Inefficient Incentive Provision in Multitask Jobs: Experimental Evidence from Big-Box Restaurants

Sacha Kapoor*

June 27, 2018

Abstract

Contract theory typically assumes that principals select the most profitable yet feasible contract. This article uses experimental evidence from large-scale restaurants to evaluate whether this assumption applies to the customary contract in the industry. The field experiment pays waiters bonuses for the number of customers they serve, in addition to the tips and hourly wages they customarily receive. Workers earn 10 percent more under this alternative contract, because of the bonuses, and because they earn more in tips. The firm earns at least 49 percent more profit in the short run. There is no discernible reduction in long-run profit. Even after acknowledging the material gains from the alternative, and months of evidence, the firm reverted to the customary contract. The article explores rationales for why this was the case.

JEL: J33, L81, M52

Keywords: Contracts, Incentives, Multitasking, Productivity, Profit Maximization

*Kapoor (kapoor@ese.eur.nl): Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. I thank my supervisors, Dwayne Benjamin, Gustavo Bobonis, and Nicola Lacetera for invaluable guidance during the course of this research. I thank Josse Delfgaauw, Robert Dur, Arvind Magesan, and Bauke Visser for detailed and insightful comments on later versions of the paper. The paper has benefited from conversations with Victor Aguirregabiria, Iwan Barankay, Michael L. Bognanno, Branko Boskovic, David P. Byrne, Yoram Halevy, Joshua Lewis, Hugh Macartney, Robert McMillan, Daniel Parent, Imran Rasul, Vincent Rebeyrol, Carlos Serrano, Aloysius Siow, Trevor Tombe, and Tom Wilkening. The research was supported by the Canadian Labour Market and Skills Research Network, and the Social Sciences and Humanities Research Council (of Canada). All omissions and errors are my own.

Standard contracting models assume that the firm or principal selects the most profitable of the feasible contracts. This article explores the validity of this assumption using (field) experimental evidence from big-box restaurants.

Big-box restaurants are full-service restaurants that belong to long retail chains, where the focus is on mass production of standardized packages of goods and services, where the revenue they generate from each customer is largely the same across customers, and where, because of this, customer volume is their primary instrument for increasing revenue. At the same time, these restaurants tend to rely on the customary wage contract in the industry, namely tips and hourly wages,[1] a contract which generates incentives for customer service that are too strong (or incentives for customer volume that are too weak) from the perspective of the firm. The inconsistency of their focus on customer volume and simultaneous use of the customary contract makes big-box restaurants well-suited to studying this basic assumption of standard contracting models, as the inconsistency suggests, in theory, that "better" alternative contracts are available.

To investigate the prospect of a better alternative, I conducted a field experiment at a big-box restaurant, wherein workers were paid bonuses for their customer volume in addition to their tips and hourly wages. The experiment is complemented by rich administrative data on the performance and activities of workers from a couple of franchises from a big-box chain, as well as information about customers, such as the tips they pay and whether they return at a later date. The experiment lets me estimate the causal effect of an alternative contract on the performance and profitability of workers, and thus investigate the implications of the experimental contract for profits in the short and longer run. Rich data lets me speak to whether and how the incentive problem is normally dealt with, and whether the firm makes good decisions more generally.

I find the experimental contract increased firm revenue and short-run profit by 10 and at least 49 percent, respectively. I find no effect on repeat business. Workers earned 10 percent more, because of the bonuses, and because they earn *more* in tips. The results suggests the experimental contract is more profitable than the customary contract, and that it is Pareto improving.

Given the superiority of the experimental contract, I explore justifications for the initial use of the customary contract. I exploit facts about the industry to explain why adoption of the experimental contract by competing firms should not erase the observed gains from the experimental contract. I exploit the richness of the data to investigate the use of informal arrangements as a supplement to the customary contract, showing informal arrangements are operational, but that the gains are nowhere near the gains from the experimental contract. Lastly, I assess whether the firm generally makes good choices, their aptitude for choices that increase profit. I show they do, and are particularly skilled at keeping costs low, 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 alter-

[1]Recent legislation tells us just how pervasive the contract is. 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.

native, the firm decided against permanent adoption of the experimental contract. 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, even though they knew 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 interests of the firm. These factors suggest that the CEO has an aversion to material spending, particularly when it pays workers more than their perceived worth. The experimental results suggest that this aversion governs contractual choice more than it should. The conclusion explores other explanations for reversion, including implementation, switching, and long-run costs.

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),[2] 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 the optimality of contracting, 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.[3] 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

Organization of Production. The franchises are part of a long chain of higher end "big-box" restaurants, which are only open for dinner and are located in the suburbs of Toronto.[4] They are isolated from competing firms, in the sense that they have their own dedicated building and

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

[3] 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).

[4] 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.

that the customer has to drive to get there, and such that customers cannot be solicited from the street. The chain sets prices and product offerings and these are fixed for all days of the week, and all stores within a broad geographic area. Each franchise has a general manager, 3-4 subordinate managers, 30-40 waiters, as well as several supporting staff (busboys, bartenders, and chefs). The experiment focuses on the relationship between waiters (henceforth workers) and managers.

Workers' primary tasks 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.[5]

A well-defined formula matches workers with customers. New customers are either seated right away or informed of the current wait time for seating. If there is no wait, matches are based on worker start times. The worker with the earliest start time gets the customer. If there is a wait, 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, *i.e.* 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 from 3:30pm until approximately 6:30pm. Worker stop times are largely based on a fixed set of rules. When line-ups for seating 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 to the best of their ability, rejecting them only on occasion, when there is a shortage or surplus of willing workers. The schedule is posted before the workweek begins.

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

Wage Contract and Incentive Problem. 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

[5] See (Braverman and Stiglitz, 1982) for a model where the production technique is a choice variable for the worker. See (Courty and Marschke, 2004), (Oyer, 1998), (Asch, 1990) for studies which examine incentive problems that relate to discretion.

the support staff at the end of each shift.

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.

Customer volume is especially important on "busy" days (Friday and Saturday evenings), where, apart from the summer holidays, there are generally more customers than seats. The franchises serve approximately 2500-3000 customers on average per week and approximately half come from busy days. The revenue loss from having 100 customers decide wait times are too long, a common occurence on busy days, is more than $4000. Much of this amounts to a profit loss because labor costs are effectively fixed.

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).[6] 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 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.

Awareness of the Incentive Problem. 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.

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

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.[7]

2 Experimental Contract

The experiment paid workers bonuses for their customer volume in addition to the tips and hourly wages they would normally receive. The bonus was equal 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.[8] 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.[9] 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 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

[7]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 is to show that a better contract exists, and not to evaluate the particulars of the chosen experimental contract.

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

[9]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.

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.

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 Data and Descriptive Evidence

Figures 1(a) and 1(b) visualizes the effect on tip earnings and revenue, our key outcomes for workers and the firm. Each figure plots a pair of time series from a base sample that includes only the treated franchise, during and before workers were offered the bonuses. The blue dashed line depicts the time series for weekends (Friday and Saturday) during the 2008-2009 season. The red solid line depicts the time series for weekends during the 2009-2010 season. The vertical black line indicate the weekend when the treatment began.

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 2008-2009. While year-over-year comparisons help account for seasonal patterns that are common 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.[10] 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 they otherwise would have. Either way, the crisis could generate an upward bias in the estimates.[11]

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.

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

Row 2 shows, unsurprisingly, workers serve more customers. Around three 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 2.6 and 3.6, depending on the sample we use. It is always statistically significant at conventional significance levels.

The effect on revenue per customer is less clear. When data from the treated franchise alone is used (Columns 3 and 4), the estimates imply workers generate less revenue from each customer

[10] 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.

[11] 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.

[12] 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.

served. When data from both franchises is used (the remaining columns), the estimates imply workers generated more. With these data, however, the estimates are statistically insignificant. The ambiguity ultimately highlights the value 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 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.

Quasi-random matches are useful for another reason. From time to time, I will 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.

Bills from the pre-experimental period are summarized in Online Appendix Table A.1.2. 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 the bill settlement and the first order on the next bill, and 20 minutes between the last dessert order and when the bill is settled (time to linger).[13] The tip rate has a mean of 14.4 percent and a variance of 4.5 percentage points.

4 Baseline Results

I assume outcomes y_{ifd} for worker i at franchise f at calendar date d is generated by:

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

α_i is a worker fixed effect, T_{fd} indicates the offer of bonuses for customer volume, 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, which can have a common influence because the franchises are located within 1 hours driving distance. Coworker days in sample proxies for the help a worker might receive from others (Drago and Garvey,

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

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 2 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 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. The remainder of this section investigates whether the experiment made workers and customers worse off.

Effort Costs. Figure 2 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 its 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 increase that happened during the first block.

Figure 2 has three implications. First, 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 relative to a preference that depends on money earnings and effort costs. Second, it implies the problem with workers is not that they unaware of the possibility that they could earn more money 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, 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.

Trade Offs. Table 3 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 sales of the base good and add on, and on the time to linger. The second and third panels are nested in accordance with the input or task the variable affects.

Workers respond to the bonus by serving 2.5 ($p < 0.01$) more customers, but generating 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 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 (under customers) 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 there was 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 4 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 3.

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.

Customer Satisfaction. To evaluate the potential cost of a reduction in service quality, I estimate

$$\tau_b = \beta_1 q_b + \beta_2 T_b + \beta_3 T_b q_b + \mathbf{X_b \Gamma} + \varepsilon_b,$$

where τ_b is the tip rate on bill b.[14] Note that bills can differ depending on the worker who handles it, the customer, franchise, date, and table. q_b is a compact representation of the effort and time use measures. T_b indicates whether the bill was handled at the treated franchise on a treated day. X_b includes fixed effects for the time of day, day of the week, calendar week, and franchise. The specification excludes fixed effects for the table, bill start time, and number of customers on the bill because these variables can be interpreted as bad controls. It excludes time between customers

[14] 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.

and time to linger because their coefficient estimates are fragile statistically. The interaction terms reflect the indirect experimental effect that operated via changes in effort and time use.

Estimates are found in Table 5: 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 10^{th} 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.

Wait Times. 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 the support staff. Second, in part because workers moved faster, the customer waited less time for a seat. The experiment generated positive spillovers, which could improve customer perceptions of the firm's overall service quality. This could spill over onto customer perceptions of the service quality of the worker who served them.

The firm tracks excess demand, the number of consumers who leave after learning the wait time. If the experiment shortened waits, then it would have reduced excess demand. Accordingly, I estimate

$$y_{fd} = \alpha + \beta T_{fd} + \mathbf{X}_{fd}\mathbf{\Gamma} + \gamma_d + \epsilon_{fd},$$

where y_{fd} is the number of customers who left. T_{fd} indicates the availability of the performance incentive. $\mathbf{X}_{fd}$ includes an indicator for tailored performance standards and total demand, the total number of customer arrivals. The latter helps us control for level differences in the demand for their goods and services 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 specification is parsimonious because we only have 2 observations per day. Estimates are found in Table 6.

The estimates all show a fifty percent reduction (from 34 to about 17) in the number of customers who leave after learning the wait time.[15] The reduction suggests wait times were indeed shorter under the experimental contract. Thus, while customers may have received worse personal service, they received better service from the firm as a whole.

Residual Benefits. 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

[15] Excess demand is measured with error. Some customers, particularly those who see the line ups before entering the building, will leave before finding out the wait time. The error could lend itself to upward or downward bias, depending on whether the experimental contract makes the line look longer (because more people decide to stay) or shorter (because customers are seated more quickly). If measured excess demand is correlated with unmeasured excess demand (customers who leave before entering the building), and the experimental contract makes lines look longer, we will overestimate its effect on the number of customers who leave.

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.[16] 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.

5 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 $\%\Delta\Pi \approx 49\%$.[17] This section elaborates on why 49% might provide a good approximation for the change in overall profit.

Repeat Business. A revealed preference argument implies we should expect a moderate effect on repeat business.[18] 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 this, when I asked if we

[16] 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.

[17] 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.

[18] 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.

could pay for customer volume on slow days, the answer, unequivocally, was no. This is not to say that the firm lacks concern for repeat business, but rather that the lack of a sizable and negative expected effect gives the firm good reason to carry out the experiment.

Table 7 provides auxiliary evidence of a moderate effect on repeat business. It corroborates the owner's claim, namely that 'diners' visit on slow days. 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.

Because the auxiliary evidence is only suggestive, I use credit card information to provide more direct evidence of a moderate effect on repeat business. The computer system includes information that allows me to group customers by a 'type' and follow types over time. The database 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.

Note that Tables 5 and 6 support the conclusion of no real effect on repeat business. Those tables show tip rates were unchanged, and that fewer customers left after learning the wait time.

Business-Stealing Incentives of Competitors Table 6 implies that in the very short run, shorter waits let the firm keep customers who would have otherwise went elsewhere. The marginal customer without performance bonuses was, for example, indifferent between waiting a couple of hours and going elsewhere. The marginal customer with performance bonuses was, for example, for whom an hour was enough. The bonuses let the firm keep customers whose point of indifference was between one and two hours. It is important to know the longer run implications of shorter waits, once customers update their beliefs, and once competing firms adjust accordingly. While I cannot study the issue empirically, I can speculate as to what the implications will be.

Shorter expected wait times for the customer should generate more foot traffic, which has competing effects. It creates opportunities for customer volume, but lengthens lineups. This pushes wait times up again. If the firm is still better off after this happens (they should be), then other firms should want to adopt similar contracts. Doing so lets them steal back business or workers from our firm. They could thus erode the gains that encouraged adoption in the first place.

It will be difficult for other firms to erode all the gains. 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 can generate excess demand for seating because it is the first choice destination for many consumers. If shorter expected waits help attract some of these consumers, particularly the ones who go elsewhere simply because of the wait, then the gains should persist even if other firms adopt similar contracts.

The longer run implications for product-market competition has a direct bearing on the longer run implications for labour-market competition. In this industry the competitiveness of any performance-based incentive contract depends on how competitive the firm is in the product market, and much less so on whether competing firms are offering comparable contracts. As long as the firm can ensure there is excess demand for seating, they can assure that workers will make at least as much as at competing firms. Workers will serve lots of customers, generate more revenue, and thus earn more in tips. In fact this is already the case. Many workers noted that this is why the retail chain is already able to attract the best service workers in local labour markets.[19]

Most importantly, this idea, that product-market competitiveness dictates labour-market competitiveness, would explain why the firm was able to use an inefficient contract in the first place. They were not eliminated from the market because theirs is a textbook example of a monopolistically-competitive market, where their market power in the product market allows for mistakes in the labour market.

6 Propensity for Good (Profit-Increasing) Decisions

This section reports supporting evidence of the idea that the firm generally tries to increase profit. It examines their allocations of workers to shifts, their use of implicit incentives, and their choice for the transfer rate. It ends with a brief discussion of the optimal contract.

Figure 3(b) plots the wage bill for workers against customer volume using data from 2006 and 2007. The figure essentially shows that labour is a fixed cost, as the wage bill eventually drops below 2 dollars per customer, and thus that the firm tries to allocate workers to shifts in a way that spreads the wage across as many customers as possible. Figure 3(a) plots revenue per customer against customer volume, showing that revenue per customer holds constant at roughly 35 dollars, and implying that there is a small opportunity cost to increasing volume. The figures suggest labour allocations are consistent with attempts to increase short-run profit.

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 reflects attempts to increase revenue. I use the ratio of the worker's historical revenue to that of their coworkers (in a shift) to measure their relative productivity. I will use this to evaluate whether the firm takes relative productivity into account when assigning start times and tables.

The top panel of Table 8 reports estimates for regressions of start times and table assignments on relative productivity. 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

[19]The idea that tipping lets the firm attract and retain workers is discussed in (Wessels, 1997).

[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

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 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.

The top panel of Table 8 raises concerns about whether the experimental results reflect managerial responses. Accordingly, the bottom panel reports the effect of the experimental contract on various assignments by managers. Most of the estimates imply non-response by managers. The one exception is with start times, which shows that, on average, workers started 15 minutes earlier under the experimental contract. Having said that, Online Appendix A.1.2 explains that controlling for start time has no effect on our main results.

Transfers. The firm told me the transfer rate helps the chain deal with growth in the minimum wage for waiters. A higher transfer rate alleviates the burden indirectly because it 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 workers. 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]

Optimal Contract. The optimal contract is complex, incorporating transfers, product pricing, wages, informal incentives, etc. Finding the optimal contract is impractical and besides the point. The point, rather, is that for the growing number of full-service big-box restaurants, there are simple alternatives that dominate the customary contract. Advances in point-of-sale software have facilitated production on a large scale, as well as a host of contractible performance measures. The software lets these firms contract on n, q, or any subset of q. But they don't. Why?

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.

[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.

7 Conclusion

The firm reverted to the customary contract after the experiment. They are willing to forgo

$$\pi(1) - \pi(0) = \sum [R(1) - R(0) - bn(1)],$$

where $\pi(1) = \sum R(1) - bn(1) - f$ is a simplistic representation of the payoff from the experimental contract, b is the bonus rate, $\pi(0) = \sum R(0) - f$ is the payoff from the customary contract, and the sum is taken across all workers. Results from the experiment imply $\sum R(1) - R(0) \approx \$2000$ and $\sum bn(1) = \$200$, so that $\pi(1) - \pi(0) \approx \$1800$. This suggests the experimental contract has a hidden cost of at least \$1800 per day. This section speculates on what that cost might be.

One explanation relates to the additional costs of implementing a new pay scheme (Ferrall and Shearer, 1999). These are costs of communicating the new scheme to workers, tracking their performance along payoff-relevant dimensions, and computing their payments. These costs are small to negligible here because: first, there was effectively a one-time cost to communicating the new scheme, and this cost was low because workers found the new scheme easy to understand. Second, the firm already tracks customer volume and shares this information, along with other performance metrics, with each worker at the end of each shift. Third, there is a small one-time cost to programming the software to compute payments and payroll deductions. The actual payment would only take a few extra seconds, as managers already distribute credit card tips to the worker, and compute and collect workers' transfers to the support staff.

A second explanation relates to the oscillation of one contract on slow days and another on busy days. If workers anticipate better opportunities on busy days, they might put in less work on slow days. Such losses should be counted against any gains on busy days. To examine the issue a third treatment was run on slow days, where workers were paid bonuses for average revenue (per customer). The three treatments allow for assessments of whether workers substitute revenue across slow and busy days, and thus about whether revenue losses on slow days are a cost of the experimental contract. To avoid lengthening the paper further, an analysis of the third treatment will be left for future work. I will simply stress how surprising it would be if the slow-day losses were anywhere close to the busy-day gains, as the busy-day elasticity of revenue to customer volume usually dwarfs the slow-day elasticity of revenue to revenue per customer.

A third explanation is that the data imperfectly captures the long-run costs of the experimental contract. Customers, for example, might tip less once they find out about the bonuses, even if there is no reduction in the service quality they perceive. They might simply feel less pressure to tip because they know the firm is covering more of worker effort costs.

Several factors cast serious doubt on this argument. First, tipping is a surprisingly robust institution. It was banned by several US states in the early parts of the 20th century, but later repealed, because the bans were ignored by consumers. Second, minimum wages for servers grew substantially for the better part of the 20th century. Their growth was hotly debated in the public sphere, and likely known to the general public. At the same time tip rates were on the rise. These two trends contradict the idea that customers will tip less once they find out workers are receiving more pay from the firm. Third, the experimental contract had a small effect on tip rates, even though it had a large effect on worker behavior. A small effect on tip rates is consistent with strict adherence to social norms on tipping. If customers find out about the bonuses for customer volume, then a lower tip rate would require deviations from what the norm tells them they should do. It is difficult to envisage this in light of the historical robustness of the institution.

While there are other potential explanations, relating to worker morale for example (Englmaier and Wambach, 2010; Fehr and Schmidt, 1999), the response of the CEO may in the end be the most informative for the decision to revert to the customary contract. After the experiment I reminded the CEO that an extra \$10 of incentive pay per worker (\$200 per shift) delivers about 100 dollars more in revenue per worker (\$2000 per shift). The CEO's immediate reaction was "is there a cheaper way to do it?" The CEO felt workers were earning a fair wage, enough to oblige them to act in the interest of the firm. His reaction implies that the hidden cost depends on an aversion to material spending, particularly when it exceeds his perception of what workers are worth. The experimental results implies that this aversion affects contractual choice more than it should. This study ultimately begs for an understanding of why this is the case, whether it is the case more broadly, and, if so, the implications for contract theory and efficiency 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.

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.

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 Hortasçu 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.

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

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.

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

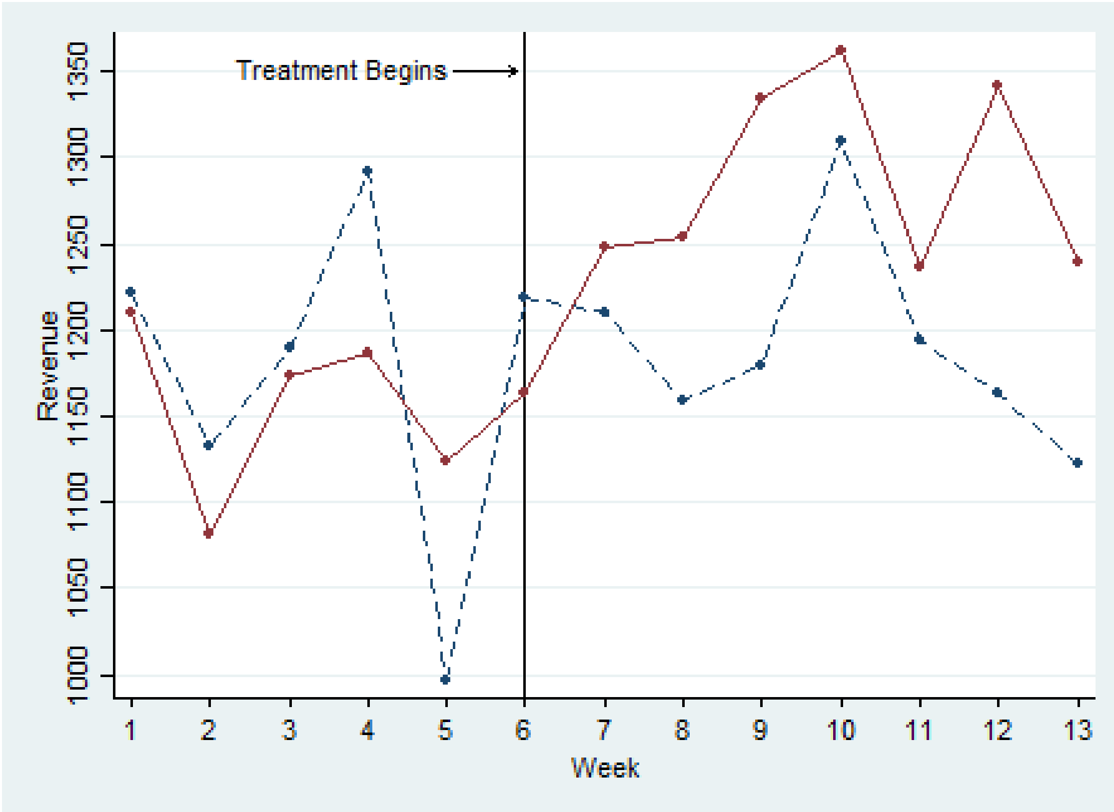
Steven Shavell. On moral hazard and insurance. *The Quarterly Journal of Economics*, 93(4): 541–562, November 1979.

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

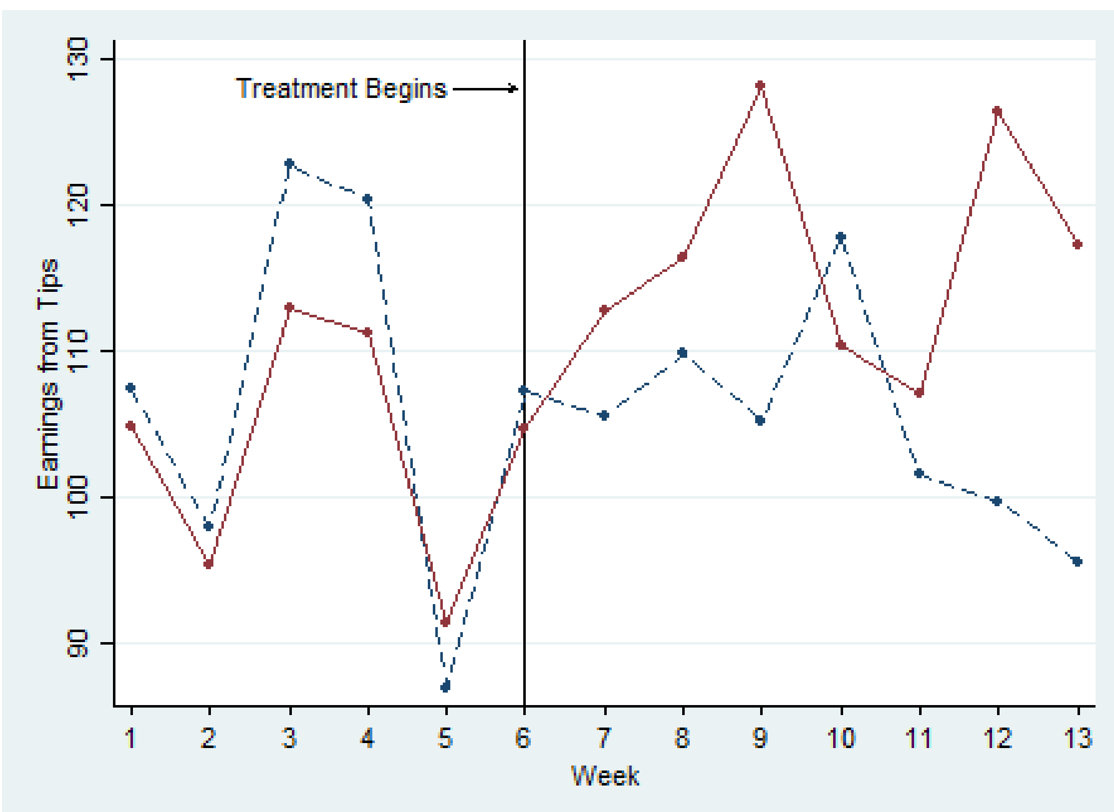
Andrew Sweeting. Price dynamics in perishable goods markets: The case of secondary markets for major league baseball tickets. *Journal of Political Economy*, 120(6):1133–1172, December 2012.

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.



(a) Effect on Revenue



(b) Effect on Tip Earnings

Figure 1: Unconditional Effects of Performance Bonuses

Table 1: **Unconditional Effects on the Performance and Earnings of Workers.** Columns (1) gives means for the 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). The fourth column compares the difference in Column (3) with the same difference from 2008-2009. The fifth column compares the difference in Column (3) with the difference at the control franchise (in 2009-2010). The sixth column and seventh columns describes the average effects when both treatments are included in the sample. The sixth column includes both treatments, but only the data from 2009-2010. The seventh column includes both treatments, and uses the data from the full sample. 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		Both Treatments		
	Oct	Nov & Jan	Difference	Treated Franchise (2008-09)	Control Franchise (2009-10)	2009-10 Only	Both Franchises (2008-2010)	Both Franchises (2006-2010)
	(1)	(2)	(3)					
Revenue per	44.09	43.55	-0.54	-2.16***	0.43	0.01	0.24	0.29
Customer	(5.44)	(5.26)	[0.49]	[0.74]	[0.43]	[0.32]	[0.01]	[0.34]
Customers	27	30	3***	3.3***	3.5***	3.6***	2.6***	2.9***
	(9)	(9)	[0.8]	[1.2]	[0.7]	[0.5]	[0.4]	[0.5]
Total Revenue	1156	1271	116***	92*	163***	158***	120***	137***
	(372)	(365)	[33.4]	[49.9]	[26.9]	[20.1]	[18.6]	[19]
Earnings from Tips	103	115	12***	15**	15***	12***	10***	-
	(46)	(52)	[4.5]	[6.6]	[3.8]	[2.9]	[2.6]	
Performance Bonus		12						
		(19)						
Share of Workers		0.52						
Earning Bonus		(0.50)						
Workers per Shift	19.4	20.6	20.1	18.6	17.2	16.9	16.7	16.9
Observations	192	328	520	942	928	2000	3084	4458

Table 2: **Production and Earnings.** All regressions include fixed effects for the worker. Standard Errors (in parenthesese) are clustered on the worker, with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	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

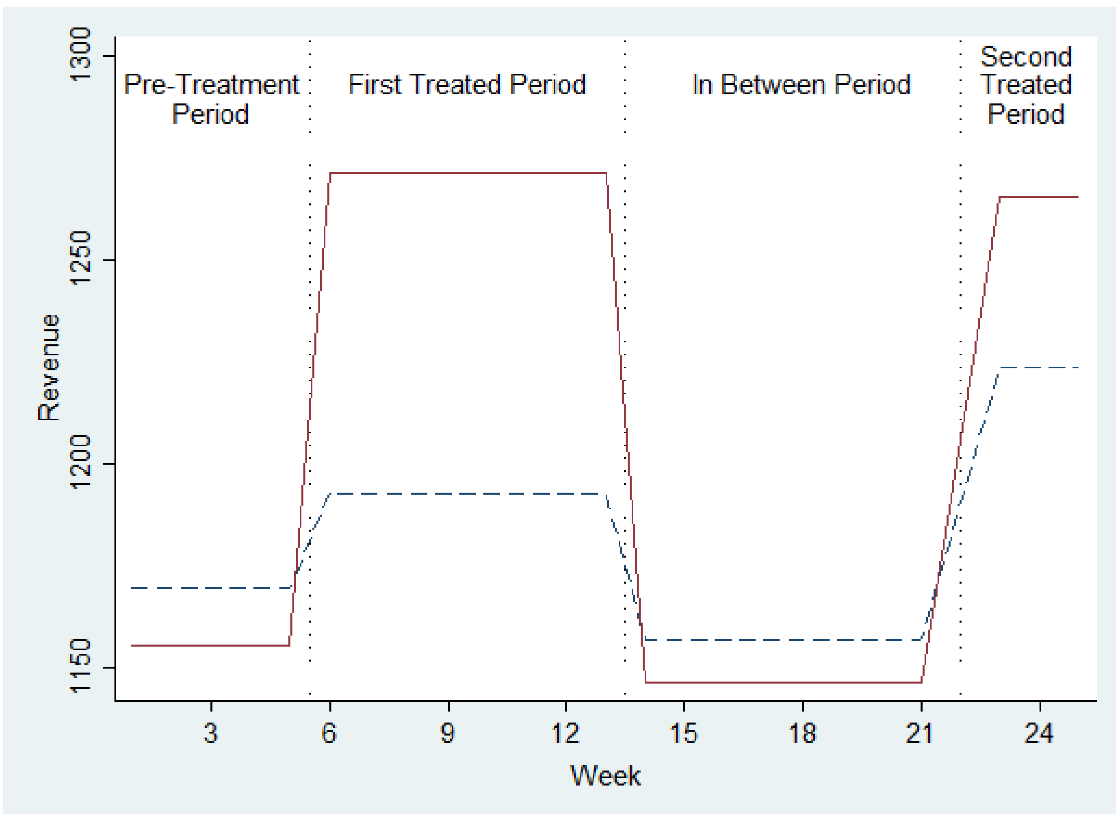


Figure 2: **Effort Costs.** The solid red line describes means for revenue during the 2009-2010 season. The dashed blue line does the same but for the 2008-2009 season.

Table 3: **Trade Offs and Inputs.** 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 4: **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 3. *** 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)

Table 5: **Effect on Tip Rates.** 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 6: **Spillover Effects of the Experimental Contract.** Excess demand is the number of customers who decide to leave upon learning the wait time for a seat. Its mean is 34 customers. All regressions include fixed effects for the day of the week, week, season, as well as the number of arrivals. Robust standard errors (in parentheses), with *** for $p < 0.01$, ** for $0.01 < p < 0.05$, and * for $p < 0.1$.

	Excess Demand			
	(1)	(2)	(3)	(4)
Performance Incentive Available	-17***	-17***	-15**	-16***
	(6)	(6)	(6)	(6)
Temperature		✓	✓	✓
Rain			✓	✓
Snow				✓
Days	156	155	155	155
R^2	0.81	0.82	0.82	0.82

Table 7: **Auxiliary Evidence for Effect on Repeat Business.** 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 (4.45)	6.14 (1.82)	7.4 (2.5)	89.7 (32.1)	20.6 (29.6)	16.2 (12.9)
Slow Days	14.50 (4.58)	6.18 (1.86)	7.5 (2.6)	88.8 (31.4)	21.0 (30.5)	25.8 (19.1)
Difference	-0.19*** (0.05)	-0.04** (0.02)	-0.06* (0.03)	0.84** (0.38)	-0.41 (0.36)	-9.6*** (0.3)

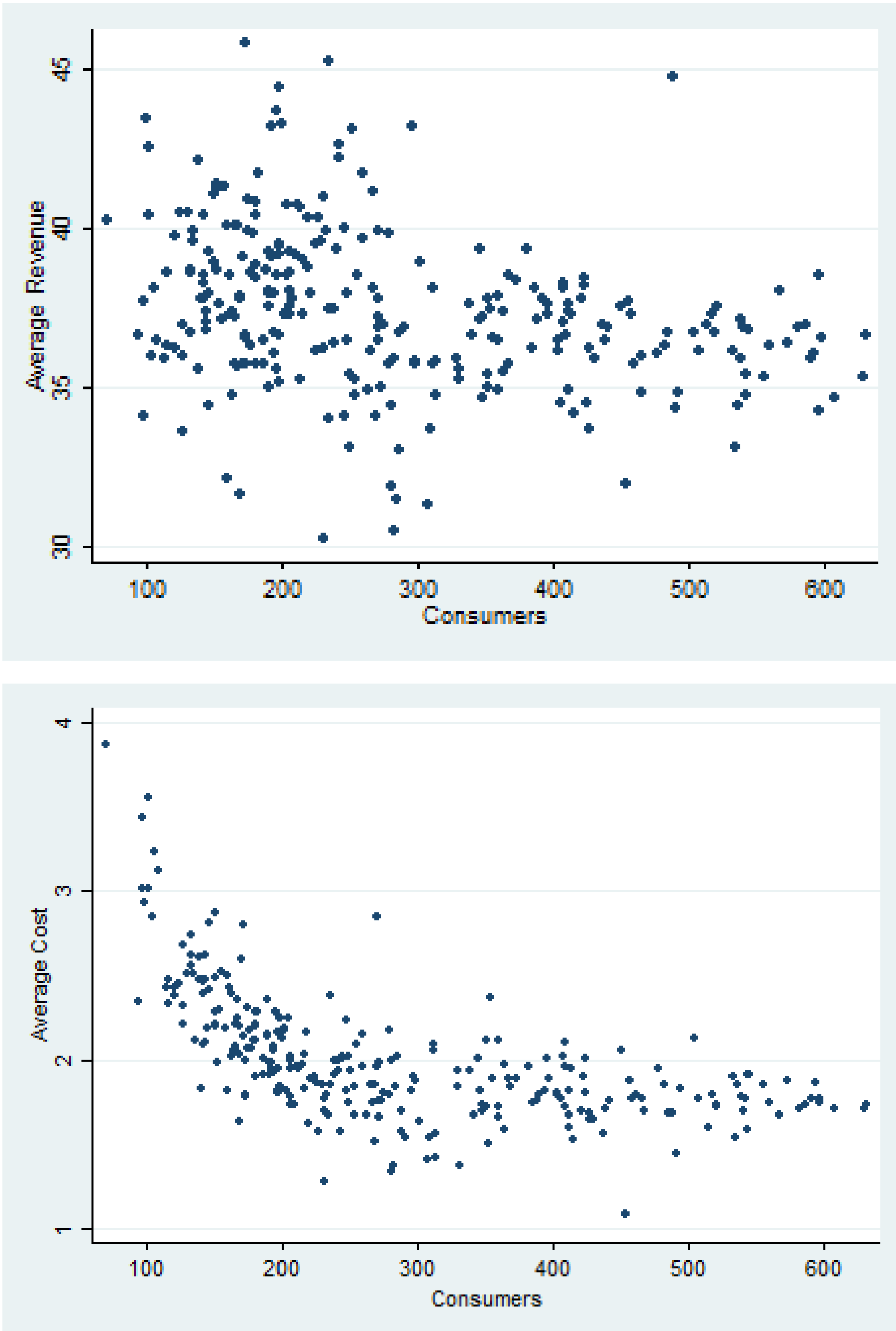


Figure 3: Average Revenue and Cost - 2006-2007 Season

Table 8: **Implicit Contracts.** 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 (Minutes)	Tables					Consumers	Workers
		Seats				Historical	Empty Seats	Work Hours
		Booth	Bench	Chair	Total	Turn Rate	(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	4:49*pm*	7.8	1.8	3.4	13.0	3.1	0.9	6.3*h*
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