Supplementary Material for Lagios, Nicolas and Méon, Pierre-Guillaume, ‘Experts, Information, Reviews, and Coordination: Evidence on How Prizes Affect Sales,’ *The Journal of Industrial Economics*, VOLUME (ISSUE), MONTH, YEAR, pp. XXX-YYY.

Appendix A. Descriptive Statistics

Table A.I. Descriptive Statistics - Awarded vs. Non-Awarded Books

	Non-awarded books ($Goncourt = 0$)				Awarded books ($Goncourt = 1$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	S.d.	Min.	Max.	Mean	S.d.	Min.	Max.
$Sales_{post}$	113,120	214,944	117	1,667,568	542,338	273,047	202,424	1,185,945
$Sales_{pre}$	25,533	30,496	222	185,486	74,560	63,816	12,356	231,030
$\log(Sales_{post})$	10.435	1.685	4.762	14.327	13.092	0.488	12.218	13.986
$\log(Sales_{pre})$	9.529	1.201	5.403	12.131	10.915	0.822	9.422	12.350
Votes	1.424	2.673	0	9	10.667	0.976	9	12
Margin	-7.493	2.874	-10	-1	1.733	1.534	0	5
Year	2011.054	4.373	2004	2018	2011	4.472	2004	2018
Movie	0.102	0.304	0	1	0.200	0.414	0	1
Other prizes	0.283	0.452	0	1	0.133	0.352	0	1
Female author	0.293	0.456	0	1	0.200	0.414	0	1
Gallimard	0.229	0.421	0	1	0.267	0.458	0	1
Grasset	0.137	0.344	0	1	0.067	0.258	0	1
Seuil	0.083	0.276	0	1	0.067	0.258	0	1
Actes Sud	0.034	0.182	0	1	0.333	0.488	0	1
Observations		205				15		

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

Appendix B. Identification Assumptions and Falsification Tests

Appendix B1. Covariate 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 V(iii) in the paper). The results are reported in Table B.I.

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 D2, which is entirely devoted to those methods.

The results show no evidence of discontinuity, except for the variables $Other\ prizes$ and $Actes\ Sud$. The former is unsurprising as $Other\ prize$ includes prizes that are directly influenced by the Goncourt.[1] 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

[1] 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.

use the other two RDD approaches, $Other\ prizes$ and $Actes\ Sud$ are never significant. This may suggest that the above results are due solely to random chance.

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

Table B.I. Covariate Balance

RDD approach	(1) Parametric	(2) Non-parametric	(3) Local randomization
$\log(\text{Sales}_{\text{pre}})$	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 prizes	-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	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 B.II.

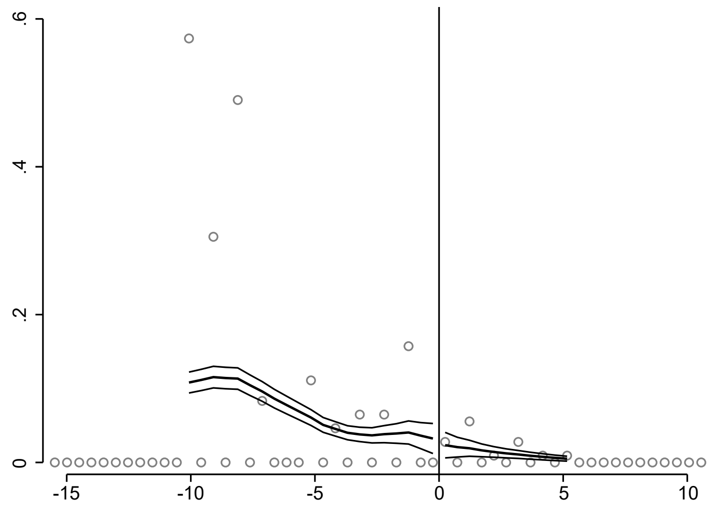
Table B.II. The Effect of the Goncourt on Book Sales – Without Books Having Won Other Prizes and Published by Actes Sud

	Outcome: $\log(Sales_{post})$			
Degree of the polynomial	Linear		Quadratic	
	(1) Without books that have won other prizes	(2) Without books published by Actes Sud	(3) Without books that have won other prizes	(4) 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. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. 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 $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Appendix B2. Manipulation of the Cut-Off

Figure B.I. McCrary [2008] Density Test



Notes: McCrary [2008] density test. The discontinuity estimate expressed as the log difference in height is -.26 and the associated standard errors .54.

Appendix B3. Placebo Cut-Offs

The RDD rests on the assumption that the cut-off at the victory margin of zero distinguishes the winners 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.

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 B.III 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.

Table B.III. Placebo Cut-Offs

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

Notes: Parametric RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, 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.

Appendix C. The Dynamic Impact of the Goncourt

In this section, we provide more insight into the dynamic impact of the Goncourt on sales by investigating how the marginal effect of the prize changes over time. To do so, we first adapt our framework to a panel setting and then estimate a treatment-covariate interaction model. Our approach can be described via the following equation:

$$\begin{aligned}\log(Sales_{iw}) = \phi \log\big(Sales_{i,w-1}\big) + \tau Goncourt_{iw} + \alpha \log(Weeks_{iw}) \\ + \mu Goncourt_{iw} \times \log(Weeks_{iw}) \\ + f(Goncourt_{iw}, Margin_{iw}) + \boldsymbol{\theta}'\mathbf{X}_i + \lambda + \epsilon_{iw},\end{aligned} \quad \text{(C1)}$$

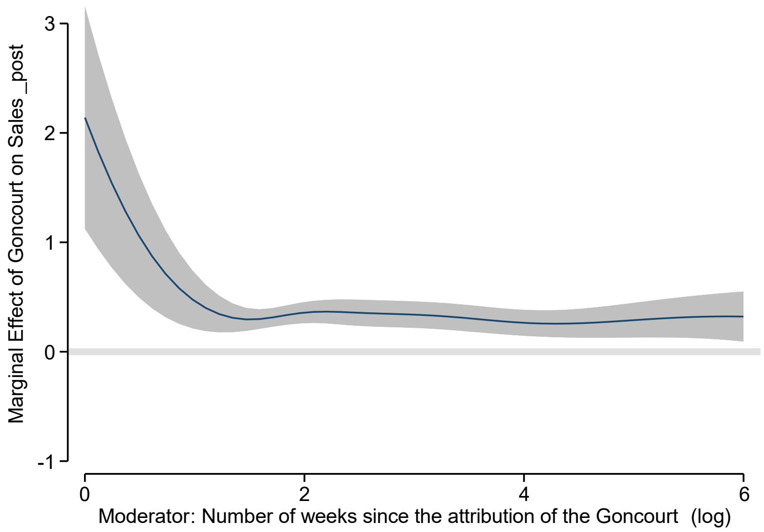
where $Sales_{iw}$ and $Sales_{i,w-1}$ are the number of sales of book i on week w and week $w-1$, respectively. The term $Goncourt_{iw}$ is a dummy equal to one if book i is awarded the Goncourt on week w and $Weeks_{iw}$ measures the number of weeks that elapsed since the attribution of the prize; both variables equal 0 before the prize awarding. Finally, $f\big(Goncourt_{iy}, Margin_{iy}\big) = \beta Margin_{iw} + \gamma Goncourt_{iw} \times Margin_{iw}$ is a function that models the impact of a book's victory margin on its sales, $\mathbf{X}_i$ is a vector of control variables (the same as in Equation (3) in the main text), and λ are fixed effects for the year of the competition.

The point of the model is to estimate the conditional marginal effect (CME) of $Goncourt_{iw}$ on $log(Sales_{iw})$, that is

$$(\Delta \log(Sales_{iw}) \,|Goncourt_{iw} = 1) = \tau + \mu \log(Weeks_{iw}). \quad \text{(C2)}$$

As in the baseline, we estimate the CMEs of $Goncourt$ on $log(Sales_{iw})$ using the kernel smoothing estimator considered in Hainmueller *et al.* [2019] to allow for non-linearities. The results are summarized in Figure C.I, which plots the marginal effect of the Goncourt as a function of the number of weeks that elapsed since the attribution of the prize. Unsurprisingly, the impact of the Goncourt is at its highest the first weeks following its attribution and then decreases over time.

Figure C.I. The Effect of the Goncourt over Time



Notes: The gray area depicts 95% confidence intervals.

Appendix D. Robustness Checks on the Effect of the Prize on Sales

Appendix D1. Alternative Way of Addressing the Fact that Older Books Have Had More Time to Sell Copies

Instead of using time dummies, an alternative way of addressing the fact that older books have had more time to sell copies is to compute pre- and post-Goncourt sales over a well-defined window around the attribution of the prize. Table D.I reports the results of doing so for a one- and a three-year window around the attribution of the prize.

Table D.I. The Effect of the Goncourt on Book Sales – Time Windows Around the Attribution of the Prize

	Outcome: $\log(Sales_{post})$			
Time window around the attribution of the Goncourt	One-year window		Three-year window	
	(1)	(2)	(3)	(4)
Goncourt	2.093***	1.472***	1.814***	1.390***
	(0.297)	(0.395)	(0.330)	(0.443)
Margin	0.0362*	0.115	0.0305*	0.0460
	(0.0171)	(0.0931)	(0.0178)	(0.0996)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Observations	220	220	220	220

Notes: Parametric RD estimates. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, and the four publisher dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Appendix D2. 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 *et al.* [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 (CIs). In particular, when the running variable is discrete, standard CIs may have poor coverage (Lee and 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.[2]

The estimates associated with each K are reported in Table D.II. 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.

[2] 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.

Table D.II. The Effect of the Goncourt on Book Sales – Non-Parametric RDD

	(1)	(2)	(3)
		Outcome: $\log(Sales_{post})$	
K	0.03	0.06	0.09
Estimate	1.507	1.413	1.413
BSD 95% CIs	[0.511, 2.504]	[0.394, 2.433]	[0.223, 2.604]
EHW 95% CIs	[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 and Kalyanaraman [2012]). EHW refers to the CIs obtained using the conventional Eicker-Huber-White standard errors. In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, 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 *et al.* [2015] and Cattaneo *et al.* [2017], who formalize and discuss the differences between the 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 $Margin = \pm 2$ in order to have 10 treated books. Therefore, if randomization holds, it must hold for the window: $W_0 = [Margin = -2,\ Margin = 2]$.

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 B show that our framework is consistent with Assumption 1.

Table D.III 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 D.III. The Effect of the Goncourt on Book Sales – Local Randomization RDD

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

Note: Local randomization RD estimates (Cattaneo *et al.* [2015]; Cattaneo *et al.* [2017]). The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. 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, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Appendix D3. Least Absolute Deviations Regressions

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 make sure that the baseline results are not driven by outliers, we re-estimate Equation (3) using the Least Absolute Deviations method, which is outlier-insensitive (Wooldridge [2010]). The results, reported in Table D.IV, are in line with baseline results.

Table D.IV. The Effect of the Goncourt on Book Sales – Least Absolute Deviations Estimates

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

Notes: Parametric RD estimates. Column (1) fits a linear polynomial while Column (2) fits a quadratic one. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Appendix D4. 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, which would imply 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 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 D.V 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 D.V. The Effect of the Goncourt on Book Sales – Alternative Victory Margin Coding Strategies

	Outcome: $\log(Sales_{post})$			
	Alternative 1		Alternative 2	
	(1)	(2)	(3)	(4)
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)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Observations	220	220	220	220

Notes: Parametric RD estimates. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

In Table D.VI, we focus only on shortlisted books, that is those reaching the final selection stage, to avoid making assumptions on the number of votes they garnered.

Table D.VI. The Effect of the Goncourt on Book Sales – Only Shortlisted Books

	Outcome: $\log(Sales_{post})$	
	(1)	(2)
Goncourt	1.507***	1.154*
	(0.448)	(0.606)
Margin	0.065	0.256
	(0.089)	(0.447)
Degree of the polynomial	Linear	Quadratic
Observations	63	63

Notes: Parametric RD estimates. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Appendix D5. 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 D.VII and D.VIII 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. Because the data is unavailable, we do therefore not control for the book's pre-Goncourt sales trend.

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

In Column (3), we introduce 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 D.VII. The Effect of the Goncourt on Book Sales – Alternative Specifications (Linear Polynomial)

	Outcome: $\log(Sales_{post})$			
	(1)	(2)	(3)	(4)
	All editions	1954-2003 editions	Publisher dummies	Sales in level
Goncourt	1.661***	1.611**	1.273***	398,662.144***
	(0.537)	(0.726)	(0.290)	(107,105.517)
Margin	0.136***	0.156***	0.050***	2,549.309
	(0.034)	(0.043)	(0.018)	(4,541.454)
Observations	854	634	220	220

Notes: Linear RD estimates. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies (except Columns (1) and (2) 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 D.VIII. The Effect of the Goncourt on Book Sales – Alternative Specifications (Quadratic Polynomial)

	Outcome: $\log(Sales_{post})$			
	(1)	(2)	(3)	(4)
	All editions	1954-2003 editions	Publisher dummies	Sales in level
Goncourt	1.829***	1.693*	1.147***	388,788.274***
	(0.646)	(0.913)	(0.381)	(145,365.572)
Margin	0.186	0.197	0.020	-27,025.801
	(0.169)	(0.208)	(0.099)	(32,811.924)
Observations	854	634	220	220

Notes: Quadratic RD estimates. The running variable is $Margin$ and refers to the victory margin with which a book has won the Goncourt. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The model specification follows Equation (3). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies (except Column (1) and (2) 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 E. Robustness Checks of the Channels of Transmission

Appendix E1. Information Effect

Table E.I. reports the raw coefficients obtained when estimating Equation (4).

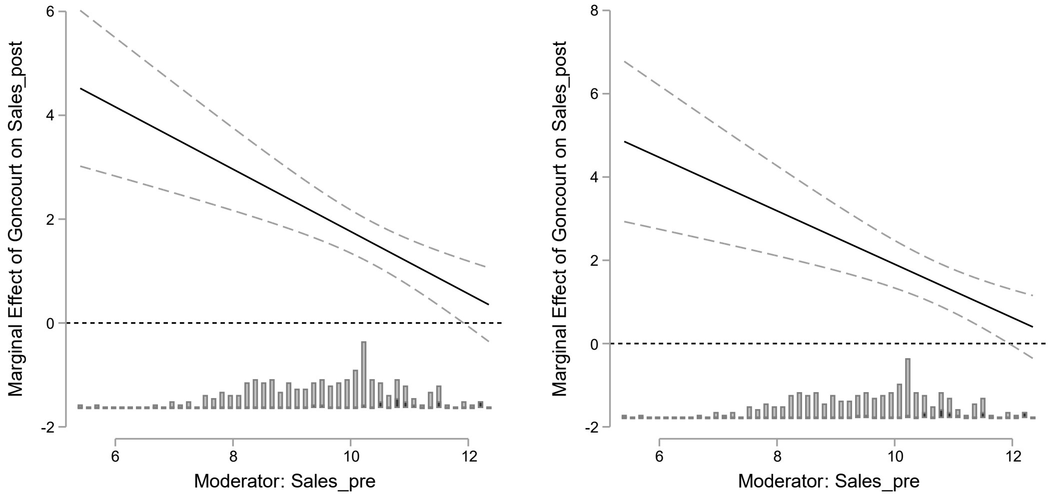
Table E.I. Interaction Between Goncourt and Pre-Goncourt Sales - Raw Coefficients

	Outcome: $\log(Sales_{post})$	
	(1)	(2)
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)
Degree of the polynomial	Linear	Quadratic
Observations	220	220

Notes: Parametric RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. Column (1) fits a linear polynomial while Column (2) fits a quadratic one. The model specification follows Equation (4). In all specifications, we control for $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant at 10% level.

Figure E.I 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 E.I. Marginal Effect of the Goncourt on Post-Goncourt Sales as a Function of Pre-Goncourt Sales – Linear Interaction

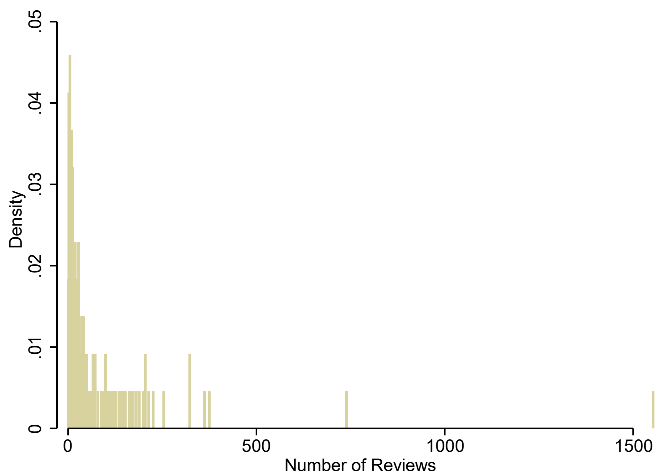


Notes: The left-hand side fits a linear polynomial while the right-hand side fits a quadratic one. The model specification follows Equation (4). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. The dash line reports 90% confidence intervals based on robust standard errors.

Appendix E2. Quality Signal

Figure E.II plots the distribution of reviews posted on Amazon.fr for the books nominated for the Goncourt between 2004 and 2018.

Figure E.II. Distribution of Reviews of the Books Nominated for the Goncourt



Tables E.II and E.III report the results of regressions excluding the books with less than 5 or 10 reviews.

Table E.II. The Effect of the Goncourt on Reviewer Sentiment – Before the Attribution of the Prize

	Outcome: *Sentiment*			
	Without books with less than 5 reviews		Without books with less than 10 reviews	
	(1)	(2)	(3)	(4)
Estimated coefficients of ordered logit model				
Goncourt	-0.229	0.162	-0.362	0.051
	(0.348)	(0.419)	(0.334)	(0.401)
#Reviews (arcsinh)	0.017	0.021	0.018	0.023
	(0.038)	(0.038)	(0.038)	(0.038)
Average marginal effect of the Goncourt on reviewer sentiment				
Negative	0.048	-0.032	0.076	-0.010
	(0.075)	(0.080)	(0.073)	(0.079)
Neutral	0.002	-0.005	0.000	-0.001
	(0.003)	(0.016)	(0.007)	(0.011)
Positive	-0.050	0.037	-0.076	0.011
	(0.073)	(0.096)	(0.067)	(0.089)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Log likelihood	-1,888	-1,887	-1,830	-1,828
Observations	1,751	1,751	1,704	1,704

Notes: RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The variable $\#Reviews$ represents the number of reviews that were already available at the time consumers posted their review. The model specification follows Equation (7). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Standard errors clustered at the book level are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Table E.III. The Effect of the Goncourt on Reviewer Sentiment – After the Attribution of the Prize

	Outcome: $Sentiment$			
	Without books with less than 5 reviews		Without books with less than 10 reviews	
	(1)	(2)	(3)	(4)
Estimated coefficients of ordered logit model				
Goncourt	-0.697***	-0.600***	-0.735***	-0.640***
	(0.202)	(0.229)	(0.206)	(0.225)
#Reviews (arcsinh)	0.281***	0.279***	0.272***	0.270***
	(0.031)	(0.031)	(0.032)	(0.032)
Average marginal effect of the Goncourt on reviewer sentiment				
Negative	0.148***	0.126**	0.155***	0.135***
	(0.044)	(0.050)	(0.045)	(0.049)
Neutral	0.013***	0.013***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
Positive	-0.160***	-0.139***	-0.169***	-0.149***
	(0.044)	(0.051)	(0.044)	(0.049)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Log likelihood	-11,062	-11,061	-10,880	-10,879
Observations	10,667	10,667	10,506	10,506

Notes: RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The variable $\#Reviews$ represents the number of reviews that were already available at the time consumers posted their review. 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.

In Table E.IV, we use an alternative proxy for consumer satisfaction. Specifically, we use the rating (number of stars) of the reviews posted on Amazon. It is a more objective and direct measure of satisfaction as it relies on no algorithm: the higher the rating, the higher consumer satisfaction.

Table E.IV. The Effect of the Goncourt on Reviewer Star Rating

	Outcome: $Rating$			
Timing of the review	Pre-Goncourt		Post-Goncourt	
	(1)	(2)	(3)	(4)
Goncourt	-0.366	-0.029	-0.625***	-0.516***
	(0.317)	(0.361)	(0.169)	(0.174)
#Reviews (arcsinh)	-0.006	-0.003	0.195***	0.192***
	(0.030)	(0.029)	(0.034)	(0.034)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Observations	1,770	1,770	10,772	10,772

Notes: RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The variable $\#Reviews$ represents the number of reviews that were already available at the time consumers posted their review. The model specification follows Equation (7). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Standard errors clustered at the book level are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

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

Table E.V. The Effect of the Goncourt on Reviewer Sentiment – Ordered Probit

	Outcome: $Sentiment$			
Timing of the review	Pre-Goncourt		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 (arcsinh)	0.007	0.009	0.164***	0.163***
	(0.023)	(0.023)	(0.019)	(0.019)
Average marginal effect of the Goncourt on reviewer 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)
63	Linear	Quadratic	Linear	Quadratic
Log likelihood	-1,908	-1,907	-11,196	-11,195
Observations	1,770	1,770	10,772	10,772

Notes: RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The variable $\#Reviews$ represents the number of reviews that were already available at the time consumers posted their review. The model specification follows Equation (7). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Standard errors clustered at the book level are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Table E.VI reports the results of a multinomial logit model, which makes no assumption on the ordering of sentiments.

Table E.VI. The Effect of the Goncourt on Reviewer Sentiment – Multinomial Logit

	Outcome: $Sentiment$			
Timing of the review	Pre-Goncourt		Post-Goncourt	
	(1)	(2)	(3)	(4)
Average marginal effect of the Goncourt on reviewer sentiment				
Negative	0.005	-0.142	0.169***	0.139**
	(0.099)	(0.092)	(0.052)	(0.059)
Neutral	0.028	0.132	-0.003	0.035
	(0.083)	(0.096)	(0.024)	(0.028)
Positive	-0.033	0.010	-0.166***	-0.174***
	(0.093)	(0.118)	(0.047)	(0.052)
Degree of the polynomial	Linear	Quadratic	Linear	Quadratic
Log likelihood	-1,854	-1,851	-11,024	-11,021
Observations	1,770	1,770	10,772	10,772

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

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

	Outcome: $\boldsymbol{Sentiment}$	
	(1)	(2)
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)
Degree of the polynomial	Linear	Quadratic
Log likelihood	-11,183	-11,182
Observations	10,772	10,772

Notes: RD estimates. The variable of interest, $Goncourt$, is a dummy that takes value one if a book has been awarded the Goncourt. The variable $\#Reviews$ represents the number of reviews that were already available at the time consumers posted their review. In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Standard errors clustered at the book level are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

Appendix E3. The Goncourt des Lycéens

To estimate the effect of 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 E.VIII 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 reviewer 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.[3]

Table E.VIII. The Effect of the Goncourt des Lycéens Prize on Book Sales and Reviewer Sentiment

Dependent variable	A. Book sales (log)		B. Reviewer sentiment	
	(1)	(2)	(3)	(4)
Estimated coefficients of the diff-in-diff model				
Estimate	1.354***	1.399***	0.009	0.027
	(0.173)	(0.174)	(0.118)	(0.118)
Average marginal effect of the Goncourt des Lycéens on reviewer 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	-	-	-12,714	-10,223
Observations	420	396	12,115	9,800

Notes: Difference-in-differences estimates. 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) of each panel uses the entire sample while Column (2) excludes Goncourt-winning books. In all specifications, we control for $log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Standard errors clustered at the book level are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

[3] The diff-in-diff specifications include the same set of control variable as in the RDD estimates. As time-varying controls may bias the diff-in-diff estimator if they are affected by the treatment, we estimate a set of regressions with no controls (Lechner [2011]). The conclusions remain similar, as shown by Table E.IX.

Table E.IX. The Effect of the Goncourt des Lycéens Prize on Book Sales and Reviewer Sentiment – No controls

Dependent variable	A. Book sales (log)		B. Reviewer sentiment	
	(1)	(2)	(1)	(2)
Estimated coefficients of the diff-in-diff model				
Estimate	1.354***	1.399***	0.174	0.146
	(0.173)	(0.174)	(0.110)	(0.108)
Average marginal effect of the Goncourt des Lycéens on reviewer sentiment				
Negative	-	-	-0.031	-0.026
	-	-	(0.020)	(0.020)
Neutral	-	-	-0.012	-0.010
	-	-	(0.007)	(0.007)
Positive	-	-	0.043	0.036
	-	-	(0.027)	(0.027)
Type of diff-in-diff model	Linear	Linear	Nonlinear	Nonlinear
Log likelihood	-	-	-12,780	-10,268
Observations	420	396	12,115	9,800

Notes: Difference-in-differences estimates. 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) of each panel uses the entire sample while Column (2) excludes Goncourt-winning books. In all specifications, we control for 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 E4. Mediation Analysis

Appendix E4.1. Formal Definition of the Mediation Analysis

Formally, our RD mediation framework is given by Equations (E1) and (E2). Equation (E1) looks at the impact of the Goncourt on the number of reviews. Equation (E2) assesses the impact of the number of reviews on sales, controlling for the prize.

$$\begin{aligned}\operatorname{arcsinh}\left(\#Reviews_{iy,\tau}\right) \\ &= \gamma_0 + \gamma_1 Goncourt_{iy} + f\left(Goncourt_{iy}, Margin_{iy}\right) \\ &+ \gamma_2 \log\left(Sales_{iy,pre,\tau}\right) + \gamma_3 \operatorname{arcsinh}\left(\#Reviews_{iy,pre}\right) + \boldsymbol{\theta}'\mathbf{X}_{iy} \\ &+ \lambda_y + \epsilon_{1iy},\end{aligned} \tag{E1}$$

$$\log\left(Sales_{iy,\tau,post}\right)$$

$$\begin{aligned} &= \delta_0 + \delta_1 Goncourt_{iy} + \delta_2 \operatorname{arcsinh}\left(\#Reviews_{iy,\tau}\right) \\ &+ f\left(Goncourt_{iy}, Margin_{iy}\right) + \delta_3 \log\left(Sales_{iy,pre,\tau}\right) \\ &+ \delta_4 \operatorname{arcsinh}\left(\#Reviews_{iy,pre}\right) + \boldsymbol{\theta}'\mathbf{X}_{iy} + \lambda_y + \epsilon_{2iy}, \end{aligned} \tag{E2}$$

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 $\log\left(Sales_{iy,\tau,post}\right)$ is the number of sales after τ months following the award. The other variables are defined as before. The parameter 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, controlling for the effect of the Goncourt (δ_1). The term $\gamma_1 \times \delta_2$ therefore measures the indirect mediation effect, that is the impact of the Goncourt on sales that runs through the number of reviews.

Appendix E4.2. Additional Time Windows

Table E.X 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 E.X. Bandwagon Effect: Mediation Analysis – Additional Time Windows

		Outcome: $\log(Sales_{post})$			
		Linear		Quadratic	
Window	Mediator	Estimate	90% BCa CI	Estimate	90% BCa CI
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 based on 5,000 sample bootstrapping.

Appendix E4.3. 2SLS Approach

To address endogeneity concerns in the mediator-outcome relationship, 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.[4] Lewbel's [2012] method can be used to obtain IV estimates when an external instrument is unavailable or too weak.[5] Therefore, in the absence of a compelling instrument for $\#Reviews_{iy,\tau}$, the method is a suitable and compelling approach.

Table E.XI 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 and Trivedi [2010]). 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

[4] For this purpose, we use the Stata command developed by Baum and Schaffer [2021].

[5] 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 and Smyth [2015]).

Lewbel's [2012] approach is less precise than conventional IV as it relies upon higher-order moments to identify the parameter of interest.

Table E.XI. Bandwagon Effect: Mediation Analysis – 2SLS Approach

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

Notes: Parametric RD 2SLS mediation estimates. 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 and Trivedi [2010]). In all specifications, we control for $\log(Sales_{pre})$, $Movie$, $Other\ prizes$, $Female\ author$, the four publisher dummies, and the time dummies. Robust standard errors are reported in parentheses. ***Significant at 1% level; **significant at 5% level; *significant 10% level.

References

Armstrong, T. B. and Kolesár, M., 2018, 'Optimal Inference in a Class of Regression Models,' *Econometrica*, 86, pp. 655-683.

Baum, C. and Schaffer, M., 2021, *IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments*.

Cameron, A. C. and Trivedi, P. K., 2010, *Microeconometrics Using Stata*, Rev. ed., (Stata Press, College Station, Texas, U.S.A.).

Cattaneo, M. D.; Frandsen, B. R. and 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, pp. 1-24.

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

Ernst, M. D., 2004, 'Permutation Methods: A Basis for Exact Inference,' *Statistical Science*, 19, pp. 676-685.

Hahn, J.; Todd, P. and Van der Klaauw, W., 2001, 'Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,' *Econometrica*, 69, pp. 201-209.

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

Imbens, G. W. and Kalyanaraman, K., 2012, 'Optimal Bandwidth Choice for the Regression Discontinuity Estimator,' *Review of Economic Studies*, 79, pp. 933-959.

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

Kolesár, M. and Rothe, C., 2018, 'Inference in Regression Discontinuity Designs with a Discrete Running Variable,' *American Economic Review*, 108, pp. 2277-2304.

Lechner, M., 2011, 'The Estimation of Causal Effects by Difference-in-Difference Methods,' *Foundations and Trends® in Econometrics*, 4, pp. 165-224.

Lee, D. S. and Card, D., 2008, 'Regression discontinuity inference with specification error,' *Journal of Econometrics*, 142, pp. 655-674.

Lewbel, A., 2012, 'Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models,' *Journal of Business & Economic Statistics*, 30, pp. 67-80.

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

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

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

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