Can Lower(ed) Expert Opinions Lead to Better Consumer Ratings?: The Case of Michelin Stars *

Xingyi Li
UCL School of Management, University College London, xingyi.li.16@ucl.ac.uk

Yiting Deng
UCL School of Management, University College London, yiting.deng@ucl.ac.uk

Puneet Manchanda
Ross School of Business, University of Michigan, pmanchan@umich.edu

Bert De Reyck
Lee Kong Chian School of Business, Singapore Management University, bdereyck@smu.edu.sg

Expert opinion exerts tremendous influence on the purchase journey, but its effect on overall consumer experience is ambiguous as it can give rise to both "expectation" and "reputation" effects. This paper explores the effect of expert opinions on consumer experience via the lens of consumer reviews in the restaurant industry, where the expert opinions are conveyed by Michelin stars. The paper uses a unique data set based on the Michelin Guide for Great Britain & Ireland from 2010-2020. The data include consumer reviews on TripAdvisor for all restaurants that were awarded Michelin stars during this period and a large pool of potential control restaurants. We apply two synthetic-control-based methods to estimate the effect of Michelin star changes on the sentiment and content of consumer reviews. We find that decreases in Michelin stars improve consumer review ratings. We examine three potential explanations for this finding. First, the positive expectation effect of lowered expert opinions outweighs the negative reputation effect. Second, there could be a change in the customer mix. Third, there may be changes on the supply side, e.g., in service levels. We find support for the first explanation, rule out the second but cannot entirely rule out the third. The analysis of review content further shows that a loss in Michelin stars leads consumers to become less focused on value and become less demanding regarding service. We discuss the implications of our findings for restaurant managers, the Michelin Guide, and other businesses that provide experience goods.

Key words: Expert Opinions, Consumer Reviews, Restaurant Industry, Michelin Star Ratings, Experience Goods, Synthetic Control Methods

---

1. Introduction

Customers look to experts and their opinions in their purchase journey as they consider them to be trustworthy sources (Chen and Xie 2005, Johnson et al. 2005, Hilger et al. 2011, Chen et al. 2012, Friberg and Grönqvist 2012, Ashenfelter and Jones 2013). In the movie industry, 60% of surveyed U.S. consumers stated that movie critic reviews can influence their decision to watch a movie (Statista 2017). In the book industry, awards such as the Booker Prize,[1] for the best novel of the year, are based on expert opinions and have been shown to have a significant impact on book sales (Ginsburgh 2003). In the restaurant industry, the Michelin Guide is one of the best-known and most prestigious expert rating systems, guiding diners in their restaurant choices (Gergaud et al. 2015). Other well-known examples of expert opinion include the American Automobile Association (AAA)'s Diamond rating in the hotel industry, Robert Parker's Wine Advocate score in the wine industry, and the J.D. Power rating in the automobile industry. Not surprisingly, it has been shown that favorable expert opinions can in general benefit product sales (e.g., Friberg and Grönqvist 2012, Ashenfelter and Jones 2013). However, research on the effect of expert opinions on consumer experience, including post-purchase interactions and consumer evaluations, is relatively sparse. A few notable exceptions include Kovács and Sharkey (2014), Gergaud et al. (2015) and Rossi (2021), who focus specifically on the effect of positive expert opinions (winning awards). In contrast, our paper explores the impact of both positive (winning awards) and negative expert opinions (losing an existing award) on consumer evaluations. Understanding the full spectrum of expert opinions on consumer evaluations is important, because consumer evaluations not only directly reflect their experiences, but also carry tangible behavioral and financial implications, including repeat purchase decisions, revenues and peer recommendations (Mittal et al. 2021, Morgan and Rego 2006).

* The authors would like to thank the editor, the associate editor, and three anonymous referees for their insightful and constructive suggestions. Special thanks are extended to Georgios Zervas for very generously sharing his data on restaurant demand in New York City. The authors would also like to thank Tat Chan, Pradeep Chintagunta, Anthony Dukes, Natasha Zhang Foutz, Yiqi Li, Jiwoong Shin, Adam Smith, Yu Song, Qingliang Wang, Jiajia Zhan, Xu Zhang, and participants at the 2022 and 2021 Marketing Science Conferences, the 2021 LBS Transatlantic Doctoral Conference, the 2022 Customer Journeys in a Digital World Conference, the 2022 AI/ML Conference, the 2024 POMS Conference, the 2024 Marketing Dynamics Conference, as well as seminar participants at Católica Lisbon School of Business & Economics, Chinese University of Hong Kong (Shenzhen), City University of Hong Kong, ESSEC, HEC Paris, Ohio State University, Peking University, Shanghai University of Finance and Economics, Shenzhen University, University of California, Irvine, University of Cambridge, University of Virginia, and Université Laval for useful comments. All errors are our own.

[1] `https://publishingperspectives.com/2022/03/awards-the-international-booker-prize-names-its-2022-longlist/`

In this paper, we investigate the effect of both favorable and unfavorable expert opinions on consumer experience. Theoretically, the impact of expert opinions on consumer experience is ambiguous. On the one hand, favorable expert opinions, seen as quality signals, enhance the reputation of the business (e.g., Hilger et al. 2011, Chen et al. 2012, Ashenfelter and Jones 2013). Consequently, business with their newly gained reputation can potentially witness improved consumer experience driven by consumer conformity, that is, individuals adjust their behaviors or beliefs to align with those of a group or social norm (Asch 1955). We refer to this positive effect of expert opinions as the *reputation effect*. On the other hand, consumer evaluations of their experiences are also based on their expectations in the sense that consumers first have expectations about an experience, then the actual experience, and then they evaluate their experience by comparing it with their expectations. As such, an experience that exceeds/meets/fails to meet their expectations is considered great/good/bad. Past work (e.g., Diehl and Poynor 2010, Fogarty 2012, Gergaud et al. 2015, Sands 2020, Rossi 2021) has shown that endorsements from experts can influence expectations and as a result influence the final experience (relative to these changed expectations). Therefore, higher expert opinions raise consumer expectations and potentially lead to disappointment as these expectations get harder to meet or exceed, while lower expert opinions can moderate expectations, which then become easier to meet or exceed, leading consumers to be more delighted with their experience. We refer to this effect of expert opinions as the *expectation effect*, noting that it is likely to be a negative effect when expert opinions become more positive.

Thus, our objective is to understand the net impact of expert opinions on consumer experience through the lens of consumer reviews. In addition to examining the relative importance of the expectation effect and reputation effect, we consider two other possible explanations that may affect consumer reviews: one being that changes in expert opinions may change the mix of customers visiting the business, and the other being that businesses may respond to changes in expert opinions by making supply-side adjustments.

Specifically, we measure the net effect of expert opinions on consumer experience in the restaurant industry, where the expert opinions are conveyed by the Michelin Guide. The Michelin Guide started evaluating restaurants in France in 1900, awarding "stars" to denote quality, and does so worldwide now. Winning Michelin stars can be can be seen as equivalent to as winning awards. We choose this setting for three reasons. First, the restaurant industry has a substantial impact on the economy. According to the National Restaurant Association, the U.S. restaurant industry is forecast to reach $898 billion in sales and provide 14.9 million industry jobs in 2022.[2] Second,

[2]https://restaurant.org/research-and-media/media/press-releases/association-releases-2022-state-of-the-restaurant-industry-report/

the Michelin Guide is updated every year in many countries or regions based on anonymous expert evaluations, with some restaurants added to the list or awarded more stars and others removed from the list or awarded fewer stars. Such annual updates provide us an opportunity to identify the effect of expert opinions summarized (via changes) in awarded Michelin stars. Third, both the reputation effect and the expectation effect have been documented in this context. For example, head chefs describe being awarded a Michelin star as akin to winning an Oscar in Hollywood.[3] In another instance, the Greenhouse restaurant in London witnessed a 25% increase in diners when it went from one to two Michelin stars.[4] However, Michelin stars can also negatively affect restaurants through heightened consumer expectations. It has been reported that Michelin receives more than 45,000 letters and 7,000 emails from customers every year, and about 25% of these are complaints about unsatisfactory experiences (Johnson et al. 2005). As a chef at a Michelin-starred restaurant stated, "Customers become more demanding, and people expect more of you and criticize things."[5] There have also been cases where the increased pressure and expectations have led chefs to "give back" Michelin stars by revamping their restaurants and food.[6] In fact, this phenomenon - the potential negative impact of Michelin stars - is labelled the "Michelin curse" in the dining industry and food media. We also see evidence in support of this in the consumer reviews we collected.[7]

We construct a unique data set based on the Michelin Guide for Great Britain & Ireland for the years 2010 to 2020. The "Great Britain & Ireland" guide covers restaurants in England, Wales, Scotland, Northern Ireland and the Republic of Ireland in one guide book every year. Our dataset consists of 262 restaurants that have been awarded Michelin stars at least once within this time period and 1,257 other "fine-dining" restaurants that never had or received Michelin stars in the same period. We collect consumer reviews for each of these restaurants from TripAdvisor to understand the consumer post-purchase experience and evaluations. We focus on TripAdvisor because it is more popular and influential than other platforms (e.g., Google, Facebook, Yelp)

---

[3] `https://www.fcsi.org/foodservice-consultant/worldwide/the-little-red-book/`

[4] `https://www.thestaffcanteen.com/Editorials-and-Advertorials/impact-michelin-stars-business`

[5] `https://www.bighospitality.co.uk/Article/2017/09/28/Michelin-Guide-chefs-discuss-is-it-still-relevant?`

[6] `https://www.bbc.com/news/world-us-canada-62854914`

[7] For example, in our consumer review data, we find that increasing Michelin stars leads to heightened consumer expectations, e.g., "Bibendum has 2 michelin stars and is very expensive-so our expectations were high..."; "Wouldn't come here again and left feeling annoyed that we had spent £260 on which we felt should have been of a higher standard for 2 michelin stars." Meanwhile, losing Michelin star(s) sometimes leads to improved consumer experiences, e.g.,"This really was the best food I have ever eaten (even compared to a Michelin starred restaurant!)";"The atmosphere is relaxed, friendly, welcoming...a real home from home (unlike some of Edinburgh's other fine dining/Michelin star establishments)."

for UK consumers.[8] We focus on two kinds of information in these reviews. First, we look at the review sentiment, measured via the five-point scale review rating on overall experience. Second, we analyze the textual content of these consumer reviews in order to gain deeper insights into the underlying factors influencing the ratings. For both kinds of review information, we control for "supply" side changes, primarily by restricting our analyses to restaurants that did not change their menu (we collect current and past menus from the restaurant websites) in response to the Michelin star changes. While our main analysis focuses on Great Britain & Ireland, the findings are expected to hold more broadly, as demonstrated by the replication study with New York City data in Section 6.5.

We apply two synthetic-control-based methods (Abadie et al. 2010, Li 2020) to identify the net effect of Michelin star(s) changes on consumer reviews. In the first method (SCM-DiD), we create a time-varying synthetic control restaurant that best matches the focal awarded restaurant, and then apply the difference-in-differences framework (Hackmann et al. 2015). The second method employs the cohort-based synthetic difference-in-differences (SynthDiD) - see Arkhangelsky et al. (2021) and Berman and Israeli (2022). In terms of the textual analysis of the review data, we extend established Latent Dirichlet allocation (LDA) methods (e.g., Tirunillai and Tellis 2014, Büschken and Allenby 2016, Puranam et al. 2017, Hollenbeck 2018) by allowing for heterogeneous hyper-parameters based on review characteristics and semantic word characteristics.

Setting wise, our work is closest to Gergaud et al. (2015), who show that Michelin stars improve consumers' perceived quality of the awarded restaurant (measured via the Zagat surveys). However, our work differs in three significant ways. First, Gergaud et al. (2015) only consider the *first* publication of the Michelin Guide in a single market (New York City) in 2005. This means that they can only examine the effect of *gaining* Michelin stars. In contrast, we consider the Michelin Guide for Great Britain & Ireland during an 11-year period (2010-2020), which allows us to identify all types of changes in Michelin stars and examine the effect of *both gaining and losing* Michelin stars. Second, we rely on consumer reviews, which arise organically rather than through responses to (survey) questions (as that paper does), reducing the potential for bias or distortion associated with survey research, such as sampling bias and non-response bias (e.g., Copas and Li 1997), and social desirability response bias (e.g., Krosnick 1999), among others. Consumer reviews also provide deeper insights as we are able to analyze both the rating and the associated text. Third, Gergaud et al. (2015) use difference-in-differences and propensity score matching methods. We are able to

[8] `https://bdaily.co.uk/articles/2019/06/26/34s-of-uk-consumers-check-online-reviews-tripadvisor-25x-more-influential-than-google`

leverage state-of-the-art methods in causal inference - two synthetic-control-based methods - that provide better identification, especially in terms of controlling for time-varying confounders (cf. Xu 2017).

Our results on review sentiment show that decreases in Michelin star(s) improve the consumer review ratings. In contrast, an increase in Michelin star(s) has no impact on the consumer review ratings. Turning to the analysis of the review content data, we find that when a restaurant loses or receives fewer Michelin stars, consumers become less demanding on service aspects and also focus less on "value for money" considerations. In addition, consumers also appear less concerned about food in their reviews. These results are consistent across both synthetic-control-based methods. We also show the robustness of these results via an analysis that uses observable restaurant characteristics to select the control group, analyses with an alternative dependent variable and an alternative time window, and an additional falsification test.

Our findings support the explanation that the positive expectation effect of lowered expert opinions outweighs the negative reputation effect, and we present evidence suggesting that the second mechanism of changes in consumer mix is unlikely to be the main driver. However, the third mechanism – supply-side changes – cannot be entirely ruled out.

Our findings go some way in terms of shedding light on the "Michelin curse." The Michelin Guide has five publicly acknowledged assessment criteria: quality of the products, mastery of flavor and cooking techniques, the personality of the chef in the cuisine, value for money, and consistency between visits.[9] In order to gain and/or keep a Michelin star, restaurants need to perform to satisfy these criteria. Many chefs struggle with these, especially consistency as that dampens creativity and lowers innovation. In fact, according to Hayward (2021), Michelin awards "damage" restaurants, causing them to narrow their creativity to obtain stars and to stop innovating in order to keep the stars. Overall, our paper suggests that losing Michelin stars is not necessarily bad news for restaurants, especially vis-a-vis the consumer experience. Conversely, winning Michelin stars does not seem to improve the customer experience in any material way.

To summarize, this paper makes the following contributions. First, we conduct a rigorous analysis on the (net) effect of expert opinions on consumer experiences. Our findings show that, in our setting, a decrease in Michelin stars, reflecting a lower(ed) expert opinion, can lead to better consumer review ratings. To the best of our knowledge, this is the first instance of the documentation of this outcome. We conducted a comprehensive series of analyses to demonstrate that the alternative explanations are less likely to be the main drivers of our main finding. Second,

[9] `https://guide.michelin.com/en/article/news-and-views/how-to-get-michelin-stars`

by analyzing consumer review text data, we identify key drivers of the customer experience, further enriching managerial insights on the value of receiving favorable or unfavorable expert opinions. Third, by adopting two synthetic-control-based causal inference methods and an augmented LDA model, we provide a rigorous and general empirical framework for analyzing consumer responses to external shocks using review data. Finally, we provide a data-based explanation for the "Michelin curse," offering implications for chefs, restaurant managers, the Michelin Guide, and other businesses that provide experience goods.

The rest of the paper is organized as follows. We describe the data and present descriptive statistics in Section 2, followed by the empirical strategy in Section 3 and empirical results in Section 4. We test multiple alternative explanations in Section 5 and conduct robustness checks in Section 6. Finally, we conclude in Section 7 with a discussion of the managerial implications, limitations and potential future extensions.

2. Data

2.1. The Michelin Guide and Awarded Restaurants

The Michelin Guide evaluates restaurants via the use of a group of anonymous inspectors that operate worldwide. Inspectors are anonymous when visiting the potential restaurants in order to guarantee that restaurants treat them as regular consumers.[10] Every decision relating to Michelin stars is decided by multiple inspectors from different global regions who take turns to visit a restaurant in order to ensure that the final outcome is based on a consensus (among inspectors). In other words, no single inspector can assign or remove Michelin stars for a restaurant.

We construct a comprehensive data set based on the Michelin Guide for Great Britain & Ireland from the year 2010 to 2020. We denote restaurants which received Michelin stars at least once within this time frame as *awarded restaurants*. For each awarded restaurant, we extract the restaurant's characteristics (e.g., official website URL, address, postcode, price level and cuisine type) from the restaurant's TripAdvisor page, and use a postcode checker to identify whether the restaurant is in an urban or a rural area.[11] In total, our data cover 262 awarded restaurants that received Michelin stars at least once.[12] Among these awarded restaurants, 91 (34.7%) are located in London,

[10] See, for example, `https://www.forbes.com/sites/karlaalindahao/2019/10/23/the-secret-life-of-an-anonymous-michelin-restaurant-inspector-2019/?sh=230efd5135c9`

[11] The rural/urban classification is based on offices of national statistics in the UK (i.e., England, Wales, Scotland, Northern Ireland) and Ireland. In England and Wales, the rural/urban classification was developed by the Office for National Statistics. In Scotland, the rural/urban classification was developed by the Scottish government's Geographic Information Science & Analysis Team. In Northern Ireland, the rural/urban classification was published by the Northern Ireland Statistics and Research Agency (NISRA). In Ireland, an interactive map, "area type classification," was developed by the Central Statistics Office.

[12] We exclude twelve (out of 278) restaurants that did not have a TripAdvisor page, and four (out of 278) that did not have an official website.

235 (89.7%) are associated with the highest price level (as labeled by TripAdvisor), 234 (89.3%) specialize in European cuisines (e.g., British, European and French, etc.), and 170 (64.9%) are located in urban areas.

As our goal is to analyze the effect of expert opinions (conveyed via Michelin stars) on consumer reviews, it is crucial to observe Michelin star changes (either an increase or a decrease in the number of Michelin stars, or an addition to or a deletion from the Michelin Guide). We define the *guidebook year* as the period between the publication dates of two consecutive guides. For example, the 2019 Michelin Guide was published on October 1, 2018 and the 2020 Michelin Guide was published on October 7, 2019, so the period between these two dates corresponds to guidebook year 2019. During these 11 guidebook years, 207 (79.0%) awarded restaurants experienced Michelin star changes at least once, and the remaining 55 (21.0%) restaurants kept the same Michelin stars throughout.

Table 1 lists the Michelin star awards and Michelin star changes by guidebook year. Every guidebook includes more than one hundred awarded restaurants, most of which are one-star restaurants. Michelin star increases can be new additions to the Michelin list (e.g., from no-star to one-star) or gaining more stars (e.g., from one-star to three-star), and Michelin star decreases can be removals from the Michelin list (e.g., from one-star to no-star) or losing stars but remaining on the list (e.g., from three-star to one-star). In total, there are 269 star changes, with 174 star increases and 95 decreases. In this paper, we do not separate the cases within Michelin star increases and Michelin star decreases, because we only observe 15 (out of 174 star increases) instances where an awarded restaurant gained more stars, and 6 (out of 95 star decreases) instances where an awarded restaurant lost stars but remained on the list.

2.2. Pool of Control Restaurants

For identification purposes, we further construct a large pool of *control restaurants*, which never received Michelin stars during the data period, are located in the cities with at least one awarded restaurant and are categorized as "fine-dining" on TripAdvisor.[13]

Specifically, we take the following steps to construct the pool of control restaurants. First, we check the city information for each awarded restaurant on Google, and then collect the city's TripAdvisor restaurant page. The 262 awarded restaurants are located in 73 cities in Great Britain and Ireland. Second, from the city's TripAdvisor restaurant page, we scrape the URL links of all "fine-dining" restaurants that are not in the list of the 262 awarded restaurants. We are able to

[13]On TripAdvisor, restaurants are assigned one of the three labels: "Cheap-eats," "Mid-range," and "Fine-dining." Fine-dining restaurants are typically associated with the price-level symbol "££££", though in rare instances, some are marked with the symbol "£££" or "££". Specifically, 27 (out of 262) awarded restaurants and 60 (out of 1,257) control restaurants are fine-dining with "££" or "£££" price labels.

Table 1 **Summary of Michelin Stars (2010 to 2020)**

Guidebook year	# Michelin restaurants				# Michelin star changes				
	total	one-star	two-star	three-star	total	increase–additions to the guide	increase–awarded but gaining more stars	decrease–removals from the guide	decrease–losing stars but remaining on the guide
2010	117	99	14	4	–	–	–	–	–
2011	123	103	16	4	15	10	1	4	0
2012	132	110	18	4	19	13	2	4	0
2013	141	117	20	4	21	14	2	5	0
2014	152	127	21	4	20	15	1	4	0
2015	159	134	21	4	19	13	0	6	0
2016	163	137	22	4	32	17	1	13	1
2017	170	146	20	4	33	18	2	11	2
2018	171	146	20	5	26	13	1	12	0
2019	180	155	20	5	39	22	1	13	3
2020	187	159	23	5	45	24	4	17	0

Note: We do not consider Michelin star changes in guidebook year 2010 because guidebook year 2010 provides the initial star levels for the period under investigation.

collect the TripAdvisor URLs of 1,803 “fine-dining” restaurants in total, which includes 262 awarded restaurants and 1,541 “fine-dining” restaurants that never received Michelin stars during the data period (i.e., control restaurants). Third, similar to data collection for awarded restaurants, we collect information on each of these 1,541 control restaurants, including their official website URL, characteristics (e.g., address, postcode, price level and cuisine type), and rural/urban classifications. 284 (out of 1,541) control restaurants did not receive consumer reviews during the period of our study, and are removed from the control pool. Thus, we have a pool of 1,257 control restaurants, of which 1,197 (95%) are associated with the highest price level (as denoted by TripAdvisor).

2.3. Consumer Review Data

We scrape TripAdvisor consumer reviews for each of the 262 awarded restaurants and the 1,257 control restaurants. As discussed earlier, TripAdvisor is chosen because it is more popular and influential than other platforms (e.g., Google, Facebook, Yelp) for UK consumers. The consumer reviews include the review text and an overall evaluation of the dining experience on a five-point scale, with a higher rating indicating a better experience. Our sample includes 889,660 consumer reviews.

Table 2 reports key statistics on the review data by Michelin star level. Note that a single awarded restaurant can appear with different Michelin star levels in different years. Overall, holders of higher Michelin stars have more consumer reviews on TripAdvisor. This is likely due to the reputation effect of Michelin stars: consumers are more likely to visit, review, and indicate their satisfaction (or not) with an awarded restaurant. While the consumer review ratings for the awarded restaurants are somewhat higher than those for the control restaurants, the differences are not statistically significant.

Table 2 Summary Statistics of the Review Data (by Michelin star level)

	Awarded Restaurants				Control
	No-star	One-star	Two-star	Three-star	Restaurants
Number of restaurants	252	241	31	6	1,257
Number of reviews	46,044	146,683	35,445	7,521	653,967
Average number of reviews per restaurant	183	609	1143	1,254	520
Mean of restaurant-level average review rating (s.d.)	4.50 (0.40)	4.47 (0.28)	4.58 (0.22)	4.63 (0.25)	4.25 (0.60)

Note: "No-Star" refers to awarded restaurants in guidebook years when they did not receive a Michelin star. "Control Restaurant" refers to restaurants that never received Michelin stars in the data period.

2.4. Use of Menus as Supply-Side Controls

Changes in Michelin star status could result in restaurants adjusting various aspects such as food, decor, service, etc. As mentioned earlier, we control for these via our sample construction. First, we retrieve all available historical menus for each awarded restaurant and control restaurant since the publication of the Michelin Guide 2010, using the Wayback Machine (`https://archive.org/web/`) to access archived versions of the restaurants' official websites. Then, for each restaurant, we check menus on each date that the website has been archived,[14] and determine whether there have been any changes compared to the last archived menu. Any modifications to the menu, such as adding items, deleting items, changing prices, or altering item descriptions, are classified as menu changes. Over the 11-year period (2010 - 2020), we find that the number of menu changes is quite modest, averaging 15.8 changes for an awarded restaurant and 5.1 for a control restaurant. In order to control for menu changes, we restrict our data to include only those awarded restaurants and control restaurants *without* menu changes in the 180-day period around the Michelin Guide release (90 days before and 90 days after the publication date). As a result, we exclude 17 (out of 269) star change observations in the awarded group and 110 (out of 1,257) restaurants in the control group.[15]

Second, the restriction of the time window to just 90 days post the Michelin Guide release makes it unlikely that restaurants can successfully make major (non-menu) changes, e.g., decor and/or re-training the staff to deliver a different service level. In addition, we carry out a detailed analysis

[14]Note that the Wayback Machine does not archive all websites on a daily basis.

[15]In robustness checks not reported in the paper, our main findings remain consistent without controlling for menu changes at the awarded and control restaurants. Results are available upon request from the authors.

on the trends in "service-related" review topics in the twelve-month period following the Michelin Guide updates, and find that the attention paid to service (in the reviews) stays stable. Section 5.1 provides the relevant details on supply-side explanations.

2.5. Final Sample in Main Analyses

After making the above selections, the final sample we use in the following empirical analysis includes 252 star changes (denoted as *treated unit*) and 1,147 control restaurants. Table 3 shows the number of awarded restaurants gaining Michelin stars, the number of awarded restaurants losing Michelin stars, and the number of control restaurants in the pool, by guidebook year. Note that not all restaurants have received consumer reviews every year, so the number of control restaurants varies by year and generally increases over time because of consumer review accumulation.

Table 3 Summary of the Number of Restaurants in Empirical Analyses (by Guidebook Year)

Guidebook year	# Awarded restaurants with Michelin star increases	# Awarded restaurants with Michelin star decreases	# Control restaurants
2011	11	4	427
2012	14	3	539
2013	16	5	594
2014	13	4	643
2015	13	5	705
2016	17	13	756
2017	20	12	812
2018	14	11	898
2019	23	13	978
2020	28	13	1,020
Total	169	83	7,372

Note: As a control restaurant can be included in the control pool for multiple guidebook years, the sum of the control restaurants units exceeds the total number of 1,147.

2.6. Additional Reviewer-level Data

To analyze whether changes in Michelin stars change the mix of consumers who visit the restaurant (e.g., Bondi et al. 2023), we further collect comprehensive data about the reviewers, as outlined below.

First, to understand if the restaurant attracts different types of consumers after the Michelin star change, we collect the TripAdvisor profile pages of reviewers who have reviewed an awarded restaurant within the 90-day guidebook windows. The TripAdvisor profile page contains reviewer-level information, such as their location of registration, registration time, and all of the reviews they have posted (not limited to those for the awarded restaurants). We collected TripAdvisor profile pages for 52,210 unique reviewers, who have written 1,617,923 reviews from 2010 to 2020.

Second, we collect restaurant information associated with these 1,617,923 reviews. These reviews are associated with 327,852 unique restaurants. For each of these 327,852 restaurants, we access its TripAdvisor page to collect restaurant characteristics and all consumer reviews. We were able to locate TripAdvisor pages for 279,359 (out of 327,852) restaurants. These restaurants have been reviewed by 45,274 (out of 52,210) reviewers in the data, and have received a total number of over 79 million reviews. These review data will enable us to assess whether changes in Michelin stars led consumers to visit a different type of restaurant.

3. Empirical Strategy

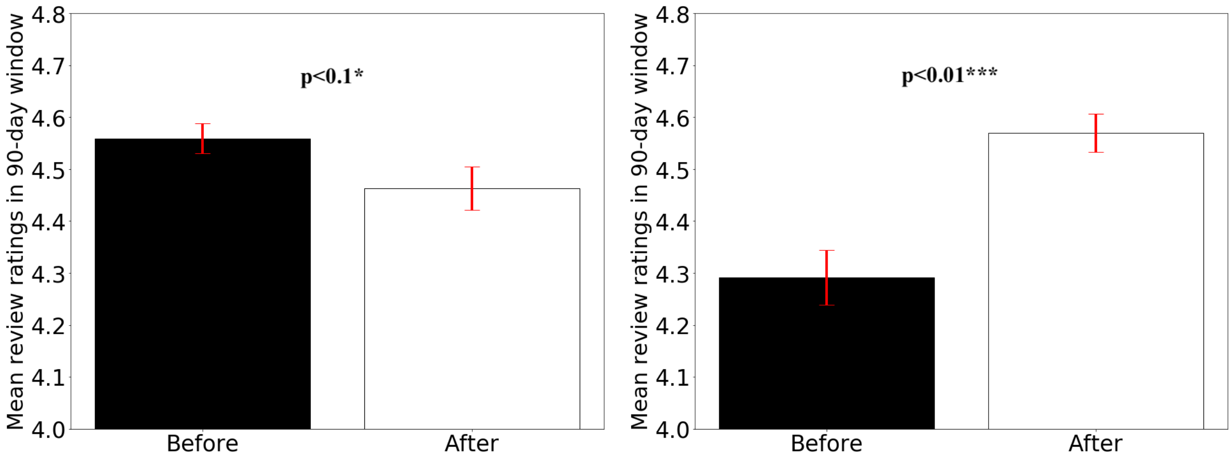
This section proceeds as follows. First, to provide model-free evidence, Section 3.1 shows the mean review ratings for treated units in the 90-day windows before and after Michelin star changes, respectively for those gaining stars and for those losing stars. Next, we describe two variants of the synthetic control method (SCM) for estimating the causal impact of Michelin star changes. Section 3.2 describes the first method SCM-DiD (Hackmann et al. 2015) and Section 3.3 describes the second method SynthDiD (Arkhangelsky et al. 2021, Berman and Israeli 2022). Both SCM-DiD and SynthDiD have been shown to provide clean identification and aid in causal inference. Each method has its own advantages. SCM-DiD creates a time-varying synthetic control restaurant to best match each treated unit and then applies the difference-in-differences framework, allowing controls for fixed effects at the restaurant level and the time level. SynthDiD separates treated units into guidebook-specific treated cohorts and then estimates the cohort-based synthetic difference-in-differences model, which relaxes the strong parallel-trend assumptions for all units and all time periods. We apply both methods to ensure the robustness of the results.

3.1. Model-Free Evidence

Figure 1 shows the mean review ratings received by the awarded restaurants in the 90-day windows before and after the Michelin star changes. Clearly, the restaurants with Michelin star increases (left panel) received lower consumer review ratings after the Michelin star changes, and the restaurants with Michelin star decreases (right panel) received higher consumer review ratings after the Michelin star changes. The initial model-free evidence suggests a relationship between the Michelin star changes and the consumer review ratings, which is in line with the expectation effect.

This pattern obviously does not control for potential confounding factors. Table 4 summarizes identification challenges, possible confounding factors and alternative explanations for the observed pattern(s), along with our approach to dealing with these.

Figure 1 **Mean Review Ratings in a 90-day Window Before/After Guidebook Release for Michelin Star Increases (left) and Michelin Star Decreases (right)**



Note. Error bars represent standard deviations.

3.2. Synthetic Control Method and Difference-in-Differences Framework (SCM-DiD)

After the release of a new guidebook, a restaurant is either treated (i.e., with Michelin star changes) or untreated (i.e., without Michelin star changes). In order to predict the potential outcomes of a treated unit "as if" there were no Michelin star changes, we employ the synthetic control method (SCM, Abadie et al. 2010, 2015) to create a best-matching control restaurant. The synthetic control method allows us to capture any possible trends that might affect identification of the effect of the Michelin star change.

For each treated unit (i.e., an awarded restaurant with a Michelin star change in a specific guidebook year), we create a donor pool which consists of all available control restaurants offering the same type of cuisine. Then, for the focal awarded restaurant and each control restaurant in the donor pool, we construct a "restaurant-guidebook year" panel of consumer reviews with the following variables: yearly average review ratings, yearly variance of review ratings, and yearly cumulative number of reviews.

Based on the "restaurant-guidebook year" panel data, we construct a synthetic control restaurant for each treated unit as a weighted combination of the donor restaurants, with weights chosen so that the resulting synthetic control restaurant best-approximates the treated unit in the pre-treatment period in terms of the relevant characteristics. On average, a synthetic control restaurant is constructed from a pool of 404 control restaurants. The outcome variable is the yearly average review rating. The predictors include the yearly variance of review ratings, yearly cumulative number of reviews, and price level. In addition, we follow Abadie et al. (2011) to include as a special

Table 4 Identification Challenges, Alternative Explanations and Proposed Solutions

Type			Solutions and Empirical Models
Identification Challenges		General trend	*SCM-DiD* (Section 3.2) and *SynthDiD* (Section 3.3). Robustness check using placebo guidebook publication dates (Section 6.4).
		Different panel lengths across restaurants before treatment	*SynthDiD* Use 18-month review data for each treated and control unit (Section 3.3).
		Control restaurants selected based on SCM may not be fully comparable with treated restaurants	*SCM-DiD* Robustness check using manually selected control restaurants based on location, price, and cuisine type (Section 6.1).
Alternative Explanations	Supply-side changes	Restaurants change menus	*SCM-DiD* and *SynthDiD* Focus on restaurants without menu changes in the 180-day period around the Michelin Guide publication date (Section 2.5).
		Restaurants make changes to serving size (food remains the same)	*SCM-DiD* and *SynthDiD* Robustness check using subset of restaurants evidencing consistency (Section 5.1.2).
		Restaurants make major non-food changes (e.g., decor, service)	*SCM-DiD* Focus on the short time window around the Michelin Guide publication date (Section 3.2). Robustness check using an alternative time window (Section 6.3). *Other Analysis* Focus on “service-related” topic metrics, and analyze probabilities of relevant topics over the twelve-month period between guidebook releases (Section 5.1.3).
	Demand-side changes	Restaurant demand changes	*SCM-DiD* and *SynthDiD* Use log-transformed normalized Google search intensity as the dependent variable. Use Farronato and Zervas (2022)’s OpenTable reservation as the dependent variable (Section 5.2.1).
		Consumers show sympathy towards restaurants losing stars	*SCM-DiD* and *SynthDiD* Use review volume as the dependent variable (Section 5.2.2). Replication with restaurants serving British cuisine (Section 5.2.2).
		Changes in the mix of customers visiting the restaurant	Analyze whether a restaurant attracts different types of consumers after the Michelin star change, and whether a Michelin star change led consumers to visit a different type of restaurant. (Section 5.2.3).
		Michelin star changes may change the proportion of extreme reviews	*SCM-DiD* and *SynthDiD* Use the percentage of 5-star reviews to measure restaurant-level sentiment (Section 6.2).

predictor, the average review rating in the 90-day pre-treatment period, to ensure that the synthetic restaurant is similar to the treated unit right before the treatment. For every restaurant, the outcome variable and predictors are calculated with an average number of 601 reviews. Therefore, we constructed 223 synthetic control restaurants corresponding to 223 (out of 252) treated units. The remaining 29 treated units do not have enough reviews on at least one side of the treatment time and therefore are dropped in the SCM procedure. For the 223 pairs of treated and synthetic control restaurants, Table A.1 in Online Appendix A presents the comparison results of their review characteristics during the pre-treatment period, which show that the treated and synthetic controls are comparable across all measured dimensions.

Next, we undertake an event study approach and focus our analysis on a window of 90 days before (pre-treatment window) and 90 days after (post-treatment window) the release of the new guidebook. The SCM procedure described above results in 223 pairs of treated and synthetic control restaurants. For each pair of treated unit and its synthetic control, we aggregate the reviews and retain observations in the pre- and post- windows, so that there are four observations on each pair: treated-pre, treated-post, control-pre, and control-post.

Finally, we estimate the effects of Michelin star changes on consumer reviews in a stacked difference-in-differences framework (Hackmann et al. 2015). When examining the effect on review sentiment, we use the mean consumer review rating as the dependent variable. When analyzing the effect on review content, we first extract topics from textual reviews and then use the mean probability of each topic as the dependent variable. The stacked difference-in-differences model is specified as follows:

$$\begin{aligned} Y_{it} = \; & \beta_1 After_{it} + \beta_2 After_{it} \times Increase_{it} + \beta_3 After_{it} \times Decrease_{it} \\ & + \beta_4 OneStar_{it} + \beta_5 TwoStar_{it} + \beta_6 ThreeStar_{it} \\ & + \beta_7 X_{it} + \beta_8 Z_{it} + \alpha_{p(i)w(t)} + \gamma_i + \varepsilon_{it}, \end{aligned} \tag{1}$$

where i denotes restaurant, t denotes guidebook year, $p(i)$ denotes the pair of restaurant i and its synthetic control, and $w(t)$ denotes the guidebook window defined as a window of 90 days before and 90 days after the release of guidebook for year t ($t \in \{2011, ..., 2020\}$). Therefore, the guidebook window $w(t)$ includes observations in guidebook year $t-1$ and observations in guidebook year t.

The dependent variable, Y_{it}, is the outcome of interest (e.g., mean review rating in sentiment analysis, mean topic probabilities in content analysis) for restaurant i in the part of the guidebook window belonging to guidebook year t. $After_{it}$ is an indicator variable which takes the value of 1 if the observation is in the post-treatment window, and takes the value of 0 otherwise. We include dummy variables - $Increase_{it}$ and $Decrease_{it}$ - to denote two treatment groups, indicating the changes in Michelin star level (i.e., increase, decrease, or unchanged).

Specifically, $Increase_{it}$ ($Decrease_{it}$) takes the value of 1 if restaurant i gained (lost) stars in guidebook year t compared with guidebook year *t–1*. The interaction term between $After_{it}$ and $Increase_{it}$ ($Decrease_{it}$) therefore measures the treatment effect above and beyond the general trend. Corresponding to the three-star rating system in the Michelin Guide, we add three indicator variables, $OneStar_{it}$, $TwoStar_{it}$, and $ThreeStar_{it}$, to control for the current Michelin star level of restaurant i in guidebook year t. X_{it} is a vector of cumulative review characteristics for restaurant i in the window belonging to guidebook year t, constructed based on all available reviews prior to the window. These characteristics are: the logarithm of the total number of reviews, the cumulative average review rating, and the variance of previous ratings. Z_{it} is a measure of average demand of restaurant i in the 90-day window, proxied by the normalized search intensity collected from Google Trends. We include pair–window fixed effect $\alpha_{p(i)w(t)}$ to control for unobservable factors specific to the restaurant pair $p(i)$ during the window $w(t)$. Restaurant fixed effect γ_i controls for unobservable time-invariant restaurant characteristics such as the restaurant's general decoration style, and ε_{it} is an idiosyncratic error term.

3.3. Synthetic Difference-in-Differences (SynthDiD)

The SCM-DiD model presented above includes 223 synthetic control restaurants, one for each treated unit. As our data span 11 guidebook years, these synthetic control restaurants may have different panel lengths before treatment, depending on the guidebook year of Michelin star changes. Different pre-treatment panel lengths in SCM may bias the estimates, thus we address this potential issue with the synthetic difference-in-differences (SynthDiD) approach (Arkhangelsky et al. 2021). SynthDiD allows both unit and time weights, where the unit weights are selected in a similar way as SCM, and time weights are added so that within a unit, the weighted average outcomes across pre-treatment periods approximate those in the post-treatment period.

The SynthDiD is designed for a balanced panel where the treated units have the same treatment time. In our setting, treatment time varies by restaurant. Therefore, we follow Berman and Israeli (2022) to adapt the SynthDiD method to the staggered treatment time by separating treated units into guidebook-specific treated cohorts, estimating the treatment effect for each cohort separately, and then aggregating them into an overall average treatment effect. We do this in four steps. First, for each guidebook year t, we create three cohorts: treated cohort $r_t^{increase}$ consisting of treated units with an increase in Michelin stars; treated cohort $r_t^{decrease}$ consisting of treated unites with a decrease in Michelin stars; and control cohort $r_t^{control}$ consisting of control restaurants. We denote the number of restaurants in the three cohorts respectively by $N_t^{increase}$, $N_t^{decrease}$, and $N_t^{control}$.

Second, for each restaurant in the guidebook-specific treated or control cohort, we extract review data in the period of one year before treatment to six months after treatment. We then divide the 18-month data into nine consecutive two-month blocks, and calculate the restaurant-level mean outcome (e.g., review rating in sentiment analysis, topic probabilities in content analysis) in each two-month block. A restaurant is excluded from the cohort if it does not have the full nine blocks of data, or if it is an awarded restaurant but has more than one change of Michelin stars within this 18-month period (i.e., changed Michelin stars in two consecutive years). As a result, we retain 148 (out of 252) treated units, including 95 treated units for gaining Michelin stars ($Increase$), and 53 treated units for losing Michelin stars ($Decrease$). Correspondingly, there are 4,334 control units. Table 5 summarizes the treated and control cohorts in the data constructed above.

Table 5 Summary of Treated and Control Cohorts in SynthDiD

	Treated Cohorts		Control Cohorts
	Increase	Decrease	
Total number of units	95	53	4,334
Avg. number of units in a guidebook-specific cohort	10.6	5.9	481.6
Avg. number of reviews per unit (within 18 months)	163.9	143.2	194.0

Note: As a control restaurant can be included in the control pool for multiple guidebook years, the sum of the control units exceeds the total number of control restaurants 1,147.

Third, for each guidebook year t, we estimate the cohort-level treatment effect of gaining Michelin stars, $ATT_t^{increase}$, using treated cohort $r_t^{increase}$ and control cohort $r_t^{control}$. Similarly, we estimate the cohort-level treatment effect of losing Michelin stars, $ATT_t^{decrease}$, using treated cohort $r_t^{decrease}$ and control cohort $r_t^{control}$. Standard errors are estimated using bootstrapping (Algorithm 2 of Arkhangelsky et al. (2021)), or the placebo method (Algorithm 4 of Arkhangelsky et al. (2021)) if a cohort includes only one treated restaurant.

Lastly, we aggregate the cohort-level treatment effects into the overall treatment effect (ATT) by taking the weighted average as follows:

$$ATT^{increase} = \frac{\sum_t N_t^{increase} \cdot ATT_t^{increase}}{\sum_t N_t^{increase}} \tag{2}$$

$$ATT^{decrease} = \frac{\sum_t N_t^{decrease} \cdot ATT_t^{decrease}}{\sum_t N_t^{decrease}} \tag{3}$$

Standard errors for the overall ATT are computed as a weighted average of the cohort-level standard errors.

4. Results

This section reports the estimation results from the SCM-DiD and SynthDiD analyses. Section 4.1 reports the effects of Michelin star changes on sentiment of consumer reviews. Section 4.2 reports the results on content of consumer reviews, where we first extract topics of consumer reviews using the Latent Dirichlet Allocation (LDA) model (Section 4.2.1), and then estimate the effect on topic probabilities (Section 4.2.2).

4.1. Effects of Michelin Stars on Sentiment of Consumer Reviews

Table 6 presents the results of the SCM-DiD model (Section 3.2), using the mean consumer review rating as the dependent variable in Equation (1). Column (1) controls only for Michelin star levels and fixed effects, and Column (2) adds the full set of controls. The estimated coefficient for $After$ is significantly negative, suggesting a declining trend in online ratings, which is consistent with prior literature (e.g., Moe and Trusov 2011, Li and Hitt 2008). The estimated coefficient for $After \times Increase$ is insignificant, suggesting that gaining Michelin stars does not lead to changes in consumer review ratings. However, the estimated coefficient for $After \times Decrease$ is significantly positive and the magnitude is larger than that of $After$, suggesting an increase in the consumer review ratings for restaurants that lost Michelin stars. This is likely driven by the expectation effect of Michelin stars: consumers lower their expectations for restaurants with Michelin star decreases, and as a result, tend to be more satisfied with the dining experience.

Table 7 presents the results of the SynthDiD model (Section 3.3).[16] The overall ATT for Michelin star increases remains insignificant, and the overall ATT for Michelin star decreases is significant with a value of 0.311. Based on the 18-month data around the treatment time, we construct Figures 2 and 3, which show the dynamic treatment effect estimates (with their 95% confidence intervals) for gaining and losing Michelin stars respectively.[17] Time 1 represents the first two-month block after the Michelin star change, and other times represent two-month blocks relative to the Michelin star change. In the pre-treatment periods (i.e., Time –5 to Time 0), the estimated ATT values are approximately zero in both figures, confirming the parallel pre-trends. In the post-treatment periods (i.e., Time 1 to Time 3), Figure 2 shows that the confidence intervals on the ATT for gaining Michelin stars contain zero in all periods, suggesting that consumer review ratings do not change after a restaurant gained Michelin stars. In Figure 3, it is evident that the ATTs for losing Michelin stars are positive, indicating an increase in consumer review ratings for restaurants that lost Michelin stars. Both plots are consistent with the overall ATTs reported in Table 7.

[16] The cohort-level ATT estimates are reported in Online Appendix C.

[17] See Online Appendix A.1 in Berman and Israeli (2022) for details on the methodology used to create the plot.

Table 6 Effects of Michelin Star Changes on Sentiment of Consumer Reviews by SCM-DiD

	Synthetic control + Difference-in-Differences	
	(1)	(2)
After	–0.069***	–0.045**
	(0.014)	(0.020)
After × Increase	0.040	0.049
	(0.052)	(0.056)
After × Decrease	0.318***	0.281***
	(0.058)	(0.064)
One Star	–0.060	–0.072
	(0.048)	(0.057)
Two Star	0.002	–0.024
	(0.081)	(0.102)
Three Star	0.998***	0.955***
	(0.196)	(0.196)
ln(number of reviews+1)		–0.056**
		(0.028)
Cumulative average rating		0.654*
		(0.365)
Cumulative rating variance		0.108
		(0.353)
ln(normalized search volume+1)		–0.017
		(0.103)
Pair–window FE	Yes	Yes
Restaurant FE	Yes	Yes
Observations	892	892
Number of pairs	223	223
R^2	0.851	0.856

Note: Robust standard errors clustered at pair level are in parentheses, and they are Diff-in-Diff regression-based clustered standard errors. In Online Appendix B, we report bootstrapped standard errors following Arkhangelsky et al. (2021) and Adalja et al. (2023). The results remain consistent. *** p<0.01,** p<0.05,* p<0.1.

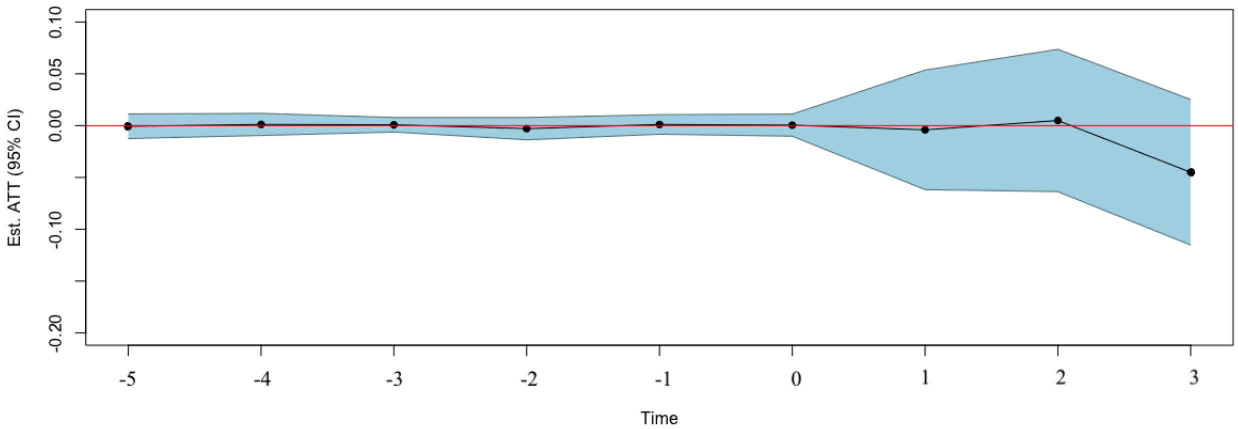
Table 7 Effects of Michelin Star Changes on Sentiment of Consumer Reviews by SynthDiD

	Aggregated Synthetic Difference-in-Differences	
	(1) Increase	(2) Decrease
Overall ATT	–0.015	0.311**
	(0.064)	(0.138)
Total number of treatment units	95	53
Total number of control units	4,334	4,334

Note: Aggregated standard errors are in parentheses. ** p<0.05.

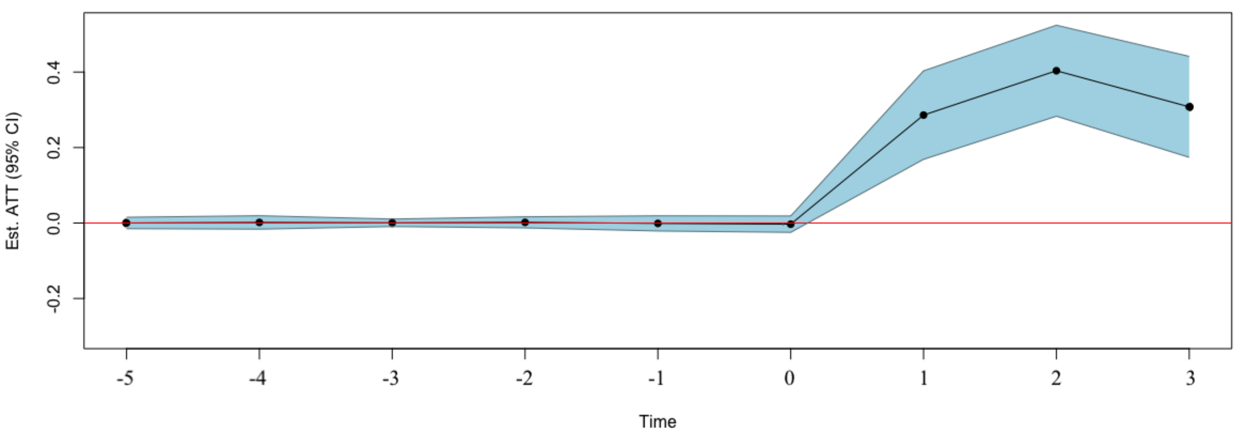
Together, our results show that when a restaurant loses stars, the positive expectation effect outweighs the negative reputation effect, leading to higher consumer ratings. In contrast, when

Figure 2 **SynthDiD Treatment Effects for Gaining Michelin Stars**



Note. Time –5 to Time 0 correspond to the one-year pre-treatment period (i.e., six consecutive two-month blocks), and Time 1 to Time 3 are post-treatment periods (i.e., three consecutive two-month blocks). Time 1 denotes the first two-month block after the Michelin star increase.

Figure 3 **SynthDiD Treatment Effects for Losing Michelin Stars**



Note. Time –5 to Time 0 correspond to the one-year pre-treatment period (i.e., six consecutive two-month blocks), and Time 1 to Time 3 are post-treatment periods (i.e., three consecutive two-month blocks). Time 1 denotes the first two-month block after the Michelin star increase or decrease.

a restaurant gains stars, it is possible that the positive reputation effect negates the potential negative effects of higher expectations, leading to an overall null effect. However, we are unable to separate the expectation effect from the reputation effect given the observational nature of our data. Interestingly, because of the asymmetric effects on gaining and losing Michelin star(s), these results potentially suggest that a restaurant can achieve more favorable consumer ratings if it first gains Michelin star(s) and then loses it/them. However, a lot more data with such changes is needed before this statement can be made definitively.

We further check whether the results vary with the restaurant location, price level or cuisine type, but do not find any such differences.

4.2. Content of Consumer Reviews

Having demonstrated the effect of Michelin star changes on the consumer review ratings, we next delve into the content or the reviews to understand the mechanisms behind the effects. We apply an LDA model to extract topics from textual reviews (Section 4.2.1), and estimate the effects of Michelin star changes on the identified topics using SCM-DiD and SynthDiD (Section 4.2.2).

4.2.1. LDA Model We estimate an LDA model to extract topics from textual reviews, extending the standard LDA model by allowing for heterogeneous hyper-parameters based on review characteristics and semantic word characteristics. We choose five topics based on the topic coherence score. Details on the model and estimation are provided in Online Appendix D.1.

Table 8 displays the top 20 words in descending order in terms of the posterior probability to be associated with each topic. It appears that Topic 5 discusses the general dining experience with an overall evaluation, whereas Topics 1–4 discuss the dining experience in four different aspects. Topic 1 relates to value for money. Generally, consumers think the experience is good but might be overpriced, as evidenced by the words “price,” “bite,” and “expensive.” Topic 3 centers around the menu and food, as evidenced by the use of words such as “starter,” “dessert,” “steak,” “beef,” “fish,” “cheese,” etc. Topic 5 includes words describing the general experience in various aspects (e.g., “experience,” “wine,” “food,” “dining,” “meal”). Both Topic 2 and Topic 4 relate to service but are associated with different valence. Topic 2 relates to complaints about services, such as issues regarding time (“time,” “wait,” “minute”), as well as interactions with service personnel (“waiter,” “staff,” “ask,” “come”), which possibly relate to attempts to resolve issues.[18] Topic 4 relates to positive service encounters, because most adjectives for this topic have a positive valence (“great,” “excellent,” “amazing,” “friendly,” and “attentive”). Our descriptions of topics continue to hold

[18] As an example, a representative review that has a high probability ($\theta > 0.85$) for Topic 2 (issues with order) is presented below. It was posted after the restaurant gained Michelin stars, which provides further evidence that consumers might have higher expectations after a restaurant gains Michelin stars. *“Been several times prior to the changes and the Michelin star award so maybe expectations were too high. On arrival seated ourselves in the bar, staff were busy in and out of restaurant no welcome smile or will be with you soon. Totally ignored. After about 10 minutes someone came to take drinks order was very pleasant and hospitable. Nice table taken to on time, extremely disappointed to be told on seated that there was only one lamb left which we immediately reserved. On taking our order we did politely express our disappointment that of only two meat choices one was not available, the response from the waitress was a shrug and well they are closed for the next two days! One of our party of 4 was very disappointed with the roast potatoes, tasted not fresh but rather as if been keep warm for hours. When paying the bill, a very reasonable bill for Michelin star, we did raise our complaints they were not received very well. Whatever business one is in, how complaints are treated gives an insight on the company and their standards, flitch of bacon came up wanting in this area more than in any other. Poor defensive excuses of new staff not properly trained, well they should have been.”*

when only considering words that are unique to each of the five topics (see Table D.1 in Online Appendix D.2).

Table 8 Top 20 Words Under the LDA Model (K = 5)

Rank	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
	Value for Money	Issues with Order	Menu and Food	Service and Staff	Overall Experience
1	food	table	main	food	menu
2	good	ask	starter	service	course
3	service	take	dessert	great	wine
4	price	order	cook	staff	experience
5	place	time	steak	excellent	food
6	great	get	good	recommend	dish
7	nice	book	dish	visit	tasting
8	menu	arrive	course	lovely	well
9	really	drink	delicious	amazing	star
10	wine	come	beef	friendly	every
11	bite	tea	meal	good	dining
12	get	staff	order	place	meal
13	like	wait	fish	time	chef
14	expect	say	cheese	lunch	eat
15	better	waiter	taste	delicious	staff
16	quality	minute	sauce	atmosphere	time
17	staff	bar	chocolate	definitely	visit
18	little	service	bread	attentive	win
19	expensive	leave	serve	birthday	michelin
20	quite	tell	menu	fantastic	best

We verify our interpretation of topic valence by checking the correlations between a review's overall rating and the probability of being associated with each of the five topics (value for money, issues with order, menu and food, service and staff, and overall experience).[19] In Table 9, Topic 5 (overall experience) is positively correlated with the overall rating, so the higher the Topic 5 probability, the higher the review rating. We find a correlation of –0.521 ($p < 0.0001$) between Topic 1 (value for money) and the overall rating. This is intuitive: consumers may be more likely to mention value for money when it is low, which may make them less satisfied. Surprisingly, Topic 3 (menu and food) has a negative correlation of –0.174 ($p < 0.0001$) with the overall rating, possibly because consumers tend to complain about food when mentioning it. While both Topic 2 and Topic 4 relate to service, Topic 2 (issues with order) is negatively correlated with the overall rating, whereas Topic 4 (service and staff) is positively correlated with the overall rating. These correlations are consistent with our topic interpretation above.

[19]The topic probabilities for each review add up to one.

Table 9 Correlations Between Overall Review Rating and Topic Probabilities

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
	Value for Money	Issues with Order	Menu and Food	Service and Staff	Overall Experience
Overall Rating	–0.521***	–0.551***	–0.174***	0.261***	0.513***

Note: *** $p<0.01$.

4.2.2. Effects of Michelin Stars on Topics of Consumer Reviews Given the topic distributions obtained from the LDA model, we aggregate the review-level topic distributions to restaurant level in the 90-day guidebook windows for SCM-DiD and in the two-month blocks for SynthDiD. Then, we analyze the effect of Michelin star changes on the topics of consumer reviews with models described in Section 3, using as dependent variable the mean probability of each of the five topics. Tables 10 and 11 respectively report the SCM-DiD and SynthDiD estimation results.

Column (5) of Table 10 shows that a decrease in Michelin stars is associated with an increase in the discussion of overall experience (Topic 5). As overall experience (Topic 5) is positively correlated with the review's overall rating, the result is consistent with our prior findings on sentiment of consumer reviews. We find that consumers are more likely to discuss value for money (Topic 1, Column (1)) when a restaurant gains Michelin stars, and are less concerned about it when a restaurant loses Michelin stars. This is consistent with reference dependence (Gerstner 1985, Winer 1986, Rao and Monroe 1989, Almenberg and Dreber 2011) and expectation effect: consumers raise their expectations and become more demanding with recommendations from experts. Regarding menu and food (Topic 3, Column (3)), we note that consumers tend to mention these aspects less frequently when a restaurant loses Michelin stars, possibly because they have lower expectations about food in such cases. Finally, for the two service-related topics (Topic 2 and Topic 4), an increase in Michelin stars is associated with an 8.8 percentage point increase in the proportion of Topic 2 and a 17.4 percentage point decrease in the proportion of Topic 4. In contrast, a decrease in Michelin stars is associated with a 9.2 percentage point decrease in the proportion of Topic 2 and a 17.5 percentage point increase in the proportion of Topic 4. This suggests that consumers become more demanding on service quality when restaurants gain Michelin stars, and less demanding when restaurants lose Michelin stars. The results from the SynthDiD estimation in Table 11 show a similar pattern. Note that the results are not driven by menu changes, because we focus on restaurants without menu changes in the guidebook windows.

Together, our results on the content of consumer reviews shed light on the underlying factors behind the changes in review sentiment following changes in Michelin stars, providing support on the expectation effect. Service and "value for money" are crucial to customer satisfaction. This finding is highly relevant to practitioners as they navigate the impacts of expert opinions. As our results show, receiving a favorable expert opinion can put more pressure on the business due to heightened customer expectations. Thus, practitioners need to be proactive in terms of anticipating this and preparing accordingly.

Table 10 Effects of Michelin Star Changes on Topics of Consumer Reviews by SCM-DiD

	(1) Topic 1	(2) Topic 2	(3) Topic 3	(4) Topic 4	(5) Topic 5
	Value for Money	Issues with Order	Menu and Food	Service and Staff	Overall Experience
After	0.006	–0.004	0.000	–0.003	0.001
	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)
After × Increase	0.131***	0.088***	–0.018	–0.174***	–0.027
	(0.021)	(0.015)	(0.011)	(0.018)	(0.025)
After × Decrease	–0.160***	–0.092***	–0.098***	0.175***	0.175***
	(0.016)	(0.014)	(0.012)	(0.023)	(0.024)
One Star	0.004	–0.001	–0.006	–0.002	0.005
	(0.019)	(0.014)	(0.011)	(0.019)	(0.025)
Two Star	–0.016	–0.008	–0.021	0.042	0.002
	(0.041)	(0.026)	(0.022)	(0.031)	(0.054)
Three Star	–0.133	–0.054	0.032	0.144***	0.010
	(0.125)	(0.086)	(0.025)	(0.046)	(0.216)
ln(number of reviews+1)	0.003	0.010**	0.009	–0.009	–0.013
	(0.009)	(0.005)	(0.008)	(0.011)	(0.012)
Cumulative average rating	–0.034	–0.067	–0.070	0.010	0.162
	(0.098)	(0.076)	(0.078)	(0.099)	(0.126)
Cumulative rating variance	0.014	–0.036	–0.132**	0.059	0.095
	(0.075)	(0.069)	(0.063)	(0.065)	(0.071)
ln(normalized search	–0.023	0.012	–0.029*	0.014	0.027
volume+1)	(0.027)	(0.022)	(0.016)	(0.042)	(0.041)
Pair–window FE	Yes	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes	Yes
Observations	892	892	892	892	892
Number of pairs	223	223	223	223	223
R^2	0.860	0.827	0.785	0.885	0.901

Note: Robust standard errors clustered at pair level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. After applying a Bonferroni correction for multiple comparisons on five topics, significance levels are adjusted as follows: *** p<0.002, ** p<0.01, * p<0.02, and results remain consistent.

Table 11 **Effects of Michelin Star Changes on Topics of Consumer Reviews by SynthDiD**

	(1) Topic 1	(2) Topic 2	(3) Topic 3	(4) Topic 4	(5) Topic 5
	Value for Money	Issues with Order	Menu and Food	Service and Staff	Overall Experience
Increase	0.071***	0.038***	–0.015	–0.111***	0.018
	(0.021)	(0.013)	(0.012)	(0.020)	(0.017)
Decrease	–0.063***	–0.054**	–0.044***	0.144***	0.105***
	(0.024)	(0.024)	(0.017)	(0.039)	(0.030)

Note: Overall ATT reported. Aggregated standard errors are in parentheses. *** p<0.01, ** p<0.05.

5. Alternative Explanations

Although two variants of the synthetic control method allow us to capture possible trends that might affect identification of the effect of the Michelin star change, as summarized in Table 4, there still exist potential supply- and demand-side factors that may lead to the observed effects. We address concerns related to supply-side factors in Section 5.1 and concerns related to demand-side factors in Section 5.2.

5.1. Supply-side Factors

There are three supply-side changes that may affect consumer reviews: menu changes, changes in serving size given the menu, and changes in restaurant decor and/or service. We discuss each in turn.

5.1.1. Menu Changes One alternative explanation to the finding is that restaurants may have changed their menu following the Michelin star change. Recall that our sample excludes restaurants that have changed their menus during the window around the Michelin Guide release time, thus it is unlikely that the effects are driven by menu changes.

5.1.2. Serving Size Changes Although we have controlled for menu offerings and focused on a short-time window, one concern is that restaurants can modify serving sizes or the quality of their dishes without changing the menu. As discussed in Section 2.1, the Michelin star selections are confirmed through repeated visits by different inspectors within a year, ensuring consistency. Should there be changes in serving size or food quality post a Michelin star status change, it would likely be noted by the inspectors during their consistency assessments and could result in an adjusted star rating the following year. Thus, restaurants that retain their new Michelin star level in the next guidebook year (e.g., sustaining a 1-star status after an increase from 0-star) are presumed to uphold consistent food quality and serving sizes. We replicate the analysis with this subset of "highly consistent" restaurants, and the results in Table E.1 of Online Appendix E are consistent with prior results.[20][21]

5.1.3. Non-food Changes The third concern related to the supply-side is that restaurants may have made major changes in their decor or service. In SCM-DiD, we focus on a window of 90 days before and 90 days after the Michelin Guide release, a timeframe likely too short for significant changes. We revisit this issue by conducting a robustness check with a shorter window period in Section 6.3. In SynthDiD, Figure 3 indicates a very significant positive treatment effect in an even shorter period (60 days) after the new Michelin Guide release. This makes it even less likely that changes in decor and service could be causing our results. We further check the trends in "service-related" review topics in the twelve-month period following the Michelin Guide updates in Online Appendix E.2. We observe that the probability of service-related topics being mentioned remains stable.

This being said, we acknowledge that with sufficient commitment from management, there is a possibility of relatively swift improvements in service quality. Because consumer reviews reflect both objective service quality and subjective perceptions influenced by expectations, we cannot completely rule out the potential impact of unobserved service quality adjustments.

5.2. Demand-side Factors

There are three demand-side changes that may affect the interpretation of our results: changes in restaurant demand, consumers showing sympathy for restaurants losing stars, and changes in the mix of consumers visiting a restaurant. We discuss each in turn.

5.2.1. Restaurant Demand Michelin star changes may induce changes in consumer interest and restaurant demand. To see how Michelin stars affect restaurant demand, we estimate Equations (1), (2) and (3) with the log-transformed normalized Google search intensity (collected from Google Trends) as the dependent variable (see Table E.2 in Online Appendix E). Our findings reveal that changes in a restaurant's Michelin star status do not significantly change its search volume, suggesting that our results are unlikely to be primarily driven by changes in restaurant demand.

This being said, Google search intensity only includes searches originated from Google, and it is possible that changes in Michelin stars lead to changes in searches on review and booking websites such as TripAdvisor and OpenTable. We next examine restaurant demand using daily OpenTable

[20]Note that our data period ends at guidebook 2020 and does not cover Michelin star levels in guidebook 2021, thus this analysis includes star changes before guidebook year 2020.

[21]Note that, for brevity, in Table E.1 and subsequent tables, we do not report the estimates associated with other control variables in SCM-DiD, which are qualitatively similar to those in Column (2) of Table 6.

reservation data collected by Farronato and Zervas (2022) on New York City restaurants. This dataset contains information on the daily availability of tables for two between 18:30 and 19:30 at each restaurant in the period of April 2013 to March 2017. Within this time period, we first check five New York City Michelin Guides (guidebook 2013 to guidebook 2017), and identify 117 awarded restaurants that received Michelin stars at least once. Among these awarded restaurants, there are 54 instances of Michelin star increase and 39 instances of star decrease during guidebook 2014 to guidebook 2017 (with guidebook 2013 serving as our baseline). Second, we match these awarded restaurants with the restaurants in Farronato and Zervas (2022)'s OpenTable reservation data, and identify 70 (out of 117) awarded restaurants with OpenTable records. Third, we denote each "restaurant-guidebook year" as a unit, and keep units for which we observe booking information immediately before and after the guidebook release. In the end, we retain a total number of 222 units, with 27 units associated with Michelin star increases, 13 units associated with Michelin star decreases, and 182 units where star status remained unchanged. We then estimate the following regression model analogous to Equation (8) from Farronato and Zervas (2022), using data in a short window around the guidebook release dates:

$$
\begin{aligned}
Soldout_{id} = \; & \beta_1 After_d + \beta_2 After_d \times Increase_{id} + \beta_3 After_d \times Decrease_{id} \\
& + \beta_4 OneStar_{id} + \beta_5 TwoStar_{id} + \beta_6 ThreeStar_{id} \\
& + \alpha_i + \gamma_d + \varepsilon_{id} \qquad (4)
\end{aligned}
$$

where i denotes restaurant, and d denotes day. The outcome variable $Soldout_{id}$ is an indicator variable which equals 1 if restaurant i is sold out between 18:30 and 19:30 on day d. $After_d$ is an indicator variable which takes the value of 1 if day d is in a window after the Michelin guidebook update. $Increase_{id}$ ($Decrease_{id}$) takes the value of 1 if restaurant i gained (lost) stars in the corresponding new guidebook. We control for restaurant fixed effect α_i and day fixed effects γ_d. Table E.3 in Online Appendix E shows the results.

The results indicate that compared to restaurants that maintained the same Michelin star level, restaurants gaining Michelin star(s) experience an increase in demand, whereas those losing Michelin star(s) do not experience a significant change in demand.[22] Although this sample of restaurants differs from our main sample, we posit that the relationship between Michelin stars and restaurant demand applies generally. This implies that the observed effects of Michelin star decreases on

[22]We acknowledge the possibility that the lack of statistical significance may be attributable to the small number of observations.

consumer reviews (Section 4) are unlikely to be driven by changes in restaurant demand. For restaurants gaining Michelin star(s), we conjecture that an increase in demand could potentially compromise the dining experience (possibly due to overcrowding etc.). Another explanation is that this increased demand could cause the restaurants to become overwhelmed and unable to maintain their usual high standards, indicating supply-side changes as discussed in Section 5.1.3. Nonetheless, the fact that the consumer review ratings in our main sample remained stable despite increased demand further suggests that the results are unlikely to be driven by changes in restaurant demand.

5.2.2. Consumer Sympathy One alternative explanation for our results is that consumers show their sympathy to underdogs (i.e., restaurants losing Michelin stars) and thus try to defend them in reviews. If this were the main mechanism, we would expect an increase in review volume for restaurants losing Michelin stars. To test if this is the case, we estimate Equations (1), (2) and (3) with the volume of consumer reviews as the dependent variable. Table E.4 in Online Appendix E shows the results. We do not find significant changes in review volumes for restaurants gaining or losing Michelin stars,[23] suggesting our results are unlikely an outcome of the consumer sympathy to underdogs.

5.2.3. Consumer Mix Potential changes in Michelin stars might change the mix of consumers who visit the restaurant. There are two possible mechanisms that could lead to the change in customer experience.

First, a change in the Michelin star ratings does not change the consumer mix visiting the restaurant. Thus any change in experience is driven by the change in expectations. Second, a change in the Michelin star ratings does change the consumer mix visiting the restaurant. Thus any change in experience is driven by a combination of selection and the change in expectations.

To ensure a clear identification, it is important to provide evidence that the second mechanism discussed above is unlikely to be at play. To do this, we use the reviewer-level data described in Section 2.6 for two sets of analyses. First, we look at the characteristics of all the reviewers who have reviewed the focal restaurant *before* the star change and the characteristics of those who reviewed *after* the star change. Second, we look at the reviewers of the focal restaurant, and examine their behavior in terms of the characteristics of *all* the restaurants i.e., not just the ones in our sample, that they visit before and after the star change.

[23] Consumer sympathy could potentially be more evident for British cuisine restaurants, as consumers might be inclined to support their national cuisine. To explore this, we replicate our analysis with restaurants serving British cuisine and the results are consistent. The results are available upon request.

Restaurant-level Analysis: Reviewer Characteristics As described in Section 2.6, we have collected the TripAdvisor profiles of 52,210 unique reviewers, who have provided 1,617,923 reviews spanning from 2010 to 2020, of which 52,224 reviews are for awarded restaurants. Based on this dataset, we construct a series of reviewer-level characteristics. We then compare characteristics of reviewers who reviewed the restaurant *before* the Michelin star change against those who reviewed *after* the change. If we observed no significant changes in these characteristics, it would provide us with greater confidence that the change in consumer mix is not the main driver behind our findings. We detail the steps below.

First, we introduce four variables to describe reviewer characteristics based on their profile: (i) local consumer, (ii) picky consumer, (iii) cumulative number of restaurant until each awarded restaurant, and (iv) cumulative mean review rating until each awarded restaurant. Table F.1 in Online Appendix F provides detailed definitions and examples of these four variables. To illustrate the construction of these variables, consider a reviewer who is registered in the United States and has provided eight reviews. Among the eight reviews sorted in chronological order, the fifth and eighth reviews are for two awarded restaurants, each receiving a “5-star” rating. The remaining six reviews have “4-star” ratings. The “Example” column in Table F.1 shows the values of the four variables for this reviewer. The variable “Local consumer” takes the value of 0 because she is not registered in the United Kingdom or Ireland. The variable “Picky consumer” takes the value of 0 because she has given 5-star review ratings. The cumulative number of restaurants until the two awarded restaurants are respectively 4 and 7. The cumulative mean review rating until the first awarded restaurant is 4 because the previous four reviews all have 4-star ratings. The cumulative mean review rating until the second awarded restaurant is 4.14 because among the seven previous reviews, six have 4-star ratings and one has a 5-star rating.

Next, for each of these 52,224 reviews, we extract the reviewer’s characteristics at the time of the review. For illustration, Table F.2 in Online Appendix F shows the reviewer characteristics associated with the two reviews by the example reviewer presented in Table F.1.

Lastly, similar to the data preparation step of SCM-DiD, we aggregate the reviews at the restaurant level for both the pre- and post-treatment windows, and then use the mean consumer review rating (or the mean topic probability) as the dependent variable. In this specific analysis, we aggregate the reviewer characteristics constructed at the review level – the four variables listed in Table F.2 – at the restaurant level. The resulting average of the review-level reviewer characteristics for each restaurant is the dependent variable. Essentially, when aggregating the variables “local consumer” and “picky consumer” at the restaurant level, we are measuring the percentage of “local (picky) consumers" associated with the restaurant. When aggregating the other two variables at the restaurant level, we are measuring the mean value of those variables (averaged across reviewers) for the restaurant.

In this analysis, we have 217 awarded restaurants and 1,040 "restaurant-guidebook year" units. Among these 1,040 units, 71 are associated with Michelin star increases, 53 are associated with Michelin star decreases, and the remaining units represent cases where the Michelin star status remained unchanged. Each of these units comprises two observations: one for the period before the Michelin star rating change and the other for the period after the Michelin star rating change. Note that we use a basic regression model without including control restaurants, as this additional reviewer-level dataset is derived from awarded restaurants in our dataset.

The results of the analysis are shown in Table 12. Columns (1) and (2) indicate that the percentage of local consumers and the percentage of picky consumers do not change significantly after Michelin star changes. Columns (3) and (4) suggest that the review intensity and average review rating are similar between reviewers who reviewed the restaurant prior to the change and those who reviewed it afterwards. Together, these results suggest that Michelin star changes do not have a significant impact on the types of consumers who visit awarded restaurants. Consequently, it implies that a change in the consumer mix is unlikely to be the primary driving factor behind our findings.

Reviewer-level Analysis: Restaurant Characteristics We have shown that consumers who reviewed a restaurant before the Michelin star change are not fundamentally different from consumers who reviewed the restaurant after the Michelin star change. Next, we examine whether changes in Michelin stars led consumers to visit different types of restaurants, drawing on all the reviews across every restaurant that has been reviewed by a reviewer in our dataset.

As discussed in Section 2.6, we located the TripAdvisor pages for 279,359 (out of 327,852) restaurants that have been reviewed by 45,274 (out of 52,210) reviewers who have reviewed an awarded restaurant within the 90-day guidebook windows. In total, these reviewers have provided 1,101,305 reviews. For each of these reviews, we collect both time-invariant and time-varying characteristics of the corresponding restaurant at the time of the review. Specifically, the time-invariant characteristics, the price level and cuisine type, were extracted from the restaurant's Tripadvisor page. Then, we calculate the cumulative review characteristics (number of reviews, mean star rating, standard deviation of star rating) for the restaurant up to the review date, leveraging the dataset of 79 million reviews we have collected. We illustrate this process with an example in Table F.3 in Online Appendix F.

After computing restaurant characteristics at the time of each review, we aggregate these restaurant characteristics at the reviewer level. Specifically, for each review within a reviewer's

Table 12 Effect of Michelin Star Changes on Reviewer Characteristics (90-day Window)

	Percentage (restaurant-level)		Mean (restaurant-level)	
	Local Consumer (reviewer-level)	Picky Consumer (reviewer-level)	# of restaurants until each awarded restaurant (reviewer-level)	Mean review rating until each awarded restaurant (reviewer-level)
	(1)	(2)	(3)	(4)
After	0.044***	–0.001	1.349**	–0.010
	(0.008)	(0.004)	(.551)	(0.016)
After × Increase	0.028	–0.001	–2.360	0.016
	(0.022)	(0.010)	(1.727)	(0.042)
After × Decrease	–0.025	–0.005	–0.839	0.060
	(0.033)	(0.014)	(2.362)	(0.054)
Other control variables (Michelin stars, the number of reviews, cumulative average review rating, cumulative rating variance, Google search volume)	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes
Guidebook year FE	Yes	Yes	Yes	Yes
Observations	2,080	2,080	2,080	2,080
"restaurant-guidebook year" units	1,040	1,040	1,040	1,040
Number of restaurants	217	217	217	217
R^2	0.040	0.016	0.278	0.021

Note: Robust standard errors clustered at restaurant level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

profile, we calculate the cumulative restaurant characteristics of her previously reviewed restaurants. Again, we illustrate this process with an example in Table F.4 in Online Appendix F.

We then construct four variables to describe whether the restaurant is different from the reviewer's previously reviewed restaurants, as presented in Table 13. First, we check if the price level and cuisine type of the restaurant differ from those restaurants the reviewer had previously reviewed. Next, we look at whether the review rating stands out from the reviewer's previous ratings. To define what counts as "standing out", we look at whether the rating falls within a normal range, calculated as the average plus or minus one standard deviation ($mean \pm SD$). Lastly, we calculate the difference in ratings ($\Delta Rating$) by comparing the rating of the current review against the reviewer's average rating up to that point.

To analyze whether a review in the reviewer's profile corresponds to a restaurant that differs from those she had reviewed previously, we use these four variables as dependent variables and estimate difference-in-differences models at the review level. The results are presented in Table 14. We control for both restaurant-level cumulative characteristics (such as average review rating, total number of ratings, and rating variance) and reviewer-level cumulative characteristics within their profile (such as average ratings, number of ratings, number of unique price levels, and number of unique cuisine

Table 13 **Reviewer-level Restaurant Characteristics Measurement and Definition**

Variable	Definition
Whether new price level	Equals 1 if the restaurant has a different price level from those previously reviewed. Otherwise, equals 0.
Whether new cuisine type	Equals 1 if the restaurant has a different cuisine type from those previously reviewed. Otherwise, equals 0.
Whether rating out of range of $mean \pm SD$	Equals 1 if the review rating for the focal restaurant is out of the range of previous ratings ($mean \pm SD$). Otherwise, equals 0.
Rating difference, $\Delta Rating$	Difference between the focal review rating and the cumulative mean review rating

types). We also add fixed effects on price level, cuisine type, reviewer, month, and guidebook year. Note that our analysis includes reviews starting from the third one in the reviewer's profile, because the initial two restaurants serve as a basis for computing rating variances and provide baseline price levels and cuisine types. Table 14 shows that, across all four columns, there are no significant changes in restaurant characteristics at the reviewer-level after Michelin star changes. This suggests that consumers maintain their usual dining preferences, and thus, changes in Michelin stars do not appear to significantly influence consumers' decisions to visit these restaurants.

Overall, based on observables in a large amount of reviewer and review data, we find that the pool of reviewers at a focal restaurant does not change, and that reviewers of the focal restaurant do not exhibit any change in their restaurant choices/preferences, before and after Michelin star changes. This provides strong supportive evidence that the second mechanism i.e., a change in consumer mix after a Michelin star change, is not driving our results.[24]

6. Robustness Checks

We conduct a battery of robustness checks, including a difference-in-differences analysis with control restaurants manually selected based on location, price, and cuisine type (Section 6.1), an alternative dependent variable to measure review sentiment (Section 6.2), an alternative window in SCM-DiD (Section 6.3), a falsification test with placebo guidebook publication dates in SynthDiD (Section 6.4), and a replication study with New York City data (Section 6.5).

6.1. Rule-based Control Restaurants

The SCM-DiD model in Section 3.2 employs the SCM to create a time-varying synthetic control restaurant that best matches the focal awarded restaurant, which is a data-driven approach. We

[24] An obvious caveat to this analysis is that we do not have data on restaurant visitors who do not write reviews at all (or write only on less prominent sites than TripAdvisor). Hopefully, the large sample sizes in both analyses conducted in this section mitigate this concern.

Table 14 **Effect of Michelin Star Changes on the Characteristics of Reviewed Restaurants**

	Whether new price level	Whether new cuisine type	Whether rating out of range $mean \pm SD$	$\Delta Rating$
	(1)	(2)	(3)	(4)
After	0.002	0.006***	–2.140e-05	0.001
	(0.001)	(0.002)	(3.213e-04)	(0.001)
After × Increase	–0.019	–0.005	–2.247e-04	–0.001
	(0.013)	(0.009)	(4.204e-04)	(0.004)
After × Decrease	0.022	–0.013	–9.870e-05	–0.004
	(0.018)	(0.009)	(4.450e-04)	(0.006)
One Star	0.070***	0.004**	2.639e-04	0.002***
	(0.003)	(0.002)	(2.579e-04)	(0.001)
Two Star	0.052***	0.018***	9.149e-04*	0.003**
	(0.005)	(0.004)	(5.041e-04)	(0.002)
Three Star	0.015	0.050***	1.366e-03	0.002
	(0.009)	(0.009)	(1.009e-03)	(0.003)
Cumulative average rating	0.002***	0.001	–6.427e-03***	1.011***
(Restaurant-level)	(0.001)	(0.001)	(4.813e-04)	(0.001)
Cumulative # of ratings	–0.002**	0.004***	4.096e-04*	–0.002***
(Restaurant-level)	(0.001)	(0.001)	(2.388e-04)	(0.001)
Cumulative rating variance	–0.002*	0.003**	–6.253e-03***	–0.001
(Restaurant-level)	(0.001)	(0.001)	(5.928e-04)	(0.001)
Cumulative average rating	–0.012***	–0.002	–4.187e-04	–0.198***
(Reviewer-level)	(0.002)	(0.002)	(8.555e-04)	(0.003)
Cumulative # of ratings	–0.109***	–0.112***	–3.823e-03***	0.020***
(Reviewer-level)	(0.005)	(0.006)	(4.095e-04)	(0.002)
Cumulative # of price levels	0.057***	–0.005**	–4.849e-04**	–0.003***
(Reviewer-level)	(0.002)	(0.002)	(1.952e-04)	(0.001)
Cumulative # of cuisine types	0.001***	0.001***	3.900e-05***	–0.000***
(Reviewer-level)	(0.000)	(0.000)	(1.150e-05)	(0.000)
Restaurant price level FE	Yes	Yes	Yes	Yes
Restaurant cuisine type FE	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Guidebook year FE	Yes	Yes	Yes	Yes
Observations	883,589	883,589	883,589	883,589
Number of reviewers	17,775	17,775	17,775	17,775
R^2	0.190	0.246	0.085	0.968

Note: Robust standard errors clustered at reviewer level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

check the robustness with a rule-based control restaurant selection, which explicitly selects control restaurants that closely match the awarded restaurants in terms of location, price level and cuisine type. Specifically, for each of the 262 awarded restaurants, we select from the pool of 1,147 control restaurants a control restaurant that satisfies the following criteria: (1) the control restaurant needs to be geographically close to the focal awarded restaurant: in urban areas within 0.5 miles and in rural areas within 10 minutes driving distance;[25] (2) the control restaurant has the same price level on TripAdvisor as the awarded restaurant; and (3) the control restaurant has the same cuisine type on TripAdvisor as the awarded restaurant. If more than one restaurant satisfies the above criteria, we give preference to the one that appears on the "best nearby restaurants" page recommended by TripAdvisor. Note that we allow each control restaurant to be matched with at most one awarded restaurant, that is we use matching without replacement, to ensure that the results are not driven by a small group of control restaurants which are matched with many awarded restaurants. In the end, 227 (out of 1,147) control restaurants are selected, leading to 227 treated-control pairs. The remaining 35 awarded restaurants without identified control restaurants are either located in rural areas without nearby restaurants, or located in urban areas but do not have nearby restaurants with the same price level and cuisine type. Out of the 227 restaurant pairs identified, 156 are located in urban areas and 71 are located in rural areas. On average, the distance between the focal awarded restaurant and the selected paired control is 0.12 miles (s.d. = 0.29) in urban areas and 7.01 miles (s.d. = 8.48) in rural areas.

We then estimate the difference-in-differences model (Equation (1)) with the treated-control restaurant pairs where both restaurants have received reviews in the 90-day pre- and post-treatment windows. This results in 143 (out of 227) restaurant pairs. We report the results on review sentiment (Column 1) and review content (Columns 2-6) in Table 15. The results are qualitatively similar to those in Tables 6 and 10.

6.2. Alternative Sentiment Measure

One concern is that the proportion of extreme reviews (i.e., 5-star-rating and 1-star-rating reviews) has changed, but the mean review rating may not change. Following Shin et al. (2023), we replicate the review sentiment analysis with the percentage of 5-star-rating reviews at the restaurant level, instead of the mean review rating, as the outcome variable. Column (1) in Table 16 and Column (1) in Table 17 respectively replicate Column (2) in Table 6 (SCM-DiD) and Table 7 (SynthDiD). Results are consistent with our prior findings: decreases in Michelin stars improve consumer review ratings.

6.3. Alternative Window

As mentioned earlier, our main analysis with the restriction of the 90-day time window around the Michelin Guide release makes it hard for restaurants to have the time and/or resources to pull off major changes in decor and/or service levels. We further shorten the period to 60-day time window

[25]The distance is calculated by two restaurants' geocoded longitudes and latitudes. The travel time is estimated with Google Maps.

Table 15 **Robustness Checks: DiD with Control Restaurants Selected via Rule-Based Criteria**

	(1) Overall Rating	(2) Topic 1	(3) Topic 2	(4) Topic 3	(5) Topic 4	(6) Topic 5
		Value for Money	Issues with Order	Menu and Food	Service and Staff	Overall Experience
After	–0.097	0.014	0.011	–0.004	–0.025	0.003
	(0.064)	(0.012)	(0.012)	(0.010)	(0.019)	(0.010)
After × Increase	0.100	0.133***	0.066***	–0.032	–0.183***	0.016
	(0.112)	(0.032)	(0.024)	(0.022)	(0.039)	(0.036)
After × Decrease	0.261**	–0.171***	–0.086***	–0.079***	0.184***	0.153***
	(0.115)	(0.026)	(0.026)	(0.023)	(0.042)	(0.033)
Other control variables (Michelin stars, the number of reviews, cumulative average review rating, cumulative rating variance, Google search volume)	Yes	Yes	Yes	Yes	Yes	Yes
Pair–window FE	Yes	Yes	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	572	572	572	572	572	572
Number of pairs	143	143	143	143	143	143
R^2	0.708	0.811	0.699	0.737	0.801	0.900

Note: Robust standard errors clustered at pair level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

around the Michelin Guide release. Column (2) in Table 16 replicates Table 6 with a 60-day window in SCM-DiD,[26] and the results are robust.

6.4. Falsification Test

The SynthDiD model relaxes the strong parallel-trends assumption for all units and all time periods. However, it assumes that there exist unit and time weights such that the averaged treated unit and the weighted average of the control units satisfy a parallel trends assumption for the averaged post-treatment period and the weighted average of the pre-treatment periods (Arkhangelsky et al. 2021). In other words, the selection of weights on control units and pre-treatment periods depends on the actual treatment time. One possible concern regarding this design is that we may be measuring a general trend among the treated restaurants instead of a causal effect of the Michelin star changes. To alleviate this concern, we conduct a falsification test by generating a "placebo" guidebook publication date that is 90 days before the actual publication date. We then replicate Table 7 with the placebo guidebook date.[27] Results are presented in Column (2) of Table 17. The insignificant ATTs indicate that our results are unlikely driven by a general time trend.

[26] The re-construction of SCM results in 208 (out of 252) synthetic control restaurants corresponding to 208 treated units. 44 units were dropped because they do not have enough reviews on at least one side of the 60-day pre- or post-treatment window.

[27] The re-construction of 18-month data with nine consecutive two-month blocks around "placebo" guidebook publication date results in 136 (out of 252) treated units, including 83 units for gaining Michelin stars and 53 units for losing Michelin stars, and 4,307 control units.

Table 16 Robustness Checks: SCM-DiD with Alternative Sentiment Measure and Alternative Window

	Alternative dependent variable (Section 6.2)	Alternative window (Section 6.3)
	(1)	(2)
After	–0.019**	–0.079***
	(0.008)	(0.029)
After × Increase	0.010	0.019
	(0.034)	(0.076)
After × Decrease	0.089***	0.349***
	(0.033)	(0.075)
Other control variables (Michelin stars, the number of reviews, cumulative average review rating, cumulative rating variance, Google search volume)	Yes	Yes
Pair–window FE	Yes	Yes
Restaurant FE	Yes	Yes
Observations	892	832
Number of pairs	223	208
R^2	0.821	0.821

Note: Robust standard errors clustered at pair level are in parentheses. *** p<0.01,** p<0.05.

Table 17 Robustness Checks: SynthDiD with Alternative Sentiment Measure and Falsification Test

	Alternative dependent variable (Section 6.2)	Falsification test (Section 6.4)
	(1)	(2)
Increase	–0.012	0.041
	(0.034)	(0.075)
Decrease	0.159***	0.052
	(0.062)	(0.133)
Total number of treated units	148	136
Total number of control units	4,334	4,307

Note: Overall ATT reported. Aggregated standard errors are in parentheses. *** p<0.01.

6.5. Replication with NYC Restaurants

To investigate whether the effects generalize to other countries, we conduct a replication study in the context of New York City (NYC) using data described in Section 5.2.1. The detailed data construction process is described in Online Appendix G. Using the mean consumer review rating as the dependent variable, we replicate the analysis of review sentiment in Equation (1). There are two key differences compared to our main analysis. First, instead of using Google search intensity as a proxy of restaurant demand Z_{it}, we measure the average demand for each restaurant within the 90-day window by calculating the percentage of fully-booked days during that period. Second, we replace the pair–window fixed effect $\alpha_{p(i)w(t)}$ with the window fixed effect $\alpha_{w(t)}$ because we do not match a restaurant with its control due to the limited sample.

The estimation results are presented in Table 18. Column (1) controls only for Michelin star levels and fixed effects, and Column (2) adds the full set of controls. Both columns reveal that the estimated coefficient for $After \times Increase$ is not statistically significant, indicating that gaining Michelin stars does not lead to changes in consumer review ratings. However, the estimated coefficient for $After \times Decrease$ is significantly positive, suggesting an increase in consumer review ratings for restaurants that lost Michelin stars. These results align with our main analysis. It is worth noting that the NYC data set has a smaller sample size (20 Michelin star increases and 8 Michelin star decreases), which impacts the level of statistical significance. Nevertheless, the overall trend and direction of the effects remain consistent.

Table 18 NYC Replication Results: Effects of Michelin Star Changes on Sentiment of Consumer Reviews

	DV: mean review rating	
	(1)	(2)
After	0.000	–0.022
	(0.040)	(0.047)
After × Increase	–0.146	–0.131
	(0.117)	(0.130)
After × Decrease	0.236**	0.210*
	(0.116)	(0.108)
One Star	–0.043	–0.062
	(0.069)	(0.068)
Two Star	0.086	–0.002
	(0.096)	(0.124)
ln(number of reviews+1)		0.026
		(0.038)
Cumulative average rating		0.629***
		(0.201)
Cumulative rating variance		0.105
		(0.214)
Percentage of fully booked days		0.124
		(0.130)
Window FE	Yes	Yes
Restaurant FE	Yes	Yes
Observations	252	252
Number of units	126	126
R^2	0.612	0.629

Note: Robust standard errors clustered at restaurant level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

7. Discussion and Conclusion

Expert opinion exerts tremendous influence on the consumer journey, but its effect on consumer experience is ambiguous as it can give rise to both expectation and reputation effects. Favorable expert opinions can enhance the reputation of a business, and potentially improve consumer experience by guiding consumer opinions, but they may also harm consumer experience by raising consumer expectations. Likewise, while unfavorable expert opinions may harm the reputation of a business, they also have the potential to improve consumer experience by lowering consumer expectations. We investigate the tension between the expectation effect and the reputation effect as a result of expert opinion through the lens of consumer reviews in the context of the restaurant industry and Michelin stars.

We apply two synthetic-control-based methods to identify the effect of Michelin star changes on the sentiment and content of consumer reviews. We find consistently that decreases in Michelin stars improve consumer review ratings. Analyses on review content further show that service and "value for money" appear to be the key drivers of customer satisfaction, and when a restaurant is removed from the Michelin Guide or loses stars, consumers tend to become less demanding on service, and focus less on value for money. As noted earlier, prior work has never documented the fact that a lowered expert rating can lead to a better consumer experience.

There are three potential key explanations for the finding that decreases in Michelin star(s) improve consumer review ratings. First, the positive expectation effect of lowered expert opinions outweighs the negative reputation effect. Second, changes in Michelin stars change the mix of customers visiting the restaurant. Third, restaurants respond to Michelin star changes by making supply-side adjustments. Our findings support the first explanation, and we present evidence suggesting that the second mechanism is unlikely to be the main driver. However, the third mechanism – changes in supply-side factors such as service level – cannot be ruled out with complete certainty.

Our results reveal potential explanations for the "Michelin curse," i.e., the downside(s) of gaining Michelin stars. We offer substantive managerial insights for restaurant managers, the Michelin Guide, and other firms providing experience goods as a whole.

For restaurant managers, understanding the impact of Michelin stars allows them to better navigate the impact of Michelin ratings on their business. Our evidence suggests that losing a Michelin star can potentially improve consumer review ratings. This insight is transformative because it challenges the conventional perception that losing a star is detrimental. Managers can use this information to strategically manage customer expectations and focus on delivering consistent service quality rather than solely chasing the coveted Michelin star(s). Additionally, our findings inform practical marketing strategies in response to Michelin stars. First, our analyses of the topics of consumer reviews indicate that consumers pay more attention to service than to food or menu. Restaurants can leverage this by streamlining their menus, offering fewer unique dishes but ensuring a variety that balances with service efficiency. This approach can enhance operational efficiency and customer satisfaction. Second, since consumers are less concerned about value for money when a restaurant loses Michelin stars, these restaurants have an opportunity to introduce premium dishes with expensive ingredients (e.g., caviar, truffles, saffron, and wagyu beef). This strategy can increase revenue without negatively impacting consumer evaluations. Conversely, restaurants gaining Michelin stars should be cautious with price increases as value for money remains critical for their customers. Third, our analysis identifies "wine" and "sommelier" as significant aspects of the dining experience (Topic 5 in Table 8 and Table D.1). Therefore, restaurants can benefit from putting more effort in the wine list, and hiring professional sommeliers to recommend wines to complement customers' tastes and to pair with their menu choices. Not only does this cater to consumer experience, but it also taps into a substantial profit margin area, as alcohol sales contribute more than 80% of the profit for most fine-dining restaurants.[28] Fourth, the increased focus on service after Michelin star changes highlights the importance of investing in staff training. Restaurants should allocate resources to enhance service quality, ensuring that staff are well-trained to meet high standards. This investment can lead to sustained positive reviews and customer satisfaction, even if the restaurant loses a Michelin star.

For the Michelin Guide, given the controversy on "consistency" as a criterion and the lack of transparency in award decisions, it can balance consistency and innovation (Ospina 2018) in their evaluation criteria. In addition, the Michelin Guide was established in the early 20th century and began to award stars for fine dining establishments in 1926. In the age of social media, consumer reviews and feedback can potentially be a valuable consideration in the assessment process.

For other businesses providing experience goods, our research offers valuable managerial insights. Companies tend to invest money and time with the purpose of being recommended by experts or showing better results in expert based rating systems. This often leads to businesses spending more on features and/or attributes that are not necessarily relevant for the customer experience.[29] However, our findings reveal that winning such endorsements and/or awards does not always lead to improved consumer evaluations, and that losing an award may turn out to be a blessing in

[28] `https://www.thebalancesmb.com/restaurant-fine-dining-2888686`

[29] See, for example, `https://www.forbes.com/sites/forbesbusinesscouncil/2022/03/30/how-hotel-art-affects-ratings/`

disguise. Essentially, businesses should be open to the understanding that favorable expert opinions can be a double-edged sword. As a result, they need to devote resources in a manner that balances "pleasing" experts (by playing to the criteria they use) and managing customer expectations and delivering fulfilling experiences.

There are several limitations to the present study that represent opportunities for future research. First, our study focuses on online reviews to assess the impact of Michelin stars. This approach inherently captures only the experiences and opinions of customers who choose to post reviews. Existing literature (e.g., Fradkin and Holtz 2023, Tadelis 2016) suggests that individuals who write online reviews may not be representative of the broader customer base. Incorporating other social media and offline word-of-mouth into the research framework would broaden our understanding of how consumer opinions are influenced by expert opinions. Second, due to the lack of access to sales and revenue data for restaurants in the UK and Ireland, we are unable to analyze the economic impact of (the change in) Michelin stars. Third, our analyses show that service is unlikely to be main driver of the observed effects. However, we cannot completely rule out the potential impact of unobserved service quality adjustments. We hope that future research will able to address this. Fourth, we relied on the Wayback Machine to collect restaurant menus, which may result in potential missing menu updates since the Wayback Machine does not archive all websites on a daily basis. Finally, this research mainly focuses on the Michelin Guide for Great Britain & Ireland with a replication study on New York City's Michelin Guide, and future research can extend the scope of the analyses to other countries and/or industries.

References

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association* 105(490):493–505.

Abadie A, Diamond A, Hainmueller J (2011) Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software* 42(13), URL `https://www.jstatsoft.org/issue/view/v042`.

Abadie A, Diamond A, Hainmueller J (2015) Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2):495–510.

Adalja A, Liaukonytė J, Wang E, Zhu X (2023) Gmo and non-GMO labeling effects: Evidence from a quasi-natural experiment. *Marketing Science* 42(2):233–250.

Almenberg J, Dreber A (2011) When does the price affect the taste? results from a wine experiment. *Journal of Wine Economics* 6(1):111–121.

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic difference-in-differences. *American Economic Review* 111(12):4088–4118.

Asch SE (1955) Opinions and social pressure. *Scientific American* 193(5):31–35.

Ashenfelter O, Jones GV (2013) The demand for expert opinion: Bordeaux wine. *Journal of Wine Economics* 8(3):285–293.

Berman R, Israeli A (2022) The value of descriptive analytics: Evidence from online retailers. *Marketing Science* (forthcoming).

Bondi T, Rossi M, Stevens R (2023) The good, the bad and the picky: Reference dependence and the reversal of product ratings. *Available at SSRN 4453681* .

Büschken J, Allenby GM (2016) Sentence-based text analysis for customer reviews. *Marketing Science* 35(6):953–975.

Chen Y, Liu Y, Zhang J (2012) When do third-party product reviews affect firm value and what can firms do? the case of media critics and professional movie reviews. *Journal of Marketing* 76(2):116–134.

Chen Y, Xie J (2005) Third-party product review and firm marketing strategy. *Marketing Science* 24(2):218–240.

Copas JB, Li H (1997) Inference for non-random samples. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 59(1):55–95.

Diehl K, Poynor C (2010) Great expectations?! assortment size, expectations, and satisfaction. *Journal of Marketing Research* 47(2):312–322.

Farronato C, Zervas G (2022) Consumer reviews and regulation: Evidence from NYC restaurants. Technical report, National Bureau of Economic Research.

Fogarty JJ (2012) Expert opinion and cuisine reputation in the market for restaurant meals. *Applied Economics* 44(31):4115–4123.

Fradkin A, Holtz D (2023) Do incentives to review help the market? evidence from a field experiment on airbnb. *Marketing Science* 42(5):853–865.

Friberg R, Grönqvist E (2012) Do expert reviews affect the demand for wine? *American Economic Journal: Applied Economics* 4(1):193–211.

Gergaud O, Storchmann K, Verardi V (2015) Expert opinion and product quality: Evidence from New York City restaurants. *Economic Inquiry* 53(2):812–835.

Gerstner E (1985) Do higher prices signal higher quality? *Journal of Marketing Research* 22(2):209–215.

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

Hackmann MB, Kolstad JT, Kowalski AE (2015) Adverse selection and an individual mandate: When theory meets practice. *American Economic Review* 105(3):1030–66.

Hayward T (2021) What's wrong with the Michelin guide? *Finantial Times*. URL `https://www.ft.com/content/e622ec53-ea9f-487a-a434-747f13f5ffa0`.

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.

Hollenbeck B (2018) Online reputation mechanisms and the decreasing value of chain affiliation. *Journal of Marketing Research* 55(5):636–654.

Johnson C, Surlemont B, Nicod P, Revaz F (2005) Behind the stars: a concise typology of michelin restaurants in Europe. *Cornell Hotel and Restaurant Administration Quarterly* 46(2):170–187.

Kovács B, Sharkey AJ (2014) The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly* 59(1):1–33.

Krosnick JA (1999) Survey research. *Annual review of psychology* 50(1):537–567.

Li KT (2020) Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association* 115(532):2068–2083.

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Information Systems Research* 19(4):456–474.

Mittal V, Han K, Lee JY, Sridhar S (2021) Improving business-to-business customer satisfaction programs: Assessment of asymmetry, heterogeneity, and financial impact. *Journal of Marketing Research* 58(4):615–643.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *Journal of Marketing Research* 48(3):444–456.

Morgan NA, Rego LL (2006) The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing science* 25(5):426–439.

Ospina D (2018) How the best restaurants in the world balance innovation and consistency. *Harvard Business Review.* URL `https://hbr.org/2018/01/how-the-best-restaurants-in-the-world-balance-innovation-and-consistency`.

Puranam D, Narayan V, Kadiyali V (2017) The effect of calorie posting regulation on consumer opinion: a flexible latent dirichlet allocation model with informative priors. *Marketing Science* 36(5):726–746.

Rao AR, Monroe KB (1989) The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research* 26(3):351–357.

Rossi M (2021) Quality disclosures and disappointment: Evidence from the Academy Awards. *Proceedings of the 22nd ACM Conference on Economics and Computation*, 790–791.

Sands D (2020) The role of third parties in value creation and capture: Why Michelin stars may not be a good thing. *Academy of Management Proceedings*, volume 2020, 21105 (Academy of Management Briarcliff Manor, NY 10510).

Shin M, Shin J, Ghili S, Kim J (2023) The impact of the gig economy on product quality through the labor market: Evidence from ridesharing and restaurant quality. *Management Science* 69(5):2620–2638.

Statista (2017) How much do movie critic reviews influence your decision to see a movie? `https://www.statista.com/statistics/682930/movie-critic-reviews-influence/`, [Online; accessed 08-September-2020].

Tadelis S (2016) Reputation and feedback systems in online platform markets. *Annual Review of Economics* 8:321–340.

Tirunillai S, Tellis GJ (2014) Mining marketing meaning from online chatter: Strategic brand analysis of big data using Latent Dirichlet Allocation. *Journal of Marketing Research* 51(4):463–479.

Winer RS (1986) A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research* 13(2):250–256.

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25(1):57–76.

Online Appendix for

Can Lower(ed) Expert Opinions Lead to Better Consumer Ratings?: The Case of Michelin Stars

Xingyi Li (University College London), Yiting Deng (University College London),
Puneet Manchanda (University of Michigan), Bert De Reyck (Singapore Management University)

A. Review Characteristics for Treated and Synthetic Control Restaurants

Table A.1 shows the review characteristics for treated and synthetic control restaurants during the pre-treatment period. Clearly, the treated and synthetic controls are comparable across all measured dimensions.

Table A.1 Review Characteristic for Treated and Synthetic Control Restaurants in Pre-treatment Period

	Treated	Synthetic Control	T-test statistics	p-value
	(1)	(2)	(3)	(4)
Mean of yearly average star rating	4.469	4.434	1.239	0.216
Mean of yearly variance of review ratings	0.805	0.813	–0.215	0.830
Mean of yearly cumulative number of reviews	4.895	4.991	–0.717	0.474
Mean of average star rating in 90-day pre-window	4.468	4.468	–0.016	0.987
Mean of price level	3.897	3.928	–1.209	0.227

B. Bootstrapped Standard Errors for SCM-DiD

We use the approach outlined by (Arkhangelsky et al. 2021) and (Adalja et al. 2023) to calculate bootstrap standard errors for the SCM-DiD analysis reported in Table 6. For each treated unit, we independently resample the donor pool consisting of control units 1,000 times. For each bootstrap sample b, the estimator $\hat{\delta}^b$ is obtained following the procedure described in Section 3.2. The bootstrap variance is calculated as $\hat{\mathcal{V}}_{\delta}^b = \frac{1}{1000}\sum_{b=1}^{1000}(\hat{\delta}^b - \frac{1}{1000}\sum_{b=1}^{1000}\hat{\delta}^b)^2$. The results shown in Table B.1 are consistent with those in Table 6.

Table B.1 Bootstrap Treatment Effect and Standard Errors

	Increase	Decrease
Estimation	0.086	0.363***
	(0.034)	(0.038)
p-value	0.227	0.000

Note: The table presents bootstrap mean, standard errors (in parentheses), and average p-value of the treatment effect among 1000 iterations. p<0.01, ** p<0.05, * p<0.1.

C. Additional Details on SynthDiD

Table C.1 shows the cohort-level SynthDiD ATT estimates for review sentiment.

Table C.1 Cohort-level Estimates by SynthDiD

Guidebook window	(1) Increase	(2) Decrease
2012	0.185	1.560***
	(0.122)	(0.332)
2013	–0.032	0.099
	(0.073)	(0.086)
2014	–0.048	0.236
	(0.084)	(0.186)
2015	–0.019	0.375
	(0.056)	(0.305)
2016	–0.047	0.169**
	(0.058)	(0.157)
2017	–0.012	0.192***
	(0.045)	(0.056)
2018	0.134*	0.557***
	(0.078)	(0.191)
2019	–0.021	0.336***
	(0.063)	(0.108)
2020	–0.094*	0.280***
	(0.049)	(0.104)

Note: Standard errors for each guidebook window calculated with bootstrap or placebo are in parentheses. We do not observe available treated units in guidebook window for the 2011 guidebook. *** p<0.01, ** p<0.05, * p<0.1.

D. The LDA Model

D.1. Technical Details on the LDA Model

To facilitate textual data analysis, we preprocess all reviews by splitting the text into its component words, eliminating punctuations, lemmatizing words into dictionary form, transforming plurals to singular, removing stop words ("a", "an", "the," etc.) and all words that occur in less than 1% of the reviews in the data set (Griffiths and Steyvers 2004, Büschken and Allenby 2016, Puranam et al. 2017, Berger et al. 2020). After preprocessing, there are 785 unique words in the vocabulary, and the average length of the textual reviews is 37 words (s.d. = 30.83). Note that this step is at the review level, so we preprocess all available 889,660 consumer reviews for the 262 awarded restaurants and 1,257 control restaurants.

The LDA model assumes a certain data-generating process for the review text: When consumers write reviews, they can choose words to express their opinions about multiple dimensions (i.e., topics) of the dining experience, such as food and service. Thus, each review $d \in \{1, ..., D\}$ includes a mixture of K topics, and each topic $k \in \{1, ..., K\}$ is characterized by a probability distribution over a vocabulary of V words $v \in \{1, ..., V\}$. The standard LDA model assumes the same Dirichlet prior for all of the per-review topic distributions (α) and the same prior for all of the per-topic word distribution (β). In other words, it ignores review and word characteristics that might affect the distribution of topics. To account for potential heterogeneity, we allow the hyperparameter α_d to be a function of the length and rating of the textual review. If two reviews have few characteristics in common, their Dirichlet prior α_d should be different, resulting in the different topic distribution $\boldsymbol{\theta}_d$. For instance, a longer review might discuss more topics and therefore have more evenly spread topic distributions. Similarly, we allow the hyperparameter β_k to be a function of latent semantic word characteristics. If two words have different semantic characteristics (e.g., they are antonyms rather than synonyms), we expect that these two words will have different probabilities of appearing in the same topic k. In other words, if a topic "prefers" a certain word v, we expect that it will also prefer other words with similar semantic characteristics to v.

At the review level, we segment the review characteristics into quintiles based on the length of the review (measured by the number of words before preprocessing) and its rating. Thus, the review characteristics are represented by two categorical 5-level variables, which are further converted into L_{doc} dummy variables, each corresponding to one level of the $5 \times 5 = 25$ combinations of the review length quintile and the star rating, so $L_{doc} = 25$ in our model. In addition to the observable review characteristics, we add an intercept term to capture the characteristics that are unrelated to the L_{doc} binary variables. Therefore, the characteristics of review d are defined by an (L_{doc}+1)-dimensional binary vector $\boldsymbol{f_d} = \{1, f_{d,1}, f_{d,2}, ..., f_{d,l}, ..., f_{d,L_{doc}}\}^T$, where $f_{d,l}$ equals 1 if review d has the characteristic indicated by label l and 0 otherwise (Zhao et al. 2017).

At the word level, frequency counts of word occurrences in a corpus are the primary data to all unsupervised methods for learning word representations. However, standard LDA approaches do not consider word characteristics, presenting challenges with short texts, where word co-occurrences are too sparse to provide meaningful context. For example, it is possible that topics associated with synonyms have a prior tendency to be similar, so that when one synonym is rare but the other is common within the corpus, the topics estimates for the rare one can be improved. A global log-bilinear regression model GloVe provides an effective measure for the linguistic or semantic similarity of word representations (Pennington et al. 2014). Under GloVe representations, each word is represented by a high dimensional vector that is pre-trained on some large external corpus, e.g., Wikipedia, Twitter, and Google News. Accordingly, we choose a set of 50-dimensional word embeddings pre-trained on Twitter[30] as our original word characteristics. Similar to the review characteristics, we convert the continuous-valued word characteristics into binary values, following Guo et al. (2014). Let $\boldsymbol{M''}$ be a $V \times 50$ matrix, where V is our vocabulary size. Each row $v \in \{1, ..., V\}$ of $\boldsymbol{M''}$ represents a 50-dimensional embedding of vocabulary word v. For the j^{th} dimension ($j \in \{1, ..., 50\}$) of word embeddings, we divide the corresponding column vector $\boldsymbol{M''_{\cdot j}}$ into two parts, with one part including all positive elements ($M''_{\cdot j+}$) and the other including the negative elements ($\boldsymbol{M''_{\cdot j-}}$). Next, we transform $\boldsymbol{M''}$ to a same-dimension matrix $\boldsymbol{M'}$ as follows:

$$M'_{v,j} = \begin{cases} 1 & if\, M''_{v,j} > mean(\boldsymbol{M''_{\cdot j+}}), \\ -1 & if\, M''_{v,j} < mean(\boldsymbol{M''_{\cdot j-}}) \\ 0 & otherwise \end{cases},$$

where mean$(\cdot)$ denotes the mathematical mean. The insight behind this transformation is that we only consider the word embeddings with strong positive or negative values on each dimension j and omit the values that are close to 0. Finally, we use two dummy variables to encode each column j in $\boldsymbol{M'}$ and transform $\boldsymbol{M'_{v,j}} \in \{-1, 0, 1\}$ to binarized word characteristics. Thus, the original continuous-valued word labels are converted to L_{word} unique binary labels ($L_{word} = 100$ in this case). The labels of each word $v \in \{1, ..., V\}$ are defined by an (L_{word}+1)-dimensional binary vector $\boldsymbol{g_v} = \{1, g_{v,1}, g_{v,2}, ..., g_{v,l'}, ..., g_{v,L_{word}}\}^T$, where $g_{v,l'}$ equals 1 if word v has the characteristic indicated by label l' and equals 0 otherwise.

The LDA model describes the joint probability distribution over both the observable data (words in the review) and the hidden variables (topics of the review). In our LDA model, we allow the Dirichlet prior $\boldsymbol{\alpha}_d$ to be a function of review characteristics $\boldsymbol{f_d}$, and the Dirichlet prior $\boldsymbol{\beta}_k$ to be a function of word characteristics $\boldsymbol{g_v}$, specified as follows

$$\alpha_{d,k} = exp(\sum_{l=1}^{L_{doc}} f_{d,l}\, \lambda_{l,k}) = exp(\boldsymbol{f_d}^T \boldsymbol{\lambda_k}), \quad \lambda_{l,k} \sim F(\alpha_{d,k}) \tag{5}$$

$$\beta_{k,v} = exp(\sum_{l'=1}^{Lword} g_{v,l'}\, \delta_{l',k}) = exp(\boldsymbol{g_v}^T \boldsymbol{\delta_k}), \quad \delta_{l',k} \sim G(\beta_{k,v}) \tag{6}$$

where F$(\cdot)$ and G$(\cdot)$ denote a function of parameters inside (Zhao et al. 2017). We initialize the value of $\alpha_{d,k}$ as $1/K$, i.e., equal probability for K topics per review. After $\lambda_{l,k}$ is sampled, we can update the value of $\alpha_{d,k}$ and iterate over the (L_{doc}+1)-dimensional vector $\boldsymbol{f_d}$. Similarly, we initialize the value of $\beta_{k,v}$ as 0.01

[30]The word embedding was pre-trained on 2 billion tweets with 1.2 million unique words by Pennington et al. (2014).

(i.e., equal probability for 100 words per topic), and update $\beta_{k,v}$ by iterating over the $(L_{word}+1)$-dimensional vector $\boldsymbol{g_v}$.

We vary the number of topics between two and ten,[31] and estimate the LDA model incorporating both review-level and word-level characteristics by Markov Chain Monte Carlo (MCMC).[32] We find that the LDA model with five topics yields the highest topic coherence score, a measurement that has been shown to make the resulting topics more interpretable (Chang et al. 2009, Röder et al. 2015, Zhang and Luo 2023). We therefore set the number of topics $K = 5$, and estimate the LDA model with both review-level and word-level characteristics.

D.2. Unique Words under the LDA Model

Table D.1 displays the words that are unique to each of the five topics in decreasing order of the posterior probability.

Table D.1 Unique Words under the LDA Model (K = 5)

Topic number	Topic name	Unique words
Topic 1	Value for Money	price, bite, expect, better, quality, little, expensive, quite, much, disappoint, small, high, overall, find, portion, though, dont, although
Topic 2	Issues with Order	ask, arrive, tea, wait, waiter, minute, leave, tell, give, seat, sit, didnt, afternoon, bill, offer, waitress, another, people, glass, hour, day, bring, show
Topic 3	Menu and Food	main, starter, dessert, cook, steak, beef, fish, cheese, sauce, chocolate, bread, lamb, start, chicken, side, duck, pudding, meat, tasty, cream, chip, roast, pork, scallop, follow, share, crab, salad, potato
Topic 4	Service and Staff	recommend, amazing, friendly, definitely, attentive, fantastic, highly, love, thank, beautiful, worth, perfect, night, return, cocktail, treat, always, welcome, superb
Topic 5	Overall Experience	tasting, star, every, dining, chef, win, michelin, present, year, kitchen, room, list, sommelier, ever, without, work, choice

[31] A larger number of topics is less preferred because it generates topics with significant overlap.

[32] We maximize the likelihood of the topic assignments for each word in the corpus with respect to the parameters $\lambda_{l,k}$ and $\delta_{l',k}$, and obtain the review-level topic proportions $\boldsymbol{\theta}_d$. We run the MCMC chain for 15,000 iterations, with the first 1,500 iterations as burn-in. The hyperparameters $\boldsymbol{\alpha}_d$ and $\boldsymbol{\beta}_k$ are estimated and optimized every 100 iterations.

E. Additional Results on Mechanisms

E.1. Serving Size Changes

Table E.1 shows the results with the subset of “highly consistent” restaurants discussed in Section 5.1.2.

Table E.1 **Subset of Restaurants Evidencing Consistency**

	(1) SCM-DiD	(2) SynthDiD
Increase	0.001	0.002
	(0.075)	(0.068)
Decrease	0.340***	0.331**
	(0.079)	(0.160)

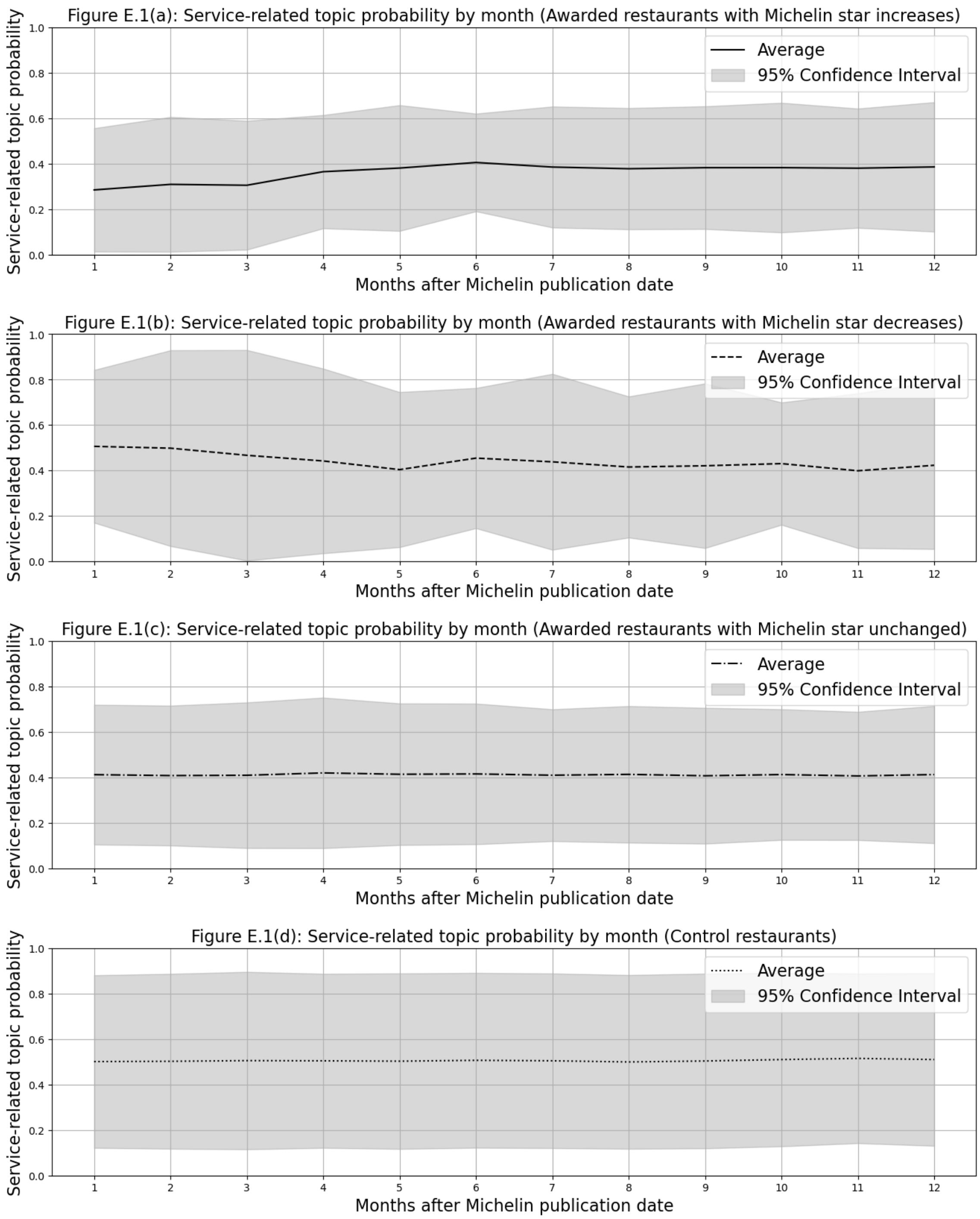
Note: Regression coefficients on Equation (1) are reported in Column (1), and overall ATTs estimated by Equations (2) and (3) are reported in Column (2). Robust standard errors clustered at pair level (Column 1) and aggregated standard errors (Column 2) are in parentheses. *** p<0.01. ** p<0.05.

E.2. Non-food Changes

We analyze trends in “service-related” review topics following the Michelin Guide updates. To do so, we focus on “service-related” metrics, based on topics 2 and 4 (cf. Section 4.2.1). This involves aggregating the probabilities of relevant topics over the twelve month period between guidebook releases. We follow a three-step procedure: First, we categorize all the restaurants by guidebook years into four groups: awarded restaurants with Michelin star increases (169 units); awarded restaurants with Michelin star decreases (83 units); awarded restaurants whose Michelin star status remained unchanged (2,091 units); and control restaurants (7,517 units). Second, for each unit within each of these four restaurant groups, we aggregate the reviews by month following the Michelin Guide publication date. Third, for each of these four restaurant groups, we plot the average service-related topic probability, aggregated across restaurants and Michelin Guidebook years, along with their 95% confidence intervals, as shown in Figure E.1. This additional analysis expands the period in our previous analyses from a maximum of 90 days to a full year.

If the star changes led to adjustments in service levels, we should expect to observe a corresponding shift in the probability of this topic being mentioned for restaurants that experienced a Michelin star change. However, as shown in Figure E.1, for all four groups, the probability of service-related topics being mentioned remains stable over the twelve month period. The trends in the third and fourth groups are more stable, because of the substantially larger numbers of observations.

Figure E.1 Service-related Topic Probability by Month



Note. To plot this figure, we follow a three-step procedure. First, we categorize all the restaurants by guidebook years into four groups: awarded restaurants with Michelin star increases (169 units); awarded restaurants with Michelin star decreases (83 units); awarded restaurants whose Michelin star status remained unchanged (2,091 units); and control restaurants (7,517 units). Second, for each unit within each of these four restaurant groups, we aggregate the reviews by month following the Michelin Guide publication date. Third, for each of these four restaurant groups, we plot the average service-related topic probability, aggregated across restaurants and Michelin Guidebook years, along with their their 95% confidence intervals.

E.3. Restaurant Demand

Table E.2 shows the results of restaurant demand using the log-transformed normalized Google search intensity (collected from Google Trends) as the dependent variable. Table E.3 shows the results of restaurant demand using daily OpenTable reservation data collected by Farronato and Zervas (2022) on New York City restaurants. Column 1 (2) is based on a window of 60 (90) days before and 60 (90) days after the guidebook release dates.

Table E.2 Google Trends Search Volume

	DV: log-transformed normalized Google search intensity	
	(1) SCM-DiD	(2) SynthDiD
Increase	0.020	0.023
	(0.064)	(0.075)
Decrease	0.004	–0.082
	(0.059)	(0.088)

Note: Regression coefficients on Equation (1) are reported in Column (1), and overall ATTs estimated by Equations (2) and (3) are reported in Column (2). Robust standard errors clustered at pair level (Column 1) and aggregated standard errors (Column 2) are in parentheses.

Table E.3 Effect of Michelin Stars Changes on Restaurant Demand (New York City)

	DV: whether 18.30 – 19.30 slot sold out on OpenTable	
	60-day window (1)	90-day window (2)
After	–0.081	0.275***
	(0.062)	(0.076)
After × Increase	0.084**	0.078**
	(0.034)	(0.031)
After × Decrease	–0.018	–0.030
	(0.050)	(0.041)
One Star	–0.056	–0.046
	(0.042)	(0.037)
Two Star	0.188**	0.208***
	(0.073)	(0.060)
Three Star	0.206***	0.222***
	(0.078)	(0.064)
Restaurant FE	Yes	Yes
Day FE	Yes	Yes
Observations	24,987	35,375
Number of units	222	222
R^2	0.510	0.507

Note: Robust standard errors clustered at restaurant level are in parentheses.
*** p<0.01,** p<0.05.

E.4. Consumer Sympathy

Table E.4 presents the estimation results for Equations (1), (2), and (3), using the volume of consumer reviews as the dependent variable.

Table E.4 Volume of Consumer Reviews

	DV: review volume	
	(1) SCM-DiD	(2) SynthDiD
Increase	8.042	0.312
	(5.426)	(1.985)
Decrease	–7.971	–2.412
	(4.939)	(2.996)

Note: Regression coefficients on Equation (1) are reported in Column (1), and overall ATTs estimated by Equations (2) and (3) are reported in Column (2). Robust standard errors clustered at pair level (Column 1) and aggregated standard errors (Column 2) are in parentheses.

F. Additional Details on the Reviewer-level Analyses

F.1. Reviewer Characteristics

Table F.1 Reviewer Characteristics

Variable	Definition	Example
Local consumer	Equals 1 if the reviewer is registered in the "United Kingdom" or "Ireland". Otherwise, equals 0.	0
Picky consumer	Equals 1 if the reviewer has never given a "5-star rating" in their profile. Otherwise, equals 0.	0
Cum. # of restaurants until each awarded restaurant	The number of restaurants that a reviewer has reviewed until each awarded restaurant.	4 and 7
Cum. mean review rating until each awarded restaurant	Mean review rating across all previously reviewed restaurants.	4 and 4.14

Table F.2 Reviewer Characteristics at the Time of the Review (Example)

Order of review	Local consumer	Picky consumer	Cum. # of restaurants until each awarded restaurant	Cum. mean review rating until each awarded restaurant
5	0	0	4	4
8	0	0	7	4.14

F.2. Restaurants Characteristics

We illustrate the process of computing restaurant characteristics at the time of the review. Table F.3 shows an example of a reviewer with eight reviews. Columns (1) and (2) show the restaurant ID and review date. For each restaurant, we extract its time-invariant characteristics (i.e., price level and cuisine type) from the corresponding TripAdvisor page, as presented in Columns (3) and (4). Then, leveraging the dataset of 79 million reviews we have collected, we calculated the cumulative review characteristics for each restaurant up to the review date. For instance, for the first review in Table F.3, we calculated the review characteristics for restaurant "105866" until 13 August 2014. Columns (5) to (7) illustrate these characteristics, including the logarithm of the total number of reviews, the mean and the standard deviation of previous ratings.

Assuming that the first five reviews in Table F.3 are written by the same reviewer. For each review within her profile, we calculate the cumulative restaurant characteristics that have been reviewed up to that point. Table F.4 provides an illustrative example. Columns (1) to (4) display the TripAdvisor profile of this reviewer, with each review indicating a review rating, the reviewed restaurant, and a specific date. Columns (5) to (10) describe the cumulative restaurant characteristics within the reviewer's timeline.

Table F.3 Restaurant Characteristics at the Time of the Review (Example)

	Restaurant ID	Date	Price Level	Cuisine Type	ln(Cum. number of reviews+1)	Cum. mean rating	Cum. rating s.d.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# 1	105866	2014-08-13	$$$$	French, European	6.864	3.889	1.056
# 2	033473	2014-08-25	$$ - $$$	European	5.357	3.923	0.946
# 3	086008	2014-12-03	$$$$	Seafood	3.996	4.450	0.828
# 4	018994	2015-02-11	$$$$	Bar, British	5.234	4.232	1.008
# 5	008075	2015-05-06	$$$$	Steakhouse	6.768	4.591	0.843
# 6	025005	2015-05-16	$$$$	European	7.001	4.535	0.827
# 7	037418	2015-08-15	$$$$	America	6.743	4.215	1.026
# 8	140968	2015-08-18	$$$$	Japanese	5.727	3.478	1.142

For instance, in the case of the first review in the profile, no cumulative restaurant characteristics exist. For the third review, the cumulative restaurant characteristics would incorporate information from the antecedent two restaurants. As detailed in Table F.3, the third review corresponds to a "seafood" restaurant with "$$$$" price level. Before this entry, the reviewer had visited a "$$$$" priced restaurant and another priced at "$$ - $$$ ". Thus, by the third review, there are two unique price levels, shown in Column (5) of Table F.4. In terms of cuisine type, uniqueness is determined based on specific word. For example, the first restaurant is labeled as "French, European," whereas the second is simply as "European" which is a subset of prior cuisine type. Therefore, up to the third review, the cumulative number of unique cuisine types is one (Columns (6) in Table F.4). Moreover, Columns (7) to (9) in Table F.4 compute the average review characteristics for those restaurant that have been reviewed so far. These calculations are derived from the information in Columns (5) to (7) of Table F.3, respectively. We also determined the range of review ratings up to each respective review. Columns (10) in Table F.4 present the rating range as "$mean \pm SD$" with ranges derived using columns (8) and (9).

Table F.4 Cumulative Characteristics of Restaurants Reviewed at the Reviewer Level (Example)

TripAdvisor Profile				reviewer-level cum. restaurant characteristics					
Order of Review	Review Rating	Restaurant ID	Date	Cum. number of unique price levels	Cum. number of unique cuisine types	Cum. average number of reviews	Cum. mean rating	Cum. rating standard deviation	Cum. rating range $mean \pm SD$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	4	105866	2014-08-13	–	–	–	–	–	–
2	4	033473	2014-08-25	1	1	6.863	3.889	1.056	[2.833, 4.945]
3	4	086008	2014-12-03	2	1	6.111	3.906	1.001	[2.905, 4.907]
4	4	018994	2015-02-11	2	2	5.405	4.087	0.943	[3.144, 5.000]
5	5	008075	2015-05-06	2	3	5.363	4.124	0.960	[3.164, 5.000]

Note: As review rating is in 5-point scale, the right boundary of the rating range is "minimum(5, $mean + SD$)" in Column (10).

G. Data Construction for Replication Study with NYC Restaurants

We construct the dataset for the replication study through five steps, outlined in Table G.1. Steps 1-3 are conducted in the same manner as described in Section 5.2.1: focusing on the Michelin guidebooks in NYC from 2013 to 2017, matching the awarded restaurants with (Farronato and Zervas 2022)'s OpenTable reservation data, and retaining units that had reservation information available for both the pre-guidebook and post-guidebook windows. Moving on to the fourth step, we proceed to collect OpenTable reviews specifically related to these awarded restaurants. As a result, our NYC replication dataset consists of 73,229 reviews for 52 (out of 70) awarded restaurants from 2013 to 2017. The remaining 18 awarded restaurants are excluded from the dataset due to the absence of an OpenTable page. In the fifth step, we focus on restaurants where we observed reviews within both a 90-day window before and after the guidebook release date. In the end, the data set for NYC replication includes 126 "restaurant-guidebook year" units that correspond to 47 awarded restaurants. Among these 126 units, 20 experienced increases in Michelin stars, 8 experienced decreases in Michelin stars, while the remaining units maintained the same Michelin stars.

Table G.1 Data Construction Steps in New York City Replication

Steps	Data sources	# Awarded Restaurants	# Michelin star increases	# Michelin star decreases	# "restaurant-guidebook year" unit
Step 1	Michelin Guides in NYC from 2013 to 2017	117	54	39	468
Step 2	OpenTable reservation data (Farronato and Zervas 2022)	70	28	13	230
Step 3	Reservations available on both sides of the guidebook release date	70	27	13	222
Step 4	OpenTable review data from 2013 to 2017	52	21	8	141
Step 5	Reviews available on both sides of the guidebook release date	47	20	8	126