Inefficient Incentives and Nonprice Allocations: Experimental Evidence from Big-Box Restaurants

Sacha Kapoor[*]

January 10, 2020

Abstract

Queues are puzzling to economists because they are consistent with wasted profit in equilibrium. Standard rationales trace the puzzle to the pricing of goods. This article uses field experimental evidence from large-scale restaurants to trace the puzzle to the pricing of labor. The customary wage contract in these settings fosters congestion and longer queues because it can encourage workers to emphasize the quality rather than quantity of output. To study this problem, the field experiment pays waiters bonuses for customer volume on days with excess demand, in addition to the tips and hourly wages they customarily receive. The experimental contract shortens queues substantially, generating surplus gains for consumers with no discernible cost in terms of perceived service quality. Workers earn more via the bonuses and because they earn more in tips. Short-run profits increase by at least 49 percent. There is no discernible reduction in long-run profit. The firm reverted to the baseline contract on excess demand days after many months of evidence, even after acknowledging the gains from the experimental contract. The evidence suggests the puzzle may partly be explained by inefficient wage contracting.

JEL: J33, L81, M52

Keywords: Uniform Pricing, Congestion Externalities, Nonprice Rationing, Queues, Wage Contracts, Profit Maximization, Restaurants

Why do businesses use queues to allocate goods among consumers? Queues are suggestive of excess demand, where the goods are priced below market clearing levels, and where a capacity constrained business can increase profit simply by raising prices. Gary Becker (1991) famously identified this situation as puzzling because it fit poorly with conventional profit-maximizing pricing models at the time. Becker rationalized queues via a model of social influence, where consumers value goods more when the goods are valued by other consumers, aggregate demand is consequently *increasing* in price over some range, and prices are relatively high rather than relatively low. In his model queues can coexist with high profit-maximizing prices in equilibrium. While his argument stimulated scholarship into the role of social interactions in markets, its applicability outside of markets for fashionable luxury goods seems limited.

An alternative argument is consumers perceive price increases during supply shortages as unfair relative to queues, and that businesses keep prices fixed because aggrieved consumers will punish them later on (Rotemberg, 2005). This argument applies to a broader class of goods because it supports the coexistence of queues and low profit-maximizing prices in equilibrium. While the argument is intuitive, empirical support for it is mixed to nonexistent. For example, in surveys of restaurant customers, Kimes and Wirtz (2003) found consumers perceive time-of-day pricing as fair and are neutral on the fairness of day-of-week pricing. Furthermore, as Becker (1991) himself notes, the long run implications of these perceptions are unclear. If prices are situational and consumers know this, they should be able to sort into their strategically preferred situation with relative ease.

A third argument recognizes the opportunity costs of the space used to serve customers (Lott and Roberts, 1991). Longer service times per customer lengthen queues and expected wait times for prospective customers. The congestion externality increases the opportunity cost of serving pre-existing customers and the implicit rent paid to these customers for every extra minute they stay. The implicit rental rates are then capitalized into goods prices.[1] This justification for queues is appealing because it better suits the realities of these production settings. It is problematic because it implies goods prices track implicit rental rates, varying smoothly therefore with queues and expected wait times. But this is opposite to what is observed in many retail segments, as many businesses practice uniform pricing (DellaVigna and Gentzkow, 2019), charging the same price for the same good in all situations.

The above arguments all tie the existence of queues to the pricing of goods. Neither of the arguments ties their existence to the pricing of the labor used in serving consumers. This is surprising because average service times are outcomes of repeated interactions between customers and workers, making worker behavior a potentially important driver of congestion externalities and queues in turn. From this perspective, the puzzle may be as much about why wage contracts are not used to better exploit queues as it is about why goods pricing is not used. The present article explores this possibility empirically, specifically considering the role of wage contracts towards queues and efficiency using (field) experimental evidence from big-box restaurants.

Big-box restaurants are full-service restaurants that belong to retail chains where the focus

---

[1]See (Lott and Roberts, 1991) for a discussion of the costs of charging customers explicitly for the service time.

is on mass production of standardized packages of goods and services. Standardization implies relatively small differences in revenue per customer across customers and that customer volume is consequently the primary instrument for increasing revenue. The context is useful for several reasons. First, queues are commonplace. Second, the customary wage contract - tips and hourly wages - generates incentives for customer volume that are too weak (or incentives for customer service that are too strong) from the perspective of the firm.[2] There are *ex ante* reasons therefore to suspect the customary wage contract as a primary driver of long queues and inefficiency.

To investigate whether this is the case, I engineered an exogenous change in the customary wage contract at a big-box restaurant on excess demand days, wherein waiters were paid bonuses for their customer volume in addition to their tips and hourly wages. The experiment is complemented by rich administrative data from a couple of franchises from a big-box chain. The data facilitates an in depth investigation into the causal effect of the alternative contract on consumers, workers, and the firm. The customer data includes a proxy for queue length, the tips they pay, and whether they return at a later date. The worker data includes rich detail about their activities and performance. The firm data includes profit metrics for both the short and longer run.

The evidence shows a sharp reduction in excess demand under the experimental contract. This is suggestive of surplus gains for both low and high value consumers. Low value consumers gain because of reallocation from their outside option (another restaurant, eating at home *e.g.*) to a more preferred option. High value consumers, who would have stayed under the customary contract, gain because they save on costly waiting time. While I find some suggestive evidence of a reduction in the service quality workers deliver, there is no discernible effect on tip rates, consistent with a low cost to the surplus gains in terms of overall service quality. No tip rate effect may alternatively be attributable to shorter wait times or faster service, which can compensate customers for marginal reductions in other aspects of personal service quality.

The experimental contract increased worker earnings by 10 percent. Workers earned more via the bonuses and because they earn more in tips. I exploit some dynamics in the experimental design to provide evidence of the experimental contract compensating them for costly effort. The evidence implies nonnegative surplus gains for workers during the experiment.

The experimental contract increased firm revenue by 10 percent. It increased short-run profit by at least 49 percent. I find no evidence of an effect on repeat business (up or down) and thus on profit in the months following the treatment. Taken together, the results suggests that the experimental contract may be Pareto improving.

If this is the case, then why use the customary contract on excess demand days initially? I explore several justifications. I investigate whether the customary contract was initially supplemented by informal arrangements. I show informal arrangements are operational, but that the gains are small relative to the experimental contract. I also assess whether the firm generally

---

[2]Recent legislation tells us just how "customary" is the contract. In 2008 the U.S. federal government introduced legislation that integrates tips in the calculation of the minimum wage. Under the new legislation, the firm need only compensate the worker up to the point where tips plus the hourly wage equals the mandated minimum. Before the new legislation, the hourly wage had to equal the mandated minimum.

makes good choices. I show they do, and are particularly skilled at keeping costs down, suggesting that something other than poor decision making explains the initial use of the customary contract.

Despite generally making good choices, and becoming fully aware of a more profitable alternative, the firm decided against permanent adoption of the experimental contract on excess demand days. What explains this decision? There are two factors, according to the CEO. The first is the money. The dollar cost was $200 per shift (20 workers $\times$ $ 10 per worker), which was financed externally by a grant that aims to bridge the gap between academia and the business community. The firm would rather not finance it themselves, despite knowing that it generates $2000 more in revenue per shift. The second factor is "workers earn enough already"; they earn a fair wage that obliges them to act in the firm's interests. These factors suggest the CEO has an aversion to material spending, particularly when it pays workers more than their perceived worth.

The baseline experimental results contribute to a literature that examines the effects of individual incentive pay (Bellemare and Shearer, 2013; Copeland and Monnet, 2009; Jayaraman et al., 2016; Lazear, 2000; Paarsch and Shearer, 1999; Shearer, 2004),[3] and that focuses more specifically on the effects in jobs where workers carry out multiple tasks (Muralidharan and Sundararam, 2011), (Dumont et al., 2008), and (Brickley and Zimmerman, 2001). Unlike previous studies, this article examines the effects of a scheme that effectively pays workers to *lower* the quality of their output. It shows that such a scheme can make workers, the firm, and *consumers* no worse off.

The decision to not pay for customer volume in the pre-experimental period alone begs questions about contractual efficiency, as the decision violates the informativeness principle (Holmstrom, 1979, 1982; Shavell, 1979), wherein all easy-to-observe signals of worker effort are written into wage contracts.[4] Such concerns are reinforced by the decision to not pay for customer volume in the post-experimental period. In this regard, the article contributes to a literature that questions the optimality of decisions by the firm. Early contributions debated the assumption that firms maximize profit (Hall and Hitch, 1939; Machlup, 1946). Recent studies have examined the optimality of behavior by professional athletes or their coaches (Abramitzky et al., 2012; Romer, 2006), by sellers of tickets to sporting events (Sweeting, 2012), cable companies (Byrne, 2015), firms in spot markets for electricity (Hortaçsu and Puller, 2008), rental car companies (Cho and Rust, 2010). This article examines the optimality of wage contracting in big-box restaurants.

1   Context

The franchises are part of a long chain of higher end big-box restaurants. They are only open for dinner and located in the suburbs of Toronto.[5] The franchises are isolated from competing firms,

---

[3]For prominent contributions to the literature that focuses on team-based incentives, see (Bandiera, Barankay, and Rasul, 2005) and (Hamilton, Nickerson, and Owan, 2003).

[4]The economics literature has several explanations for the use of contracts that seem to violate the principle, including custom (Allen and Lueck, 2009; Young and Burke, 2001), imitation of what works (Alchian, 1950), and concerns for fairness (Fehr and Schmidt, 2004; Fehr, Klein, and Schmidt, 2007).

[5]Patrons are drawn from a relatively homogeneous population, consisting mostly of married people (nearly two-thirds) of homogenous origins (more than two-thirds). **Source:** Statistics Canada.

having their own dedicated building, such that consumers must drive to get there. Consumers cannot be solicited from the street therefore. The franchises serve approximately 2500-3000 customers on average per week.

Prices and product offerings are set centrally by the chain. They are the same for all days of the week. They are fixed over an extended periods of time, usually lasting more than a year. Prices and product offerings are the same for all stores within the same broad geographic area.

Approximately half the weekly customer volume comes from busy days (Friday and Saturday evenings). Busy days almost always have queues for seating during the high season, which begins in September and ends in May. Upwards of 100 customers can walk away in a single night after learning the expected wait time for a table. The revenue loss from having 100 customers walk away is substantial. Much of this amounts to lost profit because labor costs are effectively fixed.

Figures 1(a) and 1(b) uses data from busy and slow days during 2006 and 2007 (pre-experiment) to illustrate the importance of customer volume. Figure 1(a) plots revenue per customer against customer volume, showing it holds constant at roughly 35 dollars, and implying a relatively small opportunity cost to increasing volume. Figure 1(b) plots the wage bill per customer against customer volume, showing labor is essentially a fixed cost, as the average wage bill declines steeply before stabilizing below 2 dollars per customer. It pays therefore to allocate workers to shifts in a way that spreads the wage across as many customers as possible.

Organization of Production. Each franchise has a general manager, 3-4 middle managers, 30-40 waiters (henceforth workers), and a support staff (greeters, busboys, bartenders, and chefs).

The primary tasks of workers are assisting customers with the menu, physically delivering menu items, removing clutter from tables when necessary, and handling payment. The chain gives workers guidelines for interactions with customers, including time limits on when they should be greeted, the delivery of products, and delivery of the bill. The chain also has guidelines on the order in which items should be sold: drinks, then appetizers and salads, the main course, and then desserts. Outside the guidelines, workers have discretion over their interactions with customers.[6]

A well-defined formula matches workers with customers. New customer arrivals are seated immediately if there is no queue. They are assigned to workers on the basis of worker start times. The worker who starts earliest gets the first customers. The next earliest starter gets the second customers. This continues until the restaurants are at full capacity. Once there is a queue, the customer is matched to the first available worker. In this case, matches implicitly depend on worker table assignments, as some tables turn over faster than others. The formula makes customer-worker matches quasi-random conditional on worker start times and table assignments.

Workers are assigned table sections by managers before the restaurants open. Table sections are generally fixed for the duration of a shift and are not shared among workers. Table sections usually consist of 2-4 tables with 10-16 seats. Worker start times are assigned by managers before the start of each workweek. Start times are usually staggered at 15-30 minute intervals, ranging

---

[6]See (Braverman and Stiglitz, 1982) for a model where workers can choose the production technique. See (Courty and Marschke, 2004), (Oyer, 1998), (Asch, 1990) for studies which examine incentive problems relating to discretion.

from 3:30pm until approximately 6:30pm. Worker stop times are largely based on a fixed set of rules. Once queues disappear, some workers get to leave. Workers leave in the same order that they started. The remaining workers stay until the last customer has left.

Workers declare their preferences over shifts in advance of each work week. Managers accommodate these preferences, rejecting them only on occasion, when there is a shortage or surplus of willing workers. Shift and start time assignments are posted before the workweek begins.

There are 2-4 managers on duty per shift. Table and customer assignments are usually done in consultation with other managers and support staff (greeters, *e.g.*). Managers have direct authority over busboys, telling them which tables need to be bussed and when. Managers rotate through these duties. They earn a percentage of the firm's accounting profits.

Worker Incentives. Workers are paid tips and a fixed hourly wage, equaling the minimum wage for servers in Ontario. Tips are at the discretion of customers and effectively proportional to the revenue each worker generates. Tips are not shared with other waiters, but they are shared with the support staff. Workers transfer 4 percent of the revenue they generate to the support staff at the end of each shift.[7]

To fix ideas while keeping things simple, assume tips equal $\tau r n$, where τ is the tip rate, r is revenue per customer, and n is customer volume. Workers have several channels for increasing tip earnings at a cost: they can try to increase the tip rate by, for example, spending more time socializing with customers; they can try to generate more revenue from customers by, for example, convincing them to purchase add-on goods; they can try to serve more customers by, for example, moving faster or discouraging purchases of desserts.

In the opinion of managers, the busy-day losses stem from worker tendencies to focus on customer service. In theory, the problem can stem from the formal wage contract. The contract is problematic either because it overemphasizes personal service (at the expense of customer volume) or because it provides weak incentives for customer volume. The contract can overemphasize personal service because better service can yield higher tip rates and more revenue from each customer. Better service means less to the firms because they do not accrue, at least not fully, the gains from higher tip rates. Because better service often comes at the expense of customer volume, the worker serves fewer customers than the firms would prefer. The contract can provide weak incentives if tips are simply insufficient for encouraging the customer volume the firm prefers.

Our problem departs from classic incentive problems in an important way. In the classic multitasking incentive problem Holmstrom and Milgrom (1991), the contract pays the worker for performance in the easy-to-measure task (*e.g.* quantities), unless it comes at the expense of performance in the hard-to-measure task (*e.g.* quality).[8] In our setting the customary contract pays the worker relatively more for their performance in the hard-to-measure task: service quality. This could come at the expense of their performance in the easy-to-measure task: customer volume. In the classic problem the firm wants more focus on the hard-to-measure task. They would pay for it

---

[7]It is worth noting that tipping helps firm attract and retain workers. See (Wessels, 1997) for details.

[8]See (Baker, 1992) for a different take on multitasking.

5

but, because the task is hard measure, are unable to. In our setting the firm wants more focus on the easy-to-measure task. They could easily pay for it, but choose not to.

The firm has incentivized customer volume historically but has typically avoided schemes which involve large outlays of money. They often run contests where the prize is in-kind (a free drink) or reward workers with earlier start times and better tables (for customer volume) if they have proclivity for generating customer volume without sacrificing much service quality. Further to this point, the firm regularly reminds workers of the importance of customer volume, even going so far as to discourage them from offering customers the dessert menu, encouraging them instead to politely deliver the bill once the main course is done.

That the goal is well-known, and even incentivized, reinforces the research design. A standard concern with experiments inside firms relates to whether treatment effects reflect implicit incentives to appease bosses, rather than the explicit incentives experiments typically offer. If the worker believes treatment non-response can cost them the job, for example, then treatment effects will reflect career concerns as well as the response to the bonus. This is less of a concern in our setting because workers knew the goal well before the experiment began.[9]

2 Experimental Contract

The bonus amounted to $b(n-n^*)I(n > n^*)$, where b is the bonus rate, n^* is a performance standard for customer volume, and I is the indicator function. The bonus rate was chosen so that a worker who exceeded the performance standard by one standard deviation would earn between \$20 and \$30, or more than 10% of average daily earnings.[10] The performance measure and standard, n and n^*, were both adjusted for average shift length (hours worked) and average section size (the number of seats a worker is responsible for). The performance standard was calculated using historical data.[11] It helps prevent workers from earning the bonus without changing their behavior.

The experiment was run at one of the franchises during the high season (September until May). The experiment had two treatment blocks. In the first block, workers at the treated franchise were paid bonuses on every Friday and Saturday in November and January of the 2009-2010 season. In the second block, treated-franchise workers were paid bonuses on every Friday and Saturday in late April and May of 2010. The second treatment block differs from the first in that the performance standard was tailored to the individual worker on the basis of the table section they were assigned that evening. The standard was increased if the worker's tables facilitated customer

---

[9]The contests help with another standard concern, placebo effects, where any explicit incentive would generate the same results. The contests are useful because they always come with explicit incentives, but are relatively ineffective at increasing customer volume. That said, placebo effects are a minor concern, as the primary goal here has little relation to the intricacies of the chosen experimental contract.

[10]The rate was set at \$3 for every tenth of a point above the performance standard.

[11]For each table, the number of customers served per shift was calculated and then divided by the average work hours for workers. The calculation yields values of 2.11 customers per hour for Fridays and 2.72 customers per hour for Saturdays. These numbers were divided by the number of seats at the table, and then averaged over all tables. This calculation yields values of 0.4 customers per hour per seat for Fridays and 0.41 for Saturdays. Ultimately, the experiment pays workers for the rate at which they turn over the seats in their section.

volume historically, and decreased if the tables impeded customer volume historically.

In early October of 2009, the CEO informed workers that someone would be conducting interviews at the firm. The ethics review board requested that I identify myself as an unpaid researcher from the University of Toronto, inform workers that the general purpose of the interviews was to understand factors underlying wait times for customers, and request their participation in the study. The CEO asked me to handle administration of the experiment.

At the time of the interviews, workers were unaware that they were to receive bonuses for good performance starting in November of 2009. During this month, I also made daily appearances at the firm. The purpose was to familiarize workers with the experimenter and to reduce the chances of experimental outcomes reflecting efforts to appease the experimenter. My presence during the control period helps because if workers were trying to appease me, by for example undertaking activities that usually reduce wait times, they were probably already doing it before the experimental contract was implemented.

In November of 2009, workers were informed about the performance bonuses on the day of (after they arrived) their previously scheduled shift. To smooth the transition to the experimental contract, the bonuses were introduced as an otherwise typical contest. To ensure that workers understood the scheme, they were asked to demonstrate their understanding in the context of several hypothetical examples before the start of their shift. To ensure that one dollar in tip earnings (which are paid immediately) was equivalent to one dollar in experimental earnings, workers were paid (privately) once the shift was completed. To minimize the influence of sorting on the empirical results, the length of the treatment period was not revealed to workers. Each worker's experience followed a similar pattern on subsequent treated days.

While workers were unaware of the start or end date for the experimental contract, they may form and revise expectations concerning whether and when the contract is available. These expectations can affect worker and manager decisions pertaining to the shifts workers work, and thereby compromise estimates of the pure incentive effect of the experiment. Online Appendix Table A.1.1 uses data on worker bids for shifts and their acceptance by the firm to speak to the concern that the experimental contract altered the matching of workers to shifts (*cf.* Ackerberg and Botticini (2002)). It shows that the treatment effect on bids is weak, in line with what I was told by workers, namely that other commitments (other jobs, school, family) drive decisions about whether and when to work. It also shows a weak effect on accepted bids, in line with what the CEO instructed managers to do during the study, namely to go about their business as they normally would.

The transfer of tips to the support staff raises questions about the effect of the experimental contract on the support staff, as well as the implications for the experimental results. There are two issues. The first relates to whether the experimental contract elicited a direct response from the support. They had incentive to respond because the contract can increase revenue, and because more revenue implies more transfer income. The second relates to whether workers paid support staff side payments in exchange for help with customer volume.

The support staff showed little to no interest in the study, many were completely unaware of

the particulars of the experiment, having said things like "whatever you're doing with the waiters, it is working". They at best responded indirectly to the extra volume the experimental contract generated.[12] The scale makes side payments implausible in practice because it would lead to significant congestion. The firm would moreover fire any employee caught doing this.

Note that the research design either treats all workers in a shift or none. Complementarity in production stopped us from randomizing within shifts. If the treatment motivates workers to move faster, and the extra speed alleviates congestion, control workers would find it easier to serve more customers. The congestion would then lead to underestimates of the true effect of the bonuses.

3 Descriptive Statistics

Managers and greeters log detailed records of the number of customers who leave after learning the expected wait time. I use this variable to measure excess demand. The mean and standard deviation are 33.9 and 34.8, respectively. The maximum is 180. Note that the number of leavers measures excess demand with error, as consumers may leave before learning the expected wait time, perhaps because they infer lengthy wait times upon observing long queues. Later I will discuss the implications on treatment effect estimates for the number of leavers.

I will sporadically make use of tip rate information from bill-level data, which is based on every bill paid by either credit or debit (more than 75 percent of bills). I will estimate the relationship between tip rates and worker effort and time use. These regressions can be problematic if the payment method varies systematically with personal attributes of customers. This is a problem if, for example, the attributes are used by the firm to assign pay-by-cash customers to poor-service workers, and pay-by-cash customers generally tip less. Then we could overstate the relationship between tip rates and worker efforts and time use. Quasi-random matching stops these selective assignments from happening.

Tip rates and service quality measures from the pre-experimental period are summarized in Online Appendix Table A.1.2. The tip rate mean is 14.4 percent with a standard deviation of 4.5 percentage points. More than 7 items are sold at a price per item of just over 6 dollars. Bills take about 90 minutes on average, with about 20 minutes between bill settlement and the first order on the next bill, and 20 minutes between the last dessert order and bill settlement (time to linger).[13] As I will explain, these five variables (average price, items sold, time with and between customers, and time to linger) are reasonable measures of worker inputs when aggregated to the daily level.

Figures 2(a) and 2(b) visualizes the effect on tip earnings and revenue, our key outcomes for workers and the firm. Increases in tip earnings and revenue are present in comparisons of outcomes from just the 2009-2010 season (the red solid line), and in comparisons of differences in outcomes (between October and the other months) from 2009-2010 with the same differences from 2008-

---

[12]In part this is because for the support staff the pace of the work is outside their control. A computer tells the kitchen, for example, which goods to produce first.

[13]The time use measures are noisy by definition. Ideally, we would know when the customer sits down, and when the bill is delivered.

2009. While year-over-year comparisons help account for common seasonal patterns across years, it is unclear that they allow for a causal interpretation. They do if unobserved changes in outcomes, from October to November and January, are the same in 2009-2010 as they were in 2008-2009.[14] The economic crisis that began in 2008 raises doubts about whether this was the case. The crisis might have led to a smaller change in 2008-2009 relative to 2009-2010, as big spenders may have stopped visiting the franchise in November of 2008 and January of 2009, or the customers who continued to visit spent less than usual. Either way, the crisis could bias the estimates upwards.[15]

Data from a comparable franchise helps with the concern. Both franchises were opened by the same ownership group. The franchises have identical menus, and thus prices, compensation schemes, variable costs, procedures, organizational structures, and even similar physical layouts. The franchises are located in adjacent subdivisions, being about 30 minutes apart by car. The main difference between the franchises is scale, as the treated franchise is bigger. Online Appendix A.1.1 explores the similarities and difference in greater detail, explaining and showing how the scale only generates intercept differences between the franchises. Ultimately, with the control franchise, the treated franchise this year and last (and 2006-07), our full sample consists of about 120 workers, and more than 4400 worker-calendar date observations.[16]

Table 1 summarizes the unconditional effects explicitly. Rows 1 and 2 describe the effects on the trade off between customer volume and revenue per customer. Rows 3 and 4 describe the effect on the money payoffs for workers and the firm. Rows 5 and 6 describe earnings from the experiment, as well as the share of workers who earned the bonus. Moving left to right, the table summarizes the sources of variation we will use to interpret the effects causally. Note that the treated franchise targets 20 workers per shift on days with excess demand. The control franchise has fewer workers per shift because, as noted elsewhere, the primary difference between the franchises is scale.

Table 1 shows the effects on revenue and tip earnings are robust. The revenue increase ranges from $92 and $163. The increase in tip earnings ranges from $12-15 dollars. The estimates imply a more than 10% gain in money payoffs for workers and the firm.

Row 2 shows workers serve more customers. Around 3 more in November 2009 and January 2010 than they did in October of 2009. Moreso when the difference is compared with the difference for 2008-2009. The estimate varies between 3.3 and 3.5, depending on the sample we use. It is always statistically significant. The effect on revenue per customer is less clear. Columns 3 and 4 imply workers generate less revenue from each customer served. Column 5 shows an increase in revenue per customer. Two of the estimates are statistically insignificant. The ambiguity ultimately highlights the importance of data from multiple franchises.

Outcomes by the calendar date provide good measures of worker behavior because customer idiosyncrasies are averaged out at this level of aggregation. The idiosyncrasies are averaged out

---

[14]The study omits data from December of 2009 because waits for seating are unusual during this month. In December, demand is more evenly spread across all days of the week.

[15]Employee turnover was low for the period under study (before, during, and after the treatment). More than 75 percent of employees have shifts in both years.

[16]Our data from the 2006-07 season is not as rich as the other data. It excludes tip rates, table assignments, as well as detailed information about worker inputs.

because of quasi-random matching of customers with workers, and because on days with excess demand there are lots of customers. The matching process implies that two workers, with open tables, are equally likely to draw a good customer. Because each worker sees lots of customers, differences in draws balance out over the course of a shift. The average customer is less likely to differ systematically across workers in the same shift.

4   Consumer Surplus

I assume the number of leavers at franchise f on date d are generated by

$$Leavers_{fd} = \alpha + \beta(Volume\ Bonus)_{fd} + \mathbf{X}_{fd}\mathbf{\Gamma} + \gamma_d + \varepsilon_{fd},$$

$(Volume\ Bonus)_{fd}$ indicates the availability of the performance incentive. $\mathbf{X}_{fd}$ includes an indicator for tailored performance standards and the total number of customer arrivals. Arrivals helps us control for level differences in consumer demand across franchises, as depicted in Online Appendix Figure A.1.1. γ_d includes fixed effects for the day of the week, week, season, as well as weather controls. The influence of the weather is the same for the two franchises because they are located within 1 hours driving distance of each other.

$Leavers_{fd}$ measures the true number of leavers with error. It counts the number of consumers who left after hearing the wait time, not the number who left before hearing the wait time. Miscounts are common. Both types of measurement error are more likely during busy service periods. While the error should correlate positively with the true number of leavers, the implications for estimates of β is ambiguous, differing depending on how the error varies with the experimental contract. However, if the experimental contract decreased miscounts or the number who left before hearing the wait time, perhaps because because consumers are seated more quickly, then this measurement error will lead to underestimates of the effect on leavers.

Estimates are found in Table 2. Columns 2 differs from Column 1 in that it includes fixed effects for the franchise. A comparison of the columns tells us how much of the overall effect comes from within the treated franchise (because the control franchise is never treated) and from differences between the treated and untreated franchises. The comparison provides suggestive evidence on the role of scale in the franchise-level response to the experimental contract because, as noted in the last section and Online Appendix A.1.1, scale is the primary difference between the franchises. I base inference on robust standard errors, but the inferences are robust to several other approaches, including the wild cluster bootstrap clustered at the franchise level.

Column 1 shows 15.88 fewer leavers under the experimental contract ($p < 0.01$), amounting to a 47 percent reduction relative to a mean of 34. This evidence is consistent with the idea that the customary contract is an important driver of congestion and queues. Column 2 shows 7.48 fewer leavers once we condition on the franchise ($p > 0.1$). Approximately half the overall decrease can be attributed therefore to the within treated franchise response. The other half comes from between franchise variation, generated by differences in scale for example.

10

Shorter queues should generate surplus gains for low value consumers, as some of these were reallocated from their outside option (another restaurant, eating at home *e.g.*) to a more preferred option. Shorter queues should also generate surplus gains for high value consumers, who would have stayed even with longer queues. These consumers gain via savings on costly waiting time.

To evaluate the potential cost of the surplus gains, I estimate

$$(Tip\ Rate)_b = \beta_1 q_b + \beta_2 T_b + \beta_3 T_b q_b + \mathbf{X}_b \mathbf{\Gamma} + \varepsilon_b,$$

where b indexes the bill.[17] Note that bills can differ depending on the worker who handles it, the customer, franchise, date, and table. T_b indicates whether the bill was handled at the treated franchise on a treated day. X_b includes fixed effects for time of day, day of the week, calendar week, and franchise. q_b is a compact representation of effort and time use measures. The interaction terms reflect the indirect experimental effect that operated via changes in effort and time use. Note that the specifications exclude time between customers and time to linger in the interest of parsimony and because their coefficient estimates are fragile statistically.

Estimates are found in Table 3. Column 1 gives the base effect, Column 2 includes effort and time use, and Column 3 their interaction with the treatment. The time and effort measures are standardized, their coefficients measure the effect of a one standard deviation increase. Tip rates increased by about a 10th of a percentage point, but this increase is not statistically different from 0. Column 3 suggests customers became less sensitive to the service measures. The coefficient on average price went from -0.14 to 0.03 (a statistical 0). The coefficient on quantities (items sold) fell from 0.24 to 0.13. There was no change in the coefficient for time with the customer.

A couple of factors explain the negligible effect on tip rates and customers' diminished sensitivity to prices and quantities. First, the bonuses made workers move faster. This is consistent with the reduction in the time between customers and observations made by managers and support staff. Second, customers appeared to spend less time waiting. By shortening wait times, the experiment may have improved customer perceptions of the overall service quality, and may even have improved customer perceptions of the service quality of the worker who served them. In either case, the results suggest that the cost to the surplus gains for consumers was marginal at most.

5 Worker Surplus

I assume outcomes y_{ifd} for worker i is generated by:

$$y_{ifd} = \alpha_i + \beta(Volume\ Bonus)_{fd} + \mathbf{X}_{ifd}\mathbf{\Gamma} + \gamma_d + \varepsilon_{ifd}, \tag{1}$$

---

[17]A natural concern with tip rates as a dependent variable relates to whether normal distributions are reasonable approximations for the distributions of our test statistics. Accordingly, I ran the residuals through several diagnostics tests. None of the tests implied normality. Because of this, I estimated the specification at various quantiles of the conditional tip rate distribution. The estimates have no substantive effect on my conclusions.

α_i is a worker fixed effect, γ_d includes fixed effects for the day of the week, the calendar week, and season, and $\mathbf{X}_{ifd}$ includes a count of the days worked in our sample, the average days in sample for coworkers who work that day, the worker's start time, and fixed effects for the tables they were assigned. Note that days in sample excludes days from after the start of the treatment.

γ_d reflects trends that influence the franchises in similar ways, such as the weather. Coworker days in sample proxies for the help from others (Drago and Garvey, 1998; Itoh, 1991). It presumes seasoned coworkers are better at helping while, at the same time, balancing their own responsibilities. Start times and table fixed effects help account for differential opportunities to produce and earn more. Early starts and better tables (*e.g.* ones near the kitchen) give the worker better opportunities to serve more customers. Some tables afford customers greater privacy and comfort.

The parameter of interest is β. β should be positive when the dependent variable is customer volume. β should be negative when the dependent variable is revenue per customer or service quality, if there is a trade off with customer volume. β can be positive or negative when the dependent variable is total revenue.

Table 4 reports estimates for revenue and tip earnings. Column 1 reports results for specifications that only include worker fixed effects. The remaining columns illustrate the influences of controls. Columns 1 through 4, and 6 through 9, make no distinction between the two treatments, even though earning a bonus was more difficult when workers had individual performance standards. Columns 5 and 10 separate the effects of the two treatments, and thus of the standard and bonus rate.

Columns 5 and 10 have three notable patterns. First, making no distinction draws down estimates of the incentive effect. From Column 4 to 5, the effect on revenue increases from \$88 to \$113. From Column 9 to 10, the effect on earnings increases from \$6 to \$9. Second, tailoring the standard decreases output and earnings. It decreases revenue by \$78, and tip earnings by \$10. That said, the estimates imply workers and the firm were no worse off monetarily.

My preferred specification for the remainder of this section includes binary variables for the performance incentive and tailored standard, fixed effects for the worker, day, week, and season. It excludes inconsequential variables, like days in sample (own and peers). It excludes start times and table fixed effects because they are inconsequential and bad controls, reflecting the direct treatment effects on workers and indirect effects that operate through managerial behavior.

Table 5 examines the effects on more direct measures of worker behavior: the top panel reports the effect on revenue per customer and customer volume; the middle the effect on revenue per item (average price), items sold (quantities), time with and between customers; the third the effect on base good and add on sales, and on the time to linger. The second and third panels are nested in accordance with the input or task the variable affects.

Workers served 2.5 ($p < 0.01$) more customers under the experimental contract, but generated 68 cents less ($p > 0.1$) revenue from each. The left nest of the middle panel (under Revenue per customer) shows workers sell 0.1 fewer items to each customer they serve ($p < 0.1$). The same nest of the third panel shows workers sell 0.7 fewer items in the part of the bundle that most customers

12

order ($p < 0.05$), and 0.3 fewer dessert items ($p < 0.1$). The estimates imply the trade off from serving more customers includes a small change in the revenue they generate from each.

The right nest of the second panel shows the experimental contract induces workers to spend 3.4 fewer minutes with customers ($p < 0.01$). The right nests of the second and third panels show less time between customers, and less time for customers to linger, but that $p > 0.1$ in both cases. The estimates imply a negative relationship between customer volume and time per customer.

Table 6 evaluates whether these patterns are consistent with reductions in service quality. The bottom row correlates the tip rate with the time and effort measures. The estimates come from Column 3 of Online Appendix Table A.1.3, which shows the estimates are largely robust to the inclusion of various controls. The estimates there help to some extent with concerns that "time" and "effort" are equilibrium outcomes. To make the comparison obvious the top row replicates relevant treatment effects from Table 5.

Customers tip less when they pay higher prices. They tip more when they consume more items, the bill takes more time, there is more time to linger, and when there is more time between customers. The signs are the opposite of the signs for the treatment effects on these inputs. Opposite signs support a reduction in service quality under the experimental contract.

Figure 3 visualizes the evolution of revenue in the treated franchise in the treated and control seasons (2009-2010 and 2008-2009). In between the two treatment blocks, revenue returned to pre-treatment levels, and to its levels from the same period last year. There was then a large increase during the second treatment block, similar to the first block increase.

Figure 3 has three implications. First, it implies the results are not a consequence of transitory responses to treatment, including responses that arise because of Hawthorne, placebo, or experimenter demand effects. Second, it implies the problem with workers is not that they unaware of greater earnings opportunities via less time with individual customers. If they were previously unaware, and learned it from the experiment, then they should continue to deliver worse service after the contract is taken away. Third, in suggesting that the experimental contract is compensating workers for their effort costs, it implies workers are no worse off under the experimental contract, at least in relation to a worker preference that depends on money earnings and effort costs.

6   Profits

The profits from a shift are $\Pi = (1 - \psi)(\sum R) - F$, where the sum is taken over all workers, R is revenue, F are fixed costs, and ψ denotes the share of revenue allocated to variable (food) inputs. Our primary interest is in measuring the percentage change in profits

$$\%\Delta\Pi = 100\frac{N[(1 - \psi)\Delta R - I] - \Delta F}{(1 - \psi)NR - F} \approx 100\frac{(1 - \psi)\Delta R - I}{\delta R}$$

where I is the experimental incentive cost per worker, δ is the share of every dollar earned that becomes profit (after accounting for fixed costs). The owner told me the variable margin ψ_p on each

product category p. We will replace ψ with the maximum across all categories. The owner also told me some recent values for δ. We will take the minimum of these values because the minimum is a lower bound on the profitability of each dollar earned. The values for δ were similar across product categories - the choice for the lower bound has little to no effect on the calculation. After plugging in the reported margins, we obtain $\%\Delta\Pi \approx 49\%$.[18] This section elaborates on why 49% might provide a good approximation for the change in overall profit.

A revealed preference argument implies that we should not expect a significant negative effect on repeat business.[19] Why would the owner let us conduct the experiment if he expected otherwise? I asked him about it. This is what he said

> Consumers are smart. They sort themselves into days that best suit their needs. Regulars avoid busy days, as they prefer more attention from the worker. On busy days we get one-timers, consumers who dine out once a year and who prefer a place that is lively and busy.

His statement has two messages. First, busy-day customers are unlikely to return, at least not for a while, no matter the service quality. Second, because of their inexperience, busy-day customers find it more difficult to detect slight changes in service quality. In line with his statement, when I asked if we could pay for customer volume on slow days, the answer, unequivocally, was no.

Table 7 corroborates the owner's claim that 'diners' visit on slow days and thus that the negative effect on repeat business should be moderate. Slow-day customers pay more in tips, pay more for each item they buy, and buy more items. They are also more likely to linger at the table. It also takes longer to reseat their tables. Somewhat surprisingly, workers spend more time overall with busy-day customers (Column 4). Column 6 makes this less surprising, as it shows extra time with customers is more than balanced out by the extra time it takes to reseat the table. It is also less surprising because slow days have less congestion in the kitchen.

In the online appendix I use credit card information to provide more direct evidence of a moderate effect on repeat business. The data allows me to group customers by a 'type' and follow types over time. The data identifies the customer by the type of card (Gold, Platinum, etc.), the financial institution or bank that issued the card, as well as the financial service provider (Visa, Mastercard, etc.). Online Appendix Figure A.1.2 shows that these repeat types follows a similar evolution, across the franchises, after the end of the first treatment.

Several factors can explain the moderate effects on repeat business. First, the surplus gains to consumers under the experimental contract may offset or even outweigh the losses from reductions in some aspects of personal service quality. This could explain the moderate effect and could even increase repeat business. Second, the time horizon for the data may be inadequate for properly capturing the long-run effects on repeat business. The fact that the busy-day customers visit infrequently alone supports this possibility, as in some cases it can take years before the customer

---

[18]Ours is an underestimate of the effect on short-run profit. It ignores transfers to the support staff, which offset the incentive cost by about \$3 per worker.

[19]Customers can punish the firm by going somewhere else, either to another firm, or home. Alternatively, they can tell their friends or post an unfavorable online review. I checked Yelp reviews for the period of my sample. There were very few. None complained about being hurried to leave.

returns. Third, and relatedly, it may be difficult to fully capture the long-run costs of the experimental contract. For example, consumers might tip less once they find out about the bonuses, even if there is no reduction in perceived service quality. They may simply feel less pressure to tip because they know the firm is covering more of worker effort costs.

I can say little about the long run effects of adjustments by competing firms in the absence of exogenous variation in competitive pressure. It is important to note however that for various reasons - including have the strong brand name and loyal customer base of a big-box restaurant, and (perceived) product differentiation - the firm is a leader in local product markets. It generates excess demand on busy days because it is a choice destination for many consumers. Shorter queues allow the firm to keep the lower value consumers among this group. From this perspective, it will be difficult for other firms to erode all the gains from the experimental contract.

7 Propensity for Good Decisions

The evidence implies nonnegligible surplus gains from the experimental contract on excess demand days. Why then would the firm initially use the customary contract on these days? The simplest answer might be that the firm is run poorly. I present evidence against this argument. However, I also explain that despite having a propensity for decisions which enhance productive efficiency, and despite becoming fully aware of a more profitable alternative that benefits consumers without imposing significant costs on workers, the firm reverted to using the customary contract on excess demand days following the experiment. I briefly discuss the rationale for reversion.

Implicit Contracts. Managers can assign early start times to workers who tend to serve more customers. They can also assign high-volume workers to more tables or tables that turn over more quickly. I evaluate whether the historic use of these instruments on excess demand days reflects attempts to increase revenue. I will use the ratio of the worker's historical revenue to that of their coworkers (in a shift) to measure their relative productivity. I will then evaluate whether the firm takes relative productivity into account when assigning start times and tables on excess demand days. If they do then it would support productive efficiency as a central goal.

The top panel of Table 8 reports estimates for regressions of start times and table assignments on relative productivity during the pre-experimental period. Columns 2 to 5 examines the effects on the quality and quantity of the seats. Booth, bench, and chair seats measure quality in the sense that customers tend to prefer booth seats to benches to chairs. Columns 6 examines the effects on how, given the table assignment, easy it is for workers to serve more customers. The historical turn rate is based on the ratio of the historical customer volume of the table to the number of seats.[20] Column 7 examines the effect on whether the firm fills the table of the worker. Specifically, it reports the effect on the share of seats that go ununused. This column provides a check on the

---

[20]The specifics are as follows. I calculated an average turn rate for each table for the last year. I then ranked the tables by their one-year averages. Finally, for each shift, I computed an average ranking for the tables assigned to the worker. The ranking was used because it reduces concerns that our measure reflects the productivity of workers assigned the table, rather than the table's true capacity to facilitate customer volume.

exogeneity of the procedure that matches customers with workers. Column 8 looks at work hours.

Columns 1, 5, and 8 shows high-productivity workers are assigned earlier start times, more seats, and end later. The evidence in Columns 2-4 and 6 show relative productivity has modest effects on the quality of seating assignments, and on the extent to which the assignments facilitate volume. The estimates suggest start times, table assignments, and hours are instruments for increasing revenue. The top panel Table 8 ultimately has two takeaways. First, it seems that the firm was trying to use implicit contracts to increase profit. Second, informal incentives are insufficient for solving the firm's problem. If they were, the experiment would not have had the effect that it did.

This raises concerns about whether the experimental results reflect managerial responses. Accordingly, the bottom panel reports the experimental contract's effect on various assignments by managers. Most estimates imply non-response by managers. The lone exception is with start times, which shows workers started 15 minutes earlier under the experimental contract. Having said that, Online Appendix A.1.2 explains that the main results are unaffected by controls for start time.

Transfers. The firm told me the transfer rate helps the chain deal with growth in the minimum wage for waiters. A higher transfer rate allows the chain to delay wage hikes for the support staff (not all of whom are paid the minimum). The firm thus uses minimum wage increases to pass support staff costs onto waiters. This is why the transfer rate is positive, and has risen steadily over the past few years. A few years ago it was three percent. In our sample it was four.

The increase is consistent with the response of a profit-maximizing principal who faces a limited liability constraint (Sappington, 1983). The introduction or increase of the minimum wage constrains the set of feasible transfers.[21] It forces the firm to pay workers more in every state of nature. The best response of the firm is to increase the transfer rate, so that workers transfer more in good states, and to therefore alter worker incentives on the margin. The response moves the relationship away from the first best, even if workers are risk neutral.[22]

Reversion. Given their propensity for efficiency, and despite becoming aware of a surplus enhancing alternative, the firm reverted to the customary contract on excess demand days. Why? While there are several explanations, including implementation costs (Ferrall and Shearer, 1999), costs to switching contracts across slow and busy days, worker morale costs (Englmaier and Wambach, 2010; Fehr and Schmidt, 1999), the most informative may be the one offered by the CEO. After the experiment I reminded him that an extra $10 of incentive pay per worker ($200 per shift) delivers about 100 dollars more in revenue per worker ($2000 per shift). His immediate reaction was "is there a cheaper way to do it?" He added that he felt workers were already earning a fair wage. My takeaway from this conversation was that this particular CEO is very careful with money, especially when it came to spending more than his reference point of what something is worth.

---

[21] Azar (2004) lays out some historical cases where there was a fee to work at certain upscale restaurants.

[22] Azar (2012) argues, interestingly, that increases in the minimum wage can reduce welfare. Higher minimum wages increase the firm's incentive to introduce service charges (instead of tips). Workers lose because they normally earn more with tips. The firm loses because the customer becomes a less effective monitor of service quality.

8 Conclusion

Queues are puzzling for economists because they are consistent with wasted profit in equilibrium. The puzzle has inspired a generation of research that tries to rationalize queues via conventional profit-maximizing pricing models. These rationales almost always trace the presence of queues to some feature of the goods market, such as a consumer preference for goods that are valued by other consumers (Becker, 1991). The present article traces the presence of queues to a feature of the labor market, namely the wage contracts of workers who facilitate transactions with consumers. In doing so, it accounts for an important reality of production settings where queues are common, their susceptibility to externalities due to congestion.

The article does this using field experimental evidence from big-box franchise restaurants, where an experimental contract paid workers bonuses for customer volume in addition to the tips and hourly wages they normally receive. The treatment was run on busy days when excess demand is predictable. Queues were shorter under the experimental contract. Shorter queues generate surplus gains for high value consumers who were already willing to wait and for low value consumers who would have otherwise pursued their outside option. There was no discernible effect on tip rates, suggesting no significant cost to the surplus gains in terms of service quality. This is not so surprising because shorter waits may have compensated customers for marginal reductions in service quality and because workers tended to move faster during the experiment. The treatment increased worker earnings by 10 percent, and the evidence suggests that they experienced nonnegative surplus gains. The treatment increased revenue by 10 percent and profit by at least 49 percent. I find no evidence of an effect (positive or negative) on profit in the longer run. Altogether the evidence suggests that the experimental contract may be Pareto improving.

It would be of practical interest to compare different mechanisms for exploiting queues. Conventional wisdom among industry insiders attributes long queues to sales of items like dessert, which increase revenue and profit marginally while increasing service times substantially. By this token, an alternative to paying workers for customer volume would be to increase dessert prices sufficiently during predictably busy service periods. Of course this raises questions about how high the price would have to be to generate gains comparable to the ones found here. Comparable gains through price increases may be prohibitive from the perspective of repeat business or reputation among consumers more broadly. To this end, and to properly compare mechanisms, better long run measures of customer satisfaction are needed.

This article ultimately complicates rather than solves the puzzle. It raises questions about why neither wage contracts nor goods prices nor some combination thereof are used to deal with queues. It begs for an answer that explains why a well-functioning firm would use the customary contract on excess demand days initially, why such a firm would revert to the customary contract on these days even after becoming aware of a more profitable alternative, as well as the prevalence and practice of uniform contracting more generally.

References

Ran Abramitzky, Liran Einav, Shimon Kolkowitz, and Roy Mill. On the optimality of line call challenges in professional tennis. *International Economic Review*, 53(3):939–964, August 2012.

Daniel A. Ackerberg and Maristella Botticini. Endogenous matching and the empirical determinants of contractual form. *Journal of Political Economy*, 110(3), 2002.

Armen A. Alchian. Uncertainty, evolution, and economic theory. *The Journal of Political Economy*, 58(3):211–221, June 1950.

Douglas W. Allen and Dean Lueck. Customs and incentives in contracts. *American Journal of Agricultural Economics*, 91(4):880–894, November 2009.

Beth Asch. Do incentives matter? the case of navy recruiters. *Industrial and Labor Relations Review*, XLIIIs:89S–106S, 1990.

Ofer H. Azar. The history of tipping - from sixteenth-century england to united states in the 1910s. *The Journal of Socio-Economics*, 33(6):745–764, December 2004.

Ofer H. Azar. The effect of the minimum wage for tipped workers on firm strategy, employees and social welfare. *Labour Economics*, 19(5):748–755, October 2012.

George P. Baker. Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3):598–614, June 1992.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962, August 2005.

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

Charles Bellemare and Bruce Shearer. Multidimensional heterogeneity and the economic importance of risk and matching: evidence from contractual data and field experiments. *The Rand Journal of Economics*, 44(2):361–389, Summer 2013.

Avishay Braverman and Joseph Stiglitz. Sharecropping and the interlinking of agrarian markets. *American Economic Review*, 72(4):695–715, 1982.

James A. Brickley and Jerold L. Zimmerman. Changing incentives in a multitask environment: Evidence from a top-tier business school. *Journal of Corporate Finance*, 7(4):367–396, December 2001.

David P. Byrne. Testing models of differentiated products markets: Consolidation in the cable tv industry. *International Economic Review*, 56(3):805–850, August 2015.

Sungjin Cho and John Rust. The flat rental puzzle. *The Review of Economic Studies*, 77(2):560–594, April 2010.

Adam Copeland and Cyril Monnet. The welfare effects of incentive schemes. *Review of Economic Studies*, 76(1):93–113, 2009.

Pascal Courty and Gerald Marschke. An empirical investigation of gaming responses to explicit performance incentives. *Journal of Labor Economics*, 22(1), 2004.

Stefano DellaVigna and Matthew Gentzkow. Uniform pricing in u.s. retail chains. *The Quarterly Journal of Economics*, 134(4):2011–2084, 06 2019.

Robert Drago and Gerald Garvey. Incentives for helping on the job: Theory and evidence. *Journal of Labor Economics*, 16(1):1–25, January 1998.

Etienne Dumont, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics*, 27 (6):1436–1450, December 2008.

Florian Englmaier and Achim Wambach. Optimal incentive contracts under inequity aversion. *Games and Economic Behavior*, 69:312–328, 2010.

Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and competition. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

Ernst Fehr and Klaus M. Schmidt. Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics*, 106(3):453–474, November 2004.

Ernst Fehr, Alexander Klein, and Klaus M. Schmidt. Fairness and contract design. *Econometrica*, 75(1):121–154, January 2007.

Christopher Ferrall and Bruce Shearer. Incentives and transactions costs with the firm: Estimating an agency model using payroll records. *The Review of Economic Studies*, 66(2):309–338, April 1999.

R. L. Hall and C. J. Hitch. Price theory and business behavior. *Oxford Economic Papers*, (2): 12–45, May 1939.

Barton H. Hamilton, Jack A. Nickerson, and Hideo Owan. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *The Journal of Political Economy*, 111(3):465–497, June 2003.

Bengt Holmstrom. Moral hazard and observability. *The Bell Journal of Economics*, 10(1):74–91, Spring 1979.

Bengt Holmstrom. Moral hazard in teams. *The Bell Journal of Economics*, 13(2):324–340, Autumn 1982.

Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7(Special Issue):24–52, January 1991.

Ali Hortaçsu and Steven L. Puller. Understanding strategic bidding in multi-unit auctions: a case study of the texas electricity spot market. *Rand Journal of Economics*, 39(1):86–114, Spring 2008.

Hideshi Itoh. Incentives to help in multi-agent situations. *Econometrica*, 59(3):611–636, May 1991.

Rajshri Jayaraman, Debraj Ray, and Francis de Véricourt. Anatomy of a contract change. *American Economic Review*, 106(2):316–358, February 2016.

Sheryl E. Kimes and Jochen Wirtz. Has revenue management become acceptable?: Findings from an international study on the perceived fairness of rate fences. *Journal of Service Research*, 6(2): 125–135, 2003.

Edward P. Lazear. Perfomance pay and productivity. *American Economic Review*, 90(5):1346–1361, December 2000.

John R. Lott and Russell D. Roberts. A guide to the pitfalls of identifying price discrimination. *Economic Inquiry*, 29(1):14–23, 1991.

Fritz Machlup. Marginal analysis and empirical research. *The American Economic Review*, 36(4): 519–554, September 1946.

Karthik Muralidharan and Venkatesh Sundararam. Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, 119(1):39–77, February 2011.

Paul Oyer. Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *The Quarterly Journal of Economics*, 113(1):149–185, February 1998.

Harry J. Paarsch and Bruce S. Shearer. The response of worker effort to piece rates: Evidence from the british columbia tree-planting industry. *The Journal of Human Resources*, 34(4):643–667, Autumn 1999.

David Romer. Do firms maximize? evidence from professional football. *Journal of Political Economy*, 114(2):340–365, April 2006.

Julio J Rotemberg. Customer anger at price increases, changes in the frequency of price adjustment and monetary policy. *Journal of monetary economics*, 52(4):829–852, 2005.

David Sappington. Limited liability contracts between principal and agent. *Journal of Economic Theory*, 29:1–21, 1983.

Steven Shavell. On moral hazard and insurance. *The Quarterly Journal of Economics*, 93(4): 541–562, November 1979.

Bruce Shearer. Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies*, 71:513–534, 2004.

Andrew Sweeting. Price dynamics in perishable goods markets: The case of secondary markets for major league baseball tickets. *Journal of Political Economy*, 120(6):1133–1172, December 2012.

Walter John Wessels. Minimum wages and tipped servers. *Economic Inquiry*, 35(2):334–349, April 1997.

H. Peyton Young and Mary A. Burke. Competition and custom in economic contracts: A case study in illinois agriculture. *American Economic Review*, 91(3):559–573, June 2001.
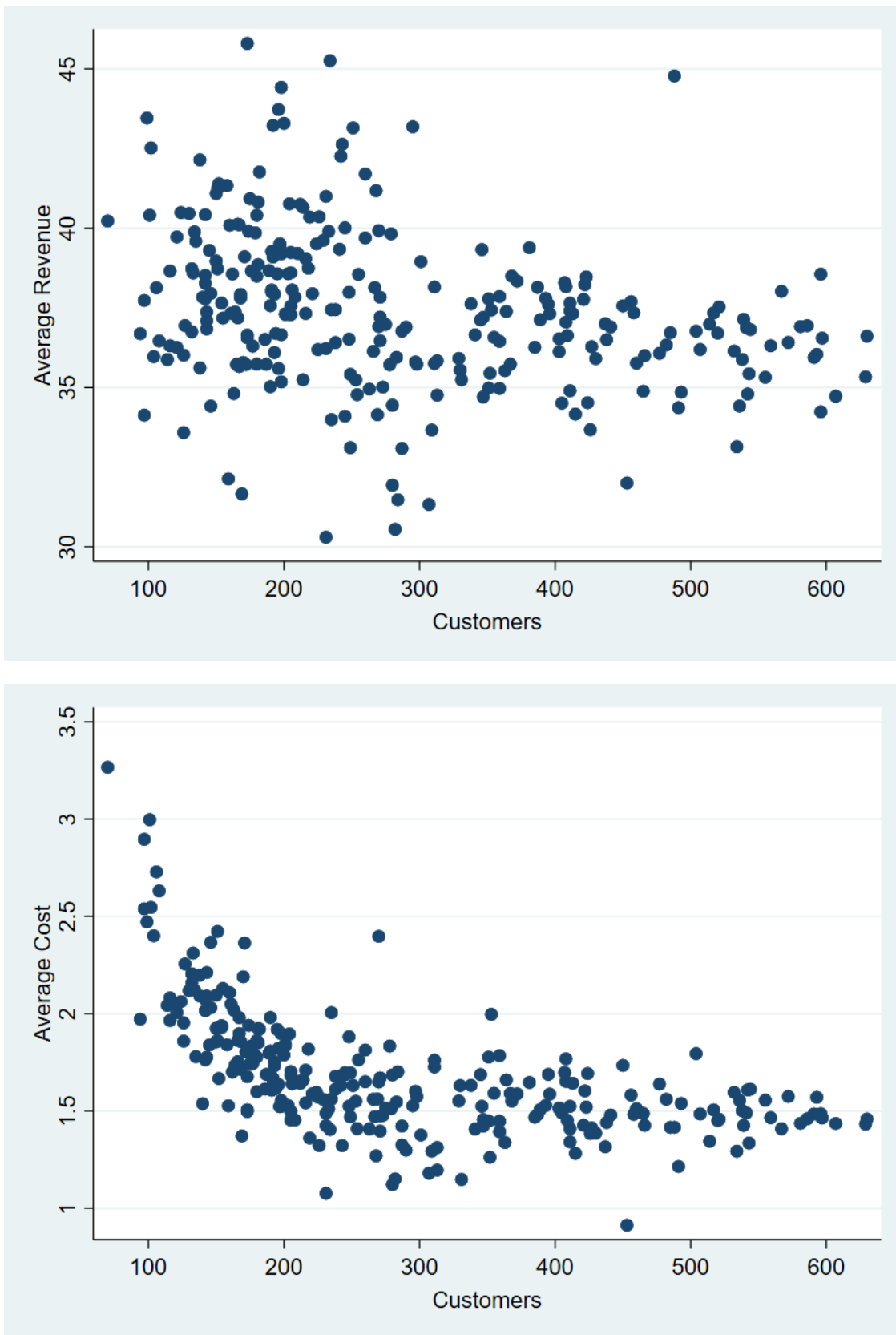
Figure 1: **Average Revenue and Cost - 2006-2007 Season**. Each dot represents the average revenue or cost at one of the franchises on a particular date. Average revenue is the ratio of total revenue for the day to the number customers served that day. Average cost is the ratio of the wage bill to the number of customers. The wage bill is total amount spent on the hourly wages of waiters.
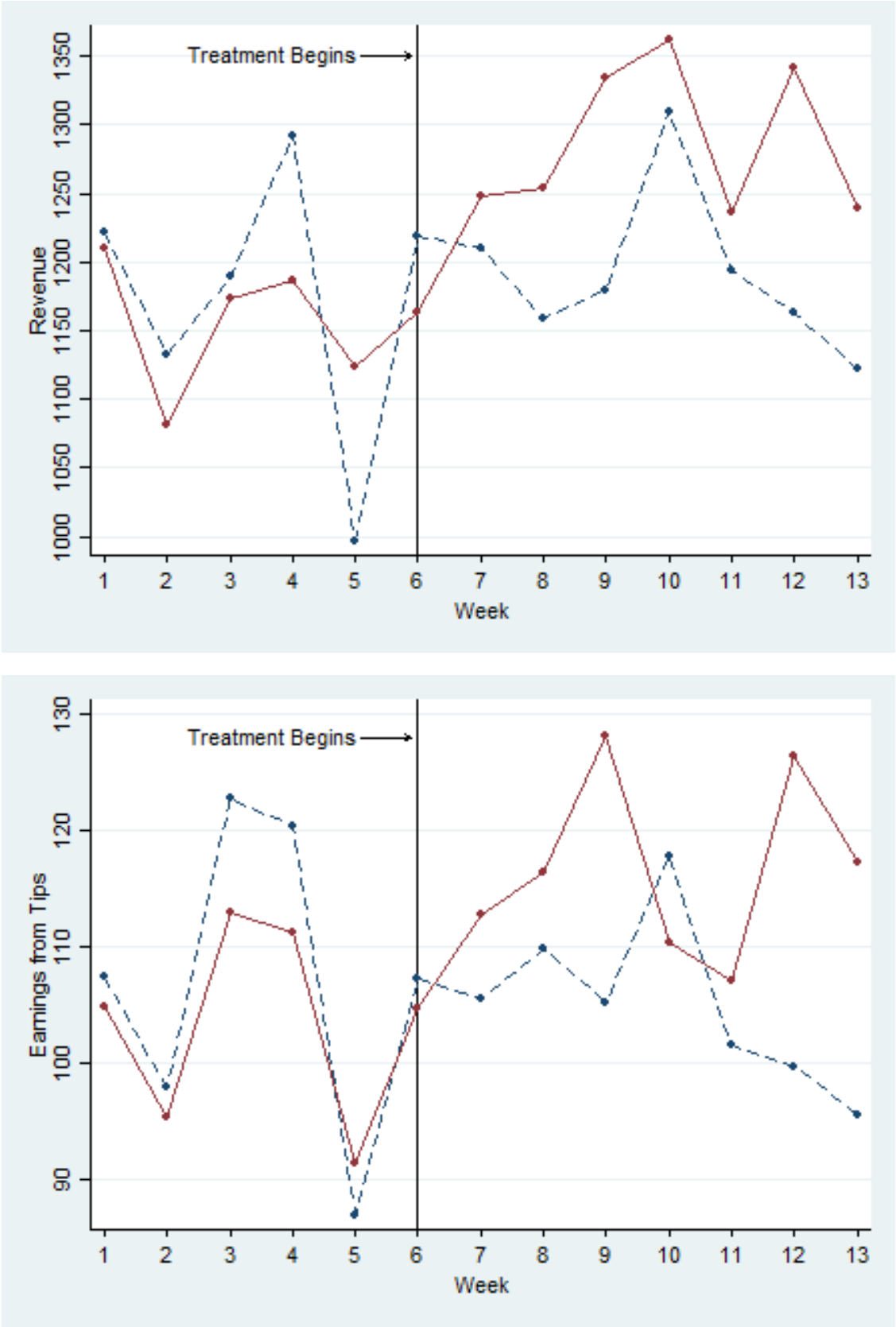
Figure 2: **Worker Output and Tip Earnings.** Each dot is the average revenue or tip earnings across all workers who worked at the treated franchise on the excess demand days (Fridays and Saturdays) of a particular week. The red solid line plots the averages for the treated season 2009-2010. The blue dashed line plots the averages for the prior season 2008-2009. The vertical black line denotes the beginning of the treatment period in the treated season.

Table 1: **Summary Statistics.** Columns (1) gives means for worker-date level outcomes in October of 2009. Column (2) does the same for November of 2009 and January of 2010. Column (3) gives the difference between Columns (2) and (1). Column (4) gives the year over year difference-in-differences estimate within the treated franchise. Column (5) takes the difference between Column (3) and the analog for the control franchise. Estimates of the standard deviation are in round parentheses. Standard errors for the differences are in square parentheses, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

| | Treated Franchise in 2009-10 | | | Comparing (3) with | |
| | Oct | Nov & Jan | Difference | Treated Franchise (2008-09) | Control Franchise (2009-10) |
	(1)	(2)	(3)	(4)	(5)
Revenue per Customer	44.09 (5.44)	43.55 (5.26)	-0.54 [0.49]	-2.16*** [0.74]	0.43 [0.43]
Customers	27 (9)	30 (9)	3*** [0.8]	3.3*** [1.2]	3.5*** [0.7]
Total Revenue	1156 (372)	1271 (365)	116*** [33.4]	92* [49.9]	163*** [26.9]
Earnings from Tips	103 (46)	115 (52)	12*** [4.5]	15** [6.6]	15*** [3.8]
Performance Bonus		12 (19)			
Share of Workers Earning Bonus		0.52 (0.50)			
Workers per Shift	19.4	20.6	20.1	18.6	17.2
Observations	192	328	520	942	928

Table 2: **Experimental Contract and Excess Demand.** The unit of observation is the franchise-date combination. Leavers is the number of customers who decide to leave upon learning the wait time for a seat. The mean for this variable is 34 customers. All regressions include fixed effects for the day of the week, week, season, outside temperature, rainfall, snow, and the number of arrivals. Robust standard errors (in parentheses), with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	Leavers	
	(1)	(2)
Performance Incentive Available	-15.88***	-7.48
	(6.02)	(5.98)
Franchise Fixed Effects		✓
Days	155	155
R^2	0.82	0.85

Table 3: **Changes in Perceived Service Quality.** The unit of observation is the bill. Tip rates are tips as a percentage of the after tax bill. Tip rates are measured on a 100 point scale. All regressions include fixed effects for the worker, day of the week, week, season, franchise, and a dummy indicating the marginal incentives to earn the bonus. Service quality measures are standardized, so that their coefficients indicate the effect of one standard deviation changes. Standard Errors (in parenthesese) are clustered on the worker, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

| | Tip Rate | | |
	(1)	(2)	(3)
Performance Incentive Available	0.09	0.10	0.09
	(0.13)	(0.12)	(0.12)
Average Price		-0.12***	-0.14***
		(0.04)	(0.04)
Items Sold		0.22***	0.24***
		(0.03)	(0.03)
Time With Customer		0.09**	0.08**
		(0.03)	(0.03)
Interaction of Performance Incentive and			
Average Price			0.17*
			(0.10)
Items Sold			-0.11*
			(0.07)
Time With Customer			0.03
			(0.07)
Observations	41899	40903	40903
R^2	0.01	0.02	0.02

Table 4: **Production and Earnings under the Experimental Contract.** The unit of observation is the worker-franchise-date combination. All regressions include fixed effects for the worker. Standard Errors (in parentheses) are clustered on the worker, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	Revenue					Earnings				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Performance Incentive Available	163***	141***	84***	84***	107***	10***	11***	6*	6*	9**
	(23)	(13)	(17)	(17)	(20)	(3)	(2)	(3)	(3)	(4)
Tailored Performance Standard					-73*					-10*
					(43)					(6)
Tables and Start times		✓	✓	✓	✓		✓	✓	✓	✓
Day, Week, and Season			✓	✓	✓			✓	✓	✓
Days Worked (Own and Peers)				✓	✓				✓	✓
Observations	4458	3081	3081	3081	3081	3084	3081	3081	3081	3081
R^2	0.02	0.30	0.42	0.42	0.42	0.01	0.14	0.20	0.20	0.20

26

Table 5: **Trade Offs and Inputs.** The unit of observation is the worker-franchise-date combination. The base good combines the appetizer, main course, and beverage. The add-on good is desserts. Regressions include fixed effects for the worker, day of the week, week, and season. Standard errors (in parentheses) are clustered on the worker, *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	TASKS			
	Revenue per Customer		Customers Served	
Performance Incentive Available	-0.68		2.5***	
	(0.49)		(0.6)	
Observations	4447		4458	
R^2	0.05		0.20	

	INPUTS			
	Average Price	Items Sold	Time With	Time Between
Performance Incentive Available	0.08	-0.10*	-3.4**	-0.6
	(0.05)	(0.06)	(1.3)	(0.6)
Observations	3066	3064	3066	2990
R^2	0.03	0.03	0.04	0.04

	DEEP INPUTS				
	Base Good Price	Add-on Price	Base Good Quantity	Add-on Quantity	Time to Linger
Performance Incentive Available	0.08	0.00	-0.07**	-0.04*	-1.1
	(0.05)	(0.01)	(0.03)	(0.02)	(1.3)
Observations	3066	3066	3064	3064	3026
R^2	0.02	0.04	0.05	0.10	0.02

Table 6: **Changes in Personal Service Quality.** This table shows that the experiment induced behaviors which are typically bad for tip rates. The top row reports estimates from a single regression of tip rates and measures of worker effort and time use. The bottom row reports estimates from 5 regressions: of the average price on a treatment dummy, items sold on a treatment dummy, time with customers on a treatment dummy, etc. The bottom-row estimates are taken from Table 5. *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

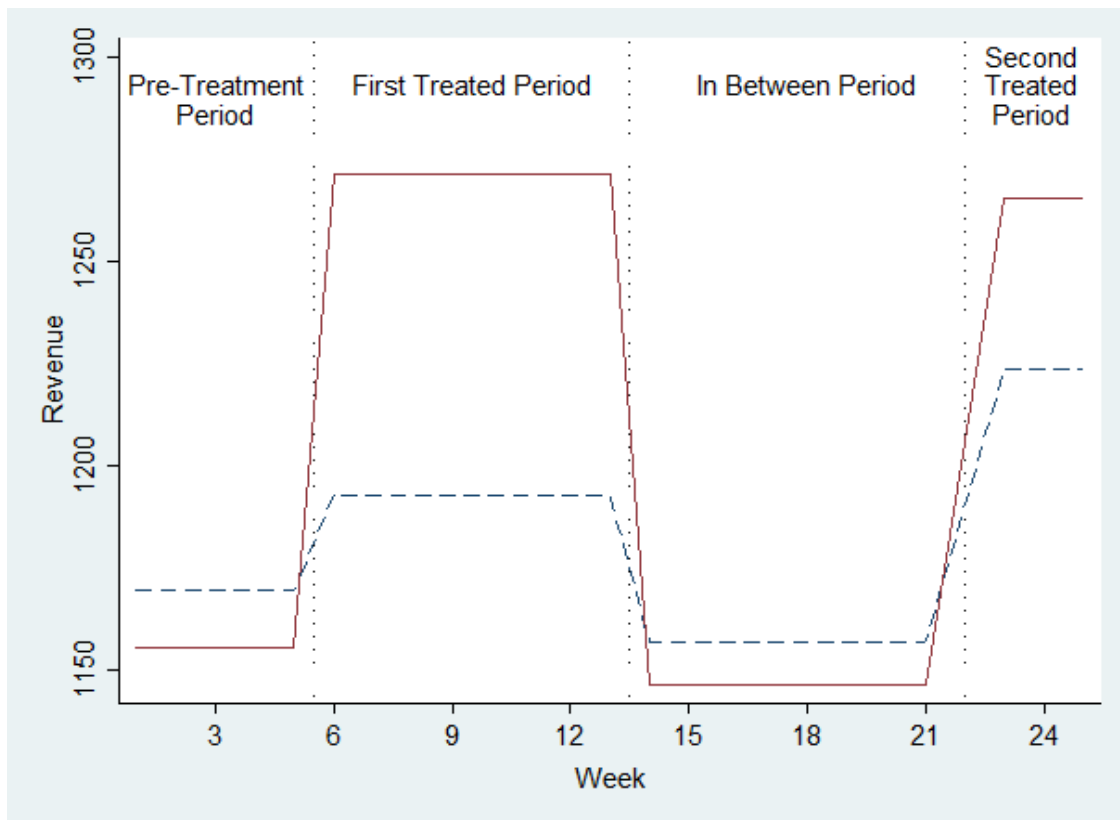|	Average Price	Items Sold	Time With Customers	Time Between	Time to Linger
Effect of					
Input on Tip Rate	-0.22***	0.14***	0.39***	0.08*	0.06
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
Experiment on Input	0.08	-0.10*	-3.4**	-0.60	-1.1
	(0.05)	(0.06)	(1.3)	(0.6)	(1.3)

Figure 3: **Effort Cost Compensation.** The solid red line describes means for revenue during the 2009-2010 season. The dashed blue line does the same but for the 2008-2009 season.

Table 7: **Auxiliary Evidence for Effect on Repeat Business.** The unit of observation is the bill. Statistics are based on bills from treated franchise before the experiment and from the control franchise. The top two rows are means and standard deviations. The bottom row is the difference in means and the standard error for the difference. *** for $p < 0.01$, ** for $p < 0.05$, and * for $p < 0.1$.

	Tip Rate (1)	Average Price (2)	Items Sold (3)	Time With (4)	Time to Linger (5)	Time Between (6)
Busy Days	14.31	6.14	7.4	89.7	20.6	16.2
	(4.45)	(1.82)	(2.5)	(32.1)	(29.6)	(12.9)
Slow Days	14.50	6.18	7.5	88.8	21.0	25.8
	(4.58)	(1.86)	(2.6)	(31.4)	(30.5)	(19.1)
Difference	-0.19***	-0.04**	-0.06*	0.84**	-0.41	-9.6***
	(0.05)	(0.02)	(0.03)	(0.38)	(0.36)	(0.3)

Table 8: **Implicit Contracts.** The unit of observation is the worker-franchise-date combination. The top panel only uses data from the treated franchise in the pre-experimental period and the control franchise. All regressions include days in the sample, its average among coworkers, fixed effects for the worker, day of the week, week, and season. Standard Errors (in parenthesese) are clustered on the worker, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	Assignments by Managers Before Shift Begins						After Shift Begins	
	Start Time	Tables				Historical	Consumers	Workers
	(Minutes)	Seats			Total	Turn Rate	Empty Seats	Work Hours
		Booth	Bench	Chair			(Share)	(by minutes)
Productivity Relative to Coworkers	-16.8*	0.98	0.04	0.51	1.52***	0.12	0.03	12*
	(9.6)	(0.62)	(0.18)	(0.38)	(0.50)	(0.18)	(0.03)	(7.2)
Observations	2109	2112	2112	2112	2112	1420	2112	1977
R^2	0.11	0.04	0.03	0.01	0.04	0.17	0.07	0.22
Performance Incentive Available	-15.6***	-0.41	0.04	-0.14	-0.50	-0.05	0.01	6
	(5.4)	(0.51)	(0.18)	(0.28)	(0.36)	(0.10)	(0.04)	(5.4)
Mean Outcome	4:49pm	7.8	1.8	3.4	13.0	3.1	0.9	6.3h
Observations	3081	3084	3084	3084	3084	2389	3084	2915
R^2	0.21	0.01	0.00	0.01	0.02	0.17	0.03	0.21

A.1   Online Appendix (Not for Publication)

A.1.1. Comparability of Control Franchise. The treated franchise is located in a subdivision where the median age is 36 years old, 48% are male, a bit more than 80% are not a visible minority, and the median household income is just over $85000. The control franchise is in a subdivision where the median age is 37 years old, 49% are male, just under 75% are not a visible minority, and the median household income is around $82000.[1]

   The main difference pertains to scale. The treated franchise is designed to handle more customer volume because it is located in a populous region. Figure A.1.1 shows this. It depicts the number of customer arrivals for the treated (solid red line) and control (dashed blue line) franchises in the 2009-2010 season, where the number of customer arrivals includes serviced customers as well as customers who left because wait times were too long. The figure shows the treated franchise has about 200 more customers than the control franchise.

   However, Figure A.1.1 also implies the differences stay roughly the same throughout the season. Accordingly, it implies that we can use regression analysis, or more specifically fixed effects, to account for the level differences across the franchises.

A.1.2. Robustness of Experimental Results to Start Times. Table A.1.4 shows start times have modest effect on the experimental results. It reports estimates that condition on the start times of workers. The estimates show start times draw down the effect of the experimental contract, but not to the extent that it changes any of our conclusions.

---

[1]The statistics are based on the 2006 Census, conducted by Statistics Canada.

Table A.1.1: **Sorting into Shifts.** The unit of observation is the worker-date combination. Data is from the treated franchise only. Calendar date controls include fixed effects for the day of the week, calendar week, and season. Demand for seating is the number of customers who visit the firm (including ones who decide not to stay). Robust standard errors in parentheses, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	Offers to Work (Per Open Slot)			Offers Accepted (Per Offer Made)		
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Incentive Available	0.15***	-0.02	-0.02	-0.09***	-0.00	-0.00
	(0.04)	(0.06)	(0.06)	(0.03)	(0.04)	(0.04)
Calendar Date		✓	✓		✓	✓
Demand for Seating (Same day last year)			✓			✓
Observations	50	50	50	50	50	50
R^2	0.17	0.42	0.42	0.16	0.42	0.42

Table A.1.2: **Summary Statistics for the Average Bill.**

| Variable	Description	Bills	Mean
Tip Rate (After Tax)		28364	14.4 (4.5)
Average Price	Revenue per Item Sold	28364	6.16 (1.84)
Items Sold	Items Sold per Consumer Served	27722	7.5 (2.6)
Time With Consumer	Minutes Between First Order and Bill Settlement	28364	89 (32)
Time to Linger	Minutes Between Last Dessert Order and Bill Settlement	27911	21 (30)
Time Between Consumers	Minutes Between Bill Settlement and First Order on Next Bill	27911	20 (16)
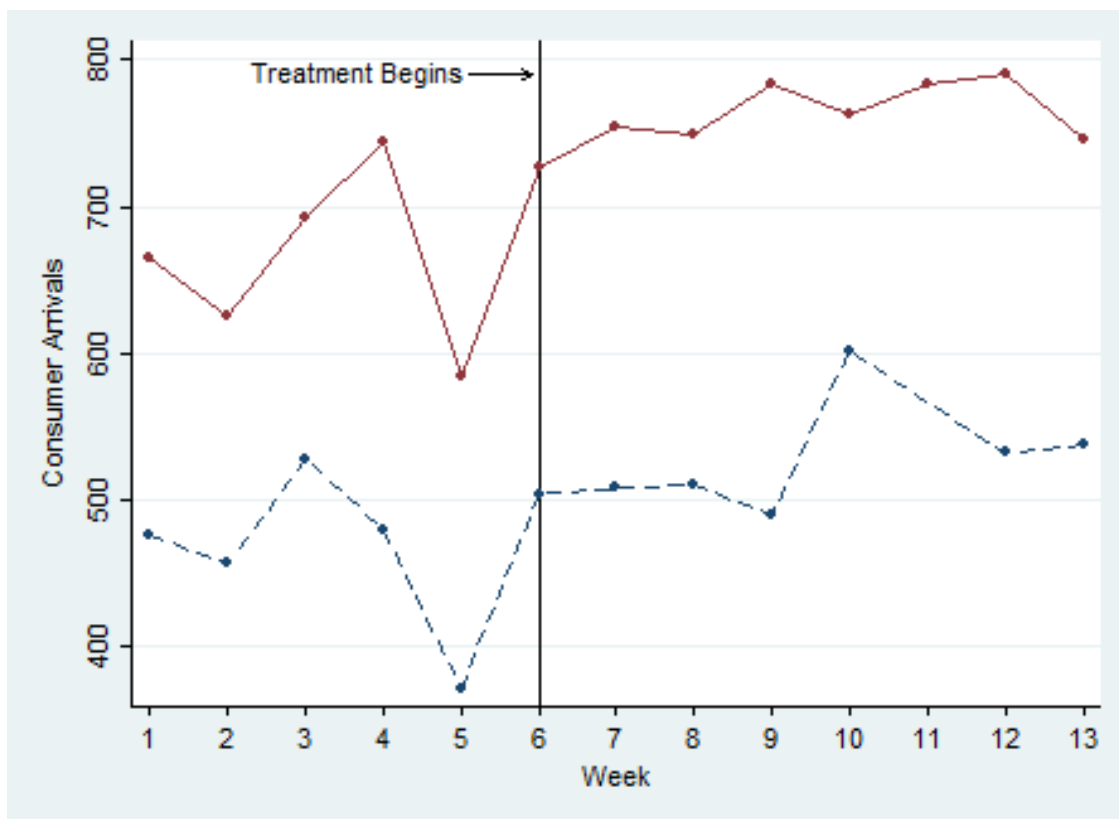
Figure A.1.1: **Arrival Rates at the Treated and Control Franchise.** The solid red line plots the number of arrivals in 2009-2010 at the treated franchise. The dashed blue line plots the number of arrivals in 2009-2010 at the control franchise. The number of arrivals includes leavers as well as stayers, where leavers are customers who leave after learning the wait time for seating.

Table A.1.3: **Tip rates and Effort and Time Use.** The unit of observation is the bill. The table reports estimates of the effect of a one standard deviation increase in service quality on the tip rate, where the mean for the tip rate is 14.11. The estimates are based on data generated before the experiment began. Standard errors are clustered on the worker, and in parentheses, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

|	Tip Rate							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Price	-0.14***	-0.25***	-0.22***	-0.20***	-0.21***	-0.17**	-0.24***	-0.23***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.07)	(0.09)	(0.06)	(0.07)
Items Sold	0.27***	0.15***	0.14**	0.14**	0.14**	0.15	0.22***	0.20***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.09)	(0.06)	(0.08)
Time With Customer	0.08**	0.22***	0.39***	0.33***	0.32***	0.25***	0.16***	0.20***
	(0.03)	(0.04)	(0.06)	(0.07)	(0.07)	(0.10)	(0.05)	(0.06)
Time Between Customers			0.08*	0.06	0.06	0.00		
			(0.05)	(0.05)	(0.05)	(0.07)		
Time to Linger			0.06	0.07	0.10**	0.12*		
			(0.05)	(0.05)	(0.05)	(0.07)		
Service Quality at Neighbors								
Average Price						0.03	-0.00	0.03
						(0.08)	(0.05)	(0.06)
Items Sold						-0.05	-0.07	-0.05
						(0.08)	(0.05)	(0.05)
Time With Customer						0.19**	0.03	0.01
						(0.09)	(0.05)	(0.06)
Time Between Customers						0.09		
						(0.08)		
Time to Linger						-0.05		
						(0.07)		
Number of Consumers		-0.26***	-0.38***	-0.36***	-0.38***	-0.31***	-0.25***	-0.25***
		(0.02)	(0.04)	(0.04)	(0.04)	(0.06)	(0.03)	(0.04)
Tip Rate on Last Bill							0.00	0.01
							(0.01)	(0.01)
Tip Rate Worker Expects								0.09***
								(0.03)
Calendar Date				✓	✓	✓	✓	✓
Franchise-Table					✓	✓	✓	✓
Bill Start Time					✓	✓	✓	✓
Observations	27722	27722	10025	10025	10025	5150	12051	7144
R^2	0.01	0.02	0.02	0.08	0.09	0.13	0.08	0.08
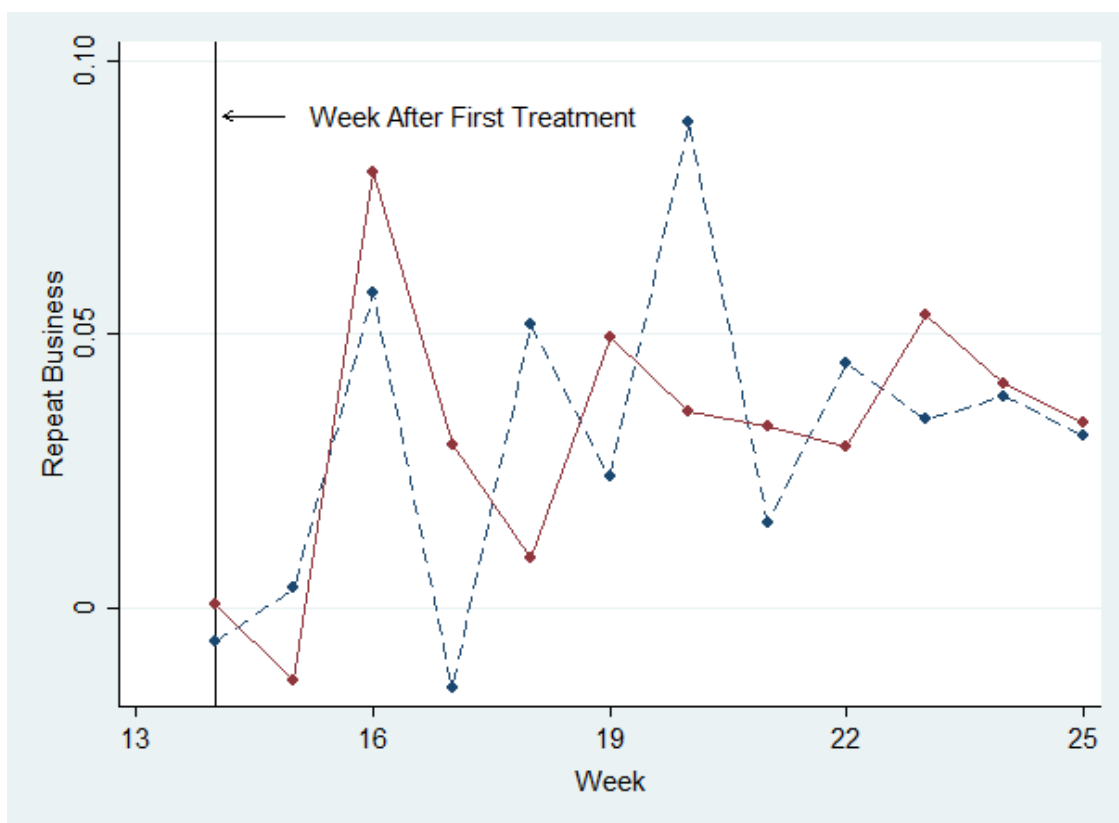
Figure A.1.2: **Repeat Business.** The repeat business measure is constructed as follows. First, for each bill, calculate the share of repeats on the bill. Second, for each shift the worker worked, average the share over the bills they handled. The figure plots the measure, after it is demeaned by a fixed effect for the franchise, as well as controls for the demand (the number of arrivals). It plots the demeaned measure after the first treatment, at both the control and treated franchises. The solid red line describes means for the measure at the treated franchise during the 2009-2010 season. The dashed blue line does the same but for control franchise.

Table A.1.4: **Experimental Contract and the Responses of Others.** All regressions include days in the sample, its average among coworkers, fixed effects for the worker, day of the week, week, and season. Standard Errors (in parenthesese) are clustered on the worker, with *** for $p < 0.01$, ** for $p < 0.05$, and * for $0.01 < p < 0.05$, and * for $p < 0.1$.

	Revenue	Earnings	Revenue per Customer	Customers	Average Price	Items Sold	Time With	Time to Linger	Time Between
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Performance Incentive Available	98.08***	8.01**	-0.04	2.29***	0.07	-0.07	-2.96**	-0.83	-0.93
	(25.35)	(3.92)	(0.47)	(0.60)	(0.05)	(0.07)	(1.28)	(1.20)	(0.67)
Observations	3081	3081	3070	3081	3081	3070	3081	3039	2990
R^2	0.21	0.14	0.04	0.25	0.04	0.03	0.05	0.01	0.04