that a book competing for the Goncourt receives 2.1 votes and sells 171,261 copies on average. Furthermore, out of the 220 books in the sample, 24 (11% of the total number) were adapted for the big or small screen, 60 (27%) won other prizes, and 63 (29%) were written by a female author.

[Place Table I about here]

V.   THE REGRESSIONS DISCONTINUITY DESIGN

V(i).   *Definition of the Cut-Off*

To provide unbiased estimates of the impact of the Goncourt on book sales, we take advantage of the discontinuity created because the prize is awarded to the nominated book receiving the highest number of votes in the last round of the decision process. More precisely, we look at the correlation between sales and the difference in the number of votes received by each book and the number of votes received by the book ranked second in the final round of the decision process. It can be argued that the books preferred by a larger number of jury members likely have characteristics that make them sell more copies. However, only the book with a positive victory margin gets the prize. Hence, there is a discontinuity in the relationship at the victory margin of 0 above which books receive the award.

We leverage this discontinuity and perform an RDD analysis where the running variable is the victory margin. Under the assumption that the conditional expectation functions of the potential outcomes given the running variable are continuous at the cut-

off, the jump around the cut-off indicates a causal effect (Cattaneo *et al.* [2019]; Lee and Lemieux [2010]).[16] The intuition behind this RDD is that, conditional on the victory margin, books directly below and above the cut-off are on average similar in terms of quality and other non-observable characteristics that may affect sales, with the difference that only those above the cut-off have been awarded the Goncourt. By comparing the outcomes of the two groups, we obtain unbiased estimates of the average treatment effect at the cutoff.

We define the running variable following Ponzo and Scoppa's [2015] strategy and normalize the number of votes given to each book competing each year for the Goncourt:

(1) $$Margin_{iy} = Votes_{iy} - \left(Votes_{Sy} + 1\right),$$

where $Votes_{iy}$ is the effective number of votes received by book i during the competition year y and $Votes_{Sy}$ is the number of votes received by the second-highest ranked book in year y. The running variable, $Margin_{iy}$, represents the victory margin by which a book won or lost. Book i wins the Goncourt in year y if its victory margin is greater or equal to 0, resulting in a discontinuity at 0. All the books below the cut-off fall in the comparison group while those above are treated. Hence,

(2) $$Goncourt_{iy} = \begin{cases} 1 & if\ Margin_{iy}\ \geq 0 \\ 0 & if\ Margin_{iy}\ < 0 \end{cases}$$

V(ii). *Empirical Model*

To identify the impact of the Goncourt on sales, we estimate the following regression:

$$\log(Sales_{iy,post})$$

(3)
$$= \alpha + \tau Goncourt_{iy} + f\big(Goncourt_{iy}, Margin_{iy}\big)$$

$$+ \phi \log(Sales_{iy,pre}) + \boldsymbol{\theta}' \mathbf{X}_{iy} + \lambda_y + \epsilon_{iy},$$

where

$Sales_{iy,post}$ is the number of post-Goncourt sales for book i competing in year y. In our baseline analyses, we use the logarithmic transformation of $Sales_{iy,post}$, as it allows us to conveniently interpret the effect of the prize as a semi-elasticity;

$Goncourt_{iy}$ is a dummy that takes value 1 if book i is awarded in competition year y;

$f\big(Goncourt_{iy}, Margin_{iy}\big) = \sum_{j=1}^{p} \beta_j Margin_{iy}^j + \sum_{j=1}^{p} \gamma_j Goncourt_{iy} \times Margin_{iy}^j$ is a polynomial function that models the impact of a book's victory margin on its post-Goncourt sales;

$Sales_{iy,pre}$ is the number of pre-Goncourt sales for book i competing in year y;

$\mathbf{X}_{iy}$ is a vector of control variables;

λ_y are time dummies coding the year of the competition;

ϵ_{iy} is the error term.

The interaction between $Goncourt_{iy}$ and $Margin_{iy}$ inside the flexible function $f(\cdot)$ allows us to account for the fact that the marginal effect of the victory margin might differ between awarded and non-awarded books.

Controlling for a book's pre-Goncourt sales ensures that the effect we observe is not driven by the fact that the prize is awarded to books that are already selling well.[17] It also controls for the fact that the Goncourt may be awarded more systematically to well-known and best-selling authors.

The set of control variables includes the dummy variables $Movie$, which is set to one if the book was given a movie or television adaptation, $Other\ prizes$, if it received another literary prize, and $Female\ author$, if it was written by a woman. If Goncourt-winning books were more likely to be adapted or win another prize, resulting in greater visibility, then the coefficient of interest would reflect the impact of the adaptation or award rather than of the Goncourt, which is why we control for the $Movie$ and $Other\ prizes$ variables. As the committee and book-buyers may be partial to a specific gender, resulting in an omitted variable bias, we also control for the author's gender.

The Goncourt jury has often been accused of favoring major publishers (Zerilli [2015]; Genova [2014]).[18] If this is the case, and if larger publishers are associated with higher sales, then the estimates may be biased. Indeed, in such a setting, the coefficient of $Goncourt$ would no longer reflect the effect of the prize itself but rather capture the fact that award-winning books are published by large publishers.[19] To control for this threat, $\mathbf{X}_{iy}$ also includes four publisher dummy variables: three corresponding to the three largest historical French publishers – Gallimard, Grasset, Seuil – and one corresponding to Actes Sud, which is a latecomer: It did not win until 2004, but has since accumulated five victories between 2004 and 2018. The reference category therefore consists of all the other publishers.

Competition year dummies, λ_y, control for the fact that books nominated in older editions of the prize have by construction more post-Goncourt sales than books nominated in more recent editions.[20] More generally, those dummies control for unobserved heterogeneity across editions of the prize to strictly focus on variations within editions. This is a useful feature as editions take place in different contexts, for instance because some are subject to more controversy than others (Assouline [2013]).

The main coefficient of interest is τ, which measures the marginal impact of winning the Goncourt, that is, the average treatment effect at the cutoff.

Due to the small sample size and the low number of support points around the cut-off, our baseline approach is a parametric RDD, which takes advantage of all available observations. The resulting larger sample makes it possible to perform extensions (e.g., making a mediation analysis, using publisher dummies, considering only shortlisted books) that would have been impossible had the sample been restricted to observations near the cut-off.[21]

Finally, the treatment is skewed because there are 205 books in the comparison group for only 15 in the treatment group, so we report confidence intervals (CIs) that are corrected for small and skewed samples.[22] More precisely, to build those CIs, we follow the recommendations of Imbens and Kolesár [2016] and apply Bell and McCaffrey's [2002] (BM hereafter) degree-of-freedom correction on standard errors in order to obtain adjusted standard errors. The 95% BM adjusted standard errors are defined as $\sqrt{\widehat{\mathbb{V}}_{\mathrm{BM}}} = \sqrt{\widehat{\mathbb{V}}_{\mathrm{HC2}}} \times (t_{0.975}^{K_{\mathrm{BM}}}/1.96)$ where $\widehat{\mathbb{V}}_{\mathrm{HC2}}$ is the variance estimator proposed by MacKinnon and White (1985) and $t_{0.975}^{K_{BM}}$ is the qth quantile of the t-distribution with K degrees of freedom (Imbens and Kolesár [2016]). We then use $\sqrt{\widehat{\mathbb{V}}_{\mathrm{BM}}}$ to construct the CIs for the parameter of interest.

V(iii). *Identification Assumptions*

In RDDs, identification requires (i) smoothly varying covariates at the cut-off, (ii) the absence of selective sorting around the cut-off, and (iii) the absence of discontinuity at points other than the cut-off (i.e. placebo cut-offs; Cattaneo *et al.* [2019]).

To assess whether covariates vary smoothly at the cut-off (i), we conduct a set of RDD analyses where $\log(Sales_{post})$ is replaced in turn by each of our control variables. In practice, we respectively regress $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, and the four publisher dummies on $Goncourt$, $Margin$, $Goncourt \times Margin$, and the time dummies. The results are presented and discussed in Online Appendix B1. Overall, we find no robust evidence of a discontinuity of the control variables at the cut-off.

Selective sorting around the cut-off (ii) is unlikely, given the mechanism generating the running variable, i.e. the votes given to each book. One may argue that publishers have an incentive to lobby the jury. However, only big publishers have significant lobbying power. Furthermore, even if publishers were to bribe the jury, it is unlikely that all ten members would be corrupt, further reducing publishers' control over the award process. As a result, publishers have, at best, only imprecise control over votes whereas only precise control invalidates the RD design (Lee and Lemieux [2010]). In addition, the finding that the three big publisher dummies are smooth functions of the treatment provides further evidence in favor of the absence of manipulation.

Another possible concern is that jury members may have an incentive to manipulate the votes, for example to award the Goncourt to books that are already selling well in order to increase the value and reputation of the prize. In Online Appendix B1,

we show this concern to be unfounded by demonstrating that sales prior to the prize are not discontinuous at the cut-off. [23]

Finally, to check that there are no jumps at placebo cut-offs (iii), we follow Imbens and Lemieux's [2008] recommendation. Specifically, we divide our sample in two subsamples, the observations at the left of the cut-off and those at the right, and we perform an RDD in each subsample by using the median of the running variable as the cut-off.[24] We find no evidence of discontinuity at either side (see Table B.III in Online Appendix B3).

VI. BASELINE RESULTS

VI(i). *A First Look at the Data*

To provide a first sense of the effect of the Goncourt on book sales, we plot the relationship between $\log(Sales_{post})$ and $Margin_{iy}$ in Figure 1 for all the books nominated between 2004 and 2018.

To reduce noise and make the discontinuity easier to identify, we present a smoothed plot (Calonico *et al.* [2015]; Lee and Lemieux [2010]). More precisely, we divide the running variable into bins and compute the average number of log sales into each bin. Since the running variable exhibits few mass points, we set the bin width to one so that the number of bins is equal to the number of different values taken by the running variable. We have 9 bins at the left of the cut-off ($Margin$ = [−10, -7]∪[-5; 0)) and 6 at the right ($Margin$ = [0, 5]).

[Place Figure 1 about here]

Figure 1 shows that a book's sales increase with its victory margin, suggesting that the margin captures sales potential. More importantly, Figure 1 displays a

discontinuity in the neighborhood of the cut-off, which provides initial proof of an effect of the Goncourt on sales.

VI(ii). *Regression Discontinuity Estimates*

Table II presents the estimated results for the 220 books nominated for the Goncourt between 2004 and 2018. The model specification follows Equation (3) while the BM CIs for the prize's coefficient are reported in brackets. In Column (1), we investigate the effect of the Goncourt on sales assuming a linear relationship between the outcome and the running variable and linearly controlling for the book's year of publication. As expected, the victory margin is positively correlated with sales. Importantly, the coefficient of $Goncourt$ is equal to 1.406 and is statistically significant beyond the one-percent level, meaning that winning the Goncourt boosts sales by more than 300% ($e^{1.406} - 1$ since we are in a log-lin specification). Since the average number of copies of a prize-winning book sold prior to the award is 74,560, this implies that the prize leads to an average increase of about 225,000.

Column (2) replaces the linear time trend by 15 time-dummies to allow for more flexibility. The coefficient of $Goncourt$ slightly decreases but remains large ($\tau = 1.172$) and significant at the one-percent level.

[Place Table II about here]

In RDDs, the inclusion of covariates is typically not mandatory for identification if the design is valid but it has the advantage of increasing the precision of the estimates. Column (3) therefore reports the results of a specification including covariates.

Predictably, sales prior to the award of the Goncourt correlate with post-Goncourt sales. Predictably, too, benefiting from a movie or television adaptation and winning other prizes are positively related to sales. It can be argued that a movie reminds the public of

the book's existence and gives it visibility with an additional audience, resulting in more sales a few months or years after publication. The positive coefficient of the $Other\ prizes$ dummy shows that other awards are associated with higher sales regardless of the causal effect of the Goncourt. We find no statistically significant effect of the author's gender. Finally, most publisher dummy variables are statistically insignificant. The two exceptions are Actes Sud, which bears a positive coefficient statistically significant at the one-percent level, and Seuil, which bears a negative coefficient statistically significant at the ten-percent level in one regression. Accordingly, Actes Sud's books on average enjoy greater commercial success, and those published by Seuil less commercial success, than other nominated books. Despite the large set of control variables, the coefficient of the $Goncourt$ dummy remains statistically significant beyond the one-percent level and is equal to 1.497, meaning that winning the prize boosts sales by nearly 350% or 260,000 copies.

Finally, one of the main risks of parametric RDDs is to interpret a potential nonlinearity as a discontinuity caused by the treatment. To make sure that our estimates are not subject to that threat to identification, Columns (4) and (5) report the outcome of the same specifications as Columns (2) and (3) but assume a quadratic relation between $\log(Sales_{iy,post})$ and $Margin_{iy}$. In those quadratic specifications, the coefficient of $Margin_{iy}$ is no longer significant, meaning that the margin of victory has no effect on sales.[25] However, what really matters when assessing the marginal impact of winning the Goncourt is the coefficient of $Goncourt$, which remains statistically significant at the one-percent level and indicates a substantial treatment effect amounting to a 300% rise in sales.

In all specifications, even when adjusting the confidence intervals for skewed samples, which are shown in brackets, inference remains similar.

In Online Appendix C, we explore the evolution of the impact of the Goncourt over time. As shown by Figure C.I, the effect of the prize is at its highest the first weeks following its attribution and then decreases over time.

VI(iii). *Robustness Checks*

To assess the sensitivity of our baseline results, we perform a set of robustness checks, whose outcomes are reported in Online Appendix D.

First, we explore the sensitivity of our parametric estimates by implementing two alternative RDD strategies (see Online Appendix D2 for more details on each approach). The first is a non-parametric approach, which has the benefit of being less sensitive to the degree of the polynomial and to observations far away from the cut-off; it also exhibits better inference proprieties (Table D.II; Gelman and Imbens [2019]). The second approach is the local randomization approach. It allows switching from a large-sample approximation framework to a finite-sample framework that is better tailored to small-sample inference (Table D.III; Cattaneo *et al.* [2017]). Reassuringly, the results are relatively insensitive to the approach, thus demonstrating the strength of our parametric framework.

Second, to make sure that our results are not driven by outliers, we re-estimate Equation (3), this time using a Least Absolute Deviations (LAD) estimator. As the LAD estimator minimizes the sum of the absolute residuals, it is insensitive to outliers (Wooldridge [2010]). The magnitude and the significance of the Goncourt are analogous to the baseline, as shown by Table D.IV in Online Appendix D3.

Third, we run two tests to show that the findings are insensitive to the way we code the number of votes for the books that did not make it to the last round of the final selection. To begin with, we use two alternative ways of coding the victory margin (Table D.V; see Online Appendix D4 for a description of each strategy). Then, we re-estimate Equation (3) by restricting the sample to shortlisted books (Table D.VI). In this way, only the effective number of votes matters. In both cases, the results remain similar.

Finally, we further explore the sensitivity of our baseline results by considering a series of alternative specifications (Tables D.VII and D.VIII in Online Appendix D5). To extend our results to books that won the Goncourt before 2004, we use the entire period for which votes are available, specifically 1954-2018, with the caveat that the sales figures for books published before 2004 do not include data prior to 2004. That test can be interpreted as both a robustness check and a way to estimate the long-run effect of winning the prize, because the estimate measures the effect of winning the prize in any year from 1954 to 2018 on the sum of sales between 2004 and 2019. Insofar as we can apply the RDD strategy to all books from 1954 to 2018, we can still interpret the estimate as causal. We now observe 854 books, including 64 laureates. The outcome is reported in Column (1) of Tables D.VII and D.VIII.

In Column (2), we focus on the 1954 to 2013 editions to address the concern that the evolution of books published before and after 2004 may be different.

In Column (3), we introduce a dummy for each publisher (28 dummies in total). In this way, we can capture unobserved heterogeneity among publishers: for example, some may systematically attract more successful authors, or spend more on advertising.

Lastly, as we have no prior on the functional form relating victory margins to sales, we use the number of sales pre- and post-Goncourt in level instead of in log in Column (4).

In all cases, the effect of the Goncourt remains positive and significant at conventional levels.

VII. MECHANISMS

In this section we investigate the drivers of the positive effect of the Goncourt on sales. To do so, we test the existence of the three mechanisms put forward in the theoretical discussion. Specifically, the prize may inform consumers as to the existence of a book, provide a quality signal, and be a coordination device for consumers. The mechanisms are not mutually exclusive and may even reinforce each other. We test them in turn.

VII(i). *Information Effect*

The information effect is the easiest to assess as it can be tested directly. To do so, we investigate how the impact of the prize varies according to the book's popularity prior to winning the Goncourt, captured by pre-award sales. If the information effect is at work, we should expect the prize to have a greater impact on the sales of little-known books than on those that are already popular. Indeed, if the prize is awarded to a relatively unknown work, it will allow many consumers to discover the book's existence. Conversely, awarding the prize to a book that is already known to many potential consumers will contribute little information.

To test this hypothesis, we extend our RD framework to allow for heterogeneity in the treatment effect (Becker *et al.* [2013]). Accordingly, we extend Equation (3) by

interacting $Goncourt_{iy}$ and $\log(Sales_{iy,pre})$ so as to estimate the following treatment-covariate interaction model:

$$
\begin{aligned}
\log(Sales_{iy,post}) \\
(4) \qquad & = \alpha + \tau Goncourt_{iy} + f(Goncourt_{iy}, Margin_{iy}) \\
& + \phi \log(Sales_{iy,pre}) + \mu Goncourt_{iy} \times \log(Sales_{iy,pre}) \\
& + \boldsymbol{\theta}' \mathbf{X}_{iy} + \lambda_y + \epsilon_{iy},
\end{aligned}
$$

where the variables are defined as previously. The point of the model is to estimate the conditional marginal effect (CME) of $Goncourt_{iy}$ on $\log(Sales_{iy,post})$, that is

$$
(5) \qquad \left(\Delta \log(Sales_{iy,post}) \middle| Goncourt_{iy} = 1\right) = \tau + \mu \log(Sales_{iy,pre}).
$$

We estimate the CMEs of $Goncourt$ on $\log(Sales_{iy,post})$ using the kernel smoothing estimator considered in Hainmueller *et al.* [2019] in order to relax the linear interaction effect assumption and the linearity assumption on the covariates, and to avoid excessive extrapolation.[26] The results are summarized in Figure 2, which plots the marginal effect of the Goncourt as a function of pre-prize sales. The left-hand side graph assumes a linear relation between the vote margin and sales while the right-hand side assumes a quadratic relation.[27]

Both graphs show that the impact of the Goncourt, though always positive, decreases with the recipient's pre-award sales. In addition, for very popular books, that is, books whose pre-Goncourt sales are higher than 163,191 copies, the marginal effect of the Goncourt becomes statistically insignificant. This finding echoes Sorensen's [2007] observation that appearing on the New York Times bestseller list has a larger effect on sales for debut authors and the finding by Cabral and Natividad [2016] that being the best-selling movie during the opening weekend increases a movie's later demand.

Like them, we interpret it as being in line with the existence of an information effect. In our case, the prize draws the attention of potential consumers to a book of which they were unaware but has little or no impact on books that are already best-sellers.

[Place Figure 2 about here]

VII(ii). *Quality Signal*

As we do not observe the utility that consumers expect from reading a book, let alone how they revise it when it receives an award, we cannot directly test the role of prizes as quality signals. However, we can test an implication of that role: If prizes are quality signals and the judgment of experts is trusted, then consumers will buy winning books because they expect more utility from reading them. However, if consumers blindly follow experts but have different tastes, some will be disappointed because the book will not be to their liking. By contrast, consumers' tastes are likely closer to those of their peers than to those of experts. Accordingly, if potential consumers can observe their peers' opinions, they will be more likely to buy a book they enjoy.

To test those hypotheses, we supplement the dataset by performing a sentiment analysis on the textual content of the reviews left by consumers on Amazon.fr. Sentiment analysis is a natural language processing technique for extracting the sentiment valence of an opinionated text. A sentiment can be either positive, neutral, or negative (Pang and Lee [2008]). To perform the sentiment analysis, we first used a Python script to scrape Amazon.fr and collect the textual content of each customer's review. We did so for each nominated book. We then created a second Python script to use the API of *Watson Natural Language Understanding*, the natural language processing service of IBM, and classify each review as positive, neutral, or negative.[28] On average, books nominated for

the Goncourt receive 58 reviews, of which 25 (43%) are positive and 17 (29%) are negative.[29]

The outcome variable is *Sentiment*, which codes the nature of the opinion, or sentiment, of a given consumer for a given book. Specifically,

(6)
$$Sentiment_{ciy} = \begin{cases} 0, & \textit{if the opinion is negative} \\ 1, & \textit{if the opinion is neutral} \\ 2, & \textit{if the opinion is positive} \end{cases}$$

where $Sentiment_{ciy}$ is the opinion of review c about book i competing in year y, which we interpret as a measure of the consumer's satisfaction from reading the book.

To make sure that our results are not driven by the specific approach used to measure sentiment, we also use the review's rating as a more direct and objective proxy for consumer satisfaction. Reviewers complement their comments by a rating ranging from one to five stars. Using the number of stars awarded by the reviewer as dependent variable leads to very similar findings, as shown by Table E.IV in Online Appendix E2.

The Variable $Sentiment_{ciy}$ is then used as the outcome variable of a model that relates the probability of leaving a positive review to whether the book received the Goncourt and to the number of reviews on Amazon.fr:

(7)
$$\begin{aligned} \Pr\big(Sentiment_{ciy} = \xi\big) \\ = F\big[\alpha + \tau Goncourt_{iy} + \psi \arcsinh\big(\#Reviews_{ciy}\big) \\ + g\big(Goncourt_{iy}, Margin_{iy}\big) + \phi \log\big(Sales_{iy,pre}\big) + \boldsymbol{\theta}'\mathbf{X}_{iy} \\ + \lambda_y + \eta_{ciy}\big], \end{aligned}$$

where $\#Reviews_{ciy}$ is the number of reviews of book i that were already available at the time review c was written. Because $\#Reviews_{ciy}$ can be equal to zero, it is transformed using the inverse hyperbolic sine (arcsinh). For sufficiently large values of the

transformed variable, the arcsinh transformation is similar to a log-transformation with the difference that it is defined at zero (Burbidge *et al.* [1988]).[30] The other variables are defined as before and ξ can take the value 0, 1 or 2 as defined in (6). The main difference between baseline Equation (3) and Equation (7) is that (3) is estimated using OLS while (7) is estimated using an ordered logit model, as the dependent variable follows a natural ordering.[31] Finally, because $Goncourt_{iy}$ is measured at the book level while $Sentiment_{ciy}$ is measured at the level of individual reviews, we cluster the standard errors at the book level to allow for arbitrary dependence between the reviews of a same book.

In line with the hypothesis that the tastes of the Goncourt jury differ from those of the average consumer, we expect τ to be negative, thus the prize will decrease the probability of a consumer writing a positive review. Conversely, we expect ψ to be positive, meaning that the number of available reviews increases the probability of a positive review, because the peers have similar tastes.

Panel A of Table III presents the results of Equation (7). To validate our identification strategy, we perform a placebo test in Columns (1) and (2) by looking at the effect of the Goncourt on the sentiment valence of the reviews written before the prize was awarded, respectively fitting a linear and a quadratic polynomial. In both cases, the coefficient of $Goncourt$ is not significant. Accordingly, before receiving the prize there is no pre-existing difference between awarded and non-awarded books in terms of consumers' opinion, which implies that differences observed after the award can be interpreted as caused by it.

[Place Table III about here]

In Columns (3) and (4) we perform the same analyses as in Columns (1) and (2), looking this time at post-Goncourt reviews. In both the linear and the quadratic cases, *Goncourt* bears a negative coefficient significant at the one-percent level, meaning that consumers are more likely to post a negative review of a Goncourt winner. More precisely, when a book is awarded the Goncourt, the probability that a consumer posts a negative review increases by 14 percentage points while the probability of writing a positive review decreases by 15 percentage points, on average. In addition, the coefficient of the number of reviews is also statistically significant at the one-percent level but positive, suggesting that a larger number of past reviews help consumers to choose a book they will enjoy.[32]

A corollary of our hypotheses is that the number of past reviews should not only affect the sentiment of reviews but also mitigate the negative effect of the Goncourt on sentiment. Indeed, since consumers' tastes are likely to be closer to those of their peers than to those of the experts, the availability of peer opinions should reduce the risk of making a wrong choice. As a result, the marginal effect of winning the prize on the probability of being disappointed by a book and posting a negative review should decrease when the number of reviews increases.

[Place Figure 3 about here]

We test that corollary by interacting the Goncourt dummy with the number of reviews posted before the prize was awarded. The results are summarized in Figure 3, which plots the average marginal effect of the Goncourt on consumer sentiment as a function of past reviews.[33] The left-hand side fits a linear specification while the right-hand side fits a quadratic one. As expected, both graphs show that the negative effect of the Goncourt on sentiment decreases with the number of past reviews, and even becomes

non distinguishable from zero beyond a certain threshold. Specifically, the probability of posting a negative review decreases with the number of past reviews while the probability of posting a positive review increases with it.

The findings of Table III and Figure 3 are in line with the hypothesis that some consumers interpret the prize as a quality signal but are subsequently disappointed because the book is too far from their tastes. Ultimately, those "unusual" consumers are more likely to dislike the book.[34] However, as the number of past reviews by peers increases, consumers are better informed about the match between the book and their tastes. As a result, they are less likely to be disappointed.

An alternative interpretation of the negative effect of the Goncourt on reviewer sentiment is that it raises expectations that are later disappointed. To discriminate this interpretation from the interpretation that the deterioration in reviews is driven by the mismatch in tastes between the committee and consumers, we study another prize awarded by a committee whose tastes are likely closer to the general public's. The gist of the test is that if prizes have a negative effect on reviews because they raise consumer expectations that are later disappointed, then all prizes should have a negative effect on reviews, regardless of the composition of the awarding committee. Conversely, if the deterioration in reviews is due to the discrepancy between the tastes of the committee and those of consumers, then only prizes awarded by experts should affect reviews. A prize bestowed by lay people may increase sales but not affect reviews, because the jury's tastes are closer to those of consumers.

Specifically, we estimate the effect on reviews of receiving the Goncourt des Lycéens prize. This award is based on the same first selection of books as the Goncourt but is awarded by around 2,000 high school students rather than professionals.

Accordingly, the awarding committee's taste is arguably closer to the average consumer's. Using a difference-in-differences model, we show in Online Appendix E3 that, despite having a positive and significant effect on book sales, the Goncourt des Lycéens has no impact on consumers' satisfaction. These findings are difficult to reconcile with the notion that prizes negatively affect reviews because they raise expectations that are later disappointed. Rather, they are consistent with the idea that prizes attract consumers who trust expert opinion but may be disappointed if the book that they read is too far from their tastes.[35]

VII(iii).   *Coordination Device*

In line with Adler's [1985] model, a prize can act as a coordination device if readers are better off reading a book that many others read. Although directly testing that role is difficult, we can test the key assumption of Adler's [1985] model, specifically that consumers prefer reading books that allow them to interact with a larger number of peers. We therefore gauge the extent to which the effect of the prize is driven by a bandwagon effect, which we measure by the number of online reviews posted on Amazon.fr. We refer to the number of online reviews as the volume of electronic word of mouth (eWOM, Babić Rosario *et al.* [2016]). The hypothesis is that the number of reviews is informative of the buzz caused by a book and therefore of the likelihood that consumers will have the opportunity to interact. If the prize operates through a bandwagon effect, we should expect it to increase the number of reviews, which in turn would increase sales.

We accordingly extend our RD framework to perform a mediation analysis. This allows to explore whether and how an independent variable of interest affects an outcome variable through a third one (Baron and Kenny [1986]; Hayes [2017]). In our case, we are interested in assessing whether the effect of the Goncourt is driven by the fact that

winning books benefit from a bandwagon effect. Specifically, we hypothesize that the Goncourt increases the number of reviews, which consequently boosts the number of sales.

[Place Figure 4 about here]

A formal definition of the RD mediation framework is provided in Online Appendix E4.1. It is intuitively summarized in Figure 4. Specifically, path γ_1 measures the impact of the Goncourt on the number of post-award reviews, measured over a six-month period.[36] Because we control for sales during this six-month period, the increase in reviews that we measure cannot be driven by higher sales during the period. Path δ_2 assesses the marginal effect of an additional review written in the six-month period on subsequent sales, controlling for the effect of the Goncourt (δ_1). The indirect effect is given by $\gamma_1 \times \delta_2$ and represents the prize's impact on sales via the number of reviews. An indirect effect significantly different from zero therefore provides evidence of a bandwagon effect.[37]

[Place Table IV about here]

Table IV reports the results of the mediation analysis. Panel A shows that, as expected, the Goncourt has a strong and significant impact on the number of reviews posted on Amazon.fr in the six months following the attribution of the prize (γ_1). Similarly, those reviews positively impact subsequent sales (δ_2). Panel B reports the estimates for the indirect effect $\gamma_1 \times \delta_2$ and the associated confidence intervals. We derive those confidence intervals using the bias-corrected and accelerated (BCa) method which is recommended for its superior power compared with other types of bootstrap tests (Hayes and Scharkow [2013]; Fritz *et al.* [2012]). We also report quasi-Bayesian Monte Carlo confidence intervals (Tingley *et al.* [2014]), which are more conservative and avoid

false positives (Yzerbyt *et al.* [2018]; Hayes and Scharkow [2013]). The indirect effect is significant and sizeable, meaning that the impact of the Goncourt on sales is partially driven by a bandwagon effect. In addition, the joint-significance test is always significant, lending additional credence to the notion that the word of mouth mediates the effect of the prize.

The upshot of this section is that being awarded the Goncourt increases the number of reviews that a book receives, which in turn increases its sales, regardless of the reviews' content. This can be interpreted as evidence that the prize generates a bandwagon effect, which is in line with the assumption that consumers read books that have been read by other consumers as they prefer cultural goods that give them a greater opportunity to interact with others (Adler [1985]). This is the key mechanism necessary for prizes to work as a coordination device.

VIII. CONCLUSION

We examine the causal effect of literary prizes on book sales using France's most prestigious award, the Goncourt. Taking advantage of the decision process for awarding the prize, we implement a regression discontinuity design to obtain unbiased estimates of the average treatment effect at the cutoff. We find that the Goncourt on average increases sales by 350% or 260,000 copies. In addition, we report evidence of three channels of transmission of the prize.

The first is that the Goncourt raises the awareness of potential consumers about the existence of winning books. In line with that channel of transmission, we observe that the effect of the prize on sales is inversely related to sales figures immediately prior to the announcement of the prize.

We also report evidence on the role of the prize as a quality signal, thanks to a sentiment analysis on customers' reviews on Amazon.fr. We observe that the Goncourt adversely affects the opinions posted by consumers. By contrast, the Goncourt des Lycéens, a prize awarded by high-school students whose tastes are arguably closer to the public's than those of the Goncourt committee, does not affect reviews but does boost sales. Those findings suggest that some consumers who buy a book because it won the Goncourt are subsequently disappointed, because the tastes of the awarding committee differ from their own. As the Goncourt des Lycéens is not bestowed by experts, it does not prompt the same disappointment. We further observe that the larger the number of online reviews, the less the Goncourt affects consumer sentiment, implying that peer reviews mitigate the influence of experts and allow consumers to read books that are closer to their tastes.

Finally, we show that word of mouth is a mediating factor for the prize. When adjusting our regression discontinuity design to a mediation framework, we find that the prize boosts the volume of reviews that a book receives on Amazon.fr, which in turn increases its sales, regardless of the reviews' content. This result is in line with one of the key hypotheses of Adler's [1985] theory of superstars, whereby consumers prefer cultural goods that give them a greater opportunity to interact with others. Accordingly, a literary prize likely operates as a coordinating device.

A word of caution may be in order before generalizing our results. We must emphasize that they pertain to one prize, in one country, and in a given period. Other prizes elsewhere may operate differently. Considering a longer time horizon may also allows determining the extent to which the apparition of the internet and social media affected the dynamic of the prize. Also, while we directly observe the effect of the prize

on sales and how prior success condition its effect, we only provide indirect evidence of the quality signal sent by the prize and of its bandwagon effect. Providing direct effect on those two mechanisms as well as on the way in which publishers and retailers may leverage them should receive closer scrutiny.

Our findings could also be complemented to understand the role of prizes in more detail. First, it may be wondered whether the channels of transmission that we observe for a literary prize are at work in other fields where prizes are awarded, such as other art forms, wine, journalism, and research, to name but a few. If success is directly observable in many fields thanks to sales figures, it is irrelevant in others, posing specific challenges. This is for example the case of scientific research, where citations are commonly used as a measure of success. Yet all citations may not serve the same purpose. If some signal a true influence on subsequent publications, others may be a form of name-dropping meant to signal the authors' knowledge of the field or connections. Distinguishing different types of citations or distinguishing citations from influence, for instance by looking at the evolution of the number of publications or of patents in the field of the rewarded author or paper might be a fruitful way to distinguish bandwagon effects from the effect of quality signals. Second, the finding that winning the Goncourt adversely affects online reviews suggests that prizes may decrease the utility of some consumers. Investigating more closely the total welfare effect of prizes should be a high priority. Third, our paper and the literature focus on the impact of prizes on sales. Yet, the effect may be broader. Prizes may affect the long-term reputation of the authors and publishers of the books they reward. If they have an impact on a publisher's reputation, their effects may spill-over to other authors related to it. Finally, artistic status goes beyond commercial success and can

even be at odds with it. A full understanding of the consequences of prizes therefore demands an assessment of their symbolic consequences.

References

Abdellatif, A.; Badran, K.; Costa, D. and Shihab, E., 2022, 'A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering,' *IEEE Transactions on Software Engineering*, 48, pp. 3087–3102.

Adler, M., 1985, 'Stardom and Talent,' *American Economic Review*, 75, pp. 208-212.

Ashworth, J.; Heyndels, B. and Werck, K., 2010, 'Expert judgements and the demand for novels in Flanders,' *Journal of Cultural Economics*, 34, pp. 197-218.

Assouline, P., 2013, *Du côté de chez Drouant : Cent dix ans de vie littéraire chez les Goncourt*, (Gallimard, Mayenne, France).

Azoulay, P.; Stuart, T. and Wang, Y., 2014, 'Matthew: Effect or Fable?,' *Management Science*, 60, pp. 92-109.

Babić Rosario, A.; Sotgiu, F.; De Valck, K. and Bijmolt, T. H., 2016, 'The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors,' *Journal of Marketing Research*, 53, pp. 297-318.

Baron, R. M. and Kenny, D. A., 1986, 'The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations,' *Journal of Personality and Social Psychology*, 51, pp. 1173-1182.

Becker, S. O.; Egger, P. H. and Von Ehrlich, M., 2013, 'Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects,' *American Economic Journal: Economic Policy*, 5, pp. 29-77.

Bell, R. M. and McCaffrey, D. F., 2002, 'Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples,' *Survey Methodology*, 28, pp. 169-182.

Bellemare, M. F. and Wichman, C. J., 2020, 'Elasticities and the Inverse Hyperbolic Sine Transformation,' *Oxford Bulletin of Economics and Statistics*, 82, pp. 50-61.

Berger, J.; Sorensen, A. T. and Rasmussen, S. J., 2010, 'Positive Effects of Negative Publicity: When Negative Reviews Increase Sales,' *Marketing Science*, 29, pp. 815-827.

Borjas, G. J. and Doran, K. B., 2015, 'Prizes and Productivity: How Winning the Fields Medal Affects Scientific Output,' *Journal of Human Resources*, 50, pp. 728-758.

Burbidge, J. B.; Magee, L. and Robb, A. L., 1988, 'Alternative Transformations to Handle Extreme Values of the Dependent Variable,' *Journal of the American Statistical Association*, 83, pp. 123-127.

Cabral, L. and Natividad, G., 2016, 'Box-Office Demand: The Importance of Being #1,' *Journal of Industrial Economics*, 64, pp. 277-294.

Calonico, S.; Cattaneo, M. D. and Titiunik, R., 2015, 'Optimal Data-Driven Regression Discontinuity Plots,' *Journal of the American Statistical Association*, 110, pp. 1753-1769.

Canonico, M. and De Russis, L., 2018, 'A Comparison and Critique of Natural Language Understanding Tools,' *Cloud Computing 2018 : The Ninth International Conference on Cloud Computing, GRIDs, and Virtualization*, pp. 110-115.

Canoy, M.; Van Ours, J. C. and Van Der Ploeg, F., 2006, 'The Economics of Books,' in Ginsburgh, V. and Throsby, D. (eds.), *Handbook of the Economics of Art and Culture*, Edition 1, Vol. 1, pp. 721-761, (Elsevier Publishing, Amsterdam, The Netherlands).

Carvalho, A. and Harris, L., 2020, 'Off-the-Shelf Technologies for Sentiment Analysis of Social Media Data: Two Empirical Studies,' *AMCIS 2020 Proceedings*.

Cattaneo, M. D.; Idrobo, N. and Titiunik, R., 2019, *A Practical Introduction to Regression Discontinuity Designs: Foundations*, (Cambridge University Press, Cambridge, UK).

Cattaneo, M. D.; Titiunik, R. and Vazquez-Bare, G., 2017, 'Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality,' *Journal of Policy Analysis and Management*, 36, pp. 643-681.

Chan, H. F.; Frey, B. S.; Gallus, J. and Torgler, B., 2014, 'Academic honors and performance,' *Labour Economics*, 31, pp. 188-204.

Chen, M. X. and Wu, M., 2021, 'The Value of Reputation in Trade: Evidence from Alibaba,' *Review of Economics and Statistics*, 103, pp. 857-873.

Clement, M.; Proppe, D. and Rott, A., 2007, 'Do Critics Make Bestsellers? Opinion Leaders and the Success of Books,' *Journal of Media Economics*, 20, pp. 77-105.

Darby, M. R. and Karni, E., 1973, 'Free Competition and the Optimal Amount of Fraud,' *Journal of Law and Economics*, 16, pp. 67-88.

Dubois, P. and Nauges, C., 2010, 'Identifying the effect of unobserved quality and expert reviews in the pricing of experience goods: Empirical application on Bordeaux wine,' *International Journal of Industrial Organization*, 28, 205-212.

Ducas, S., 2010, 'Prix littéraires en France : consécration ou désacralisation de l'auteur ?,' *COnTEXTES. Revue de sociologie de la littérature*, 7.

Ekelund, R. B.; Higgins, R. and Jackson, J. D., 2020, 'ART as meta-credence: authentication and the role of experts,' *Journal of Cultural Economics*, 44, pp. 155-171.

Eliashberg, J. and Shugan, S. M., 1997, 'Film Critics: Influencers or Predictors?,' *Journal of Marketing*, 61, pp. 68-78.

English, J. F., 2014, 'The Economics of Cultural Awards,' in Ginsburgh, V. and Throsby, D. (eds.), *Handbook of the Economics of Art and Culture*, Edition 1, Vol. 2, pp. 119-143, (Elsevier Publishing, Amsterdam, The Netherlands).

Ermakova, T., Henke, M., & Fabian, B. (2021). Commercial Sentiment Analysis Solutions: A Comparative Study. In *Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST 2021)*.

Feine, J., Morana, S., & Gnewuch, U. (2019). Measuring service encounter satisfaction with customer service chatbots using sentiment analysis. In *Proceedings of the International Conference on Wirtschaftsinformatik*.

Frey, B. S. and Gallus, J., 2017, 'Towards an Economics of Awards,' *Journal of Economic Surveys*, 31, pp. 190-200.

Friberg, R. and Grönqvist, E., 2012, 'Do Expert Reviews Affect the Demand for Wine?,' *American Economic Journal: Applied Economics*, 4, pp. 193-211.

Fritz, M. S.; Taylor, A. B. and Mackinnon, D. P., 2012, 'Explanation of Two Anomalous Results in Statistical Mediation Analysis,' *Multivariate Behavioral Research*, 47, pp. 61–87.

Gelman, A. and Imbens, G. W., 2019, 'Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs,' *Journal of Business & Economic Statistics*, 37, pp 447-456.

Genova, P. A., 2014, 'Rewarding the production of culture: Le Prix Goncourt,' *Contemporary French and Francophone Studies*, 18, pp. 150-157.

Ginsburgh, V. and Weyers, S., 2014, 'Nominees, winners, and losers,' *Journal of Cultural Economics*, 38, pp. 291-313.

Ginsburgh, V., 2003, 'Awards, Success and Aesthetic Quality in the Arts,' *Journal of Economic Perspectives*, 17, pp. 99-111.

Ginsburgh, V and Van Ours, J. C, 2003, 'Expert Opinion and Compensation: Evidence from a Musical Competition,' *American Economic Review*, 93, pp. 289-296.

Ginsburgh, V.; Radermecker, A. S. and Tommasi, D., 2019, 'The effect of experts' opinion on prices of art works: The case of Peter Brueghel the Younger,' *Journal of Economic Behavior & Organization*, 159, pp. 36-50.

Hainmueller, J.; Mummolo, J. and Xu, Y., 2019, 'How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice,' *Political Analysis*, 27, pp. 163-192.

Hayes, A. F., 2017, *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 2nd edition, (Guilford Press, New York, U.S.A.).

Hayes, A. F. and Scharkow, M., 2013, 'The Relative Trustworthiness of Inferential Tests of the Indirect Effect in Statistical Mediation Analysis: Does Method Really Matter?,' *Psychological Science*, 24, pp. 1918-1927.

Hilger, J.; Rafert, G. and Villas-Boas, S., 2011, 'Expert Opinion and the Demand for Experience Goods: An Experimental Approach in the Retail Wine Market,' *Review of Economics and Statistics*, 93, pp. 1289-1296.

Imbens, G. W. and Kolesár, M., 2016, 'Robust Standard Errors in Small Samples: Some Practical Advice,' *Review of Economics and Statistics*, 98, pp. 701-712.

Imbens, G. W. and Lemieux, T., 2008, 'Regression discontinuity designs: A guide to practice,' *Journal of Econometrics*, 142, pp. 615-635.

Lee, D. S. and Lemieux, T., 2010, 'Regression Discontinuity Designs in Economics,' *Journal of Economic Literature*, 48, pp. 281-355.

Loeper, A.; Steiner, J. and Stewart, C., 2014, 'Influential Opinion Leaders,' *Economic Journal*, 124, pp. 1147-1167.

MacKinnon, J. G. and White, H., 1985, 'Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties,' *Journal of Econometrics*, 29, pp. 305-325.

McCrary, J., 2008, 'Manipulation of the running variable in the regression discontinuity design: A density test,' *Journal of Econometrics*, 142, pp. 698-714.

Moser, P. and Nicholas, T., 2013, 'Prizes, Publicity and Patents: Non-Monetary Awards as a Mechanism to Encourage Innovation,' *Journal of Industrial Economics*, 61, pp. 763-788.

Nelson, P., 1970, 'Information and Consumer Behavior,' *Journal of Political Economy*, 78, pp. 311-329.

Pang, B. and Lee, L., 2008, 'Opinion Mining and Sentiment Analysis,' *Foundations and Trends in Information Retrieval*, 2, pp. 1-135.

Ponzo, M. and Scoppa, V., 2015, 'Experts' awards and economic success: evidence from an Italian literary prize,' *Journal of Cultural Economics*, 39, pp. 341-367.

Reimers, I. and Waldfogel, J., 2021, 'Digitization and Pre-purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings,' *American Economic Review*, 111, pp. 1944–1971.

Reinstein, D. A. and Snyder, C. M., 2005, ,The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics,' *Journal of Industrial Economics*, 53, pp. 27-51.

Sorensen, A. T., 2007, 'Bestseller Lists and Product Variety,' *Journal of Industrial Economics*, 55, pp. 715-738.

Sorensen, A. T. and Rasmussen, S. J., 2004, *Is Any Publicity Good Publicity? A Note on the Impact of Book Reviews*, (Working Paper, Stanford University). See https://www.ssc.wisc.edu/~sorensen/papers/bookreviews.pdf.

Tingley, D.; Yamamoto, T.; Hirose, K.; Keele, L. and Imai, K., 2014, 'mediation: R Package for Causal Mediation Analysis,' *Journal of Statistical Software*, 59, pp. 1–38.

VanderWeele, T., 2015, *Explanation in Causal Inference: Methods for Mediation and Interaction*, (Oxford University Press, New York, U.S.A.).

Wooldridge, J. M., 2010, *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, (MIT Press, Cambridge, Massachusetts, U.S.A.).

Yzerbyt, V.; Muller, D.; Batailler, C. and Judd, C. M., 2018, 'New recommendations for testing indirect effects in mediational models: The need to report and test component paths,' *Journal of Personality and Social Psychology*, 115, pp. 929–943.

Zerilli, S., 2015, 'Compte rendu de Ducas (Sylvie), La littérature à quel (s) prix ? Histoire des prix littéraires,' *COnTEXTES*.

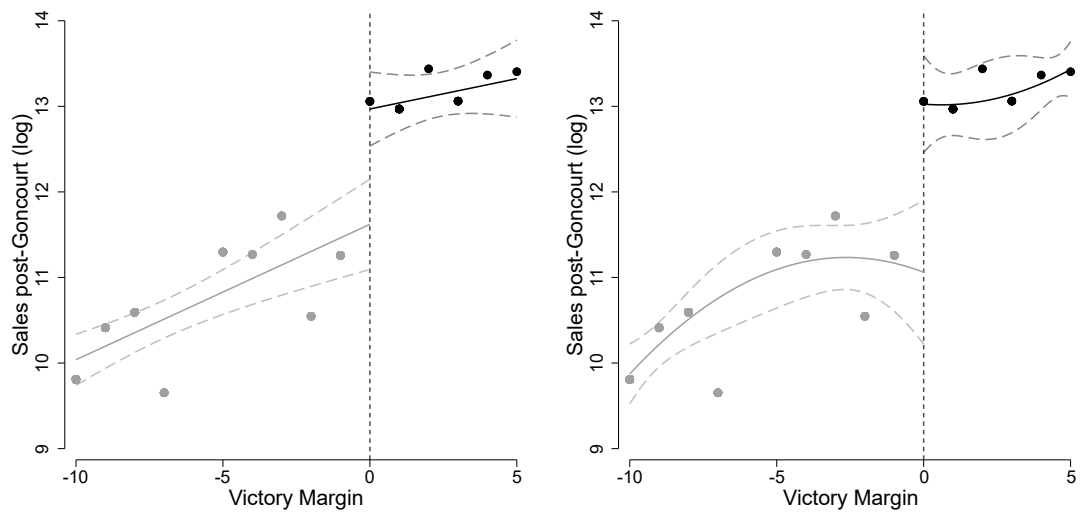**Figures**

Figure 1
Discontinuity Effect of the Goncourt on Book Sales

Notes: RD plot of the effect of the Goncourt on sales. The left-hand side fits a linear polynomial while the right-hand side fits a quadratic one. Bins represent the average number of log sales computed at each value of the victory margin. The dash line reports 95% confidence intervals based on robust standard errors.
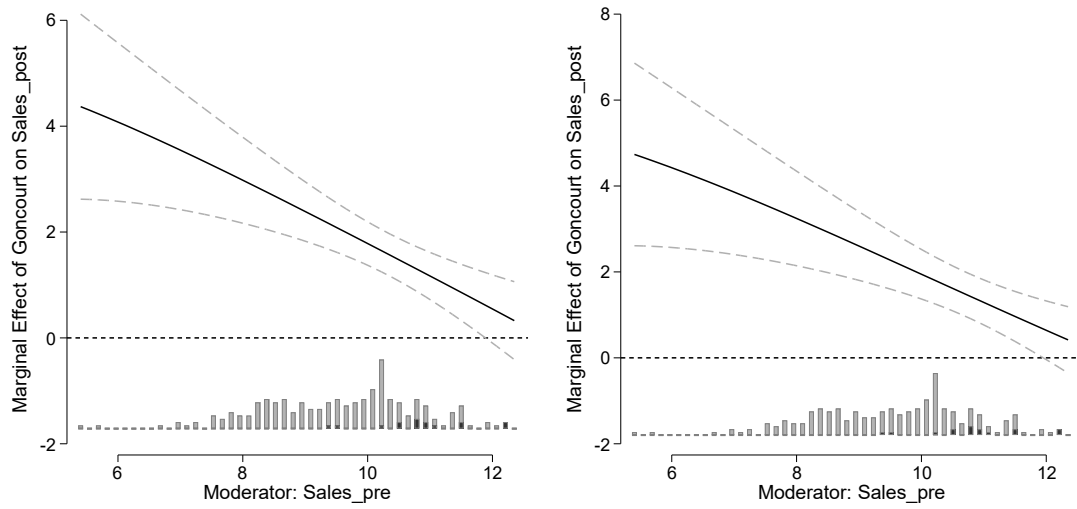
Figure 2

Marginal Effect of the Prize on post-Goncourt Sales as a Function of Pre-Goncourt sales

Notes: The conditional marginal effects are computed using the kernel smoothing estimator considered in Hainmueller *et al.* [2019]. The left-hand side fits a linear polynomial while the right-hand side fits a quadratic one. The model specification follows Equation (4). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prizes*, *Female author*, the four publisher dummies, and the time dummies. The dash line reports 90% confidence intervals based on robust standard errors.

Figure 3
Average Marginal Effect of the Goncourt on Reviewer Sentiment as a Function of the
Number of Other Reviews Available When the Review was posted

Notes: The left-hand side fits a linear polynomial while the right-hand side fits a quadratic one. In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Vertical lines report 90% confidence intervals based on standard errors clustered at the book level.
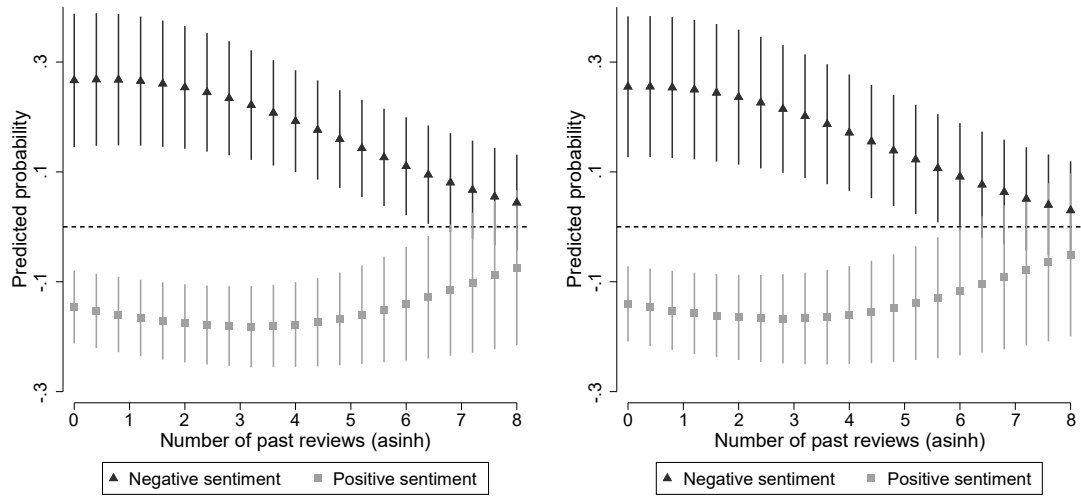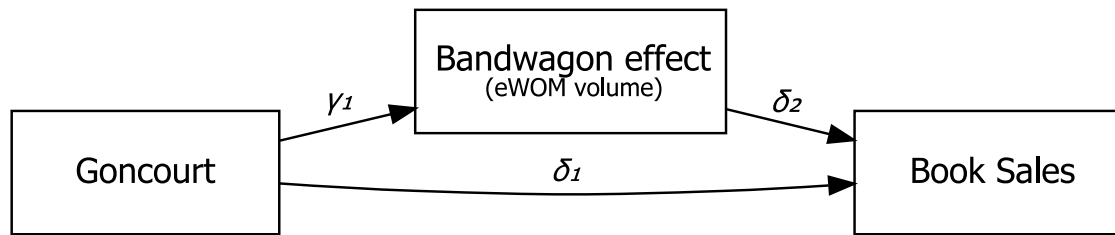
Figure 4
Assessing the Role of Word of Mouth – Mediation Analysis Framework

Notes: eWOM volume refers to the volume of electronic word of mouth, which we measure by the number of online reviews posted on Amazon.fr.

Tables

Table I

Descriptive Statistics

	(1) Obs.	(2) Mean	(3) S.d.	(4) Min.	(5) Max.
$\text{Sales}_{\text{post}}$	220	142,385	244,050	117	1,667,568
$\text{Sales}_{\text{pre}}$	220	28,876	35,778	222	231,030
$\log(Sales_{post})$	220	10.617	1.763	4.762	14.327
$\log(Sales_{pre})$	220	9.624	1.228	5.403	12.350
Votes	220	2.055	3.489	0	12
Margin	220	-6.864	3.644	-10	5
Goncourt	220	0.068	0.253	0	1
Year	220	2011.050	4.369	2004	2018
Movie	220	0.109	0.312	0	1
Other prizes	220	0.273	0.446	0	1
Female author	220	0.286	0.453	0	1
Gallimard	220	0.232	0.423	0	1
Grasset	220	0.132	0.339	0	1
Seuil	220	0.082	0.275	0	1
Actes Sud	220	0.055	0.228	0	1

Notes: The variables and the data sources are described in Section IV.

Table II

The Effect of the Goncourt on Book Sales

	(1)	(2)	(3)	(4)	(5)
	Outcome: $\log(Sales_{post})$				
Goncourt	1.406***	1.172***	1.497***	1.886***	1.375***
	(0.329)	(0.430)	(0.305)	(0.488)	(0.368)
	[0.680,	[0.230,	[0.827,	[0.708,	[0.472,
	2.131]	2.115]	2.166]	3.064]	2.279]
Margin	0.152***	0.166***	0.047**	-0.168	0.007
	(0.036)	(0.037)	(0.018)	(0.205)	(0.094)
Margin×Goncourt	-0.090	-0.026	-0.297***	-0.074	0.047
	(0.066)	(0.127)	(0.094)	(0.465)	(0.379)
Year	-0.028				
	(0.025)				
Margin squared				-0.030	-0.004
				(0.018)	(0.008)
Margin squared×Goncourt				0.128*	-0.062
				(0.071)	(0.067)
$\log(Sales_{pre})$			0.852***		0.857***
			(0.066)		(0.067)
Movie			0.802***		0.786***
			(0.240)		(0.243)
Other prizes			1.172***		1.169***
			(0.124)		(0.127)
Female author			0.025		0.019
			(0.102)		(0.103)
Gallimard			-0.028		-0.026
			(0.128)		(0.129)
Grasset			-0.142		-0.141
			(0.180)		(0.182)
Seuil			-0.340*		-0.323
			(0.202)		(0.204)
Actes Sud			0.503***		0.502***
			(0.178)		(0.178)
Time dummies		✓	✓	✓	✓
Adjusted R-squared	0.197	0.217	0.817	0.220	0.816
Observations	220	220	220	220	220

Notes: Parametric RD estimates. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). Conventional robust standard errors are reported in parentheses. Brackets report 95% CIs adjusted for small samples and skewed by using Bell and McCaffrey's [2002] degree-of-freedom correction (see Imbens and Kolesár [2016]). ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Table III

The Effect of the Goncourt on Reviewer Sentiment

| Timing of the review | Outcome: *Sentiment* | | | |
| | A. Pre-Goncourt | | B. Post-Goncourt | |
	(1)	(2)	(3)	(4)
Estimated coefficients of ordered logit model				
Goncourt	-0.214	0.153	-0.695***	-0.619***
	(0.346)	(0.412)	(0.204)	(0.228)
#Reviews (arcsinh)	0.016	0.019	0.274***	0.272***
	(0.038)	(0.038)	(0.031)	(0.031)
Average marginal effect of Goncourt on reviewers' sentiment				
Negative	0.044	-0.030	0.147***	0.130***
	(0.074)	(0.079)	(0.045)	(0.050)
Neutral	0.002	-0.005	0.013***	0.013***
	(0.002)	(0.016)	(0.002)	(0.002)
Positive	-0.046	0.035	-0.160***	-0.143***
	(0.072)	(0.094)	(0.044)	(0.050)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Log likelihood	-1,908	-1,907	-11,187	-11,186
Observations	1,770	1,770	10,772	10,772

Notes: RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. Column (1) of each panel fits a linear polynomial while Column (2) fits a quadratic one. The model specification follows Equation (7). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Standard errors clustered at the book level are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Table IV

Bandwagon Effect – Mediation Analysis

Degree of the polynomial	(1) Linear	(2) Quadratic
Panel A. Joint-significance test		
Goncourt -> Reviews (γ_1)	1.208***	1.002***
	(0.406)	(0.366)
Reviews -> Sales (δ_2)	0.198**	0.205**
	(0.087)	(0.088)
Panel B. Indirect effect		
$\gamma_1 \times \delta_2$	0.239	0.205
95% BCa CI[a]	[0.022, 0.597]	[0.003, 0.557]
95% MC CI[b]	[0.025, 0.540]	[0.004, 0.510]
Observations	220	220

Notes: Parametric RD mediation estimates. See Figure 4 for a description of the framework. The variable of interest, *Goncourt*, is a dummy that takes value one if a book has been awarded the Goncourt. Column (1) fits a linear polynomial while Column (2) fits a quadratic one. BCa CI = bias-corrected and accelerated bootstrap confidence interval. [a] Based on 10,000 sample bootstrapping. MC CI = quasi-Bayesian Monte Carlo confidence interval (Tingley *et al.* [2014]). [b] Based on 15,000 simulations. In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prizes*, *Female author*, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Footnotes

---

[1] A testable implication of the information effect is that prizes should be of greater benefit to books and authors that were less successful before the award than to those that are already familiar to consumers, because the prize will carry less information for the latter group than for the former. We explore that possibility in Section VII(i). The information effect may be magnified by bookstores and the media. Stores typically devote more space to award-wining books and single them out by displaying them in a special and visible place, using stickers, distinctive jackets, and various signs. Likewise, prizes draw media attention while publishers may concentrate their promotion efforts on award-winners.

[2] The "Lang Act", named after Minister of Culture Jack Lang and passed on August 10, 1981, is still in force today.

[3] An implication of that mechanism is that if consumers buy prize-winning books whereas the tastes of the experts who award the prize differ from those of consumers, the latter group may be disappointed when they read those books. We test that possibility in Section VII(ii).

[4] We assess the role of word of mouth in Section VII(iii).

[5] Because members of the jury must have had a literary career, they tend to be at least middle-aged. As of 2021, their average age is 69.5 years.

[6] The first selection is based on the suggestions of jury members, who can suggest several books. After initial first debate, the jury determines a first list of 15 books. For a book to be eligible for the Goncourt, it must be written in French, published by a French-language publisher, distributed in bookstores, and be sent by its publisher to each member of the jury before September 10.

[7] For example, this means that for a book published in 2000, the sales record does not begin until 2004.

[8] It may happen that books not included in the initial selection receive votes in the final one. To be consistent with the voting process and maximize the number of observations, we include those books in the sample. However, very similar estimates are obtained when excluding them for the analyses.

[9] https://www.edistat.com. Those sales figures are based on a sample group.

[10] In addition, to account for the fact that older books are more likely to have accumulated more sales than new ones, our specifications include a linear time trend or a set of dummy variables coding the year of competition, which corresponds to books' year of publication.

[11] https://www.anobii.com is a platform for book lovers that allows them to list the books they own.

[12] For example, $Votes$ is set to 5 for a book that was shortlisted but did not make it to the final round, as well as for a book that managed to get to the last round but then received no votes, and 7 for a book that received two votes in the last round.

[13] See the *Journal*'s editorial website for further details. In addition, when we restrict the sample to the books which reached the final selection and for which the coding strategy is thus redundant, we find quantitatively and qualitatively similar results (see Table D.VI in Online Appendix D4).

[14] Those are: Prix Renaudot, Prix Femina, Prix Interallié, Prix Médicis, Grand Prix du Roman de l'Académie française (awarded to a novel by the Academy), Prix du Livre Inter (awarded by a committee of listeners to France's main public radio channel, France Inter), Grand Prix des Lectrices de Elle (awarded by a committee of the women's magazine "Elle"), Prix des libraires (awarded by booksellers), and Prix Goncourt des Lycéens (awarded by high-school students from a list of books chosen by Académie Goncourt), and organized by the Ministry of Education; Ducas [2010]).

[15] Table A.I in Online Appendix A moreover reports descriptive statistics separately for awarded and non-awarded books.

[16] Formally, it means that if the functions $\mathbb{E}(Y_i(1)|X_i = x)$ and $\mathbb{E}(Y_i(0)|X_i = x)$ are continuous at $x = c$, where $Y_i(1)$ is unit's i outcome if it is exposed to the treatment, $Y_i(0)$ is unit's i outcome if it is not exposed to the treatment, X is the running variable, and c the cut-off, then the average causal effect of the treatment at the cut-off c is given by $\lim_{x \downarrow c} \mathbb{E}(Y_i|X_i = x) - \lim_{x \uparrow c} \mathbb{E}(Y_i|X_i = x) = \mathbb{E}(Y_i(1) - Y_i(0)|X_i = c)$.

[17] In Online Appendix B1, we show that $\log(Sales_{pre})$ is a smooth function of the Goncourt, suggesting that the concern is unfounded.

[18] As for $\log(Sales_{pre})$, this worry seems implausible since the big-publisher dummies are smooth functions of the Goncourt, as shown in Online Appendix B1.

[19] It can be argued, for example, that best-selling authors tend to work with larger publishers, which have bigger advertising budgets or wider networks for distributing to a wider range of stores.

[20] An alternative way of addressing the fact that older books have had more time to sell copies is to compute pre- and post-Goncourt sales over a well-defined window around the attribution of the prize. Table D.I in

Online Appendix D1 reports the results of doing so for a one- and a three-year window around the attribution of the prize. The conclusions remain unchanged.

[21] We are aware that parametric RDDs may yield noisy estimates by giving large weights to observations far away from the cut-off; they may also be sensitive to the degree of the polynomial and lead to confidence intervals that have poor coverage (Gelman and Imbens [2019]). To show that our results are robust to such concerns, we use non-parametric and local randomization approaches as alternative estimation strategies. The results are reported in Online Appendix D2 and are quantitatively and qualitatively similar to the parametric approach.

[22] Obtaining correct heteroskedasticity-robust standard errors and CIs may be problematic in small and/or skewed samples, since traditional robust standard errors, which typically rely on asymptotic properties, can be underestimated. As a result, the associated confidence intervals may have a coverage probability that is well below the nominal one (Imbens and Kolesár [2016]).

[23] The McCrary [2008] test further suggests that there is no systematic manipulation of the running variable around the threshold (see Figure B.I in Online Appendix B2).

[24] As there are only 15 observations at the right of the cut-off, we use a time dummy for each spell of five years instead of a time dummy for each year. This avoids consuming too many degrees of freedom and thus allows us to implement the test at the right of the cut-off.

[25] Presumably, the fact that the victory margin variables are no longer significant is driven by the fact that specifications (4) and (5) exhibit a high level of collinearity.

[26] Hainmueller et al. [2019] remark that CMEs may be biased if the linear interaction effect assumption of multiplicative interaction models does not hold and if there is a lack of common support for the moderator. This is because estimates will rely on extrapolating the functional form to an area where there is low empirical support. In our case, the concern is that awarded books tend to experience higher pre-Goncourt sales than non-awarded ones, as average pre-Goncourt sales amount respectively to 74,560 and 25,533 copies. At low levels of the moderator, this means that there will be little variation in the treatment as few books with low pre-Goncourt sales have won the prize.

27 We obtain similar results when using the conventional linear interaction model. The raw coefficients of the model are reported in Table E.I while the conditional coefficients are summarized in Figure E.I of Online Appendix E1.

28 A description of the service can be found at https://www.ibm.com/cloud/watson-natural-language-understanding. Watson Natural Language Understanding is publicly available and free of charge (for limited usage) and has been found to outperform other widely used natural language understanding tools (see e.g., Abdellatif *et al.* [2022]; Canonico and De Russis [2018]; Carvalho and Harris [2020]; Feine *et al.* [2019]). In addition, in contexts where detecting a negative text is of interest, Ermakova *et al.* [2021] recommend using Watson Natural Language Understanding because its ability to correctly detect negative sentiments surpasses other popular tools.

29 In addition, 86 percent of the books have received at least 5 reviews and 72 percent have received at least 10 reviews. Figure E.II in Online Appendix E2 plots the distribution of the number of reviews in our sample. Excluding the books with less than 5 or 10 reviews from our regressions does not change the results, as shown by Tables E.II and E.III.

30 The inverse hyperbolic sine is defined as $\text{arcsinh}(z) = \log(z + \sqrt{1 + z^2})$. Bellemare and Wichman [2020] propose the value of 10 as a rule of thumb to assess whether z is sufficiently large. Since the untransformed mean of $\#Reviews_{ciy}$ is equal to 201, we can safely interpret the elasticities derived with the arcsinh transformation as we would have done with a log transformation.

31 In Table E.V of Online Appendix E2, we show that using an ordered probit model leads to very similar results. In addition, to make sure that the findings are not driven by the assumption that opinions are ordered from negative to positive with neutral in between, we estimate a multinomial logit model, which makes no assumption on the ordering of sentiments. The results are shown in Table E.VI and lead to conclusions that are in line with those of the ordered models: the Goncourt decreases the probability that a consumer posts a positive review and increases the probability of writing a negative review.

32 In addition, we observe that the correlation between including the words "Goncourt" or "goncourt" in a review and the sentiment of the review is negative and statistically significant at the one-percent level, even after controlling for the effect of the Goncourt itself. More precisely, including the name of the prize in a

review is associated with a 10% increase in the probability of the review being negative and a of 11% lower probability of it being positive. We thank an anonymous referee for suggesting us to look at that correlation.

[33] The raw coefficients of the model are reported in Table E.VII in Online Appendix E2.

[34] If quality is defined as the ability to please consumers, the finding echoes Ginsburgh and Weyers' [2014] argument that artistic contests do not always select the best candidate. Referring to a difference in tastes, however, avoids making a value judgment.

[35] A similar deterioration in reviews might be the outcome of consumers blindly following their peers. However, that possibility is less likely because peers' tastes are more similar than those of experts. Moreover, we have shown that peer reviews mitigate the negative effect of the prize on the type of review and that a prize awarded by non-experts does not affect reviews. In any case, we study the role of word of mouth in the next section.

[36] As most books have a low number of reviews, we use a 6-month window to avoid having too many books with zero reviews, which would both curb the representativeness of the results and bias the OLS estimates. In Table E.X of Online Appendix E4.2, we show that the findings are robust to alternative time-windows.

[37] The timing of measurement of the variables of interest included in our RD mediation framework implies that the temporal assumption of causal mediation is satisfied (VanderWeele [2015]). Specifically, the treatment (being awarded the Goncourt) precedes the mediator (reviews), which in turn precedes the outcome (sales). One month separates each measurement. The identifying conditions of the RD mediation model are arguably more stringent than those of our basic baseline RD. Indeed, whereas the baseline RD approach only requires path δ_1 to be causal, the identification of the indirect effect $\gamma_1 \times \delta_2$ further requires (i) path γ_1 (Goncourt-reviews relationship) to be causal, (ii) path δ_2 (reviews-sales relationship) to be causal, and (iii) that there is no confounder in the reviews-sales relationship that is affected by the Goncourt (VanderWeele [2015]). Condition (i) is fulfilled by design, as it is estimated implementing the same RD approach as in the baseline simply using this time post-Goncourt reviews as outcome instead of post-Goncourt sales. To show that condition (ii) also holds, we show in Table E.XI of Online Appendix E4.3 that using an instrumental variable approach leads to similar results. Finally, to make condition (iii) more plausible, we followed VanderWeele [2015] and used a short period (i.e., one month) between the measurement of the treatment and the mediator. As noted by the author, such a time interval makes causality more plausible while still allowing the effect of the treatment to materialize."