

The Impact of Winning a Literary Award on Goodreads Ratings of Novels, 2010-2016*

Ethan Sansom

December 16, 2025

In the 2014 paper “The Paradox of Publicity: How Awards Can Negatively Affect the Evaluation of Quality.” authors Kovács and Sharke find that, counterintuitively, winning a prestigious literary award decreased readers’ perception of prizewinning novels’ quality. To test the persistence of their findings, this analysis repeats a portion of the authors’ experiment using Goodreads reviews of prizewinning novels between 2010 and 2016 with a difference-in-differences design. Contrary to previous results, we find no evidence that winning a literary prize affects the post-award Goodreads ratings of prizewinning novels relative to other similar works. While these finding suggest that the ‘paradox of publicity’ may not hold in general, there are also several limitations to this analysis and of online review data in general which we discussed further.

1 Introduction

Several studies have shown the counterintuitive result that awards granted to cultural products have an adverse effect on consumers’ perception of said products’ quality. Kovács and Sharkey (2014) demonstrated that novels which received prestigious literary awards between 2007-2011 were rated lower than other similar novels by reviewers on Goodreads.com, a social book curation and review platform. More recently, Rossi and Schleef (2024) observed that consumer ratings of movies declined after they were nominated for an Academy Award and Rita et al. (2022) found that reviewers expressed fewer positive sentiments of restaurants’ food, service, and ambiance after they received a Michelin star. All papers suggested that an increase in consumer expectations and a broader audience resulting from the reception of an award negatively impacted consumers’ perception of quality.

*Code and data are available at: <https://github.com/EthanSansom/goodreadsratings>.

Such studies, which consider a ‘treated’ sample of award winners, have necessarily small number of treated subjects compared to a large number of potential control subjects (i.e. cultural products which did not win an award). This can present challenges to inference in common identification strategies, such as a difference-in-differences design, when few treated groups are available or there are discrepancies in the number of observations (e.g. reviews) between subjects (e.g. restaurants, movies, books) (Ferman and Pinto 2019). In this analysis, we repeat portions of the Kovács and Sharkey (2014) experiment to assess the effect of winning a literary award on the Goodreads ratings of prizewinning novels between 2010 and 2016, using a large dataset of Goodreads reviews collected in 2017 (Wan and McAuley 2018; Wan et al. 2019). The purpose is twofold, first to re-test the hypothesis that awards can have a negative impact on consumers’ perception of quality and second to comment on the methodological challenges of this and other experiments which rely on online reviews as a measure of perceived quality.

The estimand of interest is the average effect of winning a literary award on a novel’s Goodreads rating following the award announcement, compared to the counterfactual trend in ratings had the novel not won an award. Given that we cannot observe the counterfactual case (the natural trend of a novel’s ratings over time in the absence of an award shock) we instead compare the change in winning novel’s Goodreads ratings trajectories relative to that of comparable non-prizewinning novels.

Contrary to prior studies, we find no evidence of a negative (or positive) impact of award winning on the Goodreads ratings of prizewinning novels. Further, no evidence is found that prizewinning novels become more popular (receive a greater number of reviews) than other comparable novels, failing to support the hypothesis that awards draw a broader reader base which contributes to lower reviews. Potential sources of this discrepancy and limitations of this analysis are discussed in Section Section 5.1.

In the following Data Section Section 2, the sample of novels considered for this analysis is described and an overview of the Goodreads review dataset is provided (Wan and McAuley 2018; Wan et al. 2019). Two difference-in-differences models estimating the effect of award-winning on average Goodreads ratings and review volume are motivated and specified in the Models Section Section 3, followed by a discussion of the model results in the Results Section Section 4. Finally, in the Discussion Section Section 5 we summarize the results and limitations of this analysis.

This paper uses the R statistical software (R Core Team 2023) to load, clean, visualize, and model Goodreads ratings. The package `here` (Müller 2025) is used to construct relative file paths, packages `readr` (Wickham, Hester, and Bryan 2025) and `arrow` (Richardson et al. 2025) are used for reading and writing datasets and models, `dplyr` (Wickham et al. 2023), `purrr` (Wickham and Henry 2025), `stringr` (Wickham 2025), and `lubridate` (Grolemund and Wickham 2011) are used for cleaning data include character and date variables, `fixest` (Bergé 2018) is used for fitting the fixed-effects regression models described in Section Section 3, and finally `modelsummary` (Arel-Bundock 2022), `tinytable` (Arel-Bundock 2025), and `ggplot2` (Wickham 2016) are used to generate tables and plots.

2 Data

2.1 Overview

To promote comparability, we consider the same slate of literary awards studied in Kovács and Sharkey (2014), those being the Man Booker Prize, the National Book Award (fiction and non-fiction categories), the PEN/Faulkner Award, and the National Book Critics Circle Award (fiction, non-fiction, memoir/autobiography, and biography categories). These awards are chosen for their prestige and reach to ensure that they represent a significant shock to the prizewinning novels’ status. As part of the selection process, judges for each of these awards name a shortlist of 3 to 7 novels from which the winner is chosen. Shortlisted novels are announced publicly, several weeks or months prior to the announcement of a winner. A dataset of shortlisted and winning novels, as well as the announcement dates of both the shortlist and winner, was collected manually for each award in the years 2010 to 2016. Shown below in Table 1 are the novels shortlisted for the 2010 Man Booker Prize.

Table 1: 2010 Man Booker Prize Shortlist

Author	Title	Type	Year	Shortlist Date	Winner Date
Howard Jacobson	The Finkler Question	Winner	2010	2010-09-07	2010-10-12
Emma Donoghue	Room	Finalist	2010	2010-09-07	2010-10-12
Tom McCarthy	C	Finalist	2010	2010-09-07	2010-10-12
Andrea Levy	The Long Song	Finalist	2010	2010-09-07	2010-10-12
Peter Carey	Parrot and Olivier in America	Finalist	2010	2010-09-07	2010-10-12
Damon Galgut	In a Strange Room	Finalist	2010	2010-09-07	2010-10-12
Emma Donoghue	Room	Finalist	2010	2010-09-07	2010-10-12

Reviews of these novels from their publication date to late 2017 are sourced from the UCSD Book Graph, a dataset of over 15 million reviews of 2,360,655 books from around 465 thousand users scraped from Goodreads.com (Wan and McAuley 2018; Wan et al. 2019). Prizewinning and shortlisted novels were matched to novels recorded on Goodreads by title and author, with matches later verified manually. Complete book reviews in the dataset include a plain-text review of a novel alongside an integer rating of 1 to 5 stars. Inline with Kovács and Sharkey (2014), this analysis considers only ratings which are accompanied by a plain-text review.

2.2 Measurement

Goodreads ratings are user-submitted evaluations of a novel’s quality on an integer scale from 1 to 5. The ratings are collected via a web scrape of publicly available data on the profiles of

Goodreads users in late 2017 and may not contain all reviews available on Goodreads, given that information about the web scraping process is not provided (Wan and McAuley 2018; Wan et al. 2019). Further, ratings may be missing due to users deleting their profile or reviews over time. As a proxy for a users subjective evaluation of a novel’s quality, online ratings are only partially reliable. Online ratings, for example, are subject to a herding effect wherein customers’ reviews are biased towards the average of previous reviews (Lederrey and West 2018). A substantial portion of Goodreads reviews (around 9%) are estimated to be sponsored reviews, which could bias ratings upwards relative to reviewers’ actual assessment of quality (Hu et al. 2023).

2.3 Sample Selection

A challenge to identifying the effect of winning an award on a novel’s perceived quality over time is distinguishing between trends in ratings attributable to the novel’s inherent features and trends in ratings attributable to winning an award. For instance, prize-winning novels may be true classics, whose relevance and appeal do not deteriorate over time even in the absence of an award. In this case, an observed post-award increase or persistence in ratings, relative to all other books, may reflect the continuation of an existing trend in ratings rather than a causal effect of winning an award. Addressing this concern requires comparing prize-winning novels to other novels with similar pre-award rating trajectories, so that discrepancies in the post-award ratings of prize-winning novels and the comparison group can be plausibly attributed to the award decision.

Kovács and Sharkey (2014) note that the awards process itself provides such a comparison group. Several weeks or months prior to choosing a winning novel, each award’s judges publicly name three to five shortlisted novels from which the winner is selected. These shortlisted books are deemed to be similar to the prizewinning novels in terms of quality by the awards’ judging panels and share a similar publication date (typically within a year of the award being announced), meaning they are more likely than other books on Goodreads to follow ratings trajectories similar to that of their prizewinning competitor.

From 2010 to 2016, there were 55 prizewinning novels¹ and 205 shortlisted finalists from the 4 prizes included in this analysis. For each prize, category, and year, the shortlisted novel with the lowest absolute difference in mean pre-award ratings compared to the prizewinning novel was chosen as the comparison group, resulting in 55 matched-pairs of prizewinning and finalist novels. This approach is meant to limit any unobserved differences in the prizewinning novels and their comparison group which could confound the impact of award-winning on the ratings of prizewinning novels (Kovács and Sharkey 2014). These 55 pairs were reduced to a final 45 pairs of novels, after omitting novels with no reviews in either the pre-award or post-award periods and duplicate nominations.

¹The Sellout by Paul Beatty won both the 2015 Man Booker Prize and the 2016 National Book Critics Circle Award for fiction.

2.4 Variables

Table Table 2 shows a sample of 5 observations of variables relevant for this analysis. The Goodreads reviews are uniquely identified by a User ID, Review ID, and a Book ID (truncated for width). Each review is accompanied by a rating between 1 and 5 stars and is time-stamped, allowed reviews to be identified as occurring either before or after a particular award’s winner is announced.

Table 2: Sample of Goodreads Reviews

User ID	Book ID	Review ID	Rating	Review	Timestamp
60ae4...	76482...	c0cf3...	2	I will start of...	2012-05-17 04:15:33
7efef...	64337...	1c44b...	5	I love this wom...	2009-11-17 16:15:05
d6aa1...	30555...	8dcaf...	4	It’s not a fun...	2017-05-26 16:18:03
a9392...	11869...	71557...	5	Beautiful and h...	2012-01-09 15:27:48
d775e...	24612...	b88e5...	2	The whole time...	2015-09-28 01:59:53

Among the 90 selected novels, the number of ratings is highly variable, with the least reviewed book having just 4 ratings compared to a maximum of 5,356 ratings (see Table Table 3).

Table 3: Descriptive Reviews Statistics

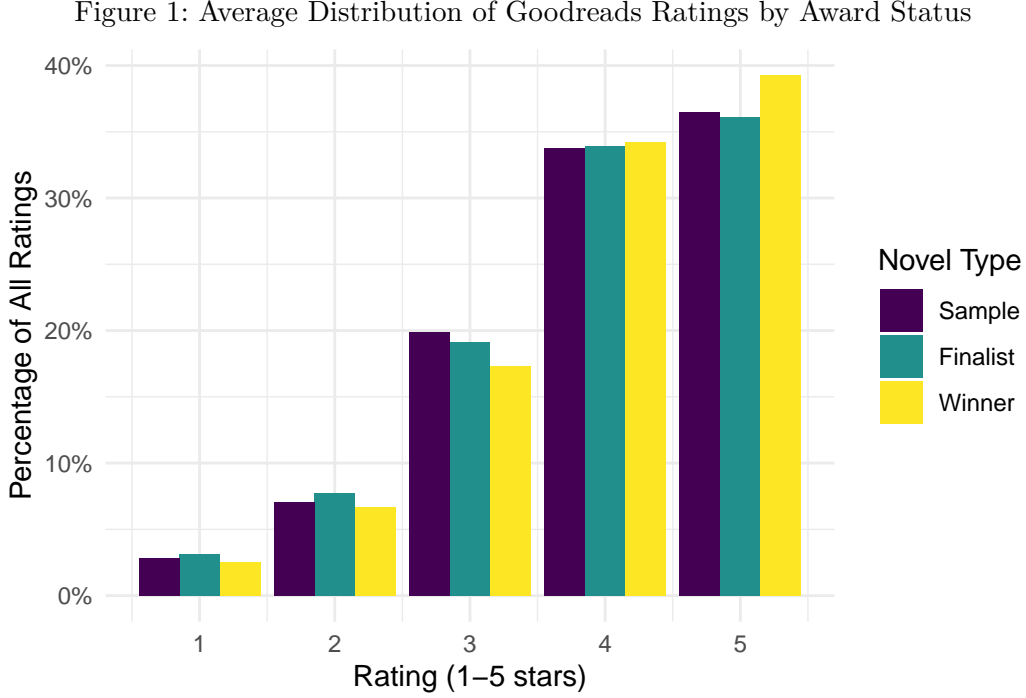
	Min	Max	Mean	SD
Reviews per Book	4.00	5356.00	539.03	818.06
Rating	1.00	5.00	4.02	1.01
Timestamp	2004-02-12 08:00:00	2017-11-03 04:30:25	2014-10-20 02:43:32.7339	

The average rating for all considered novels is around 4 stars, with prizewinning novels having slightly higher average ratings in the pre-award period (see Table Table 4) and around 50 more reviews than finalists on average.

Table 4: Balance of Ratings for Winners and Finalists Prior to Award Announcement

	Finalist (N=7426)		Winner (N=6992)	
	Mean	Std. Dev.	Mean	Std. Dev.
Reviews per Book	496.62	1305.53	547.38	854.88
Rating	4.10	0.93	4.17	0.94

Figure Figure 1 compares the distribution of ratings for prizewinning novels and finalists against that of a random sample of 10,000 novels not included in any of the award shortlists. Winners are more likely to attain a 5 star review (around 39% of ratings) compared to finalists and other novels (around 36%), but have a similar distribution of 4 and 1 star reviews. The randomly sampled novels are notably more likely to have a middle 3 star review (around 20% of ratings) compared to both finalists and award winners.



3 Models

3.1 Ratings Model

Following Kovács and Sharkey (2014), we use a difference-in-differences design with matched-pair fixed effects to estimate the effect of winning an award on the Goodreads ratings of novels.

$$\text{Rating}_{ijk} = \beta_1 \text{Winner}_i + \beta_2 \text{Post}_k + \beta_3 (\text{Winner}_i \times \text{Post}_k) + \sum_{j=1}^J \alpha_j \text{Pair}_j + \epsilon_{ijk} \quad (1)$$

The outcome Rating_{ijk} is the k th rating of the i th book in the j th matched pair of novels. 45 pairs of novels are considered ($j = 1, \dots, 45$) each containing one prizewinning novel and a shortlisted finalist ($i = 1, 2$). The number of ratings varies by novel, with a mean count of 539.03 (SD 818.06).

Covariates Pair_j are indicators for whether a rating belongs to a book in the j th match pair. The coefficients α_j are the matched-pair fixed effects, which are interpreted as the average baseline rating of novels in pair j . These fixed-effects absorb baseline characteristics shared by each pair, such as the overall quality of a shortlisted novels in a given year.

Winner_i is an indicator for whether a rating is of a prizewinning novel and Post_j an indicator for whether a rating occurred after the winner of the j th matched-pair’s award was announced. The interaction $\text{Winner}_i \times \text{Post}_j$ is equal to 1 for ratings of prizewinning novels given after their award was announced and 0 otherwise. β_1 is interpreted as the difference in average ratings between the prizewinning novel and its corresponding finalist in the pre-award period and β_2 as the average change in ratings of finalists from the pre-award to post-award period. β_3 is the difference-in-differences estimator, representing the effect of the award on ratings. In particular, β_3 measures the difference in the change of the prizewinning novel’s average ratings from the pre-award to post-award period relative to the change in the average ratings of the corresponding finalist.

The inclusion of matched-pair fixed effects α_j are important for isolating the effect of winning an award. By absorbing characteristics shared by both novels in each matched pair, these fixed-effects ensure that the difference-in-differences estimator compares each winner with its direct rival, rather than comparing all prizewinning novels to all finalists across different award years with varying baseline qualities.

ϵ_{ijk} is the idiosyncratic error term. Errors are assumed to be independent for ratings in different matched-pairs j and j' , but are arbitrarily correlated for ratings of novels within the same matched pair. Correlation is assumed at the matched-pair, rather than the novel, level to account for both autocorrelation of reviews for the same novel over time and the correlation between the winner and finalist within the same award cycle, which share a similar publishing date and are subject to the same time-specific trends in reviews. Accordingly, the standard errors of coefficient estimates are clustered at the matched-pair level.

3.2 Popularity Model

A secondary effect of winning an award on popularity is estimated using a Negative Binomial general linear model, using the same difference-in-differences covariates described in Section 3.1, with the number of reviews per novel as the outcome variable.

Let NumReviews_{ijt} denote the number of reviews of the i th novel in the j th matched-pair in the t th period, where t is either pre-award or post-award. We assume that NumReviews_{ijt} follows

a Negative Binomial distribution with a dispersion parameter θ and a conditional expectation $\mu_{ijt} = E [\text{NumReviews}_{ijt} \mid \beta, \alpha]$ modeled as:

$$\mu_{ijt} = \exp \left(\beta_1 \text{Winner}_i + \beta_2 \text{Post}_k + \beta_3 (\text{Winner}_i \times \text{Post}_k) + \sum_{j=1}^J \alpha_j \text{Pair}_j \right) \quad (2)$$

$$= \exp(\beta_1 \text{Winner}_i) \exp(\beta_2 \text{Post}_k) \exp(\beta_3 (\text{Winner}_i \times \text{Post}_k)) \exp \left(\sum_{j=1}^J \alpha_j \text{Pair}_j \right) \quad (3)$$

$$(4)$$

Note that in this model, coefficients affect the conditional mean multiplicatively rather than additively. A β_3 of 0.40 for example implies that prizewinning novels review volume increased by a factor of $\exp(0.40) \approx 1.5$ times in post-award period, relative to the change in review volume of their corresponding finalists.

A Negative Binomial distribution is chosen over the usual Poisson distribution for counts due to its flexible mean-variance relationship. In particular, the Poisson model imposes the restriction that the conditional variance of NumRatings_{ijt} is equal to the conditional mean μ_{ijt} , while the Negative Binomial variance is a function of μ_{ijt} and the dispersion parameter θ , allowing more flexibility (Hoffmann 2016).

4 Results

Table 5 displays the results of models described in Section 3.1 and Section 3.2.

Table 5: Effect of Winning an Award on Novel Ratings and Popularity

	Ratings	Count of Reviews
Winner	0.072	0.208
	0.035 (0.046)	0.316 (0.513)
Post	-0.099	0.619
	0.026 (<0.001)	0.135 (<0.001)
Winner x Post	-0.028	0.386
	0.028 (0.316)	0.217 (0.082)
Regression Type	Linear	Negative Binomial
N	48 513	180
AIC	134 180.1	2227.8
BIC	134 602.0	2381.1

Matched-pair clustered standard errors are shown below coefficient estimates (p-values in parentheses).
Both models include matched-pair fixed effects.

We do not observe the paradox of publicity noted by Kovács and Sharkey (2014) in the ratings model results. The difference-in-differences estimate of Winner x Post is near 0 ($\beta_3 = -0.028, p > 0.05$) and insignificant at the 5% level, meaning that winning an award is estimated to have no evidence of an effect on the average post-awards ratings of prizewinning novels compared to finalist novels. This is in contrast to the large and significant effect size (an estimated 0.171 star decline in average ratings) found by Kovács and Sharkey (2014).

Prizewinning novels are estimated to have a slightly higher and average rating ($\beta_1 = 0.072, p < 0.05$) than finalists in the pre-award period, although the difference is less than 0.1 stars. This pre-award difference is mechanically small, due to the matching procedure noted in section 2.3, which selected finalist novels with similar pre-award average ratings to their corresponding winning novel. The ratings of finalists are estimated to decline ($\beta_2 = -0.099, p < 0.05$) in the post-award period, although the effect on ratings is again less than 0.1 stars.

Similarly, award winning is not shown to significantly effect the number of reviews received by prizewinning novels relative to finalists. Award winning novels are estimated to be slightly more popular than finalists in the pre-award period ($\exp(\beta_1) \approx 1.32, p > 0.05$) and finalists

are estimated to receive a higher volume of reviews in the post-award period compared to the pre-award period ($\exp(\beta_2) \approx 1.86, p < 0.05$).

Given that the number of reviews per novel is highly variable, with a minimum review count of 4 and a maximum of 5,356, these findings may be skewed by a small number of volatile low-review novels. To assess the sensitivity of results to the sample composition, we re-estimate the ratings difference-in-difference model restricted to subsamples of matched-pairs in which both novels have at least 50, 100, and 200 reviews in the both the pre-award and post-award periods.

Table 6: Sensitivity Award Effects to Minimum Review Count Thresholds

	Ratings (> 0)	Ratings (> 50)	Ratings (> 100)	Ratings (> 200)
Winner	0.072	0.070	0.108	0.084
	0.035 (0.046)	0.024 (0.003)	0.030 (<0.001)	0.033 (0.011)
Post	-0.099	-0.069	-0.090	-0.080
	0.026 (<0.001)	0.021 (0.001)	0.028 (0.001)	0.031 (0.009)
Winner x Post	-0.028	-0.047	-0.050	-0.024
	0.028 (0.316)	0.027 (0.086)	0.033 (0.136)	0.037 (0.520)
N	48 513	35 558	25 721	21 998
N Matched-Pairs	45	25	18	12
AIC	134 180.1	99 427.3	72 670.1	61 963.6
BIC	134 602.0	99 664.7	72 841.4	62 083.6

Matched-pair clustered standard errors are shown below coefficient estimates (p-values in parentheses).

All models include matched-pair fixed effects.

Table Table 6 shows the results of the original ratings model compared to the restricted sample models. All estimates of the difference-in-differences coefficient remain near 0 and insignificant, suggesting that the null results observed in Table Table 5 are not an artifact of influential novels with a low volume of reviews.

5 Discussion

This paper examined the impact of winning a major literary award on the perception of a novel’s quality, using changes in Goodreads ratings over time to identify shifts in readers’ evaluations of quality after a novel publically receives an award. Following the approach of Kovács and Sharkey (2014), the ratings trajectory of winning novels is compared to that of

similar shortlisted novels using a difference-in-differences design with matched-pair fixed effects. This design aims to isolate the effect of award-winning from other ratings trends related to specific genres or periods. The results indicate that, contrary to prior work assessing the impact of awards on cultural products, found no evidence that awards negatively impacted the reviews of award-winning novels or evidence that award-winning novels received substantially more reviews in the post-award period than other finalists (Kovács and Sharkey 2014; Rossi and Schleef 2024; Rita et al. 2022).

5.1 Weaknesses and next steps

Several limitations remain in this analysis. Unlike in the analysis by Kovács and Sharkey (2014), the matched-pair design failed to eliminate statistically significant differences in the pre-award ratings of award-winning novels, suggesting that the matched-pair design may not have fully eliminated differences in the ratings trends of prizewinners and their corresponding finalists. Goodreads ratings as a measure of readers' perception of quality also face a number of shortcomings. For example, ratings are voluntary and asynchronous, meaning that changes in ratings may be attributable to changes in reviewer composition (e.g. readers who were gifted a prizewinning or shortlist novel) rather than changes in the perception of individual readers. As noted in Section 2.2, Goodreads ratings are also predisposed to tend towards the average or early ratings and contain a large proportion of sponsored reviews which may be biased.

References

- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- . 2025. *Tinytable: Simple and Configurable Tables in 'HTML', 'LaTeX', 'Markdown', 'Word', 'PNG', 'PDF', and 'Typst' Formats*. <https://doi.org/10.32614/CRAN.package.tinytable>.
- Bergé, Laurent. 2018. “Efficient Estimation of Maximum Likelihood Models with Multiple Fixed-Effects: The R Package FENmlm.” *CREA Discussion Papers*, no. 13.
- Ferman, Bruno, and Cristine Pinto. 2019. “Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity.” *The Review of Economics and Statistics* 101 (3): 452–67. https://doi.org/10.1162/rest_a_00759.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Hoffmann, John P. 2016. *Regression Models for Categorical, Count, and Related Variables: An Applied Approach*. Oakland, CA: University of California Press.
- Hu, Yuerong, Zoe LeBlanc, Jana Diesner, Ted Underwood, Glen Layne-Worthey, and J. Stephen Downie. 2023. “Complexities of leveraging user-generated book reviews for scholarly research: transiency, power dynamics, and cultural dependency.” *Int. J. Digit. Libr.* 25 (2): 317–40. <https://doi.org/10.1007/s00799-023-00376-z>.
- Kovács, Balázs, and Amanda J. Sharkey. 2014. “The Paradox of Publicity: How Awards Can Negatively Affect the Evaluation of Quality.” *Administrative Science Quarterly* 59 (1): 1–33. <https://doi.org/10.1177/0001839214523602>.
- Lederrey, Gael, and Robert West. 2018. “When Sheep Shop: Measuring Herding Effects in Product Ratings with Natural Experiments.” *CoRR* abs/1802.06578. <http://arxiv.org/abs/1802.06578>.
- Müller, Kirill. 2025. *Here: A Simpler Way to Find Your Files*. <https://doi.org/10.32614/CRAN.package.here>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Bryce Mecum, et al. 2025. *Arrow: Integration to 'Apache' 'Arrow'*. <https://doi.org/10.32614/CRAN.package.arrow>.
- Rita, Paulo, Celeste Vong, Flávio Pinheiro, and João Mimoso. 2022. “A sentiment analysis of Michelin-starred restaurants.” *European Journal of Management and Business Economics* 32 (3): 276–95. <https://doi.org/10.1108/EJMBE-11-2021-0295>.
- Rossi, Michelangelo, and Felix Schleef. 2024. “Quality Disclosures and Disappointment: Evidence from the Academy Nominations,” December. <https://doi.org/10.2139/ssrn.5073411>.
- Wan, Mengting, and Julian J. McAuley. 2018. “Item recommendation on monotonic behavior chains.” In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys2018, Vancouver, BC, Canada, October 2-7, 2018*, edited by Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O’Donovan, 86–94. ACM. <https://doi.org/10.1145/3240323.3240369>.

- Wan, Mengting, Rishabh Misra, Ndapa Nakashole, and Julian J. McAuley. 2019. “Fine-Grained Spoiler Detection from Large-Scale Review Corpora.” In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, edited by Anna Korhonen, David R. Traum, and Lluís Màrquez, 2605–10. Association for Computational Linguistics. <https://doi.org/10.18653/V1/P19-1248>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2025. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://doi.org/10.32614/CRAN.package.stringr>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://doi.org/10.32614/CRAN.package.dplyr>.
- Wickham, Hadley, and Lionel Henry. 2025. *Purrr: Functional Programming Tools*. <https://doi.org/10.32614/CRAN.package.purrr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2025. *Readr: Read Rectangular Text Data*. <https://doi.org/10.32614/CRAN.package.readr>.