CEBRIG Working Paper

Experts, Information, Reviews, and Coordination: Evidence on How Literary Prizes Affect Sales

Nicolas Lagios, Pierre-Guillaume Méon

We study the causal effect of literary awards on book sales, using France's most prestigious prize, the Goncourt. For this, we implement a regression discontinuity design, taking advantage of the fact that a committee of experts gives the prize to the book receiving the most votes. We observe that the Goncourt increases sales by 350 percent and that this effect is larger for books that sold fewer copies before the award. Additional results show that the prize results in more reviews on Amazon but increases the probability that they are negative. Finally, we report that the effect on sales is partly driven by an increase in word of mouth. These findings are consistent with a model where the Goncourt provides information on the existence of a book and where consumers use the prize as a quality signal and a coordination device but, as a result, read books that are too far from their tastes. This interpretation is backed by the finding that, despite its positive effect on sales, the Goncourt des Lycéens, a prize based on the same list of books as the Goncourt but awarded by a group of high-school students whose tastes are arguably doser to the public's than to those of experts, has no effect on the sentiment of reviews.

Keywords Awards, Literary prizes, Sales, Experts

JEL Classifications Z11, L15, L82, D12

CEBRIG Working Paper N°21-011
July 2021

Université Libre de Bruxelles - Solvay Brussels School of Economics and Management
Centre Emile Bernheim de Recherche Interdisciplinaire en Gestion
ULB CP114/03 50, avenue F.D. Roosevelt 1050 Brussels BELGIUM
cebrig@ulb.be - Tel : +32 (0)2/650.48.64

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

Nicolas Lagios[a], Pierre-Guillaume Méon[b]

This version: July 8, 2021

Abstract. We study the causal effect of literary awards on book sales, using France's most prestigious prize, the Goncourt. For this, we implement a regression discontinuity design, taking advantage of the fact that a committee of experts gives the prize to the book receiving the most votes. We observe that the Goncourt increases sales by 350 percent and that this effect is larger for books that sold fewer copies before the award. Additional results show that the prize results in more reviews on Amazon but increases the probability that they are negative. Finally, we report that the effect on sales is partly driven by an increase in word of mouth. These findings are consistent with a model where the Goncourt provides information on the existence of a book and where consumers use the prize as a quality signal and a coordination device but, as a result, read books that are too far from their tastes. This interpretation is backed by the finding that, despite its positive effect on sales, the Goncourt des Lycéens, a prize based on the same list of books as the Goncourt but awarded by a group of high-school students whose tastes are arguably closer to the public's than to those of experts, has no effect on the sentiment of reviews.

*We are grateful to Abel François, Victor Ginsburgh, Constantin Lagios, Hugues Pirotte, Judicael Poumay, Anne-Sophie Radermecker, Ilan Tojerow, Denni Tommasi, Franck Venmans, Vincenzo Verardi, and the participants of the 2021 Meeting of the of the European Public Choice Society for useful comments. We are also particularly indebted to Bernard Pivot, Françoise Rossinot, and the Nancy Municipal Archives for granting us access to the archives of the Goncourt Academy. Special thanks go to Ilias Papadopoulos who happily drove Nicolas Lagios to Nancy at 4am, helped him with the data collection during several days, and shared with him some amazing meals at Flunch. The authors claim the joint responsibility for remaining errors and omissions and will privately argue to determine which one of them is to blame.
[a] Université libre de Bruxelles (ULB), Centre Emile Bernheim, CP-114/03, avenue F.D. Roosevelt 50, 1050 Bruxelles, Belgium. (nicolas.lagios@ulb.be).
[b] Université libre de Bruxelles (ULB), Centre Emile Bernheim, CP-114/03, avenue F.D. Roosevelt 50, 1050 Bruxelles, Belgium. (p-guillaume.meon@ulb.be).

1 Introduction

If literary 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 literary 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 & Shugan, 1997, Ponzo & 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. Literary 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 & van Ours, 2003, Ginsburgh, 2003, Ponzo & 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. Prizes may accordingly work as coordination devices. 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 50$^{\text{th}}$ 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.

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 prizewinning 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 customers' reviews on Amazon.fr, and we and find that the Goncourt negatively affects their opinions. This finding is consistent with a model where consumers follow the advice of the experts on the prize committee but have tastes that differ from those of those 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 no 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 are. 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 cultural awards (Ashworth et al., 2010, Frey & Gallus, 2017; Ginsburgh, 2003, Ponzo and Scoppa, 2015). It does so by confirming the causal effect of 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 & van Ours, 2003, Hilger et al., 2011, Ginsburgh et al. 2019, Ekelund et al., 2020, Reinstein & 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 & Wu, 2020, Reimers & 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 2 discusses in more detail the three channels through which prizes can affect sales. Section 3 provides background information on the history and functioning of the Goncourt. Section 4 describes the data. Section 5 discusses the empirical strategy. Section 6 presents the baseline findings while Section 7 provides evidence on the mechanisms. Section 8 concludes.

2 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.[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,

[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 7.1. 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.

Reinstein and Snyder, 2005), and wines (Hilger et al., 2011). 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.[4] 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.[5]

3 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,

[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 7.2.

[4] Chung and Cox (1994) put forward a model leading to the same conclusion, where consumers sequentially randomly choose the artistic products they consume and are more likely to buy those that have been bought by a larger number of other consumers. The authors show that the distribution of gold records in the music industry is in line with the model.

[5] We assess the role of word of mouth in Section 7.3.

essayists, philosophers, screenwriters, etc.) on the French literary scene.[6] 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 15 books on average, then 8 and finally 4.[7] 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.

4 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.[8] We thus observe 220 books, including 15 winners.[9] 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

[6] 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.

[7] 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 fifteen 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.

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

[9] 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.

prize already has a commercial effect. For each book, we collected data on its sales, votes, and characteristics.

4.1. 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.[10] 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.[11] 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.[12] 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.

4.2. Data on votes

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

[11] 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.

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

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.[13] 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 the appendix that the results are not driven by the way that votes are coded.[14]

4.3. 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.[15]

---

[13] 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.

[14] In Appendix B2 we confirm that the findings are robust to alternative coding strategies. 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 B5 and Table B6).

[15] 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

Table 1 reports descriptive statistics. 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.

Table 1. 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 4.

5 The regressions discontinuity design

5.1 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

main public radio channel, France Inter), Grand Prix des Lectrices de Elle (awarded by a committee of 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).

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.

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:

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

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

[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 $\lim_{x \downarrow c} \mathbb{E}(Y_i|X_i = x) - \lim_{x \uparrow c} \mathbb{E}(Y_i|X_i = x) = \mathbb{E}(Y_i(1) - Y_i(0)|X_i = c)$.

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,

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

5.2 Empirical model

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

$$\begin{aligned} \log(Sales_{iy,post}) & \\ &= \alpha + \tau Goncourt_{iy} + f(Goncourt_{iy}, Margin_{iy}) + \phi \log(Sales_{iy,pr}) \\ &+ \boldsymbol{\theta}'\mathbf{X}_{iy} + \lambda_y + \epsilon_{iy}, \end{aligned} \tag{3}$$

where

$Sales_{iy,po}$ is the number of post-Goncourt sales for book i competing in year y. In our baseline analyses, we use the logarithmic transformation of $Sales_{iy,po}$, 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;

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

$Sales_{iy,pr}$ 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.

---

[17] In Appendix A1, we show that $Sales_{pre}$ is a smooth function of the Goncourt, suggesting that the concern is unfounded.

[18] As for $Sales_{pre}$, this worry seems implausible since the big-publisher dummies are smooth functions of the Goncourt, as shown in Appendix A1.

[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.

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

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.[20]

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.[21] 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 $\sqrt{\hat{\mathbb{V}}_{\mathrm{BM}}} = \sqrt{\hat{\mathbb{V}}_{\mathrm{HC2}}} \times (t_{0.975}^{K_{\mathrm{BM}}}/1.96)$ where $\hat{\mathbb{V}}_{\mathrm{HC2}}$ is the variance estimator proposed by MacKinnon and White (1985) and $t_{0.975}^{K_{BM}}$ is the q^{th} quantile of the t-distribution with K degrees of freedom (Imbens & Kolesár, 2016). We then use $\sqrt{\hat{\mathbb{V}}_{\mathrm{BM}}}$ to construct the CIs for the parameter of interest.

5.3 Identification assumptions

[20] 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 & 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 Appendix B1 and are quantitatively and qualitatively similar to the parametric approach.

[21] 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 & Kolesár, 2016).

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. Overall, we find no robust evidence of a discontinuity of the control variables at the cut-off.[22]

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 & 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 appendix A1, we show this concern to be unfounded by demonstrating that sales prior to the prize are not discontinuous at the cut-off.

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

[22] We report those results and comment on them in more detail in Appendix A1.

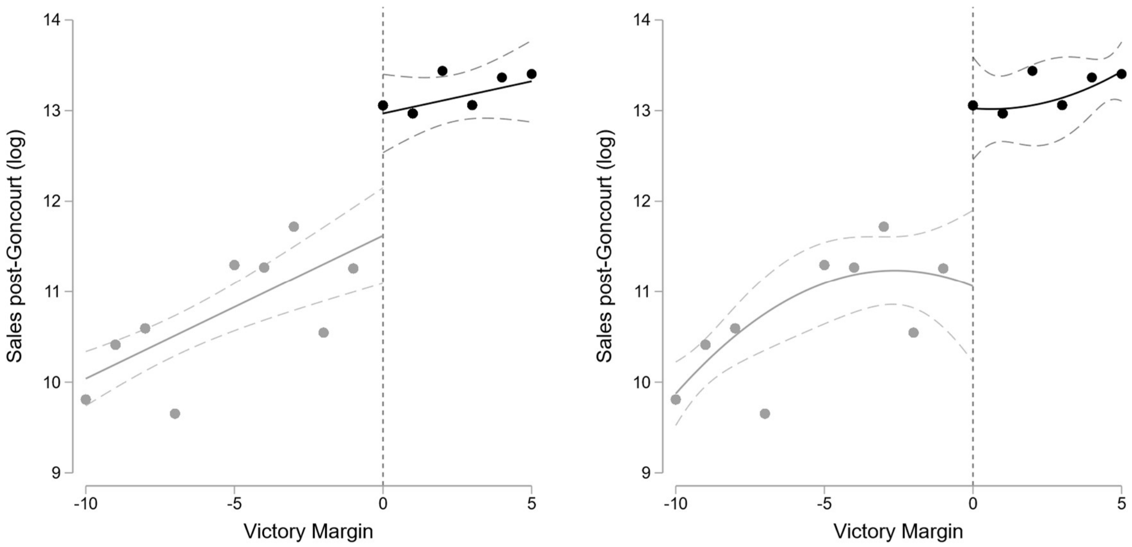
an RDD in each subsample by using as the median of the running variable as the cut-off.[23] We find no evidence of discontinuity at either side[24].

6 Baseline results

6.1 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.

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.

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

[23] The results are reported in Appendix A2.

[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.

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]).

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.

6.2 Regression discontinuity estimates

Table 2 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.

Table 2. The Effect of the Goncourt on Book Sales

	(1)	(2)	(3)	(4)	(5)
			Outcome: $\log(\boldsymbol{Sales_{post}})$		
Goncourt	1.406***	1.172***	1.497***	1.886***	1.375***
	(0.329)	(0.430)	(0.305)	(0.488)	(0.368)
	[0.680, 2.131]	[0.230, 2.115]	[0.827, 2.166]	[0.708, 3.064]	[0.472, 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)
$\log(\text{Sales}_{\text{pre}})$			0.852***		0.857***
			(0.066)		(0.067)
Movie			0.802***		0.786***
			(0.240)		(0.243)
Other prize			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 model specification follows Equation (3). The variable of interest is *Goncourt Prize* which is a dummy that takes value one if a book has been awarded the Goncourt. 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.

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 prize* 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 books on average enjoy greater commercial success, and those published by Seuil less commercial success, than other selected 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.

6.3 Robustness checks

To assess the sensitivity of our baseline results, we perform a set of robustness checks. The outcomes of these tests are reported in Appendix B.

Alternative RDD approaches

To explore the sensitivity of our parametric estimates, we implement two alternative RDD strategies: a non-parametric approach and a local randomization approach. The first 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 (Gelman & Imbens, 2019). The local randomization approach allows us to switch from a large-sample approximation framework to a finite sample framework that is better tailored to small-sample inference (Cattaneo et al., 2017).[26] Reassuringly, the results are relatively insensitive to the approach, thus demonstrating the strength of our parametric framework.

Alternative victory margin coding strategies

As only the votes of the last round of the final selection are available, we automatically add five votes to the books chosen as finalists, in order to draw a distinction between shortlisted books and the others. In an RDD framework, such a coding strategy has little influence in principle on the results, since observations far away from the cut-off are

[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] Each approach is discussed in greater detail in Appendix B1.

given little weight. However, to further show that the findings are insensitive to the coding strategy, we use alternative ways of coding the victory margin.[27] The estimates are similar, both in magnitude and in significance, which reassures us about how little the strategy impacts the results.

In addition, in Column (2) of Table B5 and Table B6, we re-estimate Equation (3), this time restricting the sample to shortlisted books. In this way, the coding strategy is redundant as only the effective number of votes matters. The results remain similar.

Regressions robust to outliers

A legitimate concern would be that the results are driven only by a subgroup of very successful prize-winning books or by a subgroup of highly unsuccessful non-prizewinners, while most awarded books do not experience higher sales. To address that concern, 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 has the advantage of being insensitive to outliers (Wooldridge, 2010). The magnitude and the significance of the Goncourt are analogous to the baseline.[28]

Alternative samples and specifications

Finally, we further explore the sensitivity of our baseline results by considering a series of alternative specifications.[29]

In the baseline, we restrict the sample to the period 2004-2018 because sales are available only from 2004 onward. To extend our results to books that won the Goncourt before 2004, we use the entire period for which votes and sales are available, specifically 1954-2018, with the caveat that the sales figures for books published before 2004 do not include data prior to 2004. We now observe 854 books, including 64 laureates. The outcome is reported in Column (1) of Table B5 and Table B6. The Goncourt effect remains

[27] Each strategy is described more precisely in Appendix B2.

[28] The results are shown in Appendix B3.

[29] The results of the linear and quadratic specifications are reported in Table B5 and Table B6 of Appendix B4, respectively.

significant and is slightly stronger than before. This stronger effect may pick up the long-run effect of the prize or simply reflect the fact that we do not control for the book's pre-Goncourt sales trend and that the data are less precise for old books.

In Column (2), we focus on shortlisted books. Again, the Goncourt is significant and has a similar impact on sales.

In Column (3), we introduce a dummy for each publisher (28 dummies in total). In this way, we are able to capture unobserved heterogeneity among publishers: for example, some may systematically attract more successful authors, or spend more on advertising. The results are similar to the baseline in terms of magnitude and significance.

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). Reassuringly, the conclusions stay similar to the baseline.

7 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.

7.1 Information effect

To explore the information effect, 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 $\log(Sales_{iy,pr}\ )$ so as to estimate the following treatment-covariate interaction model:

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

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

$$
(\Delta \mathrm{S}ales_{iy,post} | Goncourt_{iy} = 1) = \tau + \mu \log(Sales_{iy,pre}). \tag{5}
$$

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.[30] 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.[31]

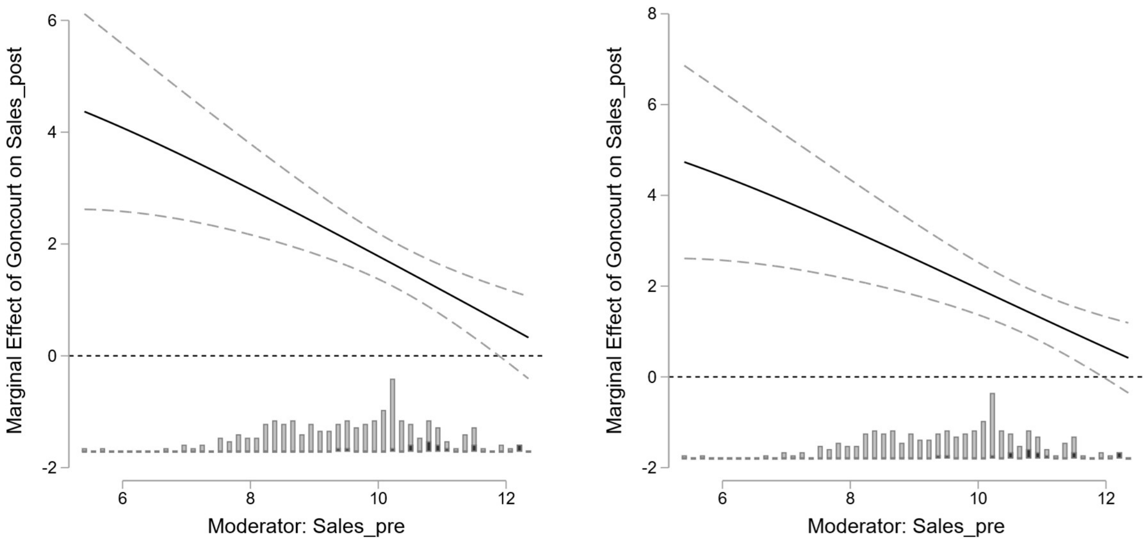
Both graphs show that the impact of the Goncourt, though always positive, decreases with the recipient's pre-award sales. In addition, for very popular books, the

[30] 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.

[31] We obtain similar results when using the conventional linear interaction model. The raw coefficients of the model are reported in Table C1 while the conditional coefficients are summarized in Figure C1 of Appendix C1.

marginal effect of the Goncourt becomes statistically insignificant. This finding is in line with the existence of an information effect whereby 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.

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 prize*, *Female author*, the four publisher dummies, and the time dummies. The dash line reports 90% confidence intervals based on robust standard errors.

7.2 Quality signal

If prizes are quality signals and the judgement 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.[32] A Goncourt-nominated book receives on average 350 reviews, of which 196 (56%) are positive and 85 (24%) are negative.

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

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

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.

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:

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

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

[32] A description of the service can be found at https://www.ibm.com/cloud/watson-natural-language-understanding.

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).[33] 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)(3) is estimated using OLS while (7) is estimated using an ordered logit model, as the dependent variable follows a natural ordering.[34] 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 3 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.

---

[33] 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.

[34] In Table C2 of Appendix C1, we show that using an ordered probit model leads to very similar results.

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.

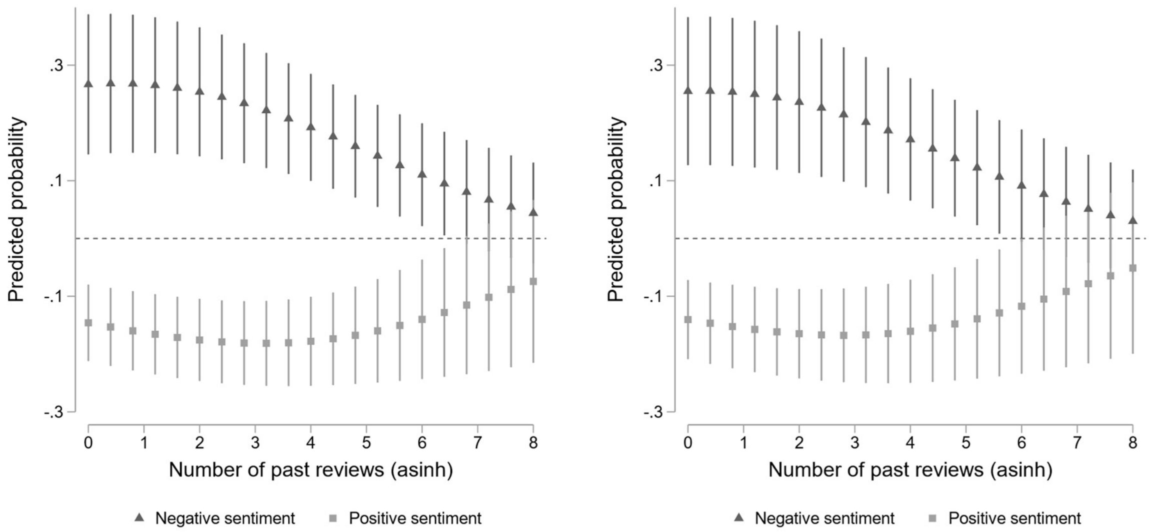
Table 3. Effect of the Goncourt on Reviewer Sentiment

	Outcome: ***Sentiment***			
Timing of the review	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	-1908	-1907	-11187	-11186
Observations	1,770	1,770	10,772	10,772

Notes: RD estimates. 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 prize*, *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.

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.

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 prize*, *Female author*, the four publisher dummies, and the time dummies. Vertical lines report 90% confidence intervals based on robust standard errors.

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.[35] 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

[35] The raw coefficients of the model are reported in Table C3.

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 3 and Figure 3 are in line with the hypothesis that 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.[36] 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

[36] 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 judgement.

difference-in-differences model, we show that, despite having a positive and significant effect on book sales, the Goncourt des Lycéens has no impact on consumers' satisfaction.[37] 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.[38]

7.3 Coordination device

To test whether the Goncourt plays the role of a coordination device, we 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.

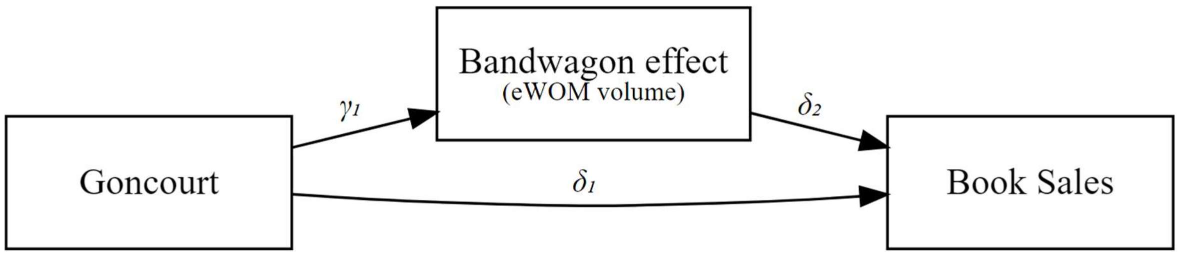
We accordingly extend our RD framework to perform a mediation analysis. This allows to explore whether and how an independent variable of interest affects an outcome variable through a third one (Baron & Kenny, 1986, Hayes, 2017). In our case, we are interested in assessing whether the effect of the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

[37] The results are reported and discussed in Appendix C4.

[38] 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.

Goncourt increases the number of reviews, which consequently boosts the number of sales.

Figure 4. Mediation Analysis



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.

A formal definition of the RD mediation framework is provided in Appendix C4. 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.[39] 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 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.[40]

[39] 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 OLS estimates. In Appendix C5, we show that the findings are robust to alternative time-windows.

[40] The timing of measurement of the variable of interest included in our RD mediation framework implies that the temporal assumption of causal mediation is satisfied (VanderWeele, 2015). Specifically, the treatment precedes the mediator, which in turn precedes the outcome. One month separates each measurement. That time interval allows the effect of the treatment to materialize without threatening the causal interpretation of the results because of unmeasured confounding factors in the mediation-outcome relationship. We moreover address this particular concern by using an instrumental variable approach. The procedure is described in Appendix C6 and the results are reported in Appendix Table C6.

Table 4 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 six months after the prize is awarded (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 & 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 & 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.

Table 4. Bandwagon Effect – Mediation Analysis

	(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. 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 prize*, *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.

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.

8 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. 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.

Those findings can be further supplemented to understand in more detail how prizes work. First, it may be wondered whether the channels of transmission that we observe in literature are at work in other fields where prizes are awarded, such as other art forms, wine, journalism, and research, to name but a few. 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

Adler, M. (1985). Stardom and talent. *American Economic Review*, 75(1), 208-212.

Ashworth, J., Heyndels, B., & Werck, K. (2010). Expert judgements and the demand for novels in Flanders. *Journal of Cultural Economics*, 34(3), 197-218.

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

Babić Rosario, A., Sotgiu, F., De Valck, K., & 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(3), 297-318.

Baron, R. M., & 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(6), 1173-1182.

Baum, C., & Schaffer, M. (2019). IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments.

Becker, S. O., Egger, P. H., & 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(4), 29-77.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28(2), 169-182.

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1), 50-61.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29(5), 815-827.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401), 123-127.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512), 1753-1769.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A practical introduction to regression discontinuity designs*: Foundations*. Cambridge University Press.

Cattaneo, M. D., Titiunik, R., & 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(3), 643-681.

Canoy, M., Van Ours, J. C., & Van Der Ploeg, F. (2006). The economics of books. In Ginsburgh, V. A., & Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, 1, 721-761, Elsevier.

Chen, M. X., & M. Wu. (2020). The value of reputation in trade: Evidence from Alibaba. *Review of Economics and Statistics*, 1-45.

Chung, K.H. & Cox, R.A. (1994). A stochastic model of superstardom: An application of the Yule distribution. *Review of Economics and Statistics*, 76(4), 771-775.

Clement, M., Proppe, D., & Rott, A. (2007). Do critics make bestsellers? Opinion leaders and the success of books. *Journal of Media Economics*, 20(2), 77-105.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics*, 16(1), 67-88.

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. & Jackson, J.D. (2020). ART as meta-credence: authentication and the role of experts. *Journal of Cultural Economics*, 44(1), 155-171.

Eliashberg, J., & Shugan, S. M. (1997). Film critics: Influencers or predictors? *Journal of Marketing*, 61(2), 68-78.

English, J. F. (2014). The economics of cultural awards. In Ginsburgh, V. A., & Throsby, D. (Eds.), *Handbook of the Economics of Art and Culture*, 2,119-143), Elsevier.

Frey, B. S., & Gallus, J. (2017). Towards an economics of awards. *Journal of Economic Surveys*, 31(1), 190-200.

Fritz, M. S., Taylor, A. B., & Mackinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47(1), 61–87.

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3), 447-456.

Genova, P. A. (2014). Rewarding the production of culture: Le Prix Goncourt. *Contemporary French and Francophone Studies*, 18(2), 150-157.

Ginsburgh, V. (2003). Awards, success and aesthetic quality in the arts. *Journal of Economic Perspectives*, 17(2), 99-111.

Ginsburgh, V., Radermecker, A.S., & 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, 36-50.

Ginsburgh, V., & Van Ours, J. C. (2003). Expert opinion and compensation: Evidence from a musical competition. *American Economic Review*, 93(1), 289-296.

Ginsburgh, V. & Weyers, S. (2014). Nominees, winners, and losers. *Journal of Cultural Economics*, 38(4), 291-313.

Hainmueller, J., Mummolo, J., & Xu, Y. (2019). How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Political Analysis*, 27(2), 163-192.

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford publications.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24(10), 1918-1927.

Hilger, J., Rafert, G., & 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(4), 1289-1296.

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98(4), 701-712.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281-355.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3), 305-325.

Mishra, V., & Smyth, R. (2015). Estimating returns to schooling in urban China using conventional and heteroskedasticity-based instruments. *Economic Modelling*, 47, 166-173.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311-329.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Ponzo, M., & Scoppa, V. (2015). Experts' awards and economic success: evidence from an Italian literary prize. *Journal of Cultural Economics*, 39(4), 341-367.

Reimers, I., & Waldfogel, J. (2021). Digitization and pre-purchase information: The causal and welfare impacts of reviews and crowd ratings. *American Economic Review*, 111(6): 1944–1971.

Reinstein, D. A., & 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(1), 27-51.

Sorensen, A. T., & Rasmussen, S. J. (2004). Is any publicity good publicity? A note on the impact of book reviews. Working Paper, Stanford University.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yzerbyt, V., Muller, D., Batailler, C., & 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(6), 929.

Zerilli, S. (2015). Compte rendu de Ducas (Sylvie), La littérature à quel (s) prix ? Histoire des prix littéraires. Paris, La Découverte, coll. « Cahiers libres », 2013. *COnTEXTES. Revue de sociologie de la littérature*.

Appendix

Appendix A. Identification assumptions and falsification tests

Appendix A1. Covariates balance

This section presents and discusses in more details the covariate balance tests to assess whether covariates vary smoothly at the cut-off (see Section 5.3 in the paper). The results are reported in Table A1.

Column (1) reports the results of the parametric RDD approach with a linear specification. Column (2) shows the estimates of the non-parametric RDD approach. For each covariate, a new optimal bandwidth is computed following Imbens and Kalyanaraman (2012). Finally, Column (3) presents the estimates for the local randomization approach; p-values are computed using the Monte Carlo permutation test. The non-parametric and the local randomization approaches are discussed in detail in Appendix B1, which is entirely devoted to those methods.

The results show no evidence of discontinuity, except for the variables *Other prize* and *Actes Sud*. The former is unsurprising as *Other prize* includes prizes that are directly influenced by the Goncourt.[41] Accordingly, this is neither a predetermined nor a placebo covariate. The fact that *Actes Sud* is unsmooth at the cut-off may be more surprising as Actes Sud is a small publisher that has won "only" five Goncourt prizes out of the 116 editions. However, because those five wins are all concentrated between 2004 and 2018, the time span of our baseline estimates, this may explain why winning the Goncourt is positively correlated with *Actes Sud* in our sample. In addition, when we use the other two RDD approaches, *Other prize* and *Actes Sud* are never significant. This may suggest that the above results are due solely to random chance.

[41] For example, the Renaudot prize is awarded immediately after the Goncourt and aims at repairing the latter's injustices. In addition, two laureates are chosen, in case the first choice has already received the Goncourt.

In any case, since we control for those variables in our estimates, we avoid any bias due to unbalancedness.

Table A1. Covariate Balance

	(1)	(2)	(3)
RDD approach	Parametric	Non-parametric	Local randomization
$Sales_{pre}$	0.286	0.443	0.684
	(0.409)	(0.530)	[0.160]
Movie	-0.021	0.189	-0.156
	(0.130)	(0.227)	[0.374]
Other prize	-0.661***	-0.406	-0.281
	(0.146)	(0.424)	[0.226]
Female author	-0.036	0.233	0.031
	(0.196)	(0.355)	[0.905]
Gallimard	-0.207	-0.536	-0.125
	(0.160)	(0.425)	[0.619]
Grasset	-0.133	-0.214	0.031
	(0.143)	(0.331)	[0.868]
Seuil	0.064	8.09e-16	5.22e-17
	(0.164)	(1.133)	[1.000]
Actes Sud	0.440**	0.403	0.313
	(0.197)	(0.452)	[0.100]
Implied bandwidth	∞	Optimal[a]	2
Observations	220	-	34

Notes: Column (1) implements a parametric linear RD; for further details see notes to Table 2. Column (2) implements a nonparametric RD with uniform kernel; the optimal bandwidth is computed following Imbens and Kalyanaraman (2012). Column (3) implements randomization tests; p-values are computed using the Monte Carlo permutation test. In all specifications, we control for time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

In addition, running a regression from which we exclude books having won other prizes or published by Actes Sud leads to very similar results, as shown by Table A2.

Table A2. Without Books Having Won Other Prizes and Published by Actes Sud

	Outcome: $\log(\boldsymbol{Sales_{post}})$			
	A. Linear		B. Quadratic	
	(1) Without books that have won other prizes	(2) Without books published by Actes Sud	(1) Without books that have won other prizes	(2) Without books published by Actes Sud
Goncourt	1.350***	1.837***	1.086***	1.827***
	(0.320)	(0.341)	(0.412)	(0.444)
Margin	0.062***	0.045**	0.102	-0.003
	(0.020)	(0.020)	(0.128)	(0.096)
Observations	160	208	160	208

Notes: Parametric RD estimates. Panel A fits a linear polynomial while Panel B fits a quadratic one. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *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 at 10% level.

Appendix A2. Placebo cut-offs

The RDD rests on the assumption that the cut-off at the victory margin of zero distinguishes the winners of the book from near-winners with identical unobserved characteristics, so that the jump in sales reflects the causal impact of winning the prize. The causal interpretation of the RDD estimates would be threatened if arbitrary cut-offs resulted in similar jumps.

Table A3. Placebo Cut-Offs

	Outcome: $\log(\boldsymbol{Sales_{post}})$	
	(1) Left of the cut-off	(2) Right of the cut-off
Goncourt	-0.171	-0.966
	(0.354)	(4.594)
Observations	205	15

Notes: Parametric RD estimates. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *Female author*, the four publisher dummies, and the time dummies (one dummy for each year in Column (1) and one dummy for each spell of five years in Column (2)). Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

We test for jumps at arbitrary cut-offs by following Imbens and Lemieux's (2008) recommendation. Specifically, we separately perform an RDD on the subsamples consisting respectively of the observations at the left and those at the right of the cut-off, using the median of the running variable in each subsample as cut-off. Table A3 shows no evidence of discontinuity at either side of the cut-off, as the coefficient of the Goncourt dummy variable turns out statistically insignificant at standard levels.

Appendix B . Robustness checks on the effect of the prize on sales

Appendix B1. Alternative RDD approaches

In this section, we implement two alternative RDD strategies: a non-parametric approach and a local randomization approach.

Alternative approach 1. Non-parametric RDD

As the first alternative to the parametric strategy, we conduct a non-parametric RDD. This consists in implementing a linear regression on both sides of the cut-off using only observations that lie within a specific window called bandwidth (Hahn, Todd, & Van der Klaauw, 2001). The running variable *Margin* exhibits few mass points whereas the RDD's conventional non-parametric framework relies on the assumption that the running variable is continuous, so special attention should be paid to the appraisal of confidence intervals. In particular, when the running variable is discrete, standard CIs may have poor coverage (Lee & Card, 2008). To address this issue, we follow Armstrong and Kolesár (2018) and Kolesár and Rothe (2018) and construct "honest" CIs by using the bounded second derivative (BSD) procedure which requires choosing a constant K that bounds the second derivative of the conditional expectation function (Kolesár and Rothe, 2018).

Following the heuristics explained in Kolesár and Rothe (2018), we view $K = 0.03$ as a good choice. Moreover, we also consider $K = 0.06$ and $K = 0.09$ in order to show the sensitivity of the results to different K choices, bearing in mind that the higher K, the more conservative the approach.[42]

The estimates associated with each K are reported in Table B1. For each value of K, the bandwidth chosen minimizes the length of the CIs. It can be seen that even in the most pessimistic case where $K = 0.09$, the coefficient of *Goncourt* is still significant. As expected, the BSD CIs are more conservative than the traditional CIs based on Eicker-Huber-White (EHW) standard errors. Overall, the estimates are very similar to the baseline, thus showing the strength of the results.

Table B1 The Impact of the Goncourt on Sales – Non-Parametric RDD

	(1)	(2)	(3)
	Outcome: $\log(\boldsymbol{Sales_{post}})$		
K	0.03	0.06	0.09
Estimate	1.507	1.413	1.413
BSD 95% CI	(0.511, 2.504)	(0.394, 2.433)	(0.223, 2.604)
EHW 95% CI	(0.599, 2.416)	(0.570, 2.257)	(0.570, 2.257)
Implied bandwidth	5	4	4
Effective # of observations	63	50	50

Note: Non-parametric RD estimates with uniform kernel. BSD refers to the bounded second derivative procedure which is used to construct "honest" CIs, as considered in Armstrong and Kolesár (2018) and Kolesár and Rothe (2018). K is the bound of the second derivative of the conditional expectation function and is fixed according to the heuristics explained in Kolesár and Rothe (2018). The bandwidth chosen minimizes the length of the CIs for each K and is computed according to Silverman's rule of thumb (Imbens & Kalyanaraman, 2012). EHW refers to the CIs obtained using the conventional Eicker-Huber-White standard errors. In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *Female author*, the four publisher dummies, and the time dummies. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Alternative approach 2. Local randomization RDD

In this section, we explicitly take into account the potential randomization nature of the RDD and its additional assumptions, which allows the use of specific randomization methods. To do that, we follow Cattaneo, Frandsen, & Titiunik (2015) and Cattaneo, Titiunik, & Vazquez-Bare (2017), who formalize and discuss the differences between the

[42] We estimate a lower bound for K by following the method described in the online supplements to Kolesár and Rothe (2018) and Armstrong and Kolesár (2018), and obtain a point estimate of 0.04. This suggests that our initial choice of $K = 0.03$ may be seen as optimistic while $K = 0.09$ may be seen as pessimistic or conservative.

randomization and continuity-based frameworks. Adopting a local randomization approach has the advantage of allowing us to switch from a large sample approximation framework to a finite sample framework, better suited for small-sample inference.

The randomization setting requires some additional assumptions to those used in the continuity-based RDD framework. The crucial feature is the existence of a window W_0 in which:

Assumption 1 (Local randomization mechanism) *Placement above or below the cut-off does not depend on the potential outcomes.*

Assumption 2 (Local stable unit treatment value assumption) *The potential outcomes do not depend on the running variable except through the treatment assignment.*

If the randomization assumption holds, it must hold for the smallest window possible. Thus, in a discrete setting the window W_0 will be the interval containing the first mass point at the left of the cut-off and the first one at the right. In our setting, this implies a window that includes the books for which $Margin = -1$ (control group) and $Margin = 0$ (treatment group). However, since there are only two books in the treatment group within this window whereas Cattaneo et al. (2015) recommend at least 10 observations at each side of the cut-off, we expand the right window to $Margin = 2$ in order to have 10 treated books. Therefore, if randomization holds, it must hold for the window: $W_0 = [Votes_n = -2; Votes_n = 2]$.

Inside W_0, since the votes differ by only a small amount, it is no heroic assumption to consider that the books included in W_0 have a similar quality, meaning that $Margin$ cannot have an impact on sales (Assumption 2). In addition, the falsification tests in Appendix A1 show that our framework is consistent with Assumption 1.

Table B2 reports the results for the randomization inference. The estimates are obtained using difference-in-means with a uniform kernel. To estimate p-values given our small sample size, we use the Monte Carlo sampling method. With a sufficient number of permutations, this method leads to the estimation of exact p-values (Ernst, 2004). The sample consists of 34 observations, including 10 treated units. It can be observed that the Goncourt has a high and statistically significant effect on sales, with a magnitude similar to the parametric and nonparametric approaches.

Table B2. Impact of the Goncourt on Sales – Local Randomization RDD

	Outcome: $\log(\boldsymbol{Sales_{post}})$
Goncourt	1.534**
Observations	34
Window	[-2, 2]

Note: Local randomization RD estimates (Cattaneo, Frandsen, and Titiunik, 2015; Cattaneo, Titiunik, and Vazquez-Bare, 2017). The estimations are obtained using difference-in-means with a uniform kernel. P-values are computed using Monte Carlo permutation tests (10,000 repetitions). In all specification (1), we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *Female author*, the four publisher dummies, and the time dummies. **Significant at 5% level.

Appendix B2. Alternative victory margin coding strategies

In our baseline estimates, we set *Votes* equal to 0 for the books in the first and second selections while those reaching the third selection automatically receive five votes, to which we add the potential votes received in the last round of the final selection. This coding is necessary, as only the votes of the last round of the final selection are available systematically, but this implies that shortlisted books receiving no votes in the last round have the same number of votes as non-shortlisted books, that is zero. To show that our results are not driven by the way that the victory margin is coded, we use two alternative coding strategies.

Alternative 1. We set $Votes = 0$ for the books in the first selection, $Votes = 3$ instead of zero for those in the second selection, and $Votes = 5$ for the final selection, to which we add the votes received in the last round. Again, this allows us to distinguish between the different selection processes.

Alternative 2. We only use the votes that are documented, i.e. those in the last voting round of the final selection. The number of votes for the books not reaching the last round of the final selection is accordingly set to 0. Despite putting the books in the first selection on the same footing as those in the third, this alternative is the least discretionary as it does not require arbitrary votes to be assigned to books not reaching the final round.

Table B3 presents the estimates associated with these different coding strategies. The results are qualitatively and quantitatively similar to the baseline estimates, thus demonstrating the robustness of the findings to the method of coding the victory margin.

Table B3. Alternative Victory Margin Coding Strategies

	Outcome: $\log(\boldsymbol{Sales_{post}})$			
	Alternative 1		Alternative 2	
	(1)	(2)	(1)	(2)
	Linear	Quadratic	Linear	Quadratic
Goncourt	1.597***	1.391***	1.388***	1.094**
	(0.301)	(0.376)	(0.396)	(0.443)
Margin	0.039*	0.053	0.111*	0.274
	(0.020)	(0.077)	(0.064)	(0.320)
Observations	220	220	220	220

Notes: Parametric RD estimates. Column (1) of each panel fits a linear polynomial while panel B fits a quadratic one. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *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 at 10% level.

Appendix B3. Least Absolute Deviations regressions

To make sure that the baseline results are not driven by outliers, we estimate the baseline expression with the Least Absolute Deviations estimator, which is outlier-insensitive (Wooldridge, 2010). The results, reported in Table B4, are in line with baseline results.

Table B4. Least Absolute Deviations Estimates

	Outcome: $\log(\boldsymbol{Sales_{post}})$	
	(1)	(2)
	Linear	Quadratic
Goncourt	1.541***	1.412***
	(0.273)	(0.486)
Margin	0.060***	-0.034
	(0.016)	(0.081)
Observations	220	220

Notes: Parametric RD estimates. Column (1) fits a linear polynomial while panel B fits a quadratic one. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *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 at 10% level.

Appendix B4. Alternative specifications and functional forms

In this section, we further explore the sensitivity of our results to the sample and the specification of the estimated relation. Table B5 and Table B6 report the results of the linear and quadratic specifications, respectively.

In Column (1) of both tables, we expand the sample to the entire period for which votes and sales are available, that is the 1954-2018 editions of the prize, with the caveat that the sales for books published before 2004 do not include pre-2004 sales. In Column (2), we focus only on shortlisted books, that is those reaching the final selection stage. In Column (3), we introduce into the specification a dummy for each publisher (28 dummies in total) to capture unobserved heterogeneity among them. Finally, in Column (4), we use the number of sales pre- and post-Goncourt in level instead of in log, as we have no prior information on the functional form relating the victory margins to sales.

The results of all these robustness checks are in line with the baseline results.

Table B5. Linear Polynomial - Alternative Specifications

	Outcome: $\log(\boldsymbol{Sales_{post}})$			
	(1)	(2)	(3)	(4)
	All editions	Shortlisted books	Publisher dummies	Sales in level
Goncourt	1.661***	1.507***	1.273***	398,662.144***
	(0.537)	(0.448)	(0.290)	(107,105.517)
Margin	0.136***	0.065	0.050***	2,549.309
	(0.034)	(0.089)	(0.018)	(4,541.454)
Observations	854	63	220	220

Notes: Linear RD estimates. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *Female author*, the four publisher dummies, and the time dummies (except Column (1) which does not control for $\log(Sales_{pre})$). Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Table B6. Quadratic Polynomial - Alternative Specifications

	Outcome: $\log(\boldsymbol{Sales_{post}})$			
	(1)	(2)	(3)	(4)
	All editions	Shortlisted books	Publisher dummies	Sales in level
Goncourt	1.829***	1.154*	1.147***	388,788.274***
	(0.646)	(0.606)	(0.381)	(145,365.572)
Margin	0.186	0.256	0.020	-27,025.801
	(0.169)	(0.447)	(0.099)	(32,811.924)
Observations	854	63	220	220

Notes: Quadratic RD estimates. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *Female author*, the four publisher dummies, and the time dummies (except Column (1) which does not control for $\log(Sales_{pre})$). Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Appendix C. Robustness checks of channels of transmission

Appendix C1. Information effect

Table C1 reports the raw coefficients obtained when estimating Equation 4.

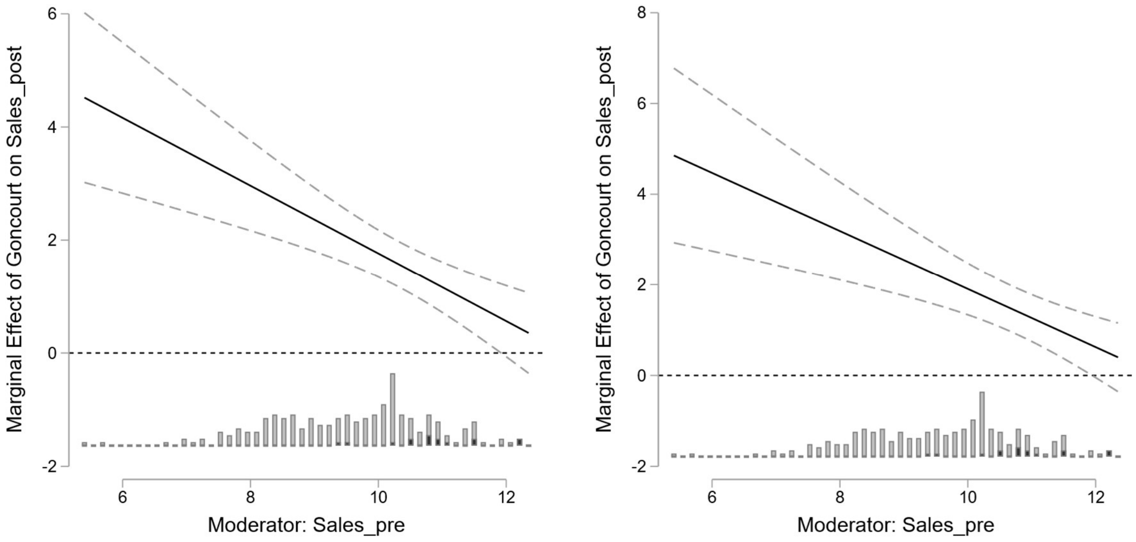
Table C1. Interaction between Goncourt and Sales pre-Goncourt - Raw coefficients

	Outcome: $\log(\boldsymbol{Sales_{post}})$	
	(1)	(2)
	Linear	Quadratic
Goncourt	7.760***	8.317***
	(1.854)	(2.290)
$\log(\text{Sales}_{\text{pre}})$	0.875***	0.875***
	(0.069)	(0.069)
Goncourt*$\log(\text{Sales}_{\text{pre}})$	-0.600***	-0.641***
	(0.178)	(0.211)
Observations	220	220

Notes: Parametric RD estimates. Column (1) fits a linear polynomial while Column (2) fits a quadratic one. The model specification follows Equation (4). In all specifications, we control for *Movie*, *Other prize*, *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 at 10% level.

Figure C1 below plots the conditional marginal effect of *Goncourt* on $\log(Sales_{post})$ using the conventional linear interaction estimator instead of the kernel smoothing estimator proposed in Hainmueller et al. (2019).

Figure C1– Marginal Effect of the Goncourt on post-Goncourt Sales as a Function of pre-Goncourt Sales



Notes: 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 prize*, *Female author*, the four publisher dummies, and the time dummies. The dash line reports 90% confidence intervals based on robust standard errors.

Appendix C2. Quality signal robustness checks

Table C2 reports the outcome of estimating Equation (7) with as an ordered probit model.

Table C2. Effect of the Goncourt on Reviewer Sentiment – Ordered Probit

	Outcome: ***Sentiment***			
Timing of the review	A. Pre-Goncourt		B. Post-Goncourt	
	(1)	(2)	(3)	(4)
Estimated coefficients of ordered probit model				
Goncourt	-0.114	0.104	-0.427***	-0.385***
	(0.222)	(0.266)	(0.126)	(0.142)
#Reviews (log)	0.007	0.009	0.164***	0.163***
	(0.023)	(0.023)	(0.019)	(0.019)
Average marginal effect of Goncourt on reviews' sentiment				
Negative	0.039	-0.034	0.149***	0.134***
	(0.078)	(0.085)	(0.045)	(0.051)
Neutral	0.002	-0.004	0.010***	0.010***
	(0.001)	(0.014)	(0.002)	(0.002)
Positive	-0.041	0.038	-0.160***	-0.144***
	(0.077)	(0.099)	(0.044)	(0.051)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Log likelihood	-1908	-1907	-11196	-11195
Observations	1,770	1,770	10,772	10,772

Notes: RD estimates. 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 prize*, *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 C3 below reports the raw coefficients obtained when estimating the model interacting the Goncourt dummy variable with the number of reviews posted on Amazon. The coefficients are interpreted in Figure 3.

Table C3. Interaction Between Goncourt and Number of Past Reviews - Raw Coefficients

	Outcome: ***Sentiment***	
	(1)	(2)
	Linear	Quadratic
Goncourt	-1.315***	-1.240***
	(0.365)	(0.379)
#Reviews	0.254***	0.251***
	(0.032)	(0.033)
Goncourt*#Reviews	0.124*	0.127*
	(0.068)	(0.068)
Log likelihood	-11183	-11182
Observations	10,772	10,772

Notes: RD estimates. 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 prize*, *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.

Appendix C3. The Goncourt des Lycéens

To estimate the effect the Goncourt des Lycéens prize on consumer satisfaction, we first estimate its effect on sales with a linear difference-in-differences model. We then turn to its impact on sentiment using a nonlinear difference-in-differences model (Puhani, 2012), as the dependent variable follows a natural ordering.

Panel A of Table C4 reports the results for the specifications with book sales as dependent variable. Column (1) uses the entire sample while Column (2) excludes Goncourt winners. In both specifications, the Goncourt des Lycéens displays a positive and significant coefficient, meaning that this prize also has a positive impact on book sales. Panel B implements similar specifications with consumer sentiment instead of sales as dependent variable. In both Columns (3) and (4), the Goncourt des Lycéens coefficient is statistically insignificant at standard levels. Accordingly, the prize has no effect on consumers' satisfaction.

Table C4. Effect of the Goncourt des Lycéens Prize on Book Sales and Reviewer Sentiment

Dependent variable	A. Book sales (log)		B. Consumer sentiment	
	(1)	(2)	(3)	(4)
Estimated coefficients of the diff-in-diff model				
Diff-in-diff	1.354***	1.399***	0.009	0.027
	(0.173)	(0.174)	(0.118)	(0.118)
Average marginal effect of Goncourt des Lycéens on consumer sentiment				
Negative	-	-	-0.002	-0.005
	-	-	(0.021)	(0.022)
Neutral	-	-	-0.001	-0.002
	-	-	(0.007)	(0.007)
Positive	-	-	0.002	0.006
	-	-	(0.029)	(0.029)
Type of diff-in-diff model	Linear	Linear	Nonlinear	Nonlinear
Log likelihood	-	-	-12714	-10223
Observations	420	396	12,115	9,800

Notes: Panel A estimates the impact of the Goncourt des Lycéens on book sales using a linear difference-in-differences model. Panel B assesses the impact of the Goncourt des Lycéens on consumer sentiment using nonlinear difference-in-differences model (Puhani, 2012). The corresponding average marginal effects are reported alongside. Column (1) uses the entire sample while Column (2) excludes Goncourt-winning books. In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *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.

Appendix C6. Formal definition of the mediation analysis

Formally, our RD mediation framework is given by Equations (8) and (9). Equation (8) looks at the impact of the Goncourt on the number of reviews. Equation (9) assesses the impact of the number of reviews on sales, controlling for the prize.

$$\begin{aligned} \operatorname{arcsinh}(\#Reviews_{iy,\tau}) & \\ &= \gamma_0 + \gamma_1 Goncourt_{iy} + f(Goncourt_{iy}, Margin_{iy}) \\ &+ \gamma_2 \log(Sales_{iy,pre,\tau}) + \gamma_3 \operatorname{arcsinh}(\#Reviews_{iy,p}) + \boldsymbol{\theta}'\mathbf{X}_{iy} + \lambda_y \\ &+ \epsilon_{1iy}, \end{aligned} \tag{8}$$

$$
\begin{aligned}
&\log(Sales_{iy,\tau,post}) \\
&\qquad = \delta_0 + \delta_1 Goncourt_{iy} + \delta_2 \operatorname{arcsinh}(\#Reviews_{iy,\ }) \\
&\qquad + \ f(Goncourt_{iy}, Margin_{iy}) + \delta_3 \log(Sales_{iy,pre,\tau}) \\
&\qquad + \delta_4 \operatorname{arcsinh}(\#Reviews_{iy,pre}) + \boldsymbol{\theta}'\mathbf{X}_{iy} + \lambda_y + \epsilon_{2iy,}
\end{aligned} \tag{9}
$$

where $\#Reviews_{iy,\tau}$ is the number of reviews measured between the awarding of the Goncourt and τ months later. In the baseline, we use a window of 6 months but let it vary as a robustness check. $\#Reviews_{iy,pre}$ is the number of reviews measured before the award and $\log(Sales_{iy,\tau,post})$ is the number of sales after τ months following the award. The other variables are defined as before. γ_1 measures the impact of the prize on the number of reviews, controlling for the number of sales, and δ_2 measures the marginal effect of an additional review on sales. $\gamma_1 \times \delta_2$ therefore measures the indirect mediation effect, that is the impact of the Goncourt on sales that runs through the number of reviews.

Appendix C5. Mediation analysis

Table C5 below reports the indirect effect of the mediation model using alternative time windows for the number of reviews on Amazon.fr, i.e. the mediator.

Table C5. Mediation Analysis – Additional Time Windows

		Outcome: $\log(\boldsymbol{Sales_{post}})$			
		Linear		Quadratic	
Window	Mediator	Estimate	90% BCa CI [a]	Estimate	90% BCa CI [a]
5 months	Number of Reviews	0.226	[0.035, 0.530]	0.191	[0.021, 0.482]
7 months	Number of Reviews	0.221	[0.043, 0.499]	0.197	[0.030, 0.461]
8 months	Number of Reviews	0.249	[0.054, 0.545]	0.228	[0.042, 0.506]

Notes: Parametric RD mediation estimates. BCa CI = bias-corrected and accelerated bootstrap confidence interval.
[a] Based on 5,000 sample bootstrapping.

Appendix C6. Mediation analysis – 2SLS approach

As explained in Section 7.3, each measurement is separated by one month. This time interval allows the effect of the treatment to materialize without threatening the causal interpretation of the results because of unmeasured confounding factors in the mediation-outcome relationship. To further address endogeneity concerns, we implement a two-stage least squares (2SLS) approach. To do so, we use the identification strategy proposed by Lewbel (2012), which exploits the presence of heteroscedasticity in the error term of the first stage to construct a valid instrument from a set of independent variables.[43] Lewbel's (2012) method can be used to obtain IV estimates when an external instrument is unavailable or too weak.[44] Therefore, in the absence of a compelling instrument for $\#Reviews_{iy,}$, the method is a suitable and compelling approach.

Table C6 reports the 2SLS estimates. The first noteworthy finding is that the instruments are strong, as shown by the Stock-Wright LM S statistic, which tests the null hypothesis of weak instruments (Cameron & Trivedi, 2009). If we now draw attention to the 2SLS estimates, they remain significant at conventional levels despite a higher magnitude than the baseline OLS estimates. This may be because OLS estimates are biased downward due to endogeneity. Alternatively, it may simply reflect the fact that Lewbel's (2012) approach is less precise than conventional IV as it relies upon higher-order moments to identify the parameter of interest.

Table C6. Bandwagon Effect – Mediation Analysis – 2SLS Approach

	Outcome: $\mathbf{log}(\boldsymbol{Sales_{post}})$

[43] For this purpose, we use the Stata command developed by Baum and Schaffer (2019).

[44] Since there is no accepted method for selecting the set of independent variables to be used in the construction of the instrument, we follow the literature and include all our variables, except the treatment and the running variable (Mishra & Smyth, 2015).

	(1) Linear	(2) Quadratic
Indirect effect		
$\gamma_1 \times \delta_2$	0.497	0.476
95% BCa CI	[0.158, 1.887]	[0.115, 2.315]
Joint-significance test		
γ_1	1.413***	1.323***
	(0.386)	(0.417)
δ_2	0.352**	0.360**
	(0.141)	(0.143)
SW LM S stat.	0.001***	0***
Observations	220	220

Notes: Parametric RD 2SLS mediation estimates. Column (1) fits a linear polynomial while Column (2) fits a quadratic one. The 2SLS estimates follow Lewbel's (2012) approach. BCa CI = bias-corrected and accelerated bootstrap confidence interval based on 10,000 sample bootstrapping. SW LM S stat. = Stock-Wright LM S statistic. The statistic tests the null hypothesis of weak instruments (Cameron & Trivedi, 2009). In all specifications, we control for $\log(Sales_{pre})$, *Movie*, *Other prize*, *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.

Additional references

Armstrong, T. B., & Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica*, 86(2), 655-683.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using stata* (Vol. 5, p. 706). College Station, TX: Stata press.

Cattaneo, M. D., Frandsen, B. R., & Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate. *Journal of Causal Inference*, 3(1), 1-24.

Cattaneo, M. D., Titiunik, R., & 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(3), 643-681.

Ernst, M. D. (2004). Permutation methods: A basis for exact inference. *Statistical Science*, *19*(4), 676-685.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.

Hainmueller, J., Mummolo, J., & Xu, Y. (2019). How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Political Analysis*, 27(2), 163-192.

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3), 933-959.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277-2304.

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2), 655-674.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1), 67-80.

Puhani, P. A. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Economics Letters*, 115(1), 85-87.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.