Expert Opinion and Product Quality: Evidence from New York City Restaurants[*]

Olivier Gergaud [a], Karl Storchmann [b] and Vincenzo Verardi [c]

Abstract

We analyze whether consumers' quality perception and/or producer investment of New York City restaurants, measured by *Zagat* scores, responds to newly appearing expert opinion, measured by *Michelin* scores. Answering this question is of general economic interest as it applies to all markets with information asymmetries. Employing a difference-in-differences approach as well as a propensity score matching approach we find significant *Michelin* treatment effects on food and décor quality. Based of these changes, we find a *Michelin*-induced price increase of approximately 30% per *Michelin* star. To examine whether the improved food and non-food quality is based on restaurant investments or is merely imagined, we analyze non-food investments by referring to *Wine Spectator* wine list awards. Our analysis suggests that *Michelin*-reviewed restaurants are significantly more likely to invest in their wine list than others. As a result, *Michelin* reviewed restaurants are more likely to improve food and non-food (esp. décor) quality leading to significant price increases. However, while restaurants that increase prices only due to décor and service improvements are more likely to go out of business, food improvements appear to secure a restaurant's survival.

Keywords : consumer preferences, food quality, décor, service, expert opinion, restaurants.
JEL Codes : D11, L15, L66.

[*] We would like to thank two anonymous reviewers and the editor, Wesley Wilson, for many helpful comments that have helped to substantially improve this paper. We are also indebted to seminar participants at the Free University of Bolzano, University of Reims, Institut Paul Bocuse, Inra Grenoble (GAEL). Olivier Gergaud gratefully acknowledges financial support from the European Institute for Social Sciences and Humanities – Lille Nord de France (Project DEAC - Dynamiques de l'Économie de l'Art et de la Culture); Vincenzo Verardi gratefully acknowledges financial support from FNRS. We also thank Maurice Gergaud for helping to gather and code the data.

[a] Department of Finance and Economics, KEDGE Business School, Bordeaux. Email: olivier.gergaud@kedgebs.com
[b] Economics Department, New York University and Department of Finance and Economics, KEDGE Business School, Bordeaux. Email: karl.storchmann@nyu.edu (corresponding author)
[c] CRED (Centre of Research in the Economics of Development), University of Namur and ECARES (European Center for Advanced Research in Economics and Statistics), Université Libre de Bruxelles. Email: vverardi@ulb.ac.be

I. INTRODUCTION

In the presence of information asymmetries consumers often rely on expert opinion to guide their purchase decision. An increasing body of economic literature analyzes the effect of critical assessments on prices and quantity consumed for a wide variety of experience goods such as wine, movies, hotel rooms or books. All of these papers analyze the outcome of influenced quality perception of consumers.

Our paper is less focused on the question whether expert opinion impacts quantity or price of the good in question but rather examines consumers' quality perceptions and their possible changes directly. We analyze whether suddenly appearing expert opinion, on a market with long-standing published consumer-assessed quality evaluations, can alter consumers' quality perception and subsequently change prices. Will consumers stick to their original assessments or will they herd towards the expert's opinion?

We investigate this question by referring to restaurants in New York City. As the undisputedly leading restaurant guide,[1] *Zagat* has rated New York City's restaurants since 1979. *Zagat* publishes its guidebook once a year by drawing on consumer surveys. It, therefore, reflects local residents' restaurant preferences, which, until 2005, had been only scantly influenced by experts. There had not been any expert guides to New York City restaurant before 2005. Nationwide expert guides such as the *Mobil Travel Guides, Fodor* or the *AAA TourBook* series, for various reasons, had never had any mentionable impact on New York City diners (Ferguson, 2008; Davis, 2012). Although the *New York Times* has published weekly reviews and assigned quality ratings to local restaurants since 1963, the number of reviews has hardly exceeded 50 per year – mostly focused on new openings. In comparison, Zagat reviews about 2,000 restaurants per year. This and the fact that the reviews are spread over about 50 *New York Times* issues substantially limited its influence and never challenged *Zagat's* position.[2]

In November 2005, however, with the first release of the red *Michelin Guide New York City*, the first one ever for the United States, *Zagat* faced serious competition. In its first year,

[1] On average, about 650,000 copies of the New York City guide are sold per year. In addition, *Zagat* reports 384,000 unique visitors to its paid online subscription service for 2008 (Davis, 2012).

[2] For a comprehensive overview of New York City restaurant reviews, their history, focus and impact, see Davis (2012).

Michelin reviewed 471 restaurants and sold more than 100,000 copies (Krummert, 2006). In contrast to *Zagat*, *Michelin* relies on experts, i.e., five anonymous professional test eaters. According to Ferguson, while Zagat is a plebiscite, Michelin is a tribunal (Ferguson, 2008).

Although the advent of the *Michelin Guide* was excitingly anticipated in New York,[3] when it finally appeared the results were met with surprise, even with dismay. Many of the city's well-regarded restaurants were not awarded a *Michelin* star while others received unexpected honors (e.g., Kurutz, 2005; Fabricant, 2005b; Cuozzo, 2005a). The press detected a bias toward French-owned venues and the *New York Post* even called the Michelin Guide the "idiot's guide" (Cuozzo, 2005b). "After learning that *Babbo* had received [only] one star, Mario Batali[4] said he didn't think New Yorkers would give much credence to the guide. He was not happy with that ranking, the same as for the *Spotted Pig*, of which he is a part-owner. 'They're blowing it,' he said. 'They can't put the *Spotted Pig* on the same level as *Babbo*'" (Fabricant, 2005b).

What credence did New Yorkers give to the *Michelin Guide*? When tackling this question we do not analyze who of the two assessments, consumer or expert ratings, are closer to (unobserved) "true quality."[5] Instead, we analyze whether *Zagat* ratings have responded to *Michelin* quality assessments and employ a difference-in-differences approach for the years 2003, i.e., two years before the first New York City Michelin edition, and 2006, one year after its publication. We refer to various ZIP-code level demographic variables, such as the number of wine stores per capita and the local incidence of the treatment (measured as percent of restaurants treated in a region) as instruments for the restaurant treatment (i.e., being Michelin reviewed). In order to assess the robustness of our findings we also employ a propensity score matching approach which is aimed at isolating the treatment effect and purge it from other confounding factors.

We find significant *Michelin* treatment effects on food quality as well as on décor. However, it is *a priori* unclear whether these effects are based on demand side imaginations or whether the reviewed restaurants have in fact invested in food, décor and service enhancements. We

[3] See, e.g., Florence Fabricant in *The New York Times* (Fabricant, 2005a).
[4] Marco Batali, a Rutgers University economics major, is the chef and owner of New York City icon restaurant *Babbo*. He is best known for his Food Network show *Molto Mario* and his role in *Iron Chef America*.
[5] In contrast to national restaurant guides, *Zagat* ratings reflect the vote of the local population and are based on a local reference. Therefore, *Zagat* ratings are not comparable across cities and rather denote a local ranking (see also Berry and Waldfogel, 2010).

analyze restaurant non-food investments by referring to *Wine Spectator* wine list awards and find that *Michelin*-reviewed restaurants are significantly more likely to invest in their wine list than others suggesting that quality improvements are real and not consumer–imagined.

We also find evidence of *Michelin*-induced price effects. When using an ordered 0-1-2-3-4 treatment variable we find a marginal price effect of 31-33% per Michelin tier (0: not in Michelin, 1: reviewed, 2: one star, 3: two stars, 4: three stars). The price increases of Michelin-reviewed restaurants are based on quality improvements for food, décor and service. However, the market seems to punish restaurants that justify price increases mainly with décor and service improvements. We find that the odds of going out of business are higher for expensive décor- and service-focused venues. In contrast, *Michelin*-reviewed restaurants that focus on food quality improvements are significantly less likely to close down in subsequent years.

The remainder of this paper is organized as follows. Section II provides a review of the related theoretical and empirical literature. In Section III we present our data and in Section IV we outline our econometric approach. Section V reports the results of our difference-in-differences approach and draws conclusions; In Section VI we employ a propensity score matching approach and compare the findings with those from section V. Section VII summarizes the main findings and discusses some implications.

II. LITERATURE

Our study is aimed at analyzing whether, against the background of well-established and relatively stable peer reviews and quality perceptions, suddenly appearing expert opinion can exert authoritative influence on consumers and/or producers.

As New York City's leading restaurant guide, *Zagat* has published consumer reviews of restaurants for more than three decades. *Zagat's* critical evaluation of a restaurant's food, service and décor are solely based on consumer reviews. Only in 2006 these consumer ratings faced the considerable competition by expert assessments, i.e., the first publication of the New York City *Michelin Guide*. Michelin only rates the food of a restaurant and oftentimes disagrees with consumer preferences. In this paper, we are not interested in whether experts are biased or fallible (see, e.g., Ashenfelter et al., 2010; Hodgson. 2008, 2009), or what "true quality" is. Instead, we examine whether the arrival of expert opinion affects perceived product quality - either due to producer efforts (supply) or due to changes in consumer evaluations (demand). Answering this question is of general economic interest as it applies to all markets with information asymmetries.

Beginning in the 1970s, there is an extensive body of literature on the impact of information on markets with asymmetric information. In contrast, analyses of consumer responses to private and public information, expert opinion and the respective framing environment are of more recent nature.

A. Asymmetric Information, Markets and Producers

Beginning with the analyzes of Nelson (1970; 1974) most of the early literature on information asymmetric information was theoretical in nature and focused on the firm and its scope of quality signaling through advertising, warranties, reputation or pricing (e.g., Schmalensee, 1978; Shapiro, 1983; Wolinsky, 1983; Milgrom and Roberts, 1986; Bagwell and Riordan, 1991; Tirole, 1996).

Bagwell and Riordan (1991) assume that the credibility and magnitude of signaling quality through prices declines as consumer become increasingly informed. When only a few consumers are informed, high quality products signal their true characteristics through prices above the full information level. However, as an increasing number of consumers becomes informed, prices converge towards the full information level. Numerous empirical papers have analyzed the relationship between firms' signaling and the consumer information from an economic and a marketing perspective for various goods (e.g., Riesz, 1978; Tellis and Wernerfelt, 1987; Curry and Riesz, 1988; Caves and Greene, 1996; Heffetz and Shayo, 2009; Schnabel and Storchmann, 2010). In particular, with respect to expert opinion, Roberts and Reagans (2007) show that the price-quality relationship of New World wines strengthens with growing critical exposure on the firm level.

Related to our analysis, Rosenman and Wilson (1991) show that consumers infer product quality on the cherry market by referring to producer characteristics, i.e., proxy variables. Likewise, consumers may assume that a restaurant's ambience or the quality of its service serves as a proxy variable for its food quality, which may justify restaurant investments in non-food characteristics -- such as décor.

B. Consumers

There is a growing body of consumer-related literature focusing on the role of peers and experts on consumer learning. All of these analyses draw on the assumption that the decisions of other consumers or the assessment of experts contain choice-relevant information. The literature on the influence of peers or "social learning" on individual decisions is based on informal approaches in the psychological literature (e.g., Deutsch and Gerard, 1955; Bandura, 1977). For instance, Becker (1991) develop a formal model in which the demand for a good, here a restaurant meal, depends positively on its aggregate quantity demanded, i.e., on peer demand. Banerjee (1992) and Bikhchandani et al. (1992 and 1998) describe localized conformity, fashions and "herd behavior" as the result of informational cascades where the decision of an individual is influenced by the actions of other individuals before him. Since, in these models, the individual is willing to give up his private information and only follows the preceding peers, the peers' actions do not contain any information and the resulting equilibrium may be inefficient.

McFadden and Train (1996) hypothesize that consumers learn from other consumers but still utilize their private information. They formalize consumer learning about a new good's quality through a rational decision process between learning from own experience and from the experience of their peers.[6] Morris and Shin (2002) show that, even when agents have private information, they might overreact to expert opinion and devalue their private information.

On the empirical side, Salganik et al. (2006) created an artificial ''music market'' in which participants downloaded previously unknown songs. When providing the treatment group of users with information about other users' music ratings, social learning is a strong determinant of a song's success. Moretti (2011) empirically examines social learning for movie sales from 1982 to 2000. He analyzes movie sales over time compared to prior expectations, measured by the number of screens dedicated to a movie in its opening

[6] Peer or social learning models are related to the earlier literature on technology adoption, where the spreading of new technologies is based on peer imitation (e.g., Griliches, 1957).

weekend, and finds a reinforcing pattern. When a movie exceeds expectations in its opening week consumers update their expectations leading to further increasing sales etc. Liu (2006) finds similar results for word-of-mouth effects on movie sales by referring to consumers' internet postings. Cai et al. (2009) set up a randomized natural field experiment in which they assess consumer choices of restaurant menu items. If provided with a (made up) list of "last week's top 5 selling dishes," consumers tend to follow their peers' consumption. Similarly, Anderson and Magruder (2012) show that positive consumer evaluations of restaurants on Yelp.com induce higher table reservations.

C. Experts

In addition to the literature on social learning from peers, there are also numerous papers that confirm the influence of experts on markets. The effect of experts on market outcomes is hard to measure since expert reviews and "true quality" are often closely correlated. Hence, most studies draw on natural (or real) experiments or make statistical inferences to disentangle the two.

For instance, Ginsburgh (2003) reports that experts significantly determine the market success of movies (through Oscars) and, to a lesser degree, of books (through the *Pulitzer Prize*). Reinstein and Snyder (2005) examine the impact of critical reviews on movie box revenues and also find positive effects of favorable reviews. Ginsburgh and van Ours (2003) analyze the Queen Elizabeth piano competition in Belgium and find that musicians who are successful in the competition will be rewarded by subsequent market success**.** Similarly, experts affect sales prices for paintings at art auctions by publishing pre-sale estimates in auction catalogues (Bauwens and Ginsburgh, 2000).

Hadj Ali, Lecocq and Visser (2008) analyze the effect of critical points awarded by wine writer Robert Parker on the *en primeur* price of Bordeaux wine. They find Parker points to have a significant but small effect on wine prices. Dubois and Nauges (2010) also study the effect of Parker points on *en primeur* prices of Bordeaux wines. They employ a structural empirical approach to disentangle the effect of experts' grades and unobserved quality on the wine price and find a significant "Parker effect." Closer related to our research, Gergaud et al. (2007) find a substantial influence of expert ratings, measured by *Guide Michelin* stars, on Paris restaurant menu prices.

In contrast to price analyses, there are only a few papers that examine the impact of expert opinion on quantity consumed. Drawing on a field experiment in wine retail stores, Hilger et al. (2011) show that favorable expert reviews have a positive influence on quantity consumed, independent of quality. On the other hand, wines that obtained below-average ratings exhibit a decrease in demand. Friberg and Grönqvist (2012) analyze the impact of expert opinion on quantity consumed by referring to the Swedish wine market. They find a substantial and long-lasting effect (more than 20 weeks) of positive reviews. In addition, they also find positive demand effects of neutral reviews and no negative effects of unfavorable reviews.

D. Framing

However, consumers' quality perception is not only influenced by own or others' experience but is also responsive to the respective consumption environment. There is plenty of evidence that consumers make contextual inferences (Kamenica, 2008) and are sensitive to the framing of the decision situation (e.g., Tversky and Kahneman, 1981). For instance, North et al. (1999) show that consumers respond to the kind of music played in a wine store. When French music was played, customers bought more than three times as many French wines than German wines. When German music was played the opposite was true. Wansink et al. (2009) report that the quantity of food we eat is only partially determined by what we were planning on consuming. Environmental factors such as package size, plate size and shape, lighting, variety etc. affect our food consumption volume far more than we realize. Plassmann et al. (2008) draw on brain scans and show that changes in the price of a product can affect neural representations of experienced pleasantness.

Similarly, experts may also be affected by framing variables. For instance, for the restaurant sector, Chossat and Gergaud (2003) show that experts, although claiming to assess food quality only, may also be influenced by non-food framing elements, such as the décor of the venue or the choice of wines in the cellar.

III. DATA

We are interested in assessing whether consumers' restaurant quality perceptions, i.e., *Zagat* ratings, have been influenced by the publication of *Michelin* expert opinion in 2005. The dataset we employ covers all New York City restaurants considered in both the 2003 and

2006 *Zagat* Surveys. These years correspond to two years before and one year after the first publication of the NYC *Michelin*. We draw on 2003 instead of 2004 data to rule out that our results are influenced by possible *Michelin* announcement effects on consumer assessments or on restaurant efforts.[7]

In the 2004 issue (which was published in 2003), Zagat published a total of 1,918 restaurant reviews based on the ratings of almost 26,000 reviewers (Zagat Survey, 2003). In the 2007 issue, published in 2006, it rated 2,014 establishments based on reports of 31,604 restaurant-goers (Zagat Survey, 2006). After removing all chain restaurants from this list, we are left with 1,518 observations. *Zagat* provides an average consumer-reported price charged for a reference dinner including one drink and tip for each restaurant. It also provides information on the consumer-perceived quality of food, décor and service on a scale ranging from 0 to 30 points separately for each category. In addition, *Zagat* lists some 90 different ethnic cuisine styles[8] that we bundled into nine broad categories to avoid singletons: Africa, Asia, Central America, Eastern Europe, Middle East, North America, South American, Western Europe, and Other.

Our treatment group consists of 471 non-randomly selected restaurants that were reviewed in the first *Michelin Guide*, 2006 edition (Michelin Travel Publications, 2005). In contrast to *Zagat*, the *Michelin Guide* claims to review the quality of food only; neither décor nor service quality should affect its rating.[9] *Michelin* rates a restaurant's food quality on a scale from zero to three stars.

[Insert Table 1 here]

[7] The publication of the first New York City *Michelin* guide was announced in February of 2005.
[8] These categories are the following: Afghan, American New, American Regional, American Traditional, Argentinean, Asian, Australian, Austrian, Bakeries, Barbecue, Belgian, Brasserie, Brazilian, Burmese, Cajun/Creole, Californian, Caribbean, Chinese, Coffeehouses/Dessert, Coffee Shops/Diners, Colombian, Continental, Cuban, Delis/Sandwich Shops, Dim Sum, Dominican, Dutch, Eastern European, Eclectic/International, Egyptian, English, Eritrean, Ethiopian, Filipino, Fish 'n' Chips, French, French Bistro, French New, German, Greek, Hamburgers, Health Food, Hot Dogs, Hungarian, Indian, Indonesian, Irish, Israeli, Italian, Jamaican, Japanese, Jewish, Korean, Lebanese, Malaysian, Mediterranean, Mexican/Tex-Mex, Middle Eastern, Moroccan, Noodle Shops, Nuevo Latino, Persian, Peruvian, Pizza, Polish, Portuguese, Puerto Rican, Russian, Sandwich Shop, Scandinavian, Scottish, Seafood, Soups, South African, South American, Southern/Soul, South Western, Spanish, Steakhouses, Swiss, Tapas, Tea Service, Thai, Tibetan, Tunisian, Turkish, Ukrainian, Vegetarian, Venezuelan, Vietnamese.
[9] The *New York Times* quotes Jean-Luc Naret, the director of the *Michelin Guides,* "*Michelin* stars refer only to what is on the plate." (Fabricant, 2005a).

Table 1 reports the descriptive statistics for food, service and décor quality as well as for prices. When looking at all restaurants, we see that all quality categories have improved from 2003 to 2006. The average meal price (including a drink a tip) has increased from $38.14 to $40.69. Table 1 reports the same data also separately for both treatment and control group for 2003 and 2006. Expectedly, the treatment group was rated higher than the control group in each category, i.e., food, service and décor. This is true before as well as after the treatment. In addition, the mean values for each group and category remained virtually unchanged between 2003 and 2006. In contrast, the average price of the restaurants in the treatment group grew significantly after the *Michelin* review. In addition, the dispersion, measured by the Coefficient of Variation (CV)[10] within each quality category tends to be slightly lower in the treatment group before and after the treatment. For 2003, this also applies to the price dispersion. After the treatment, however, the reviewed restaurants experienced a substantial increase in price dispersion: the CV of prices grew from 34.1% to 53.5%, suggesting a considerable injection of noise caused by published expert opinions.

In Table 2 we show the percentage growth rates from 2003 to 2006 in each quality category and in prices separately for treatment and control group. Although these numbers are uncontrolled for effects such as food ethnicity, they still convey a few interesting developments. First, while we find a perceived food quality improvement for non-treated restaurants of 2.99%, the treatment group exhibits a small decline. Second, and despite the lack in food enhancement, the treatment group shows a substantial price increase of 10.38% while there is only a 2.87% increase for un-reviewed restaurants.

[Insert Table 2 here]

This overview, however, disregards any influence of variables such as food ethnicity (cuisine categories), restaurant location, operating hours or payment options. In the following section we will thus employ an econometric model to analyze the *Michelin* effect on the restaurant quality categories food, service and décor, as well as on restaurant meal prices.

[10] We calculate the CV as standard deviation to the mean, $CV = \sigma / \mu$

IV. ECCONOMETRIC METHODOLOGY

Our econometric analysis relies on three difference-in-differences models, one for each category, i.e., food, service and décor, in order to assess whether the mere inclusion in the *Michelin* guide affected consumer quality assessments. We estimate the following equation:

$$Log(Q_{it}) = \beta_0 + \beta_1 \log(Q_{it-1}) + \beta_2 Mich_i + \beta_3 After_t + \beta_4 Mich_i \times After_t + X_{it}\theta + \varepsilon_{it} \qquad (1)$$

where *i* denotes individual restaurants and *t* denotes time. Q_t is a measure of quality of food, service or décor, respectively, measured in period *t* (i.e., 2006); similarly, Q_{t-1} stands for the quality variables in the prior period, i.e., 2003. Introducing the lagged dependent variable accounts for the persistence of quality over time. $Mich_i$ is a dummy variable that takes on the value one if the restaurant was reviewed in the 2006 *Michelin* guide (first edition, published in 2005) and zero otherwise. *After* is a time dummy that equals one in the period following the introduction of the guide and zero before. $Mich_i \times After_t$ is the interaction term between the two and measures whether *Q* has changed differently for those who have been introduced in the guide compared to those who have not (control group). It is also known as the difference-in-differences term. X_{it} is a matrix of control variables such as food ethnicity and some characteristics at the restaurant level (accepts credit card, open after 11pm, open on Sundays, limited number of reviews).[11]

Obviously, the treatment, i.e., being considered in the *Michelin* guide, is not random and independent of the quality of food (or service or décor, respectively) as reported by consumers in the *Zagat* guide. We, therefore, suspect an endogeneity bias. To remedy this shortcoming we instrument the treatment itself. Given the geographical clustering of *Michelin*-reviewed restaurants, we use the percentage of treated restaurants in the neighborhood as instruments.

The map provided in Figure 1 shows that all *Michelin*-reviewed restaurants are either in one of two geographical clusters in Manhattan or in two less concentrated groups in Queens and Brooklyn.[12] This spatial concentration suggests that the likelihood of being considered in the *Michelin* guide is not independent of a restaurant's geographical location. We exploit this fact

[11] *Zagat* reports if a restaurant receives only a low number of reviews.
[12] Aside from these clusters, there is only one isolated *Michelin*-reviewed restaurant in Forest Hills, Queens.

and employ a geographical location variable to instrument for being reviewed by the *Michelin* guide.

[Insert Figure 1 here]

In addition to the geographical location of the restaurant, we also explore other possible instruments for the *Michelin* treatment. We examine ZIP-code level data of various demographic and economic data that may serve as appropriate instruments for the treatment variable.[13] In particular, we employ size and racial composition of the population, per capita income, population share under the poverty line, share of full-service restaurants as well as the number of wine and liquor stores per capita. For the instrument selection we hypothesize that *Michelin* reviewed restaurants are above-average expensive and depend on a well-off clientele. We thus assume that *Michelin* restaurants predominantly locate in upscale neighborhoods with high per capita incomes and low poverty rates. Similarly, we expect the likelihood of being *Michelin*-reviewed to be positively related to the number of full-service venues (i.e., inversely related to the number of fast food outlets) and the number of wine stores. The latter draws on the fact that New York State stipulates that wine and liquor can only be sold in state-licensed wine and liquor stores. In contrast, beer is usually sold in supermarkets and convenience stores and must not be sold in wine and liquor stores.

In other words, we assume a direct relation between the regional concentration of *Michelin* restaurants and their environment (wealth/poverty and interest of the local population for fine wine and food). In addition, since all restaurants in the sample were already established when the *Michelin* guide was introduced, the instruments should be exogenous.

As will be shown later, the statistical tests tend to strongly support our intuition and show that our instruments are neither weak nor endogenous.

Defining neighbors and instruments

[13] There are 176 ZIP codes in New York City.

In order to define neighbors, we identify the geographical coordinates of all restaurants[14] and compute the distance between all pairs of observations. The smallest maximum distance between two restaurants in the dataset is 31.3 km, the largest minimum distance is 3.4 km and the general average distance between two restaurants is 5.9 km.[15] We attribute proximity spatial weights as follows:

$$w_{ij} = \begin{cases} 0 & if\ d_{ij} \notin [l_b, u_b] \\ 1/d_{ij}^f & if\ d_{ij} \in [l_b, u_b] \end{cases}$$

where (*i*,*j*) denotes a pair of locations, d_{ij} stands for the Euclidean distance between restaurant *i* and *j*, l_b and u_b denote the lower and upper bound of the specified distance band, respectively, and *f* is a positive friction parameter that is set exogenously. The friction parameter determines the rate of devaluation for neighbors compared to the geographic distance. A parameter value of one denotes that the importance of the neighborhood effect is linearly decaying in distance. A friction parameter larger than one suggests that neighborhood effects decline faster than the geographic distance and vice versa.

Since in New York City, the monetary transportation cost is virtually independent of distance traveled while time spent depends on distance, we set the friction parameter equal to 0.8 suggesting below-proportional neighbor depreciations compared to the geographic distance. However, our empirical results are not overly sensitive to different parameter values. We tried different values for this friction parameter and the estimated coefficients were mostly unaffected. We selected 0.8 since it provides the strongest and most exogenous instruments.

Finally, the values in the weighting matrix are standardized in order to ensure that the sum of all elements per row equals one. A restaurant *i* is considered a neighbor of restaurant *j* if the distance between *i* and *j* does not exceed 10 km (i.e. l_b=0 and u_b=10).[16]

We can now calculate the average number of *Michelin* restaurants in the neighborhood of

[14] The coordinates are available in decimal degrees from www.maporama.com and are converted into distances*(km)* to the equator and to the Greenwich meridian using the formula: $distance = \frac{6378.137 \cdot \pi \cdot degrees}{180}$

[15] The maximum distance between two restaurants is little informative since it merely reports the spatial spread of restaurants in New York City. Similarly, the minimum distance is virtually zero for adjacent restaurants. The largest minimum distance gives us an idea of the minimal radius needed for all restaurants to have at least one neighbor. The smallest maximum distance, on the other hand, reflects the spatial spread of restaurants compared to the central restaurant.

[16] The distance of 10 km was selected to ensure that each restaurant has at least one neighbor.

each restaurant (weighted by the distance) by multiplying the weighting matrix (*W*) by the vector identifying the *Michelin* restaurants. In other words, the frequency of "*Michelin* restaurants" in the neighborhood of each restaurant is defined by $W \cdot Mich$ (vector *WMich*). This variable is the first instrument we use for the treatment. The second instrument we consider, provided by *Zagat*, is a dummy variable that is equal to one if only a small number of customers reviewed the restaurant (*Low2003*). We hypothesize that *Michelin* can afford to disregard unknown restaurants. However, restaurants with a large number of customer reviews may enjoy an increased likelihood of being selected in the guide. Since *Zagat* refers to the number of 2003 reviews, i.e., well before the announcement of the *Michelin* launch, we deem this variable exogenous.

To summarize, the endogenous right hand side variables are *Mich* and $(Mich \times After)$. The available instruments are *WMich*, *WMich* interacted with *After* (which is exogenous) and *Low2003*. Since we employ more instruments than we have endogenous variables we test for their redundancy, over-identification (i.e., exogeneity), and weakness. We perform these tests for each model, i.e., for food, décor and service, by drawing on the Hansen J-statistic for over-identifying restrictions (exogeneity test), the Kleibergen-Paap rk LM statistic (relevance test) and the Kleibergen-Paap rk Wald F statistic (weakness test).

Note, we do not employ restaurant fixed effects. In general, difference-in-differences models may suffer from omitted variable bias (OVB) when an omitted variable is correlated with both treatment and outcome. The OVB could be significantly alleviated by including restaurant fixed effects that capture all time-invariant determinants. In our case, however, strict restaurant fixed effects pose a problem since our instruments also include a considerable time invariant component. Since we hypothesize that the treatment is crucially driven by locational reasoning (causing *Michelin* clusters), employing a fixed effect model would significantly lower the computed treatment effect (i.e., cut out some of the signal but keep the noise) and render our instruments invalid. In fact, when running a FE-IV model the Kleibergen-Paap rk (LM statistic) suggests that we cannot reject the null that our model is underidentified. We, therefore, decided to stick with the standard difference-in-differences model controlling for all characteristics that are available to us. We also include lagged dependent variables, which account for a large fraction of restaurant fixed effects, but certainly not all. Finally, we contrast the difference-in-differences approach with various other methods.

V. RESULTS

A. Impact on Quality

(1) Full Sample

Table 3 reports the OLS results of the model described in equation (1) with respect to the quality of food, service and décor. In the first three columns we model the treatment with a simple dummy variable, where 1 denotes Michelin reviewed and 0 otherwise.

Ideally, we would like to calculate the impact of each *Michelin* assessment level, i.e., no star, one star, two stars and three stars. This, however, requires more valid instruments than are available. We, therefore, resort to an ordered treatment variable that takes on the value 0 for not reviewed, 1 for reviewed but no *Michelin* star, 2 for one star, 3 for two stars and 4 for three stars. The ordered variable thus postulates a constant marginal effect of each additional *Michelin* star on the various quality variables. We report the respective results in the three right columns of Table 3.

Both model variants yield virtually identical results and show significant treatment effects (*Mich* $\times$ *After*) on food, décor and service. For instance, the dummy treatment suggests that being *Michelin* reviewed leads to a 10% increase in perceived food quality, an 18% increase in décor and a 12% increase in service quality.

However, these results are not generally supported when using instruments. As shown in Table 4, we find positive treatment effects for all quality categories. But the interacted term (*Mich* $\times$ *After*) is statistically significant only for décor, suggesting that *Michelin* reviewed restaurants invest only in their décor but not in their food quality. If consumers deem décor a proxy variable for food quality this would be in line with Rosenman and Wilson's (1991).

[Insert Table 3 here]

When comparing the OLS and 2SLS for food in Table 3 and 4 we find that the coefficient for the interacted variable we are mainly interested in (i.e., *Mich* ×*After*) is positive and significant in the OLS estimation while not significant (but still positive) in the 2SLS estimation. While the lack of significance could be due to the inefficiency of the 2SLS estimation, we also see that the point estimate decreases by approximately 50% (from 0.1 to 0.05). This suggests an upward bias in the OLS estimates which was corrected by the 2SLS model. We, therefore, conclude that the *Michelin* effect on food quality was fairly moderate (i.e., we do not reject that it is zero). We find a similar pattern for the service variable. On the other hand, the 2SLS *Mich* ×*After* coefficient for décor is significant and substantially larger than the OLS coefficient (0.18 compared to 0.39) suggesting that the OLS estimates are downward biased.

In Table 4, we also report the results of various tests for overidentifying restrictions (exogeneity), relevance and weakness of our instruments. Note that we chose different combinations of instruments depending on the resulting test statistics and that we instrument both *Mich* and *Mich*×*After*.

All first stage estimates are provided in the Appendix (Table A1). The variable "limited number of reviews" refers to a restaurant's (quantitative) unpopularity. The geography variable, as described above, denotes the regional concentration of reviewed restaurants; the number of wine stores, the share of population below the poverty line and the share of full service restaurants are by ZIP code and reflect various aspects of neighborhood desirability. In general, restaurants that do not accept credit cards, are open after 11pm and are located in a restaurant cluster are more likely to be Michelin reviewed than others. When instrumenting the ordered Michelin variable we also find adverse effect of high poverty shares in the neighborhood. In contrast, the density of wine stores per ZIP code does not exert any significant effect.

[Insert Table 4 here]

For the food variable, we calculate the Hansen J-statistic to check for overidentifying restrictions of the instruments. The resulting value of 4.674 is well below the critical χ^2 value for three degrees of freedom (7.815). We hence do not reject the null hypothesis that the instruments are exogenous. To check for the relevance of the instruments, we rely on the

Kleibergen-Paap rk LM statistic, which equals 30.75 for the food model. This value is well above the critical χ^2 value for four degrees of freedom, which is 9.488. Therefore, we reject the null that the model is underidentified. Finally, we test whether our instrument sets are weak drawing on the formal test suggested by Stock and Yogo (2005) who propose a procedure testing for the null hypothesis that the bias of 2SLS is some fraction of the OLS bias. For instance, if the bias of 2SLS is less than 10% of the OLS bias, the instrumental variable estimator has reduced the OLS bias by more than 90%. Table 5 reports the maximal IV relative bias for the 2SLS estimator. Drawing on the Kleibergen-Paap rk Wald F statistic we find a value of 7.618 for the food model which is lower (larger) than the critical value of 8.78 (5.91) tabulated by Stock and Yogo (2005) for a 20% (10%) maximal IV relative bias. We find similar results for the service and décor equations suggesting that the 2SLS estimator results in bias reductions between 80 and more than 95%. We report the results for the ordered treatment variable in the three right columns of Table 4. Overall, the findings are very similar to those when using a dummy treatment variable.

(2) Restricted Paired Sample

So far, our 2SLS results suggest that, in response to being Michelin reviewed, restaurants' perceived décor quality has significantly improved, whereas there is no discernable effect on food quality. However, conclusions need to be interpreted with care since our full sample is not only comprised of restaurants that were in business in 2003 as well as in 2006. It also includes restaurants that closed down before 2006 and of new births that did not exist in 2003.[17] In order to assess whether our results are biased due to restaurant closures or new births we also run equation (1) on a sample that is restricted to restaurants that were in business in both years, 2003 and 2006. Our further analysis is thus confined to perfect pairs.

Table 5 reports the results of the OLS models run on the restricted paired sample. Except for the number of observations, the restricted sample is almost 1000 observations smaller, we do not detect any considerable difference to the results shown in Table 3. Even all marginal effects are of almost identical size.

[17] In addition, there are restaurants that operated in both years but had one dependent or independent variable missing in one of the years.

[Insert Table 5 and 6]

The 2SLS equation for the restricted paired sample, shown in Table 6, is, however, very different from its counterpart in Table 4.[18] While the full sample 2SLS equation suggests only décor improvements (either perceived or real), the same equation run on the restricted sample suggests only food improvements. The *Michelin* effects on décor and service are insignificant. These results hold also for the ordered treatment variable. The test results provided at the bottom of the Table suggest that our instruments are neither endogenous, nor irrelevant, nor weak.

The puzzling difference between the 2SLS models run on two different samples must be due to the characteristics of restaurants that are present only in one of the two years, notably restaurants that closed down between 2003 and 2006. The difference in treatment effects for the two samples suggests that restaurants that closed down before 2006 improved their décor quality but not their food quality. In contrast, surviving restaurants show significant food improvements, but no perceived décor or service enhancements.

[Insert Table 7]

In Table 7 we show two simple probit equations on restaurant closures in 2007. While restaurants that are open on Sundays and after 11pm are less likely to go out of business, being a Middle Eastern restaurant increases the odds of shutting down. In addition, and more interesting for this study, the results in Table 8 also suggest that high food quality lowers the odds of closing down while high décor marks have the opposite effect. This seems to square with the differences between the full and restricted samples in Tables 4 and 6. The décor effect of the Michelin treatment in Table 4 dominates the food effect because the sample includes non-surviving restaurants, whereas the restricted sample does not. In addition, as shown in column (2) of Table 7, being *Michelin*-reviewed by itself lowers the odds of closing down.

Apparently, a *Michelin* review, which is *per se* good for a restaurant's odds of survival, opens

[18] The First Stage results are reported in the Appendix Table A2.

up two strategy paths. Restaurants that improve their (perceived) food quality can further secure their survival, while restaurants that only improve their décor and service quality are less likely to survive.

(3) Base Year Choice and Dynamics

We chose the base year 2003, i.e., two years before the treatment, to avoid potential disturbances by *Michelin* announcement effects. Similar to Table 6, Table 8 shows the results of a 2SLS model run on the restricted sample. However, instead of 2003, we now refer to 2004 as the base year. In general, the results of Table 8 and Table 6 are very similar for almost all variables. The only difference is that by referring to 2004 the treatment effect on décor grows and becomes significant at the 5% level. In contrast, the treatment effect on perceived food quality drops from a coefficient of 0.20 (base year 2003) to 0.15 (base year 2004). The latter is in line with an announcement effect and suggests that restaurants were on different trend paths even before the treatment. That is, between 2003 and 2004, *Michelin*-reviewed restaurants improved their food quality by more than non-reviewed venues leading to a smaller treatment effect when referring to the base year 2004.

[Insert Table 8]

The opposite is true for décor. The increase in the treatment effect for the 2004 base year suggests that the décor difference between reviewed and non-reviewed restaurants was smaller in 2004 than in 2003. This can be due to a “décoring-up“ of all restaurants, especially by non-reviewed restaurants, in expectation of the *Michelin* review.

From a restaurateur’s perspective, this strategy may be sensible as Chossat and Gergaud (2003) and Gergaud et al. (2007) show that *Michelin* evaluations in France are not solely driven by food quality but also influenced by non-food characteristics such as décor and service. Johnson and Surlemot (2005) interviewed chef-owners of *Michelin* starred restaurants in France, Belgium, Switzerland and the UK and report that receiving a *Michelin* star places enormous pressure on the owner. Massive efforts and investments are due in order to retain the recently gained (first, second or third) *Michelin* star. Since these investments include service and décor it seems to be commonly understood among restaurateurs that *Michelin* ratings – in contrast to their claim - are influenced by service and décor. These findings suggest that the higher service and décor quality may not be imagined by consumers but may rather be the result of the owner's effort.

Investments in service and décor are expensive and may only be justified if they yield higher revenue. There is some anecdotal evidence that *Michelin* stars demand a premium and are thus worth being retained. Eric Ripert, chef and owner of *Le Bernardin*, one of only three New York City restaurants that received three *Michelin* stars in 2006, reports revenue increases of at least 15% (Davis, 2012). Johnson and Surlemot (2005) find similar values for European *Michelin* starred venues. In an analysis of French *Michelin* reviewed restaurants from 1970 to 1994 Snyder and Cotter (1998) find a close relationship between investments, especially in ambience, *Michelin* stars and prices. In particular, the loss of a *Michelin* star is often predated by receding investments and lower prices. Similar findings are reported by Gergaud et al. (2007) for restaurants in Paris.

However, and as shown in Table 8, décoring and servicing-up without improved food quality may be at the expense of the future survival of the restaurant, especially for an un-reviewed venue.

B. Impact on Prices

In Table 9 we show the impact of the *Michelin* treatment on menu prices. The model specification is identical with the one for the *Zagat* quality assessments (see equation 1); we only substituted the logarithm of menu prices for the *Zagat* variable as dependent variable. The tests for overidentifying restrictions, underidentification and weak identification for the selected instruments are reported at the bottom of the Table. When employing a simple 1-0 dummy variable for the inclusion in the *Michelin Guide* we find treatment-induced price increases of approximately 40%. When using an ordered 0-1-2-3-4 treatment variable we find a marginal effect of approximately 30% per tier (reviewed, 1 star, 2 stars, 3 stars). Note that the price reported by Zagat includes a drink and tip.

[Insert Table 9 here]

We are further interested in examining whether these price increases are related to food, service or décor quality improvements and whether there is a difference between *Michelin*-reviewed and un-reviewed restaurants. We select all unique restaurants for which we have price and quality data for 2003 and 2006 and regress the nominal price difference on the respective quality difference, a constant term and various time-invariant control variables:

(2) $$(P_{2007,i} - P_{2004,i}) = \beta_0 + \beta_1 (Q_{k,2007,i} - Q_{k,2004,i}) + \sum_{m=1}^{n} \beta_m F_i$$

where *P* denotes price, *Q* quality, *k* the specific quality variable, i.e., food, décor or service, and *i* the individual restaurant. F is a vector of time-invariant variables such as food ethnicity fixed effects, open after 11am, closed on Sunday, no credit cards accepted and low response rate.

In this fashion we run 12 different regressions; Table 10 displays the results. In the columns denoted “All” we draw on 702 non-reviewed and 331 *Michelin*-reviewed restaurants. For the group of non-reviewed restaurants we do not find any significant effects of food, service and décor quality changes. In contrast, the prices of *Michelin*-reviewed restaurants exhibit significant price responses to all quality changes. The corresponding marginal effects suggest that quality improvements by one point cause price increases between \$0.38 and \$0.54. These results suggest that the price changes of *Michelin*-treated restaurants are linked to changes in the perceived quality of their food, décor or service while the prices of untreated restaurants seem to be uncorrelated to quality changes.

However, when regressing price changes only on changes in one quality dimension, we disregard possible changes in the other quality variables and may confound the respective marginal effects when the various quality changes are correlated. We, therefore, augment our analysis and restrict our sample to restaurants that exhibited a change in only one quality variable while keeping the other quality variables constant. For instance, when regressing price changes on food quality changes, we only refer to restaurants for which décor and service has not changed. The results for this “restricted sample” are also reported in Table 10. We assume that the two samples, i.e., “All” and “Restricted Sample,” are comparatively similar displaying almost identical mean prices.[19] While this procedure allows us to isolate the respective quality effects on prices, our sample size now drops by about 40% due to the fact that most restaurants experienced changes in more than one quality dimension. The corresponding results, as reported in the columns “Restricted Sample” in Table 10, confirm our prior findings for food and service; the décor effect becomes less significant. Table 10, therefore, results suggest that, while price changes of restaurants that have not been reviewed by the *Michelin* guide are detached from quality changes, menu price changes of *Michelin*-reviewed restaurants are driven by food or service quality changes.[20]

In addition, higher prices are not necessarily an indicator for a restaurant’s success. They are set by the supply side and may be triggered by cost or the wish to use prices as quality signals and may thus not reflect higher demand. In fact, we have no information about quantities.

Table 11 presents four probit equations that report the odds of a restaurant’s closure in 2007 as a function of its price. In order to control for food quality we partitioned our sample into four food quality quartiles of almost identical sample size. The top quartile (Q1) is comprised of restaurants with a food score of 23 and above, Q2 of 21 and 22-point venues, Q3 of 19 and 20-point venues and Q4 of restaurants with 18 points and less. We also included a Michelin dummy, to examine whether being *Michelin*-reviewed provides any protection, and the full set of dummy variables as listed in Table 8.

For the high food quality segment (Q1), the regressions suggest that, while being *Michelin*-reviewed provides some protection, high prices are a significant determinant for restaurants to go out of business. Note that we include the price variable in its squared form. That is, the effect of price on shutdowns is not linear but exponential. In fact, the coefficients for the first quality quartile suggest that a price of slightly above $100 (including a drink and tip) offsets the *Michelin* protection. Both price and *Michelin* effects decline with food quality.

C. Investments

[19] There are no statistically significant differences between the mean prices of the two samples.
[20] However, since prices are self-reported, we cannot rule out that the price effect results from structural changes, i.e., due to a higher restaurant rating diners may substitute more expensive meal items for less expensive ones.

We cannot be *a priori* certain whether the improved quality consumer ratings of *Michelin*-reviewed restaurants for service and décor are due to the demand or the supply side. On the one hand, experts could have influenced consumer perception. As a result, the food and décor quality of *Michelin*-reviewed restaurants would then be seen in a better light and more appreciated than before. On the other hand, the improved perceived quality can also be due to actual restaurant investments.

Ideally, we would like to regress restaurant investments in décor and service, e.g., staff per meal served, or money spent on staff training or décor, on *Michelin* points. However, since these data are proprietary and not available to us, we rely on public information to test whether *Michelin*-reviewed restaurants in fact invested more than others.

In particular, we refer to *Wine Spectator's* Restaurant Wine List Awards program, which we already mentioned in Section II. *Wine Spectator*, the largest wine magazine in the nation, has offered restaurants to compete for the "Award of Excellence" since 1981. In order to apply, a restaurant has to pay a $250 entry fee and should submit its wine list along with its menu and information on the wines' storage conditions. *Wine Spectator* then selects the winners according to their merits (for more information see Wine Spectator, 2012). Winners can be in one of three categories. The *Award of Excellence* requires wine list offers of at least 100 selections. Higher achievements are honored with the *Best of Award of Excellence* (400+ selections) or the *Grand Award* (1,500+ selections). In 2003 (2006), 9 (5) NYC restaurants received the *Grand Award*, 28 (52) the *Best of Award of* Excellence and 128 (112) the *Award of Excellence*. Building a wine collection that is sufficient to meet *Wine Spectator's* standards in quantity and quality can be a substantial investment. For instance, the *Grand Award* winner restaurant *Veritas* has a wine list with more than 3,000 selections and an inventory of 75,000 bottles. At bottle prices ranging from $25 to $10,000, this is a multi-million dollar investment even without storage cost.

We, therefore, interpret the win of *Wine Spectator Awards of Excellence* as a restaurateur's willingness to invest in non-food ambience, which may serve as a good proxy variable for investments in décor and service.[21] In addition, since restaurants with extensive wine lists, in

[21] Chossat and Gergaud (2003) use comparable data to instrument *Gault-Millau* décor ratings.

many cases, also have special wine waiters (*sommeliers*) one may even expect direct wine list-induced service improvements.

Hence, we examine the impact of the *Michelin* treatment (i.e., being included in the Guide) on the tier of the wine list award (0 = no award, 1 = *Award of Excellence*, 2 = *Best Award of Excellence*, 3 = *Grand Award*), if any, and report the results in Table 12. We estimate a two-stage difference-in-differences model that is almost identical with the one employed for the *Zagat* quality assessments and the prices as reported in equation (2) and in Table 4, 6, and 8; we only refer to *Wine Spectator* scores as dependent variable. At the first stage we regress the *Mich* and *Mich*×*After* variables on the instruments and the exogenous variables; we then estimate the *Wine Spectator* scores in a second step. Since these scores are discrete ranks, we use a ML ordered probit model with bootstrapped standard errors as this equation contains generated regressors coming from the first step. At the bottom of Table 12 we display overidentifying restriction, underidentification and weak identification tests for the selected instruments.

Table 12 reports two variants of the model. The left column considers the *Michelin* treatment as a 0-1 dummy variable, the right column distinguishes between the *Michelin* stars awarded (0-1-2-3). In both variants we find significant treatment effects suggesting that Michelin-reviewed restaurants, in fact, are more likely to receive a *Wine Spectator* award for their wine list that do others. This seems to confirm our prior assumption that higher service and décor quality ratings of *Michelin*-reviewed restaurants are, in fact, based on restaurant investment rather than on mere consumer perception.

VI. Propensity Score Matching

Since our treatment group was not selected randomly it is *per se* difficult to isolate the treatment effect since the difference in outcomes between the two groups may be due to pre-treatment characteristics. We, therefore, compare our IV results with those we receive from propensity score matching (PSM).

PSM, as first published by Rosenbaum and Rubin (1983), is a two-step approach. In a first step, we employ a binary model to calculate the conditional probability of each observation to be assigned to the treatment group given its pre-treatment characteristics (propensity score).

Then, we match treated with untreated observations based on their respective propensity scores. This way we create a counterfactual situation and can evaluate the treatment effect for matches with (almost) identical pre-treatment characteristics.

Table 13 reports the probit equation results for the treatment assignment for the base years 2003 and 2004, respectively. Both models yield similar results and show that food and décor quality are the crucial selection determinants.

These propensity scores form the base for the PSM results presented in Table 14. Note that our calculations draw on the restricted paired sample, i.e., we excluded non-surviving restaurants as well as new births. When referring to the base year 2003, we find significant treatment effects on food quality, *Wine Spectator* awards, and, especially, on prices. Overall, when referring to the base year 2003, the PSM results lend further support to our results from Section V.

However, when referring to 2004 as the base year, we observe a few changes in coefficient size and significance. First, the food quality treatment effects is almost cut in half suggesting that *Michelin* reviewed restaurants did considerably improve their food quality just before the launch of the first *Michelin* guide. We find a similar pattern for prices. The drop in coefficient size and in particular in significance suggests substantial price increases for restaurants that were subsequently *Michelin*-reviewed. We thus confirm the results from the previous chapter and find positive announcement effect for food and prices. This also underlines our conclusion that *Michelin*-reviewed restaurants were on a different trend path than were non-reviewed ones.

This gives rise to the suspicion that expert scores are not only determined by food quality but also by framing variables such as price as already suggested by Gergaud et al. (2007) for the Paris *Michelin Guide* (see also Section II.D). To analyze this for the New York guide is, however, beyond the scope of our study.

VII. SUMMARY AND CONCLUSIONS

In this paper we analyze whether consumers' quality perception and/or producer investment is influenced by newly appearing expert opinion. We investigate this question by referring to restaurants in New York City. As the leading restaurant guide *Zagat* has rated New York City's restaurants since 1979 by surveying more than 30,000 restaurant goers per year. In 2005, with the first release of the red *Michelin Guide New York City*, *Zagat* faced a serious competition. In contrast to *Zagat*, *Michelin* relies on expert eaters. Employing a difference-in-differences approach we analyze whether consumer assessments (*Zagat* ratings) have responded to *Michelin* quality assessments. Employing a difference-in-differences model for 2003 and 2006 we find significant *Michelin*-induced perceived quality increases for food and décor. However, restaurants that only improved their décor but not their food quality were more likely to go out of business. Apparently, a *Michelin* review, which is *per se* good for a restaurant's odds of survival, opens up two strategy paths. Restaurants that improve their (perceived) food quality can further secure their survival, while restaurants that only improve their décor quality are less likely to survive.

When changing the base year from 2003 to 2004, i.e., when moving closer to the 2005 treatment, we find that the treatment effect on perceived food quality drops from a coefficient of 0.20 (base year 2003) to 0.15 (base year 2004). This is in line with an announcement effect and suggests that restaurants were on different trend paths even before the treatment. That is, between 2003 and 2004, *Michelin*-reviewed restaurants improved their food quality by more than non-reviewed venues leading to a smaller treatment effect when referring to the base year 2004. The opposite is true for décor. A lager treatment effect for the 2004 base year suggests that especially restaurants that were not *Michelin*-reviewed "décored up." Both improved décor quality as well as not being *Michelin*-reviewed contribute to a higher shut down likelihood.

We also find significant *Michelin*-induced price effects. Since we are interested in knowing which quality variable induced the price change we further restricted our sample to venues that only changed one quality variable (e.g., venues that only improved food while keeping décor and service constant). Our analysis suggests that improving food leads to price increases for non- reviewed restaurants. In contrast, prices of treated restaurants only respond to changes in décor and service. However, higher prices are not necessarily reliable success indicators. In fact, we find that higher prices are associated with a higher likelihood of going out of business, especially in the top food quality segment. This may be due to the close link between décor improvements and price increases.

In order to test whether the improved food and décor quality is based on real investments or on consumer-perception only, we examine each restaurant's wine list investment by referring to *Wine Spectator* restaurant wine list awards. We assume *Wine Spectator* awards to be a good proxy variable for a restaurateur's willingness to invest in non-food ambience. Our analysis shows that *Michelin*-reviewed restaurants are significantly more likely to receive wine lists awards than do others.

When contrasting the difference-in-differences approach with a propensity score matching (PSM) model we find, generally, very similar results.

Overall, our results suggest that expert opinion on the New York City restaurant market opens up two paths for restaurants, improving food quality or improving décor only. Both strategies are costly and may raise prices. However, the market is more likely to accept food-induced price increases than non-food-induced ones. All other things equal, décor and service oriented restaurants exhibit lower survival rates than food-focused venues.

REFERENCES

Anderson, M. and Magruder, J. (2012). Learning from the crowd: regression discontinuity estimates of the effects of an online review database. *Economic Journal*, 122(563), 957-989.

Ashenfelter, O., Goldstein, R. and Riddell, C (2010). Do expert ratings measure quality? The case of restaurant wine lists. Presentation at the 4th Annual Conference of the American Association of Wine Economists at UC Davis. Mimeo.

Bagwell, K. and Riordan, M.H. (1991). High and declining prices signal product quality. *American Economic Review*, 81, 224-239.

Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.

Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3), 797-817.

Bauwens, L. and Ginsburgh, V. (2000). Art experts and auctions. Are pre-sale estimates unbiased and fully informative? *Louvain Economic Review*, 66, 131–144.

Berry, S. and Waldfogel, J. (2010). Product quality and market size. *Journal of Industrial Economics*, 58(1), 1-31.

Becker, G.S. (1991). A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy*, 99(5), 1109-1116.

Bikhchandani, S, Hirshleifer, D. and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992-1026.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1998). Learning from the behavior of others: conformity, fads, and informational Cascades. *Journal of Economic Perspectives*, 12(3), 151-170.

Cai, H., Chen, Y. and Fang, H. (2009). Observational learning: evidence from a randomized natural field experiment. *American Economic Review*, 99(3), 864-882.

Caves, R.E. and Greene, D.P. (1996). Brands' quality levels, prices, and advertising outlays: empirical evidence on signals and information cost. *International Journal of Industrial Organization*, 14, 29-52.

Chossat, V. and Gergaud, O. (2003). Expert opinion and gastronomy: the recipe for success. *Journal of Cultural Economics*, 27, 127–141.

Cuozzo, S. (2005a). A pinch of bias leaves bitter taste. *New York Post*, November 2.

Cuozzo, S. (2005b). The idiot's guide; Michelin diner book bites. *New York Post*, November 7.

Curry, D.J. and Riesz, P.C. (1988). Prices and price/quality relationships: a longitudinal analysis. *Journal of Marketing*, 52, 36-51.

Davis, M. (2012). A taste of New York: restaurant reviews, food discourse, and the field of gastronomy in America. (PhD Thesis, New York University, 2009). American Association of Wine Economists AAWE, *AAWE Working Paper* No. 106.

Deutsch, M. and Gerard, H.B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 51, 629-636.

Dubois, P. and Nauges, C. (2010). Identifying the effect of unobserved quality and expert reviews in the pricing of experience goods: empirical application on Bordeaux wine. *International Journal of Industrial Organization*, 28, 205-121.

Fabricant, F. (2005a). The tire man eats New York. *New York Times*, February 23.

Fabricant, F. (2005b). Is New York worth a trip? Oui. *New York Times*, November 2.

Ferguson, P.P. (2008). Michelin in America. *Gastronomica*, 8(1), 49-55.

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

Gergaud, O., Montano, L. and Verardi V. (2007). Stardust over Paris gastronomic restaurants. *Journal of Wine Economics*, 2(1), 24-39.

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

Ginsburgh, V. and van Ours, J. (2003). Expert opinion and compensation: evidence from a musical competition. *American Economic Review*, 93, 289-296.

Griliches, Z. (1957). Hybrid corn: an exploration in the economics of technical change. *Econometrica*, 25(4): 501-522.

Hadj Ali, H., Lecocq, S. and Visser, M. (2008). The impact of gurus: Parker grades and *en primeur* wine prices. *Economic Journal*, 118, F158-F173.

Heffetz, O. and Shayo, M. (2009). How large are non-budget-constraint effects of prices on demand? *American Economic Journal: Applied Economics*, 1(4), 170-199.

Hilger, J., Rafert, G. and Villas-Boas, S. (2011). Expert opinion and the demand for experience goods: an experimental approach in the retail wine market. *Review of Economics and Statistics*, 93(4), 1289-1296.

Hodgson, R.T. (2008). An examination of judge reliability at a major U.S. wine competition. *Journal of Wine Economics*, 3(2), 105-113.

Hodgson, R.T. (2009). An analysis of the concordance among 13 U.S. wine competitions. *Journal of Wine Economics*, 4(1), 1-9.

Johnson, C. and Surlemont, B. (2005). Behind the stars : a concise typology of Michelin restaurants in Europe. Cornell Hotel and Restaurant Administration Quarterly, 46(2), 170-187.

Kamenica, E. (2008). Contextual inference in markets: on the informational content of product lines. *American Economic Review*, 98(5), 2127–2149.

Krummert, B. (2006). Making a list, and checking it twice. *Restaurant Hospitality*, 30, (October 31, 2006).

Kurutz, S. (2005). Neighborhood report: Union Square; she’s a belle of the city, but the French are blasé. *New York Times*, November 13.

Liu, Y. (2006). Word of mouth for movies: its dynamics and impact on box office revenues. *Journal of Marketing*, 70(3), 74-89.

McFadden, D.L. and Train, K. (1996). Learning from self and others. *Journal of Political Economy*, 104(4), 684-703.

Michelin Travel Publications (2005). New York City 2006: selection of restaurants and hotels. Greenville, SC : Michelin Travel Publications

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94, 796-821.

Moretti, E. (2011). Social learning and peer effects in consumption: evidence from movie sales. *The Review of Economic Studies*, 78(1), 356-393.

Morris, S. and Shin, H.S. (2002). Social value of public information. *American Economic Review*, 92(5), 1521-1534.

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

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, 81, 729-754.

North, A., Hargreaves, D. and McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology*, 84(2), 271-276.

Plassmann, H., O'Doherty, J., Shiv, B. and Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences PNAS*, 105(3), 1050-1054.

Reinstein, D.A. and Snyder, C.M. (2005). The influence of expert reviews on consumer demand for experience goods: a case study of movie critics. *Journal of Industrial Economics*, 53(1), 27-51.

Riesz, P.C. (1978). Price versus quality in the market place. *Journal of Retailing*, 58, 15-28.

Rosenbaum, P. and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rosenman, R.R. and Wilson, W.W. (1991). Quality differentials and prices: are cherries lemons? *Journal of Industrial Economics*, 39(6), 649-658.

Salganik M.J., Dodds, P.S. and Watts, D.J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311, 854-856.

Schmalensee, R. (1978). A model of advertising and product quality. *Journal of Political Economy*, 86, 485-503.

Schnabel, H. and Storchmann, K. (2010). Prices as quality signals: evidence from the wine market. *Journal of Agricultural & Food Industrial Organization*, 8, Article 2.

Shapiro, C. (1983). Premiums for high quality products as returns to reputation. *The Quarterly Journal of Economics*, 98, 659-680.

Snyder, W. and Cotter, M. (1998). The Michelin Guide and restaurant pricing strategies. *Journal of Restaurant & Foodservice Marketing*, 3(1), 51-67.

Stock, J.H. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D.W.K. Andrews and J.H. Stock, eds. Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge: Cambridge University Press, 80–108.

Tellis, G.J. and Wernerfelt, B. (1987). Competitive price and quality under asymmetric information. *Marketing Science*, 6, 240-253.

Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, 63, 1-22.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 30(211), 453-458.

Wansink, B., Just, D.R. and Payne, C.R. (2009). Mindless eating and healthy heuristics for the irrational. *American Economic Review, Papers & Proceedings*, 99(2), 165-169.

Wine Spectator (2012). Restaurant wine lists awards. Entry Guide 2013. http://www.winespectator.com/group/show/id/rest_awards_entry_guidelines

Wolinsky, A. (1983). Prices as signals of product quality. *Review of Economic Studies*, 50, 647-658.

Zagat Survey (2003). New York City Restaurants 2004. New York: Zagat Survey.

Zagat Survey (2004). New York City Restaurants 2005. New York: Zagat Survey.

Zagat Survey (2006). New York City Restaurants 2007. New York: Zagat Survey.

Figure 1

Michelin Restaurants in New York City