

Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

The Impact of Increase in Minimum Wages on Consumer Perceptions of Service: A Transformer Model of Online Restaurant Reviews

Dinesh Puranam , Vrinda Kadiyali , Vishal Narayan

To cite this article:

Dinesh Puranam , Vrinda Kadiyali , Vishal Narayan (2021) The Impact of Increase in Minimum Wages on Consumer Perceptions of Service: A Transformer Model of Online Restaurant Reviews. Marketing Science 40(5):985-1004. <https://doi.org/10.1287/mksc.2021.1294>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Impact of Increase in Minimum Wages on Consumer Perceptions of Service: A Transformer Model of Online Restaurant Reviews

Dinesh Puranam,^a Vrinda Kadiyali,^b Vishal Narayan^c

^aMarshall School of Business, University of Southern California, Los Angeles, California 90089; ^bJohnson Graduate School of Management, Cornell University, Ithaca, New York 14850; ^cNUS Business School, National University of Singapore, Singapore 119245

Contact: puranam@marshall.usc.edu,  <https://orcid.org/0000-0002-2373-9591> (DP); kadiyali@cornell.edu (VK); bizvn@nus.edu.sg,  <https://orcid.org/0000-0001-7230-3084> (VN)

Received: October 21, 2018

Revised: March 5, 2020; November 20, 2020; January 24, 2021

Accepted: January 28, 2021

Published Online in Articles in Advance: August 25, 2021

<https://doi.org/10.1287/mksc.2021.1294>

Copyright: © 2021 INFORMS

Abstract. We study the impact of a mandated increase in minimum wages on consumer perceptions of multiple dimensions of service quality in the restaurant industry. When faced with higher minimum wages, firms might reduce the number of employees, resulting in poorer consumer service. Alternatively, higher-paid workers might be more motivated to improve consumer service. Using a combination of human annotation and several transformer models, we estimate the incidence of discussion of several service quality attributes (and their valence) in a textual data set of 97,242 online reviews of 1,752 restaurants posted over two years. We exploit a natural experiment in the County of Santa Clara, California, wherein only the city of San Jose legislated a 25% minimum wage increase in 2013. By comparing restaurant reviews in San Jose with those of synthetic controls, we find an improvement in the perceived service quality of San Jose restaurants. Specifically, we find reduced negative discussion of the courtesy and friendliness of workers. This decrease is present in independent restaurants and not in chains. This finding appears to be consistent with agency theory–based predictions of greater incentives to improve service in independent restaurants. We discuss alternative mechanisms for our results. We also discuss implications for consumers, restaurants, and policy makers.

History: Puneet Manchanda served as the senior editor for this article.

Funding: This work was supported by Johnson Graduate School of Management (Cornell University), Marshall School of Business (University of Southern California) and National University of Singapore.

Supplemental Material: Data and the online appendix are available at <https://doi.org/10.1287/mksc.2021.1294>.

Keywords: service quality • minimum wages • text analysis • transformer models • natural experiments • agency theory

1. Introduction

Most countries in the world, including the United States, China, Japan, and Germany, have long legislated a minimum hourly wage. In the United States, the new administration in 2021 is said to be interested in pursuing a \$15/hour federal minimum wage (Selyukh 2021). The goal of such legislation is to protect workers against unduly low pay and to potentially reduce inequality and poverty (International Labor Organization 2015). Across several industries, geographies, and time periods, minimum wage legislation has been shown to impact inequality (Autor et al. 2016), employment levels (Aaronson and French 2007), firm value (Bell and Machin 2018), other wages (Grossman 1983), and nonwage benefits (Bhorat et al. 2014). Despite this rich body of research, there is a lack of consensus about the directions and magnitudes of these effects. Across academia, media, policy makers, and the general public, there are strong differences in opinion about the advantages and drawbacks of such legislation. In contrast to the positive view that

minimum wages reduce poverty and inequality, improve living standards, and boost morale, opponents opine that setting minimum wages increases unemployment especially of unskilled and inexperienced workers and increases the cost of doing business.

Even though minimum wage legislation has been intensely scrutinized and debated, there is very little academic research or business-press discussion on how such legislation might impact the behavior or experiences of consumers who interact with minimum wage workers. We investigate the impact of increases in minimum wage increases by thousands of firms on various dimensions of the perceived service quality of those firms, as captured by the opinions of the firms' consumers. Especially, we view these effects as another angle to consider when measuring the economic impact of a minimum wage.

Our empirical research context is the U.S. restaurant industry. Direct labor costs are about 30% of the total operating costs of U.S. restaurants, and about 33% of restaurant workers are paid within 10% of the minimum

wage. Two-thirds of workers in the United States earning the minimum wage or less in 2016 were employed in service occupations, mostly in food preparation and serving-related jobs.¹ For these reasons, several researchers have focused on this industry to measure minimum wage effects (see Section 2 for details). We measure service quality by combining human annotation and the estimation of several transformer models on a data set of 97,242 online reviews of 1,752 restaurants posted over two years (see Section 3 for details). Traditional approaches to measure service quality, such as surveys and focus groups, might be expensive, time-consuming, and potentially subject to recall biases and demand effects (Netzer et al. 2012). Unlike such approaches, which rely on primary data collected over a short period of time and for a limited number of firms, our data are available over many months and for a large number of firms. Therefore, relative to primary data collection, our approach is especially useful for studying the impact of temporally distant events (such as past legislations) by comparing periods before and after such events.²

To analyze text, we use bidirectional transformer models (Devlin et al. 2019); to the best of our knowledge, these models are new to marketing. Traditional approaches of text classification in marketing have viewed text as its component words. Support vector machines (Vapnik 1995) and latent Dirichlet allocation (Blei et al. 2003) are inherently bag-of-words approaches that ignore word order. Recent advances in machine learning have modeled the probability of observing a word given its context words (or vice versa). These vector representations outperform previous approaches for text classification problems. The convolutional neural network (CNN) model for text classification (Kim 2014) has been very successful in improving prediction accuracy by using these representations. Timoshenko and Hauser (2019) and Liu et al. (2019a) effectively apply this technique in marketing. Our approach is based on more recent advances that replace CNNs with transformers (Vaswani et al. 2017). The word representations are learned by modeling two objective functions. The first objective function is derived from a masked language model. The model randomly hides (masks) some of the words from the input text and attempts to predict the masked word using only words that appear in its context. The model incorporates words before (left of) and after (right of) the focal word using a bidirectional transformer and accounts for word position explicitly. The second objective function is to model the probability of the next sentence. This results in word-vector representations that are context dependent. The gains in predictive performance come from the dual objective functions, size of word vectors, and incorporation of subword-level data.

Specific to our context, we first present 12,000 randomly sampled reviews for human annotation. The labeling for each review consists of mutually exclusive valence categories (positive, negative, and absent) for the discussion of each of four service attributes—leading to 12 labels per review. We then estimate several text models to predict each of these 12 labels. For a valence–attribute combination (say, “positive courtesy/friendliness”), the prediction task is to correctly label each review as discussing the courtesy or friendliness of the staff positively (or negatively). We use the best performing model (based on several performance measures) for each of the 12 labels to score all the reviews in our data set.

Given our interest in multiple service quality dimensions, we rely on the service quality literature to conceptualize and analyze several service attributes. For defining specific attributes, we consider SERVQUAL (Parasuraman et al. 1988, 1991), a general framework for service quality, and SERVPERF (Cronin and Taylor 1992) a framework for restaurants. Given our interest in wage hikes of low-wage workers, we focus on those aspects of service quality that typically involves these low-wage workers. Because frontline workers are responsible for the cleanliness of the restaurants, we investigate this attribute separately. We also investigate wait times; these include wait times for seating, receiving food, and receiving the check. Third, we investigate customer perception of workers with whom they interact, in terms of the courtesy and friendliness of workers. For robustness, we also extract discussions of price from consumer reviews.

Estimation of causal effects of minimum wage increases is difficult in part because of a lack of exogenous variations in wages (i.e., wage increases might be correlated with service quality due to unobservables), unavailability of high-quality instruments, and infeasibility of conducting field experiments with meaningful wage increases. To deal with this, we exploit the following natural experiment. The state of California requires employers to pay full state minimum wage before tips³ to tipped employees. Exclusions based on the number of employees of the business are not permitted. California is divided into 58 counties, one of which is the County of Santa Clara (henceforth, CSC). CSC has a population of 1.92 million (2015 Census), is home to Silicon Valley, and is estimated to have the third-highest per-capita gross domestic product across all cities in the world.⁴ CSC comprises eight cities: San Jose, Cupertino, Los Altos, Milpitas, Mountain View, Palo Alto, Santa Clara, and Sunnyvale. Of these cities, San Jose, a city with a population of over 1 million, increased the hourly minimum wage by 20% from \$8 to \$10 effective March 11, 2013.⁵ This is due to an ordinance that aims to raise the minimum wage to \$15 over

time. The remaining seven cities did not witness any change from the statewide rate of \$8 per hour in 2013.

Because the scope of the wage increase was limited to San Jose, we analyze data from San Jose and the remaining seven cities of CSC separately; that is, the remaining cities serve as a useful contrast and as a geographically proximate control group (see the paper by Allegretto and Reich (2018), who also use the same treatment group and control group). We use reviews posted for restaurants in the remaining seven cities to construct synthetic control groups (Abadie et al. 2010, Tirunillai and Tellis 2014, Xu 2017, Pattabhiramaiah et al. 2019). We compare reviews posted in San Jose for 12 months before and 12 months after the wage increase, with reviews in the “synthetic control unit” computed as a weighted combination of zip codes in control cities. This approach naturally satisfies the parallel trend assumption required for causal inference. The synthetic control analysis employs model-based measures of perceived service quality as dependent variables. The 12-month window of analysis leads to a data set of 32,244 reviews from San Jose before the wage increase and 35,127 reviews from San Jose after. These represent reviews of 1,230 restaurants. The seven control cities of CSC include 14,537 reviews before the wage increase and 15,334 reviews after. These represent reviews of 522 restaurants across seven cities.

Geographical proximity reduces the possibility of biased estimates due to unobservable confounds that might affect San Jose differently from the other cities in CSC. Furthermore, because our approach uses the pretreatment dependent variables for matching, it naturally conditions on both observables and unobservables (Lovett et al. 2019). In sum, we combine current methods in both text analysis and causal inference of treatment effects from panel data.

1.1. Research Questions and Agency Theory

Our first research question is whether a minimum wage increase leads to an increase or decrease in consumer opinions of service quality. The direction of this effect is a priori unclear. A restaurant might react to forced increases in minimum wages by reducing the number of employees (McBride 2017). Indeed, in 2016, McDonald’s announced the nationwide rollout of automated kiosks to replace cashiers as a way to reduce costs due to minimum wage increases (Rensi 2016). Several other chain restaurants are currently contemplating replacing minimum wage workers with robots, explicitly to deal with rising labor costs (Taylor 2018). A reduction in employee strength might lead to lower service quality. For example, with fewer employees, it might take more time for an order to be transported from the kitchen to the dining area. A kiosk, or other substitutes, might not be perceived to be as courteous and responsive as a waiter. This line of logic suggests a

negative effect of minimum wage increases on service quality. On the other hand, greater wages might lead to greater motivation of employees, and might also increase the restaurant’s ability to hire more productive or experienced workers. It has been found in at least one industry study that when employers transitioned to paying a living wage to their employees, they experienced “significantly lower rates of staff turnover, reputational benefits, reduction in sick leave, better motivated staff and an increase in productivity” (Rouse 2013)⁶ A majority of employers noticed better quality of work, and employees agreed that their work had improved after the wage increase. This suggests a positive impact on service quality. Therefore, the direction of impact on service quality is an empirical question.

We do not expect the impact of wage increases on service quality to be the same across all restaurants. An important aspect of the restaurant industry is the difference in ownership structures of chain and independent restaurants, and resulting differences in incentives and managerial abilities (Brickely and Dark 1987, Lafontaine 1992, Shepard 1993, Lafontaine and Shaw 2005). Our second research question is whether the impact of the minimum wage increase on consumer opinions of service quality differs across chain and independent restaurants. Chain restaurants have two forms of ownership: they are either corporate owned or owned by a franchisee.⁷ Independent restaurants are more likely to be owned and/or supervised by the founder. Per agency theory, corporate-owned chain outlets have the least incentive and ability to monitor service outcomes because the store manager is a salaried employee (who herself might also face incentive issues). Franchisee outlets keep residual profits and therefore are more motivated to improve service outcomes. They also have a greater ability to observe employees if they are locally based. However, to enable ease of franchising and to maintain brand quality, many national chains are built on process-driven cultures with little room for local variance in quality. In contrast, industry reports suggest independents’ success is affected by providing a better customer experience via service and differentiation (e.g., varied menu) than the more bland but reliable national chains.⁸ Independents have a greater ability to monitor employees (on visible attributes of service) compared with chains. They also have a higher incentive to monitor for higher service quality provision as a requirement for higher-paid workers, especially if accompanied by higher pass-through prices. Given their smaller size and differentiated presence, pricing changes might be easier to implement with higher wages. Based on all these insights, we distinguish between chain and independents in our analysis of minimum wage effects on perceived service quality. We expect greater effects on the service quality of independents.⁹

These questions have important implications for several stakeholders including marketers, consumers, and policy makers. A positive effect on service quality would provide an additional rationale for increasing minimum wages—one that is unrelated to employment or living standards of employees, but related to consumer satisfaction and welfare. On the other hand, a negative effect would provide robust evidence consistent with the notion that minimum wage employees are being replaced with less service-oriented substitutes. A positive effect of increased wages on service quality should give pause to managers considering replacing minimum wage employees, and provide a rationale based on business outcomes, for low-paid employees to demand wage increases. Finally, in the face of evidence that minimum wage increases could lead to price increases (Allegretto and Reich 2018), a positive effect on service quality could incentivize consumers to remain loyal to restaurants affected by the legislation, despite having to pay greater prices.

We find that negative discussion of the courtesy and friendliness of workers decreases after the wage increase in San Jose restaurants as compared with that in the control units. This change in topic discussion suggests improved service quality. Consistent with the agency theory discussed above, this service quality improvement is restricted to independent restaurants and does not extend to chain restaurants in that city. Next, we discuss the relevant streams of literature and how our work is positioned relative to each stream. Section 3 presents the textual data. In Section 4, we discuss the text model and the synthetic control analysis. Section 5 presents the results from the model and their implications. Section 6 concludes.

2. Literature Review

Our work is related to the literature on the impact of minimum wages, customer satisfaction, text mining, and service quality. The impact of minimum wage increases on employment is a politically fraught and economically complex question. We discuss a small set of papers to highlight the key issues. There are several theory models (e.g., Aaronson and French 2007) on the impact of the minimum wage increase on employment and total earnings of employed workers. These models show the employment effect varies as a function of labor and product market competition, factor substitutability, and other factors. There is a large stream of empirical work in this area with conflicting findings and a fierce debate on the reasons for these conflicting findings. Especially, there are debates on the appropriate control group when estimating the effect of the “treatment” of minimum wage increases, and controlling for heterogeneity if pooling units subjected to the regulation, across various geographies. As mentioned,

several studies have analyzed the restaurant industry with its heavy reliance on minimum-wage workers. Card and Krueger (1994) find that a minimum wage increase leads to higher employment in limited service fast-food chains. Dube et al. (2010) examine both limited-service and full-service restaurants. They also find positive employment and earnings effects. On the other hand, several researchers find negative employment effects. Among these are Neumark and Wascher (2000) and Aaronson et al. (2008), who look at limited-service restaurants. The employment effect of minimum wages is not the focus of our study. However, we draw on this literature for its distinction between types of restaurants and the importance of finding the right control group to infer causality.

Our work is more closely related to two other papers on the impact of minimum wage increases on marketing outcomes. Allegretto and Reich (2018) estimate whether minimum wage increases are passed on to consumers via increased price. They invoke the estimate of Okrent and Alston (2012) that demand for restaurant products is relatively price inelastic (-0.71). Therefore, restaurants faced with forced increases of minimum wages might prefer to not reduce employment but instead increase prices by a small amount. The alternative of reducing employment might lead to a greater reduction of profits. Using internet-based menus of restaurants tracked before and after wages increases, they find that prices increased as a result of minimum wage increases. This result holds even in border areas between higher and lower minimum wage cities, where competition might have been expected to bid away price differentials. They find that the price increase magnitude is similar to the previous markups. This implies no negative employment effects. They also do not find any competitiveness effects, based on their analysis of border areas.

More recently, Chakrabarti et al. (2020) studied the impact of minimum wage increases on a marketing-related outcome. They examined whether there was a change in restaurant hygiene (using hygiene violations data) due to the minimum wage increase. They hypothesized that higher labor costs lead to employment cuts and fewer workers, and/or cost cutting in other ways. They found an increase in less severe hygiene violations, but not in the most severe violations. They concluded that consumers face less hygienic conditions when minimum wages increase. Unlike Allegretto and Reich (2018) and Chakrabarti et al. (2020), we do not estimate price pass-through or health violations. Our interest is in consumer perceptions, and for that, we utilize data from thousands of consumer reviews. To the extent that service perceptions might be confounded by price perceptions, we jointly estimate the probability of reviews discussing price. We view our analysis as being broader and

more relevant to marketing research with its focus on service attributes. To the best of our knowledge, ours is the first paper to study the effect of any wage change on consumer behavior or experience.

The closest stream of research in marketing to our work is on the relationship between firm routines and customer satisfaction. Anderson et al. (1997) discuss whether customer satisfaction and firm costs and productivity are positively or negatively correlated. It is possible that improving customer satisfaction reduces costs such as those of returns, new customer acquisition, etc. Alternatively, improving customer satisfaction might increase costs, for example, of customer support, information technology, etc. They find that for services, in particular, customer satisfaction comes with increased costs. Note that for our context, an increase in minimum wages will increase per-worker cost. The impact on total cost is unclear because firms might reduce employment, so whether minimum wage increase leads to greater or lower service quality cannot be predicted based on this stream of research. In fact, even if total labor costs increase, higher costs are not a sufficient condition for higher customer satisfaction. So the direction of the effect of minimum wage increases is an open empirical question.

In related work, researchers have examined the link between employee satisfaction (especially front-line employees) and business outcomes, including customer satisfaction (e.g., Schneider et al. 1998, Kamakura et al. 2002, Simon et al. 2009). The predominant conclusion is a positive correlation between employee and customer satisfaction. Zablah et al. (2016) find that although attention has been focused on how satisfied front-line employees can improve customer satisfaction, there might be larger effects from satisfied customers to satisfied front-line employees. In our empirical context, it is possible that wage increases are more than offset by increased work demands, resulting in lower employee satisfaction. It is also possible that higher wages result in the hiring of more motivated current or new employees (e.g., via the efficiency wage theory in labor economics; Shapiro and Stiglitz 1984). This might result in more satisfied workers. Therefore, the link between minimum wage increase and customer satisfaction (or customer perceptions of quality) is perhaps more complex than the relationship between employee satisfaction and customer satisfaction. We differ from these papers as follows: First, we are able to extract via text analysis multiple dimensions of customer perceptions of service quality. Second, our natural experiment allows us to try to isolate causal effects on customer perceptions of service quality; we do not have confounds of reverse causality effect of customer perceptions on higher performance of front-line employees. Last, although survey data have been frequently used in the services literature, we

employ user-generated content to study this services marketing question with policy implications.

Another stream of directly relevant research is text mining methods in marketing. Our paper is related to the literature on extracting useful information from large masses of text of reviews (Archak et al. 2011, Ghose et al. 2012). Early papers have used a bag-of-words approach, where word order does not matter. Examples include Lee and Bradlow (2011), Netzer et al. (2012), Tirunillai and Tellis (2014), and Puranam et al. (2017). Recent papers have relaxed the bag-of-words approach. For example, Timoshenko and Hauser (2019) and Liu et al. (2019a) apply deep learning models that relax the bag-of-words assumption and make use of vector representations of words for improved prediction accuracy. We employ a recently developed class of deep learning models from computer science: transformers (e.g., Devlin et al. 2019). In contrast to recurrent neural networks (that rely on sequential processing and are not bidirectional) and CNNs from word embeddings (that ignore word order), transformers incorporate position and context information to generate a vector representation for each word. The approach allows each word to be related to every other word in a document, while simultaneously tracking their relative positions, leading to improved performance over previous models.

3. The Natural Experiment and the Data

We discuss the institutional details of the natural experiment setting. Although the federal government has mandated a minimum hourly wage of \$7.25 since 2009, the state of California has typically legislated a greater minimum wage. From January 1, 2008, until July 1, 2014, California has had a minimum wage of \$8.¹⁰ In San Jose, the impetus for an increase to this wage originated among sociology students of San Jose State University in 2012. These students worked with community leaders to pass a citywide minimum wage ballot item (Measure D) in November 2012 that increased the city's minimum wage to \$10. A study created for the California Restaurant Association estimated that Measure D would lead San Jose employers to cut 900 to 3,100 jobs and that it would cost these employers \$88 million to \$96 million annually in terms of wage hike costs and added employee benefits (Seipel 2012). The measure was highly contested, with considerable opposition from restaurants, the city council, and the mayor. Nonetheless, the ballot measure received 59% of the vote and went into effect in March 2013. It is estimated that the wages of 24,000 workers, including 7,100 hotel and restaurant workers, increased by 25% because of this policy.

Our substantive interest is in estimating the effect of this wage increase. As mentioned, for this purpose, we conduct synthetic control analysis on all reviews

Table 1. Summary Statistics of Reviews of San Jose Restaurants

	12 months before wage increase			12 months after wage increase			Total
	Independent	Chain	All	Independent	Chain	All	
Number of reviews	23,641	8,603	32,244	25,793	9,334	35,127	67,371
Number of unique reviewers	13,046	5,716	15,934	15,081	6,615	18,603	30,494
Number of restaurants reviewed	721	464	1,185	744	473	1,217	1,230
Rating	3.65 (1.23)	3.39 (1.31)	3.58 (1.25)	3.65 (1.26)	3.27 (1.38)	3.55 (1.31)	3.57 (1.28)
Review length in words	61.99 (45.56)	60.54 (45.36)	61.6 (45.51)	58.53 (44.37)	56.66 (42.97)	58.03 (44.01)	59.74 (44.77)
Number of mentions: Useful	1.2 (2.7)	0.99 (2.1)	1.14 (2.55)	1.18 (2.52)	0.97 (1.98)	1.12 (2.39)	1.13 (2.47)
Number of mentions: Funny	0.58 (2.18)	0.57 (1.83)	0.58 (2.09)	0.56 (1.92)	0.52 (1.52)	0.55 (1.82)	0.56 (1.96)
Number of mentions: Cool	0.58 (2.18)	0.49 (1.63)	0.56 (2.05)	0.57 (1.97)	0.47 (1.48)	0.54 (1.85)	0.55 (1.95)

of restaurants located in CSC in our data set that were posted 12 months before and 12 months after the date of the wage increase. We restrict the synthetic control analysis to only those restaurants for which (a) at least one review was posted in the 12 months before that treatment and (b) at least one review was posted in the 12 months after that treatment. In this manner, we control for any potential effect of the wage hike on restaurant entry and exit.

We now discuss the appropriateness of using the other seven cities of CSC as control units. These are geographically close to San Jose and are mostly urban, much like San Jose. More importantly, wages across San Jose and other cities are similar. For the year 2013, weekly wages were \$361 in San Jose and \$394 in other cities (Allegretto and Reich 2018). Weekly wages in San Jose and other cities for limited-service restaurants averaged \$312 and \$319, respectively. Weekly wages in San Jose and other cities for full-service restaurants averaged \$400 and \$435, respectively. Furthermore, both areas experienced similar unemployment trends (for a more detailed comparison, see Allegretto and Reich 2018). Because the impetus for minimum wage increase in San Jose was localized, we do not think unobserved factors are a problem in our research context. Also, by ensuring parallel pretreatment trends in the treatment and control group, we condition on both observables and unobservables.

Our primary source of data is a leading website that publishes online reviews of local businesses and is commonly used by consumers in the United States to post and view reviews of restaurants. Per Alexa.com, an independent research firm, the website was among the top 70 most popular websites in the United States in July 2014, the time period of the data. By the end of 2017, the website had well over 100 million reviews. For the purpose of model estimation, we utilize data on all restaurant reviews posted on this website in the 12 months before and the 12 months after March 11, 2013, the date of the wage increase.¹¹

We now describe reviews of restaurants in San Jose. Means of ratings (on a five-point scale) decreased slightly from 3.58 before the wage increase to 3.55 after.

However, because ratings are a composite metric capturing several constructs, not just service quality, this trend is not particularly insightful for our research context. The website allows readers of a review to indicate whether it is “useful,” “funny,” and “cool.” On average, a review posted before the wage increase received 1.14 mentions of being useful, 0.58 mentions of being funny, and 0.56 mentions of being cool. After the wage increase, the average (per review) number of mentions of “useful,” “funny,” and “cool” are 1.12, 0.55, and 0.54, respectively. Similar to ratings, it is unclear whether these metrics capture service quality in any discernible way. Comparing summary statistics with reviews in CSC cities other than San Jose, we find that ratings, restaurant ratings, review lengths, and indication of whether reviews are useful, funny, and cool are similar across the treatment and control group, both before and after the wage increase. Detailed summary statistics appear in Tables 1 and 2.

4. Human Annotation, Text Model, and Synthetic Controls

We first describe the process of human annotation, followed by the text analysis model, and then the synthetic control analysis that employs output from text analysis as the dependent variable.

4.1. Human Annotation

Service quality is known to be a multidimensional construct. Based on the service quality literature, we are interested in three dimensions of service quality (cleanliness of interiors; courtesy and friendliness of the staff; wait times for seating, receiving food, and check). Additionally, we are also interested in measuring price perceptions. For each of these four attributes, our objective is to obtain robust quantitative review-level estimates of whether the attribute is discussed positively, negatively, or not at all. This leads to 12 estimates per review. Descriptors of service quality are not discussed very frequently in our data. This is typical for online reviews. Puranam et al. (2017) find that as many as 200 topics are discussed in a set of

Table 2. Summary Statistics of Reviews of Restaurants in CSC Cities other than San Jose

	12 months before wage increase			12 months after wage increase			Total
	Independent	Chain	All	Independent	Chain	All	
Number of reviews	10,889	3,468	14,537	11,793	3,541	15,334	29,691
Number of unique reviewers	7,272	2,805	8,937	8,143	2,911	9,950	17,101
Number of restaurants reviewed	324	179	503	333	184	517	522
Rating	3.55 (1.24)	3.38 (1.33)	3.51 (1.26)	3.55 (1.28)	3.24 (1.38)	3.47 (1.31)	3.49 (1.29)
Review length in words	61.59 (46.23)	61.83 (47.15)	61.65 (46.45)	59.33 (45.87)	56.72 (41.22)	58.72 (44.85)	60.14 (45.66)
Number of mentions: Useful	1.17 (2.41)	1.01 (2.2)	1.13 (2.36)	1.23 (2.57)	0.94 (1.94)	1.17 (2.44)	1.15 (2.41)
Number of mentions: Funny	0.56 (1.8)	0.5 (1.7)	0.55 (1.78)	0.55 (1.8)	0.47 (1.54)	0.53 (1.75)	0.54 (1.76)
Number of mentions: Cool	0.55 (1.72)	0.45 (1.55)	0.52 (1.68)	0.52 (1.8)	0.41 (1.45)	0.49 (1.73)	0.51 (1.71)

restaurant reviews from New York City, with the most discussed topic accounting for less than 5% of the overall corpus. In our data set, several topics such as “clean,” and its antonym “unclean,” are rarely discussed. Typical topic models used in marketing are not very well suited to investigate the occurrence of rare topics of interest. To demonstrate this, we estimated two topic models from the marketing literature: latent Dirichlet allocation and nonnegative matrix factorization. We find that none of the estimated topics cleanly captures service quality. Discussion of service quality, as estimated from these models, is spread across several topics and is absent in other topics. This highlights the need for an alternate modeling approach.

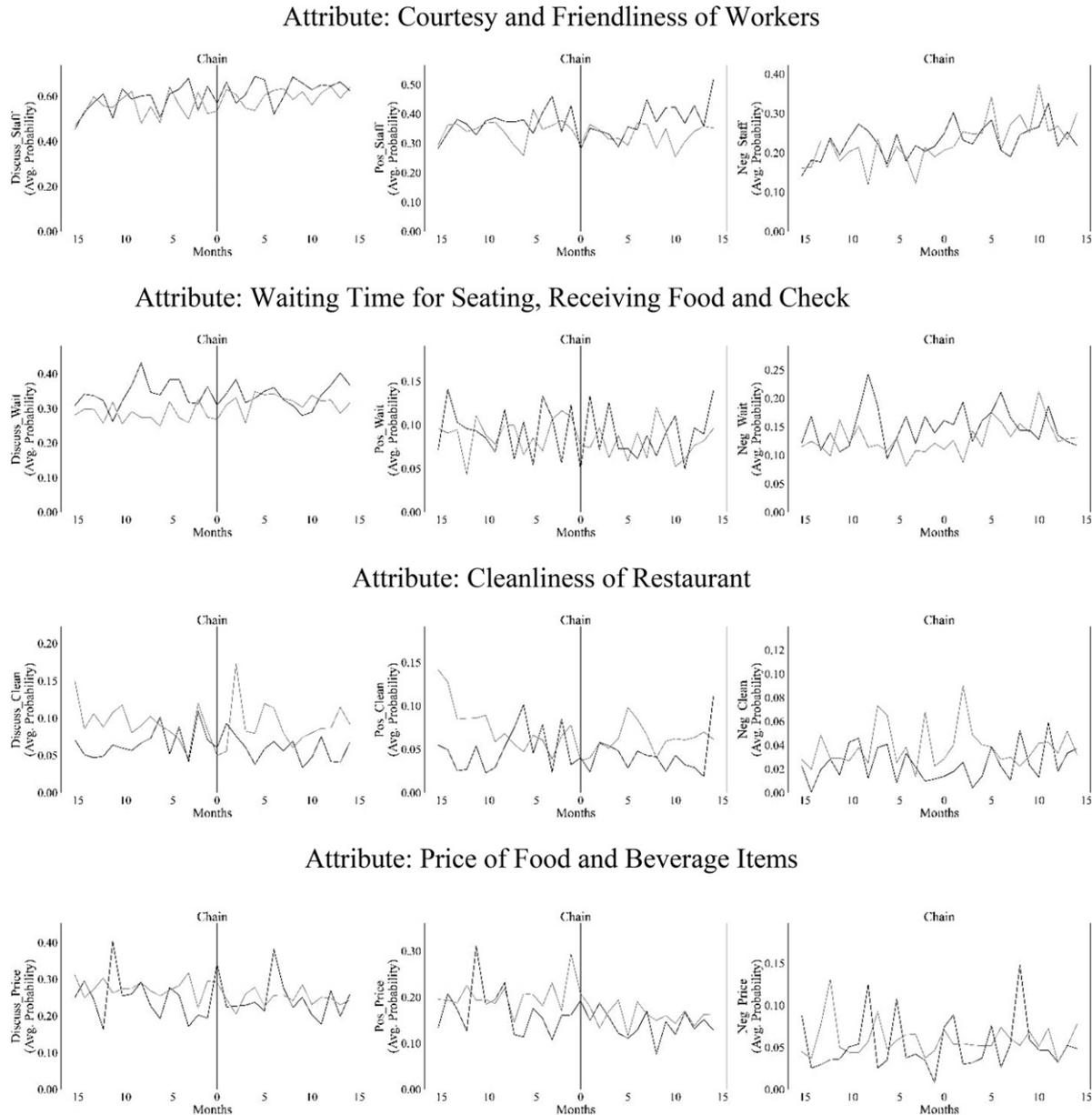
Our approach has two major steps. First, to extract measures of attributes and their corresponding valence (positive and negative) from reviews, we randomly sample 12,000 reviews from our data set and present them for human annotation (we follow the procedure of Liu et al. 2019a). Each review is viewed and labeled by three human annotators from Amazon’s Sagemaker platform.¹² The labeling consists of the following mutually exclusive valence categories for each of the four attributes: positive, negative, and absent (leading to 12 labels per review). The four attributes are the courtesy and friendliness of workers, waiting time (to be seated or to receive food or the check), the cleanliness of the interiors of the restaurant, and the price of food and beverages. For example, for the attribute about staff, the annotator was asked to tag the text of each review with one of three tags: “waiters/hostess/register clerk or any employee was spoken of positively, that is, the worker was courteous and/or friendly”; “waiters/hostess/register clerk or any employee was spoken of negatively, that is, the worker was not courteous and/or friendly”; “waiters/hostess/register clerk or any employee was not spoken of.”¹³

For each of the 12 attribute–valence combinations (or labels), we present trends of monthly mean values (i.e., the proportion of all human-annotated reviews that were labeled) over 30 months (see Figure 1 for

reviews of chain restaurants and Figure 2 for those of independent restaurants). For example, the top-right chart in Figure 1 presents the 30-month trends of the proportion of reviews in San Jose restaurants (blue line) and reviews in other CSC restaurants (orange line) that were labeled as discussing the courtesy or friendliness of workers positively. The first 15 months of this time series refers to the pretreatment period; the latter 15 months are the posttreatment period. The top-left chart presents proportions of reviews that discussed workers (irrespective of whether this discussion was positive, negative, or neutral). We find that across both chains and independents, cleanliness is the least frequently discussed attribute, and workers are the most frequently discussed. As compared with other CSC cities, the discussion of workers seems to have decreased for independent restaurants in San Jose after the wage hike (top right chart in Figure 2). Price discussion of independent restaurants in San Jose is greater than that in other cities—both before and after the wage hike, providing model-free evidence that treatment did not affect price discussion of independent restaurants. Systematic patterns in the attribute discussions of reviews of chain restaurants are less discernible from these charts, perhaps reflecting smaller effects of the treatment on chain restaurants.

One way to isolate the causal effect of the treatment could be to compute differences in means of these proportions (across the pretreatment and posttreatment periods), based solely on the annotated reviews. For each attribute–valence combination, we could test whether the differences in proportions (across the treatment and control units) change after the wage increase. This is undesirable because we want the treatment and control units to follow parallel trends in the discussions of attribute–valence combinations. Figures 1 and 2 suggest that at least for some attributes, the parallel trends assumption might not hold. We could consider restricting the synthetic control analysis (which ensures parallel trends) to human-annotated reviews. However, Xu (2017) and Pattabhiramaiah et al. (2019) recommend at least 10 time periods in the pretreatment period. Matching in each time period necessitates a sufficiently

Figure 1. Comparison of 30-Month Trends of Attribute–Valence Discussions Based on Human-Annotated Reviews of Chain Restaurants: San Jose vs. Other CSC Cities



Notes. Month zero represents the month of the wage increase in San Jose. Gray lines represent restaurants in San Jose. Black lines represent other restaurants.

large number of observations for each zip code within the treated unit and the control units for each time period. The number of annotated reviews in each time period is too small for good matching. Also, some restaurants might not receive any reviews in some time periods. So we estimate text models to predict each of these 12 labels, for reviews over a two-year period (with the wage increase at the midpoint). This enables a sufficiently long time series of reviews, with sufficient data in each month, for robust synthetic control analysis.

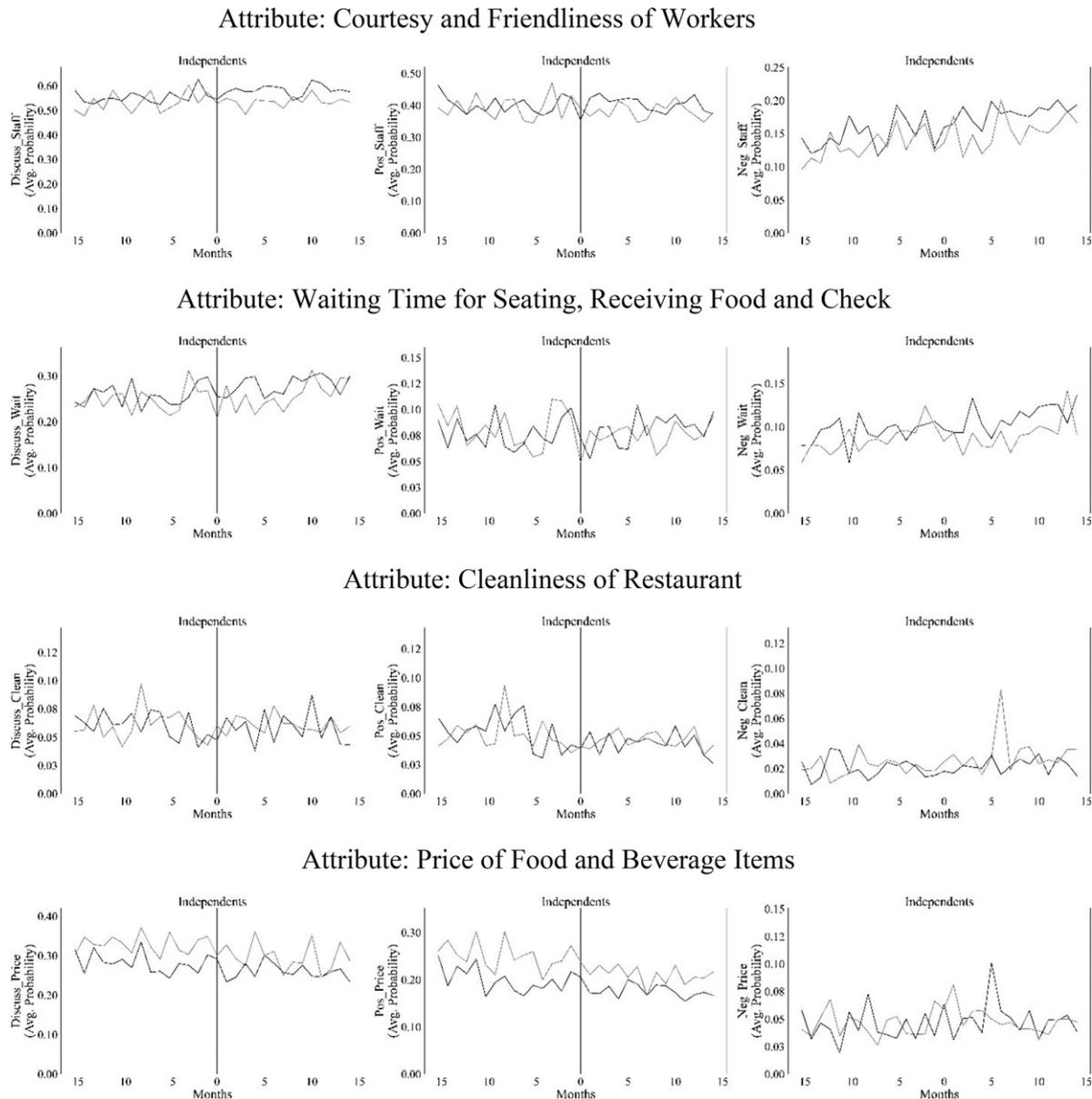
It is imperative for our text model to accurately predict the 12 labels of valence–attribute combinations.

For this, we employ bidirectional transformer models. Under a wider variety of natural language processing tasks, transformer models have been shown to significantly outperform prior state of the art. Devlin et al. (2019) report significant performance improvements over the state of the art on 11 different language tasks. We briefly describe this model.

4.2. Bidirectional Transformer Model of Text Analysis

We briefly present an overview of advances in text classification to provide context to the models used in

Figure 2. Comparison of 30-Month Trends of Attribute–Valence Discussions Based on Human-Annotated Reviews of Independent Restaurants: San Jose vs. Other CSC Cities



Notes. Month zero represents the month of the wage increase in San Jose. Gray lines represent restaurants in San Jose. Black lines represent other restaurants.

our application. In prior approaches to text classification, the text is viewed as its component words. Each review-text is represented as a vector (of size of the vocabulary) of word occurrences. Each position in the vector represents a unique word in the vocabulary. The element in that position is a measure of the occurrence of that particular word in the review text. Measures of word occurrences include binary indicators, raw frequency, and term-frequency-inverse document frequency measures. Modeling entails predicting the review label given the review texts vector representation. Support vector machines (Vapnik 1995) and naïve Bayes models are effective for the classification

of this type of text representation. Intrinsically, this is a bag-of-words approach and ignores word order.

In a breakthrough in word representations, Mikolov et al. (2013a) and Pennington et al. (2014) recover word-specific vector representations (word embeddings) of individual words. The linguistic relationship among words is captured by modeling the probability of observing a word given its context words (or vice versa). The word embeddings are learned on large corpora (a process termed “pretraining”) by predicting a focal word given its neighboring words (skip-gram) or predicting neighbors given the focal word (continuous bag of words, or CBOW). The prediction task forces

the model to learn word embeddings that capture the relationship between words. As a result, the word embeddings provide a simple way to capture relationships between words. It is then possible to transfer this understanding about the linguistic relationships between words from the large corpora to specific data sets (transfer learning). Specifically, these word embeddings serve as input for downstream tasks such as label prediction. CNNs for sentence classification (Kim 2014, Timoshenko and Hauser 2019) are an example of the application of these word vectors to label prediction tasks. CNNs were for some time the state of the art for text classification tasks. Note that the word embedding for a given word, once estimated, is static and independent of context. For example, the embedding for the word “bank” in the sentences “I drew cash from the bank” and “I stood on the river bank” is the same.

More recent advances in transfer learning (e.g., Devlin et al. 2019) not only account for word order but also allow each word to have context-dependent vector representations. Bidirectional Encoder Representations from Transformers, or BERT (Devlin et al. 2019), uses transformers (Vaswani et al. 2017), an alternative to recurrent and convolutional neural networks to capture word relationships. The word representations are learned by modeling two prediction tasks. The first prediction task is termed a masked language model. The model randomly hides (masks) some of the words from the input text and attempts to predict the masked word using only words that appear in its context. The model incorporates words before (left) and after (right) the focal word using a bidirectional transformer and accounts for word position explicitly. The prediction task is similar to the skip-gram prediction task. However, unlike the skip-gram model (Mikolov et al. 2013a), BERT makes use of a deep-learning architecture consisting of 12 layers. An additional prediction task models the probability of observing the next sentence. The model accounts for context in two key ways. First, it incorporates an “attention mechanism.” In a specific context, all the neighboring words are scored for their relevance to the focal word, and this information is encoded within the focal word’s embedding using a weighting scheme. Furthermore, rather than have just one representation of this relevance, the model allows for 12 separate relevance or attention scores. Even more, flexibility is incorporated in the model by replicating this structure in each of the 12 layers of the model. That is, for each of the 12 layers in the BERT model, there are 12 attention scores. Second, the model explicitly estimates position embeddings that capture the position of the focal word in a sentence. These position embeddings are added to the word embeddings to encode the position information. The final layer predicts the word embedding given the context and yields

word-vector representations that are context dependent. Further details of the model architecture appear in Devlin et al. (2019).

These context-dependent vectors can be used for downstream tasks. Successor models of BERT, RoBERTa (Liu et al. 2019b) and XLNet (Yang et al. 2019), further develop this framework and make adjustments to BERT on (a) the size of training data, (b) the formulation of prediction tasks, and (c) word embedding size. We summarize the differences across all word representation models in Table 3.

We experiment with several text classification models incorporating these context-dependent embeddings—linear models, CNN models, and long-short term memory models. We benchmark the performance of context-dependent embeddings with static word embeddings in prediction tasks. We find that models using context-dependent embeddings from transformer models significantly outperform the CNNs using static embeddings. The specific model for each task is selected via fivefold cross-validation. The gains in performance over Mikolov et al. (2013b) and Pennington et al. (2014) come from the dual objective functions, the size of word vectors, context dependence, and incorporation of subword-level data. We now provide details on the model specification and hyperparameter selection.

4.2.1. Model Specification. We review the alternative transformer models considered, the incremental layers added to the transformer model, the choice of loss functions, and the task-specific training procedure.

Choice of Transformer Models. The main models considered are BERT, RoBERTa, and XLNet. Each model has two versions—Base and Large. The distinction between the two is the number of parameters used to represent the model, with large versions having far more parameters. For example, BERT Large has 340 million parameters, in comparison with BERT Base, which has 110 million parameters. RoBERTa is a more robust specification of the BERT model, trained on a larger corpus than BERT, and has demonstrated significantly improved performance over BERT. XLNet is the next generation of transformer models that explicitly models the autoregressive structure of sentences. We considered both base and large versions of all three transformer models.

Classifier Specification. Each transformer model generates context-dependent embeddings for words in a document. Unlike the static word embeddings, like Word2Vec, context-dependent embeddings are dependent on the other words in the neighborhood of the focal word. The standard transformer model implementation (available on TensorFlow and PyTorch) libraries include a simple linear classifier on top of

Table 3. Comparison of Characteristics of Transformer Models

Model	Number of parameters	Pretraining data	Vocabulary size	Prediction tasks	Word representation vector size	Embedding type
Skip-gram/CBOW (Mikolov et al. 2013b)	300 × 692,000	Internal Google data set	692,000 words	Neighboring word prediction	300	Context independent
BERT Base (Devlin et al. 2019)	110M	16 GB (Books corpus + English Wikipedia)	30,000 Byte-Pair Encodings	(a) Masked language modeling, (b) next sentence prediction	768	Context dependent
BERT Large (Devlin et al. 2019)	340M	16 GB (Books corpus + English Wikipedia)	30,000 byte-pair encodings	(a) Masked language modeling, (b) next sentence prediction	1,024	Context dependent
RoBERTa (Liu et al. 2019b)	340M	160 GB = 16 GB (Books corpus + English Wikipedia + 76 GB CC-News + 38 GB Open Web text + 31 GB Stories)	50,000 byte-pair encodings	Masked language modeling	1,024	Context dependent
XLNet (Yang et al. 2019)	110M	126 GB = 13 GB (Books corpus + English Wikipedia) + 78 GB CC-News + 16 GB Giga5 + 19 GB Clue Web	50,000 byte-pair encodings	(a) Permutation language modeling, (b) next sentence prediction	1,024	Context dependent

Note. GB, gigabytes; M, Million.

these context-dependent embeddings. This is the linear classifier. As pointed out by Devlin et al. (2019), the embeddings could serve as inputs to (1) a CNN classifier, (2) an long short-term memory (LSTM) (unidirectional and bidirectional), and (3) nonlinear classifiers. We implement all of these additional specifications on the base models of BERT and RoBERTa. In all but one attribute–valence combination, the linear classifier performed well with fine-tuning.

Choice of Loss Functions. A key aspect of our data set is the class imbalance of the different attributes (i.e., attribute–valence combinations are not symmetrically distributed). Consequently, we consider two different loss functions in addition to cross-entropy loss: focal loss (Lin et al. 2017) and dice loss (Sorenson 1948).

Task-Specific Fine-Tuning. Each of the above models can be estimated in two ways. First, the transformer model embeddings are viewed as static input into the linear, CNN, LSTM, and nonlinear models. That is, the transformer model weights are not affected in the training process. This approach generally does not work for specific classification tasks (Sun et al. 2019). Consequently, we adopt a second approach termed

task-specific fine-tuning. We allow all the layers of the transformer model to learn during the training process (including the additional layers added on top of the transformer models). This substantially increases the training time, with a typical model taking approximately three hours to train over 50 epochs for a particular learning rate on a four graphical processing unit (GPU) setup machine. Larger models such as XLNet occupy more of the GPU memory and nearly treble the processing time. In exchange, we gain significant performance improvement.

4.2.2. Hyperparameter Selection. There are two critical hyperparameters in estimating transformer models for classification. The first is the learning rate, and the second is the number of epochs of training. An epoch is defined as one pass of training over all the observations in the training data. We conducted a grid search over a reasonable range of learning rates, (5e-06, 1e-05, 2e-05, 5e-05, 8e-05, 1e-04, 5e-04, and 1e-03). Both the learning rate and the number of epochs of training are learned via cross-validation. For a full fivefold cross-validation process, the processing time is approximately 25 hours for a small model (like BERT Base) versus 75 hours for a large model like XLNet. We also

considered the approach in Smith (2018) to find optimal learning rates. We incorporate early-stopping in the model to prevent overfitting on training data. We use the validation data set during cross-validation to determine whether the Matthews (1975) coefficient of correlation changes over epochs. If it has not changed over five epochs, the training loop terminates.

The model described above yields 12 review-specific estimates of the probability score of whether each of three service attributes (and price) is discussed positively, negatively, or not at all. These estimates serve as the dependent variables for the next stage of analysis. We model the incidence of positive discussion, the incidence of negative discussion, and incidence of discussion (or no discussion) separately. We do not constrain the third outcome to equal the sum of the first two outcomes. Imposing this constraint leads to a loss of predictive performance. Because maximizing predictive performance is the main objective of the text analysis, we prefer separate modeling of the three outcomes to imposing the equality constraint. Details appear in Online Appendices 1 and 2.

4.3. Synthetic Control Analysis

To identify the causal effect of the wage increase on the probability scores of attribute discussion, we employ the synthetic control method (Abadie et al. 2010, Turrillai and Tellis 2014). For a given dependent variable obtained from text analysis, this approach creates a “synthetic control unit” computed as a weighted combination of zip codes in the seven cities that form our control group. For each of 12 months preceding the treatment date (March 2013), this approach chooses weights such that the synthetic control for each zip code within the treated unit (San Jose) matches it, in terms of pretreatment trends of the dependent variable and covariates. Thus, this approach naturally satisfies the parallel trend assumption required for causal inference. The synthetic control’s posttreatment pattern is then used as the counterfactual prediction for the treated cases. Because this approach uses the pretreatment dependent variable for matching, it naturally conditions on both observables and unobservables (Lovett et al. 2019). Weights for the synthetic control unit are chosen flexibly, so that they differ for each of the 12 dependent variables, to minimize pretreatment differences between the treatment group and the synthetic control.

Following Xu (2017) and Pattabhiramaiah et al. (2019), we employ the generalized synthetic control estimator (GSCE). Given the sparseness of restaurant-level data in our setting, we aggregate the analysis at the level of zip codes. Zip codes in San Jose are matched to zip codes in the control cities. GSCE allows for multiple treatment cities. We specify an interactive fixed effects model: $y_{it} = \delta_{it}D_{it} + x'_{it}\beta + \lambda'_i f_t + \epsilon_{it}$. The subscript i indexes zip codes, and t indexes time, in

months. The term y_{it} is the outcome variable (for example positive courtesy/friendliness) of interest. It is the mean (across all reviews of restaurants in zip code i posted in time t) of probability scores of discussion of an attribute–valence combination, obtained from the text model. The term D_{it} is an indicator variable that takes a value of one for treated units in the posttreatment period. The coefficient δ_{it} measures the heterogeneous treatment effect. The term x'_{it} is a vector of observed covariates: the number of restaurants, reviews, and reviewers in zip code i at time t . The term f_t is an $(r \times 1)$ vector of unobserved common factors. The term λ'_i is a vector of unknown factor loadings.¹⁴ The number of factors is selected based on the cross-validation performance of pretreatment fit. The average treatment effect is computed as $\frac{1}{N_{tr}} \sum_{i \in T} \delta_{it}$, where T is the set of treatment units, and N_{tr} is the cardinality of T .

The GSCE automates the process of running placebo tests and provides standard errors around the estimated treatment effect while preserving the efficiency of the estimation algorithm. The estimator is robust in the presence of serial correlation (for the detailed algorithm, see Xu 2017). As further evidence of identification of the GSCE, we simulate 9,000 data sets based on different parameter values and are able to accurately recover the parameters of interest. Details appear in Online Appendix 3.

5. Model Comparison and Results

As mentioned, we estimate several models for predicting each of 12 attribute–valence combinations for reviews in the two-year window. For each attribute–valence combination, we then choose the best performing model and score all reviews based on that model. Table 4 presents the performance metrics of various models. Given the recent usage of CNNs in marketing, that serves as our primary benchmark model. Specifically, we estimate a CNN model that uses word embeddings based on Mikolov et al. (2013b); metrics appear in the column “CNN.” We also experimented with GloVe embeddings (Pennington et al. 2014), but their performance was very similar. Among transformer models, the state of the art at the beginning of 2019 was BERT (Devlin et al. 2019). RoBERTa (Liu et al. 2019b) and XLNet (Yang et al. 2019) are successors of BERT and also use transformers. Metrics for these models appear in Table 4.¹⁵ To decide which model is best, we compute three metrics: precision, recall, and F1, which are often useful for characterizing any binary classification algorithm (presence or absence of an attribute–valence combination). Precision for a class is defined as $\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$ and is interpreted as the probability that the model is correct when it says an example belongs to a class. Recall for a class is defined as $\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$ and indicates the percentage of

Table 4. Comparison of Performance of Text Models

Attribute–valence combination	Model performance metric	Transformer model with highest F1	CNN	RoBERTa	XLNet
<i>Discuss_Staff</i>	F1	0.84	0.76	0.83	0.84
	Precision	0.79	0.75	0.81	0.77
	Recall	0.91	0.77	0.86	0.93
<i>Pos_Staff</i>	F1	0.82	0.67	0.79	0.82
	Precision	0.75	0.57	0.83	0.75
	Recall	0.91	0.79	0.76	0.91
<i>Neg_Staff</i>	F1	0.79	0.56	0.79	0.77
	Precision	0.77	0.47	0.77	0.71
	Recall	0.82	0.7	0.82	0.84
<i>Discuss_Wait</i>	F1	0.74	0.67	0.72	0.7
	Precision	0.68	0.69	0.71	0.8
	Recall	0.82	0.64	0.74	0.62
<i>Pos_Wait</i>	F1	0.59	0.29	0.59	0.48
	Precision	0.71	0.5	0.71	0.37
	Recall	0.5	0.2	0.5	0.7
<i>Neg_Wait</i>	F1	0.73	0.64	0.68	0.7
	Precision	0.8	0.7	0.59	0.73
	Recall	0.67	0.58	0.79	0.67
<i>Discuss_Clean</i>	F1	0.53	0.49	0.53	0.48
	Precision	0.48	0.48	0.48	0.41
	Recall	0.58	0.50	0.58	0.58
<i>Pos_Clean</i>	F1	0.57	0.32	0.57	0.55
	Precision	0.42	0.25	0.42	0.4
	Recall	0.89	0.44	0.89	0.89
<i>Neg_Clean</i>	F1	0.58	0.42	0.46	0.58
	Precision	0.78	0.56	0.55	0.78
	Recall	0.47	0.33	0.4	0.47
<i>Discuss_Price</i>	F1	0.73	0.59	0.73	0.73
	Precision	0.68	0.61	0.72	0.68
	Recall	0.8	0.58	0.74	0.8
<i>Pos_Price</i>	F1	0.51	0.38	0.49	0.46
	Precision	0.43	0.33	0.39	0.36
	Recall	0.63	0.45	0.66	0.66
<i>Neg_Price</i>	F1	0.64	0.41	0.64	0.61
	Precision	0.73	0.42	0.73	0.74
	Recall	0.58	0.39	0.58	0.52

Notes. The term *Discuss_Staff* measures whether the courtesy and friendliness of workers was discussed (1) or not (0); *Pos_Staff* measures whether the courtesy and friendliness of workers was discussed positively (1) or not (0); *Neg_Staff* measures whether the courtesy and friendliness of workers was discussed negatively (1) or not (0); *Discuss_Wait* measures whether waiting time (for seating, receiving food and check) was discussed (1) or not (0); *Discuss_Clean* measures whether cleanliness of restaurants was discussed (1) or not (0); *Discuss_Price* measures whether price of food or beverage items was discussed (1) or not (0). Other attribute–valence combinations follow along similar lines.

examples of a class correctly identified as belonging to that class. The two metrics represent a trade-off. Increased precision comes at the cost of fewer examples of calls being correctly tagged as belonging to that class, that is, recall. The F1 captures this trade-off by computing the harmonic mean of the two metrics. Harmonic means skew toward the lower of two values, providing a conservative estimate of model performance. We use the F1 metric for choosing the best performing model.

We now discuss the average treatment effects of the wage increase on the discussions of the 12 attribute–valence combinations, as inferred from the synthetic control analysis described earlier (Table 5). We

conduct this analysis separately for independent and chain restaurants. As mentioned, our primary analysis window is 12 months (before and after). We also conduct the same analysis for 15 months (before and after) for robustness. Effects that are significant only for one time window (12 months or 15 months) are perhaps insufficiently robust; we restrict our discussion to only those treatment effects that are significant across both time windows.

We first discuss the results of the analysis of independent restaurants. We find that reviews of restaurants in San Jose discussed the courtesy and friendliness of workers to a lower extent after the wage increase, as compared with the counterfactual projection of the

Table 5. Average Treatment Effects Based on Generalized Synthetic Control Analysis

Attribute–valence combination	Independent restaurants		Chain restaurants	
	12 months	15 months	12 months	15 months
<i>Discuss_Staff</i>	−0.029 (0.009)	−0.024 (0.009)	0.007 (0.023)	−0.006 (0.020)
<i>Pos_Staff</i>	−0.015 (0.010)	−0.011 (0.009)	−0.013 (0.019)	−0.032 (0.017)
<i>Neg_Staff</i>	−0.021 (0.008)	−0.015 (0.007)	0.039 (0.023)	0.033 (0.018)
<i>Discuss_Wait</i>	−0.021 (0.007)	−0.020 (0.006)	0.056 (0.023)	0.034 (0.023)
<i>Pos_Wait</i>	−0.012 (0.006)	−0.015 (0.005)	0.004 (0.009)	−0.001 (0.009)
<i>Neg_Wait</i>	−0.012 (0.004)	−0.011 (0.004)	−0.012 (0.020)	−0.007 (0.012)
<i>Discuss_Clean</i>	0.005 (0.005)	0.006 (0.005)	0.014 (0.009)	0.008 (0.009)
<i>Pos_Clean</i>	0.004 (0.005)	0.005 (0.004)	0.012 (0.007)	−0.001 (0.008)
<i>Neg_Clean</i>	0.006 (0.004)	0.007 (0.003)	0.004 (0.006)	0.002 (0.005)
<i>Discuss_Price</i>	−0.012 (0.010)	−0.004 (0.009)	−0.047 (0.022)	−0.042 (0.021)
<i>Pos_Price</i>	−0.007 (0.008)	−0.004 (0.007)	−0.024 (0.017)	−0.019 (0.014)
<i>Neg_Price</i>	−0.004 (0.005)	−0.002 (0.004)	−0.012 (0.015)	−0.007 (0.012)

Notes. Parameters with $p < 0.05$ are in bold. The term *Discuss_Staff* measures whether the courtesy and friendliness of workers was discussed (1) or not (0); *Pos_Staff* measures whether the courtesy and friendliness of workers was discussed positively (1) or not (0); *Neg_Staff* measures whether the courtesy and friendliness of workers was discussed negatively (1) or not (0); *Discuss_Wait* measures whether waiting time (for seating, receiving food and check) was discussed (1) or not (0); *Discuss_Clean* measures whether cleanliness of restaurants was discussed (1) or not (0); *Discuss_Price* measures whether price of food or beverage items was discussed (1) or not (0). Other attribute–valence combinations follow along similar lines.

synthetic control. Specifically, there is a decrease in the negative discussion of the courtesy and friendliness of workers. This is welcome news for restaurant managers. These results are consistent with the notion that greater wages motivate workers to serve better, or that they increase the restaurant's ability to hire workers who are friendlier and/or more courteous.

We find that the discussion of waiting time (for seating, etc.) is also lower after the wage hike as compared with the discussion of the synthetic control. This could be driven by lower positive discussion, lower negative discussion, or some combination thereof. We find that both positive and negative discussion of waiting time decreases after the wage hike. This suggests that after the wage increase, waiting time becomes a less important attribute or at least less worthy of online discussion. Lower positive discussion of waiting time could be driven by fewer workers and/or replacement of workers with technology, which involves fewer positive experiences with waiting times. Lower negative discussion of waiting time (a welcome development for managers) could be an indicator of greater worker motivation post wage increase. Both effects coexist in our data, so we do not draw aggregate-level conclusions about how changes to this aspect of service quality affected restaurants and consumers.

In terms of effect size, a coefficient of -0.012 for *Neg_Wait* suggests that, compared with synthetically constructed restaurants that (a) were not subject to the wage regulation in 2013 and (b) had the same probabilities of negative discussion of waiting times for 12 months before the wage hike, reviews of independent restaurants witnessed a 1.2% point reduction in the negative discussion of waiting time for 12 months following the wage hike. Because average probabilities

of discussion of this attribute–valence combination are in the range of 7%–12% for our data period, a 1.2% point reduction represents a substantial change.

Interestingly, there is no significant change in the discussion of the cleanliness of independent restaurants. One possible reason is that workers who provide poor service (e.g., those who do not maintain cleanliness) are likely to be paid less (and minimum wages) than workers who provide good service. Because the wage increase does not depend on prior service quality, but only on existing wages, the regulation disproportionately raised the wages of workers who were providing poor service (and were paid less before the wage increase). Workers providing good service might have been paid above minimum wage before the legislated change, and hence their wages were not affected by the legislation. Consequently, their (good) service levels might have remained unaffected.

Summarizing the results for independent restaurants, we find a decrease in the negative discussion of the courtesy and friendliness of restaurants and a decrease in both positive and negative discussion of waiting times. We conclude that from a managerial standpoint, barring the decrease in the positive discussion of waiting times, the effect of the wage increase on consumer perceptions of independent restaurants was positive (negative discussions of two attributes reduced).

Our second research objective is to investigate whether the effect of wage increase varies across chain restaurants and independent restaurants. As discussed in Section 1.1, agency theory suggests that the impact of a wage increase of service quality should be greater in independent restaurants, which have greater incentive and ability to monitor service outcomes. Synthetic control analysis of reviews of chain restaurants reveals

that discussion of various service attributes after the wage increase for such restaurants are not significantly different across San Jose restaurants and the synthetic control. In other words, changes in perceptions of service quality in San Jose restaurants after the wage increase are driven by independent restaurants only. Comparison of results for the two types of restaurants provides empirical support to our theory that corporate-owned chain outlets have lower incentive and the ability to monitor service outcomes than independents. Although franchisee outlets among chain restaurants might be motivated to improve service outcomes, the imposition of process-driven service guidelines across most major U.S. chains leaves relatively less scope for improvement than independent restaurants. It is also possible that independent restaurants can improve service outcomes more easily than chain restaurants because of their smaller size and greater operational flexibility.¹⁶

We present trends of monthly means of probability scores of discussions of all attribute–valence combinations for the treatment and synthetic control units for chain restaurants (in Figure 3) and independent restaurants (in Figure 4). It is clear from these figures that the synthetic control analysis performs well in ensuring that pretreatment trends for the treatment and control units are well matched. Second, we do not find any visually discernible systematic differences in posttreatment trends between treatment and control units for chain restaurants. This provides face validity to the null effects for chain restaurants. For independent restaurants, as expected, we find that the posttreatment probability scores of negative discussion of courtesy and friendliness of workers are lower than those of the synthetic controls for at least 10 out of 12 months. This reflects in the negative estimate of the average treatment effect for this attribute–valence combination in Table 5.

5.1. Alternate Mechanisms

We discuss several other potential mechanisms that might explain our results.

It is possible that prices rise in independent restaurants, causing fewer price-sensitive customers. This compositional change results in an increase in quality discussions. However, we do not see any postpolicy change in price discussions in our data.¹⁷ We also conduct the following indirect test of price-based selection. If there had been price-based selection out of independents, we might expect an impact on entry and exit of restaurants (this might also result from wage policy, whether or not prices change). We compare estimates of two models: with all reviews, and with reviews of restaurants that have both pre- and postpolicy reviews. We find the results are unchanged.

Another alternative is as follows: with an increase in wages, there is a shift toward higher-quality

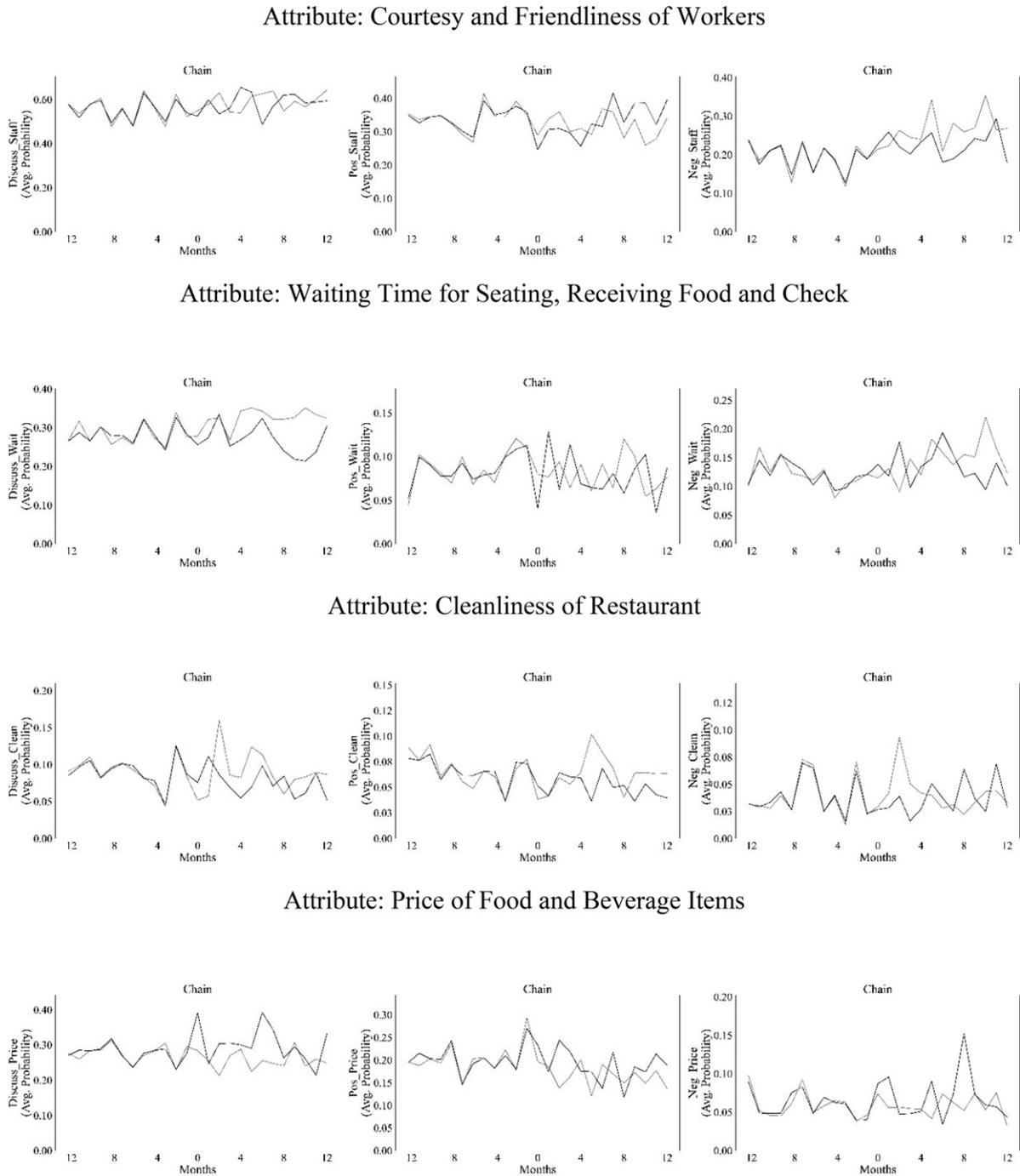
restaurants (with no changes in quality). We analyze prepolicy star ratings as a quality measure. We find that after the policy, there is no systematic difference in the number of reviews posted for higher-star restaurants (defined as those with >2.5-star rating), aggregated at the zip code level (same as for synthetic control). Note we do not have quality measures other than star ratings, and these star ratings can capture multiple dimensions of quality; our main results are focused only on service quality. Therefore, in the absence of a more fine-grained service-quality-based measure of quality, we view this evidence as only suggestive of no selection effects.

As discussed above, although there is no overall change in price discussion at independent restaurants, it is possible this overall result masks heterogeneity that supports selection effects. For example, minimum wage increases might affect lower price tier restaurants more and price rises happen here. Or higher-end restaurants might pay higher wages and here price rises might be easier to make. To measure possible heterogeneity, we gather price tier information in reviews (marked by “\$” levels in the data) captured after our data period, that is, in September 2020. That is, our price tier is fixed for a restaurant at the postpolicy level. Based on this, we see no difference in effects by price tiers (model details and parameter estimates appear in Online Appendix 4). Given the quality of the price data, we view this analysis as providing only preliminary evidence against selection/compositional effects.

Another potential confound is whether there are geographic selection effects caused by wage differentials; that is, consumers can easily travel across the treated and the nearby control cities. Consumer sorting as a result of price changes would invalidate the stable unit treatment value assumption (see Guo et al. 2020). Allegretto and Reich (2018) find that restaurants within 0.5 miles of the border (and inside San Jose) are muted in their price rise (because of lower wages on the other side of the border). Restaurants farther than this are not prone to competitive pressure from across the border. As a significantly stricter test of wider consumer search and competitive pressures from a wider set of geographic neighbors than in Allegretto and Reich (2018), we double their price- and/or quality-competition radius from 0.5 miles to 1 mile. We note that Allegretto and Reich (2018) use a difference-in-differences approach and we use synthetic control.

One possible, and blunt, way to control for border effects would be to eliminate from the sample all restaurants in San Jose within one mile from the border. A more principled and empirically driven way, and one that is better suited for our synthetic control methods, is as follows. First, we obtain the location (latitude and longitude) of each restaurant in our data set. For each restaurant, we draw a one-mile radius. We

Figure 3. Effect of Wage Increase on Attribute Discussions in Chain Restaurants: Synthetic Control Analysis



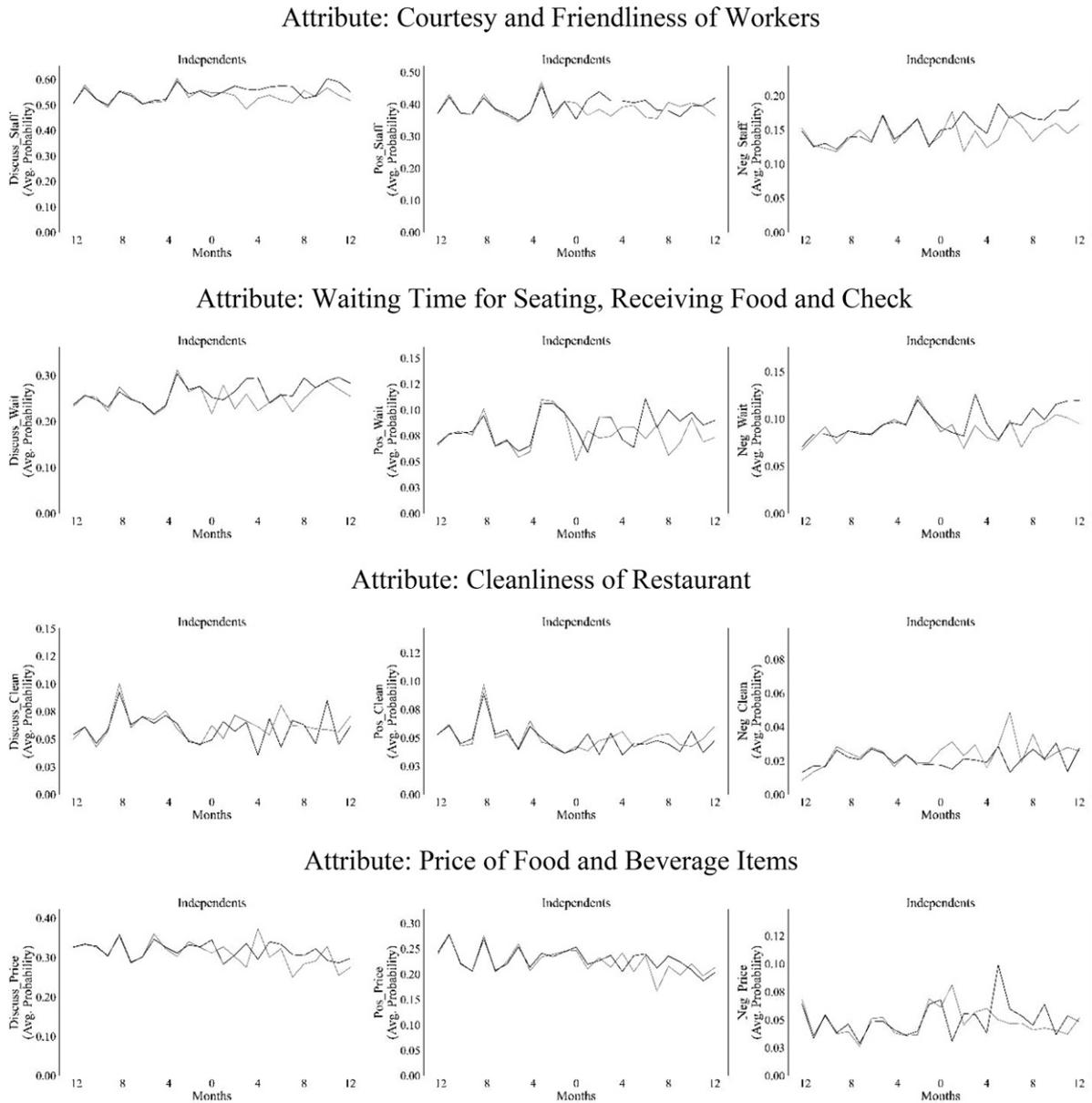
Notes. Month zero represents the month of the wage increase in San Jose. Gray lines represent restaurants in San Jose. Black lines represent counterfactuals in the absence of the treatment.

construct a synthetic group for each focal restaurant. If any restaurant in the synthetic control group for a focal restaurant is within a one-mile radius, we drop this focal restaurant from our sample. We prefer this exclusion strategy to excluding all San Jose restaurants within one mile of the border because the extent of consumer sorting depends on the distance of the San Jose restaurant from a lower-priced restaurant across

the border. It does not depend on the distance of the San Jose restaurant from the border per se. We then reestimate the generalized synthetic control model. Results (Table 6) are very similar to those reported with the full sample.¹⁸

In summary, our results suggest that agency theory-based mechanisms appear to be more consistent with our findings than the alternative mechanisms discussed

Figure 4. Effect of Wage Increase on Attribute Discussions in Independent Restaurants: Synthetic Control Analysis



Notes. Month zero represents the month of the wage increase in San Jose. Gray lines represent restaurants in San Jose. Black lines represent counterfactuals in the absence of the treatment

above. However, given the important policy and managerial question of minimum wage increases, and data limitations of the additional analyses, our findings and possible mechanism should be viewed as preliminary and deserving of further investigation.

6. Implications and Conclusions

We first discuss the implications of our research for consumers. First, our evidence suggests that consumers who value courtesy and friendliness will be well served by increases in minimum wages because these service qualities improve in independent restaurants. Second, consumers who are sensitive to other aspects

of service such as cleanliness will probably not notice a change in service levels of independent restaurants with increases in minimum wages. Third, consumers who prefer chain restaurants over independent restaurants can expect similar service levels as earlier. Wage increases have not resulted in substitution with workers or technologies that offer perceptibly lower service.

Our results bring potentially positive news for owners and managers of independent restaurants. The debate on the higher minimum wage legislation has focused solely on greater labor costs and resulting prices, without accounting for the route of service quality improvements that might soften the impact of the price increase. We find that service perceptions are more

Table 6. Average Treatment Effects Based on Generalized Synthetic Control Analysis (Excludes Border Treatment Units)

Attribute–valence combination	Independent restaurants		Chain restaurants	
	12 months	15 months	12 months	15 months
<i>Discuss_Staff</i>	−0.030 (0.010)	−0.021 (0.009)	−0.016 (0.025)	−0.025 (0.020)
<i>Pos_Staff</i>	−0.009 (0.011)	−0.004 (0.010)	−0.018 (0.019)	−0.037 (0.018)
<i>Neg_Staff</i>	−0.024 (0.007)	−0.022 (0.008)	0.027 (0.023)	0.020 (0.019)
<i>Discuss_Wait</i>	−0.033 (0.007)	−0.025 (0.008)	0.030 (0.024)	0.016 (0.023)
<i>Pos_Wait</i>	−0.021 (0.007)	−0.017 (0.006)	0.002 (0.009)	0.000 (0.008)
<i>Neg_Wait</i>	−0.013 (0.005)	−0.011 (0.005)	−0.001 (0.021)	−0.002 (0.018)
<i>Discuss_Clean</i>	0.002 (0.005)	0.006 (0.005)	0.000 (0.010)	−0.003 (0.009)
<i>Pos_Clean</i>	0.001 (0.005)	0.005 (0.004)	0.005 (0.007)	−0.007 (0.008)
<i>Neg_Clean</i>	0.004 (0.004)	0.006 (0.003)	−0.001 (0.005)	−0.004 (0.005)
<i>Discuss_Price</i>	−0.004 (0.011)	0.004 (0.010)	−0.036 (0.023)	−0.030 (0.020)
<i>Pos_Price</i>	0.006 (0.008)	0.004 (0.008)	−0.015 (0.016)	−0.015 (0.014)
<i>Neg_Price</i>	−0.003 (0.005)	−0.001 (0.005)	−0.004 (0.014)	0.003 (0.012)

Notes. Parameters with $p < 0.05$ are in bold. The term *Discuss_Staff* measures whether the courtesy and friendliness of workers was discussed (1) or not (0); *Pos_Staff* measures whether the courtesy and friendliness of workers was discussed positively (1) or not (0); *Neg_Staff* measures whether the courtesy and friendliness of workers was discussed negatively (1) or not (0); *Discuss_Wait* measures whether waiting time (for seating, receiving food and check) was discussed (1) or not (0); *Discuss_Clean* measures whether cleanliness of restaurants was discussed (1) or not (0); *Discuss_Price* measures whether price of food or beverage items was discussed (1) or not (0). Other attribute–valence combinations follow along similar lines.

likely to improve along dimensions that are easily observable to consumers—courtesy and friendliness—perhaps suggesting the need to manage other service dimensions. The limited impact of the wage increase on service perceptions of chain restaurants might reassure corporate managers and franchises of such outlets.

Policy makers might view these results as providing directionally positive evidence of minimum wage increases on consumer welfare in independent restaurants. This research could serve as an impetus to further investigate the consumer welfare implications of labor cost increases in restaurants and other service industries. They might also note that not all effects of wage increases are positive; negative discussion of waiting time increases after the wage increase in independent restaurants.

We conclude with a discussion of areas where this work can be extended. First, we were unable to obtain robust data on employee satisfaction. Establishing a text-analysis multiattribute-based link between front-line employee satisfaction and perceived customer service is a logical next step, especially given the rich literature in this area. Second, we do not have data on differences in prepolicy wages in chains and independents, nor do we have information on franchising within chains in our data location and time period. Extending data on these dimensions would allow us to test more fine-grained employee and owner incentive effects. Therefore, we view our study as an exploratory analysis of perceived service quality using deep learning methods.

Acknowledgments

The authors are grateful for the thoughtful editorial guidance of Puneet Manchanda and three anonymous reviewers.

They also thank seminar participants at the Marshall School of Business (University of Southern California), Ohio State University, Georgetown University, and the 2018 INFORMS Marketing Science Conference for their helpful feedback on previous versions of this manuscript. Varun Narayan provided valuable research assistance. The third author thanks NUS Business School for financial support.

Endnotes

- See <https://www.bls.gov/opub/reports/minimum-wage/2016/home.htm>.
- It is plausible that despite large sample sizes, user-generated textual content might also suffer from biases; indeed, no market research technique is perfect. As such, we do not propose to replace traditional approaches, but instead to augment them with data that are available for free, in larger quantities, and over longer period of times.
- See <https://www.dol.gov/whd/state/tipped.htm>.
- See https://en.wikipedia.org/wiki/Santa_Clara_County,_California.
- See <https://web.archive.org/web/20191110125010/http://www.sanjoseca.gov/DocumentCenter/View/11397>.
- See <https://www.nefconsulting.com/paying-the-living-wage-benefits-business-as-well-as-employees/>.
- There are small chain restaurants that are all owned by the founder, but these are far outnumbered by national chains like McDonald's, Burger King, Subway, Chipotle, etc.
- See <https://www.consolidatedfoodservice.com/article/independent-vs-chain-restaurants-an-uphill-battle> and <https://www.msn.com/en-us/foodanddrink/restaurantsandnews/mom-and-pop-restaurants-are-beating-out-national-chains-here%E2%80%99s-how/ar-BBBKmhq> and <https://www.forbes.com/sites/caroltice/2012/07/23/how-mom-and-pop-restaurants-can-compete-with-the-big-chains/?sh=7646c1161079>.
- To the extent that some independent restaurants have multiple outlets and are therefore “chains” by our definition, our results can be seen as a lower bound on results of chains versus independents.
- See <https://www.dir.ca.gov/iwc/minimumwagehistory.htm>.
- Later, we show robustness of our results to a 15-month time window.

¹² See <https://aws.amazon.com/sagemaker/>.

¹³ In a pretest, we included neutral as a valence category. Very few reviews were labeled neutral for any attribute.

¹⁴ We make two assumptions to enable identification of f_t and λ'_t . First, we assume that the vector of factors in each time period is uncorrelated with the vector for every other time period. Second, we assume that $\Lambda \Lambda'$ is a diagonal matrix, that is, the factor of each unit is uncorrelated with every other unit.

¹⁵ We also experimented with ALBERT (released on December 20, 2019). We do not present all of the transformer model results, but focus on the two main specifications; details of other models are available from the authors.

¹⁶ Athey et al. (2018) reformulate the synthetic controls problem as one of missing information in a matrix. They use the nuclear norm based matrix completion approach to estimate the missing treatment values (counterfactuals). We conduct synthetic control analysis using this approach and find very similar estimates of average treatment effects.

¹⁷ Recall we see a decline in price discussion in chain restaurants. There is some evidence from Allegretto and Reich (2018) that chains price lower and increase prices more after the policy. Putting these two pieces of evidence together, the postpolicy selecting out by price-sensitive customers in chains seems possible. However, we do not see a rise in quality discussions by the remaining customers. As supported by agency theory, chains are less likely to improve quality; independents have both motives and ability to improve quality.

¹⁸ Wage hikes in Seattle led to an increase in less severe hygiene violations (Chakrabarti et al. 2020). Worsening of hygiene conditions in San Jose would suggest a negative effect of wage hikes on service perceptions; our results show a positive effect. Restaurant inspection reports for 2013 are unavailable to us.

References

- Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* 105(490):493–505.
- Aaronson D, French E (2007) Product market evidence on the employment effects of the minimum wage. *J. Labor Econom.* 25(1):167–200.
- Aaronson D, French E, McDonald J (2008) The minimum wage, restaurant prices, and labor market structure. *J. Human Resources* 43(3):688–720.
- Allegretto S, Reich M (2018) Are local minimum wages absorbed by price increases? Estimates from internet-based restaurant menus. *Indust. Labor Relations Rev.* 71(1):35–63.
- Anderson E, Fornell C, Rust R (1997) Customer satisfaction, productivity, and profitability: Differences between goods and services. *Marketing Sci.* 16(2):129–145.
- Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Sci.* 57(8):1485–1509.
- Athey S, Bayati M, Doudchenko N, Imbens G, Khosravi K (2018) Matrix completion methods for causal panel data models. NBER Working Paper No. 25132, National Bureau of Economic Research, Cambridge, MA.
- Autor DH, Manning A, Smith CL (2016) The role of the minimum wage in the evolution of U.S. wage inequality over three decades: A modest re-assessment. *Amer. Econom. J. Appl. Econom.* 8(1):58–99.
- Bell B, Machin S (2018) Minimum wages and firm value. *J. Labor Econom.* 36(1):159–195.
- Bhorat H, Kanbur R, Stanwix B (2014) Estimating the impact of minimum wages on employment, wages, and non-wage benefits: The case of agriculture in South Africa. *Amer. J. Agricultural Econom.* 96(5):1402–1419.
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learn. Res.* 3:993–1022.
- Brickley J, Dark F (1987) The choice of organizational form: The case of franchising. *J. Financial Econom.* 18(2):401–420.
- Card D, Krueger A (1994) Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *Amer. Econom. Rev.* 84(4):772–793.
- Chakrabarti S, Devaraj S, Patel P (2020) Minimum wage and restaurant hygiene violations: Evidence from Seattle. *Managerial Decision Econom.* 42(1):85–99.
- Cronin JJ, Taylor SA (1992) Measuring service quality: A reexamination and extension. *J. Marketing* 56(3):55–68.
- Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Tech.* (Association for Computational Linguistics, Stroudsburg, PA), 4171–4186.
- Dube A, Lester TW, Reich M (2010) Minimum wage effects across state borders: Estimates using contiguous counties. *Rev. Econom. Statist.* 92(4):945–964.
- Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Sci.* 31(3):493–520.
- Grossman JB (1983) The impact of the minimum wage on other wages. *J. Human Resources* 18(3):359–378.
- Guo T, Sriram S, Manchanda P (2020) “Let the sunshine in”: The impact of industry payment disclosure on physician prescription behavior. *Marketing Sci.* 39(3):516–539.
- International Labor Organization (2015) Minimum wage policy guide ILO. https://www.ilo.org/global/docs/WCMS_508566/lang-en/index.htm.
- Kamakura W, Mittal V, de Rosa F, Mazzon JA (2002) Assessing the service-profit chain. *Marketing Sci.* 21(3):294–317.
- Kim Y (2014) Convolutional neural networks for sentence classification. *Proc. 2014 Conf. Empirical Methods Natural Language Processing* (Association for Computational Linguistics, Stroudsburg, PA), 1746–1751.
- Lafontaine F (1992) Agency theory and franchising: Some empirical results. *RAND J. Econom.* 23(2):263–283.
- Lafontaine F, Shaw K (2005) Targeting managerial control: Evidence from franchising. *RAND J. Econom.* 36(1):131–150.
- Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *J. Marketing Res.* 48(5):881–894.
- Lin T, Goyal P, Girshick R, He H, Dollar P (2017) Focal loss for dense object detection. *IEEE Internat. Conf. Comput. Vision (ICCV)*, 2999–3007.
- Liu X, Lee D, Srinivasan K (2019a) Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *J. Marketing Res.* 56(6):918–943.
- Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019b) RoBERTa: A robustly optimized BERT pretraining approach. Preprint, submitted July 26, <https://arxiv.org/abs/1907.11692>.
- Lovett MJ, Peres R, Xu L (2019) Can your advertising really buy earned impressions? The effect of brand advertising on word of mouth. *Quant. Marketing Econom.* 17:215–255.
- Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta Protein Structure* 405(2):442–451.
- McBride S (2017) Minimum wage hikes are causing business to cut jobs. *Business Insider* (March 16), <http://www.businessinsider.com/minimum-wage-leads-to-job-losses-2017-3>.
- Mikolov T, Corrado G, Chen K, Dean J (2013a) Efficient estimation of word representations in vector space. Bengio Y, Lecun Y, eds. *Proc. Internat. Conf. Learn. Representations*, 1–12.

- Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013b) Distributed representations of words and phrases and their compositionality. Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds. *Proc. 26th Internat. Conf. Neural Inform. Processing Systems*, vol. 26 (Curran Associates, Red Hook, NY), 3111–3119.
- Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* 31(3):521–543.
- Neumark D, Wascher W (2000) A case study of the fast-food industry in New Jersey and Pennsylvania: Comment. *Amer. Econom. Rev.* 90(5):1362–1396.
- Okrent A, Alston J (2012) The demand for disaggregated food-away-from-home and food-at-home products in the United States. Economic Research Service Report No. 139, United States Department of Agriculture, Washington, DC.
- Parasuraman A, Zeithaml V, Berry L (1988) A multiple item scale for measuring consumer perceptions of service quality. *J. Retailing* 64(1):12–37.
- Parasuraman AA, Zeithaml VA, Berry LL (1991) Refinement and reassessment of the SERVQUAL instrument. *J. Retailing* 67(4):420–450.
- Pattabhiramaiah A, Sriram S, Manchanda P (2019) Paywalls: Monetizing online content. *J. Marketing* 83(2):19–36.
- Pennington J, Socher R, Manning CD (2014) GloVe: Global vectors for word representation. *Proc. 2014 Conf. Empirical Methods Natural Language Processing* (Association for Computational Linguistics, Stroudsburg, PA), 1532–1543.
- 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 Sci.* 36(5):726–746.
- Rensi E (2016) Thanks to “fight for \$15” minimum wage, McDonald’s unveils job-replacing self-service kiosks nationwide. *Forbes* (November 29), <https://www.forbes.com/sites/realspin/2016/11/29/thanks-to-fight-for-15-minimum-wage-mcdonalds-unveils-job-replacing-self-service-kiosks-nationwide/#19fde2214fbc>.
- Rouse J (2013) Paying the living wage benefits business as well as employees. *Wired* (November 4), <https://www.wired-gov.net/wg/wg-news-1.nsf/article/Paying+the+Living+Wage+benefits+business+as+well+as+employees+05112013123000?open>.
- Schneider B, White S, Paul M (1998) Linking service climate and customer perceptions of service quality: Test of a causal model. *J. Appl. Psych.* 83(2):150–163.
- Seipel T (2012) Minimum wage: Report sees overall negative effect of San Jose’s Measure D. *Mercury News* (October 8), <https://www.mercurynews.com/2012/10/08/minimum-wage-report-sees-overall-negative-effect-of-san-joses-measure-d/>.
- Selyukh A (2021) Biden to call for raising federal minimum wage to \$15 an hour. NPR (January 15), <https://www.npr.org/2021/01/15/957371280/biden-to-call-for-raising-federal-minimum-wage-to-15-an-hour>.
- Shapiro C, Stiglitz J (1984) Equilibrium unemployment as worker discipline device. *Amer. Econom. Rev.* 74(3):433–444.
- Shepard A (1993) Contractual form, retail price, and asset characteristics in gasoline retailing. *RAND J. Econom.* 24(1):58–77.
- Simon D, Gomez M, McLaughlin E, Wittink D (2009) Employee attitudes, customer satisfaction, and sales performance: Assessing the linkages in US grocery stores. *Managerial Decision Econom.* 30(1):27–41.
- Smith LN (2018) A disciplined approach to neural network hyperparameters: Part 1—learning rate, batch size, momentum, and weight decay. Preprint, submitted March 26, <https://arxiv.org/abs/1803.09820>.
- Sorenson T (1948) A method of establishing groups of equal amplitudes in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish Commons. *Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter* 5:1–34.
- Sun C, Qiu X, Xu Y, Huang X (2019) How to fine-tune BERT for text classification? Sun M, Huang X, Ji H, Liu Z, Liu Y, eds.. *Chinese Computational Linguistics, Lecture Notes in Computer Science*, vol. 11856 (Springer, Cham).
- Taylor K (2018) America’s fast food chains are contemplating replacing minimum wage workers with robots—And it could lead to a crisis. *Bus. Insider* (January 11), <http://www.businessinsider.com/minimum-wage-increases-spur-fast-food-chains-to-consider-automation-2018-1>.
- Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. *Marketing Sci.* 38(1):1–20.
- Tirunillai S, Tellis G (2014) Mining marketing meaning from chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Marketing Res.* 51(4):463–479.
- Vapnik VN (1995) *The Nature of Statistical Learning Theory* (Springer, New York).
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. von Luxburg U, Guyon I, Bengio S, Wallach HM, Fergus R, eds. *Advances in Neural Information Processing Systems*, vol. 31 (Curran Associates, Red Hook, NY), 6000–6010.
- Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Anal.* 25(1):57–76.
- Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV (2019) XLNet: Generalized autoregressive pretraining for language understanding. Preprint, submitted June 19, <https://arxiv.org/abs/1906.08237>.
- Zablah A, Carlson B, Donovan D, Maxham J, Brown T (2016) A cross-lagged test of the association between customer satisfaction and employee job satisfaction in a relational context. *J. Appl. Psych.* 101(5):743–755.