Incentive Provision in Multitask Jobs: Experimental Evidence from the Workplace[1]

(JOB MARKET PAPER)

Preliminary and incomplete: Please do not cite.

Sacha Kapoor[2]
Department of Economics
University of Toronto

October 13, 2010

Abstract

This paper presents field evidence on the provision of incentives in multitask jobs – jobs where workers operate on more than one margin and each margin matters for productivity. I design and conduct a field experiment at a large-scale restaurant to investigate how simple incentive contracts based on some, but not all, tasks affects the firm's short-run performance. The experimental treatment pays waiters performance bonuses for the number of customers served, in addition to the tips for customer service and hourly wages already received. I find that, when paid bonuses for customer volume, the average worker earns more, is more productive, and generates higher profits for the firm. At the same time, despite the added benefits for customer volume relative to customer service, the data reveals a negligible impact on tip percentages. Overall, the findings suggest that sharpening incentive contracts to deal with incentive problems in multitask jobs has benefits for workers and the firm.

JEL Classification: J33, L81, M52
Keywords: multitasking, incentive contracts, worker heterogeneity

[1]I thank my supervisors, Dwayne Benjamin and Gustavo Bobonis, for invaluable guidance during the course of this research. This paper has benefited from discussions with Victor Aguirregabiria, Michael Baker, Stephane Bonhomme, Branko Boskovic, Loren Brandt, David Byrne, Christian Dippel, Joshua Lewis, Hugh Macartney, Arvind Magesan, Robert McMillan, Philip Oreopoulos, Carlos Serrano, Aloysius Siow, Junichi Suzuki, Trevor Tombe, Matthew Turner, and from seminar participants at the 2010 CLSRN Annual Conference. This research was supported by the Social Sciences and Humanities Research Council and the Canadian Labour Market and Skills Research Network. All omissions and errors are my own.

[2]Email: sacha.kapoor@utoronto.ca

1 Introduction

While many jobs have become more complex in recent decades, firms continue to use contracts that reward behavior in one or two highly-visible dimensions. This is puzzling because, in theory, a discrepancy between the tasks covered under employment contracts and the tasks the worker carries out can seriously distort incentives in the workplace [Holmstrom and Milgrom, 1991]. Despite the ubiquitous nature of this discrepancy, there are few empirical studies into the role of these incentive problems in the workplace.

In workplaces where workers mostly carry out one task these studies are difficult because other tasks often have a small influence on productivity. In more sophisticated settings, where workers carry out more than one task and incentive contracts have a wider scope, difficulties arise because a natural mapping between tasks and productivity usually does not exist. This paper examines the role of incentive problems in a work environment where workers are asked to carry out several tasks but, although each task matters for productivity, are only rewarded for their performance in some, and where data can be used to map individual productivity to performance in each task.

In this paper, I design and conduct a field experiment at a large-scale restaurant to investigate how the use of simple incentive contracts based some tasks, but not all, affects individual productivity, consumer satisfaction, and the firm's short-run performance. Waiters in this setting are well-suited for studying this issue because incentive problems related to multitasking arise naturally in high volume restaurants that face capacity constraints. While the firm wants workers to focus more on earnings through customer volume, because it might generate large marginal gains in revenue, the existing contract rewards workers who focus on earnings through customer service. Workers have strong incentives to focus on this margin because, though it might generate small marginal gains in revenue, it could improve earnings from tips. Since the costs of monitoring workers are large, and because consumer preferences introduce noise into performance measures, workers have the means *and* the incentive to direct their attention to customer service.

To evaluate the effect of this conflict of interest on workers, consumers, and the firm, the experiment pays workers performance bonuses for customer volume, in addition to their tips for customer service and hourly wages. The crux of the research design is to compare outcomes under the original incentive contract, where incentive problems related to multitasking are present, with outcomes under the treatment, where these problems are muted. Since these problems typically appear on days with high customer volume, where a large number of customer arrivals are not served, the experiment is restricted to Fridays and Saturdays in November 2009 and January 2010.

I base the empirical analysis on detailed transaction data from a franchise of a major North American corporation located in the Greater Toronto Area. The study complements the treatment period with four months of control-period data from October, November, and January of 2008-2009 and October 2009-2010. I use the data to construct broad within and across day measures of individual performance, such as revenue and tip percentages, and measures of individual inputs to the production process, such as customer volume, hours worked, and sales of various items. The identification strategy uses within-worker comparisons, both within and across years, based on these performance measures to investigate the role of multitasking incentive problems in the workplace. Overall, the sample includes 40 workers, 52 days, and 937 observations.

This study has three major goals. The first goal is to quantify the distortions generated by the original contract, in terms of the impact on outcomes for workers, consumers, and the firm. The second goal is to explore who responds to incentives for customer volume and how. In particular, I investigate: what input responses separate good from bad workers in multitask jobs? Do the responses explain productivity and earnings differences across workers? How do responses influence consumer satisfaction and the firm's short-run performance? The third goal is to shed light on why firms might not widen the scope of employment contracts to deal with distorted incentives in multitask jobs.

The main findings reveal that, when workers are encouraged to direct more attention to customer volume, there are large reductions in the distortions generated under the original contract. Under the experimental treatment the average worker served 2.49 more customers per day at the expense of a \$1.79 reduction in sales per customer. Over the course of a shift, the average worker sold \$66 more, which roughly corresponds to a 6% improvement in daily sales. The results suggest that, in distorting individual productivity, the original contract significantly misaligned worker behavior with the short-run interests of the firm.

I also find that the average worker earns more *and* generates larger profits for the firm. Workers earn more because of improvements in individual productivity and because, surprisingly, changes in behavior that favor customer volume had no influence on tip percentages. Under the experimental treatment the average worker earned \$16.50 more (about 10% of average earnings) per day. In addition to generating benefits for workers, the augmented contracts resulted in an 18% improvement in profits per day for the firm.

A closer examination of data reveals that, while the performance bonuses induced *all* workers to direct more attention to customer volume, the average effects for individual productivity are driven by the behavior of high-ability workers. The data shows that, in order to improve productivity, high-ability workers actually moved faster and sacrificed less in sales per customer than their coworkers. I also find that, after making the trade off, low-ability workers experienced a large reduction in tip percentages. A reduction in tip percentages suggests the responses of low-ability workers was at the expense of the overall quality of service delivered to consumers.

The heterogeneous treatment effects are used to investigate why a profit-maximizing firm might not want to sharpen employment contracts to deal with incentive problems in multitask jobs. I use the estimates to measure the sensitivity of short-run profits to the composition of workers employed at the firm. I find that these contracts are only profitable when at least 15% of the firm's workforce consists of high-ability workers.

Overall, this paper contributes to a growing literature on the role of incentives for individual productivity in the workplace.[1] In contrast with work in ([Shearer, 2004], [Lazear, 2000], and [Paarsch and Shearer, 1999]), which highlight the importance of incentive effects and sorting, this paper focuses on the importance of incentive effects *and* the allocation of effort among the various duties of the worker. This paper builds on previous work that examines this issue [Dumont et al., 2008] by considering the implications for overall individual productivity, consumer satisfaction, and the profitability of the firm. In doing so, the study provides a direct empirical connection with questions raised in [Holmstrom and Milgrom, 1991] on the importance of multitasking in agency situations.[2]

More broadly, the results speak to normative prescriptions derived in [Holmstrom, 1982], [Holmstrom, 1979], that we should expect improved outcomes from contracts that use more information about worker performance, and provides a rationale for the observation made in [Stiglitz, 1991], that such complex contracts are not often observed in firms. This paper also speaks to observations from [Gibbons, 1998] that, in combination, multiple instruments can be used to provide workers with a 'balanced package of incentives' and subsequently to reduce inefficiencies based on incentive problems in multitask jobs. In these regards, the study has particular significance for contract design in more skilled workplaces where high-ability workers have a strong presence.

The paper is organized as follows. Section 2 describes institutional features of the workplace that provide a foundation for the identification strategy used in this paper. This section also provides details about the research design. Section 3 develops a theoretical framework that generates predictions about behavior under the experimental treatment. The model also highlights the sources of conflict between worker and firm. Section 4 provides a detailed description of the data and further information about the context. Section 5 discusses the empirical findings and their theoretical implications. Section 6 presents robustness checks based on information from a control restaurant. Section 7 concludes, discussing future research, broader implications for incentives in organizations, and how the findings relate to compensation policies commonly used by governments.

[1] For a summary see [Prendergast, 1999]. More recent empirical treatments include: [Bandiera, Barankay, and Rasul, 2005], who compare productivity under relative performance evaluations with productivity under piece rates; [Hamilton, Nickerson, and Owan, 2003], who compare productivity under team production with productivity under piece rates and explore differences across workers; [Mas and Moretti, 2009], who study peer effects in a fixed wage workplace.

[2] Other significant contributions to the theoretical literature on multitasking include [Ramakrishnan and Thakor, 1991]. For empirical investigations that relate contract choice to multitasking concerns see [Fehr and Schmidt, 2004] and [Slade, 1996]).

2 Institutional Features and Research Design

In this section, I describe the process that matches consumers with workers and the research design used to study multitask-agency problems in the workplace.

An important feature of the workplace is that the firm uses a rules-based matching process. A rules-based process places limits on managerial and support worker discretion about who to match with whom. Under this process consumer-based idiosyncrasies, such as preferences, are conditionally independent across bills and subsequently averaged out when data is aggregated to the daily level.[3] As a consequence the process enables an identification strategy based mostly on worker specific heterogeneity *and* the 'average' consumer visiting the firm that day.

The main benefit of using an experimental research design is that it allows me to identify how workers adjust customer service in response to performance bonuses for customer volume. In the absence of experimental variation this exercise is challenging because these outcomes are determined simultaneously.[4]

The Matching Process

Decisions about shift allocations are based on management preferences and on the preferences of the worker. Workers publicly post shift requests (one to two weeks in advance of each workweek) and managers try to accommodate the preferences of each worker. Since a fixed number of slots are available in each shift, management will unilaterally allocate workers to over- or under-demanded shifts. The final schedule is public information. It includes information about start times, which range from 3:30pm - 6:30pm and are generally staggered at 0-15-30 minute intervals.

Conversely, end times are at the discretion of the manager. When there are no consumers left waiting for a table a subset of workers are sent home. The number sent home will depend on expectations about the number of late customer arrivals. The order of finish is generally the same as the start order.[5]

Before each shift, each worker is assigned their own section (of tables). Sections range in size from 10 to 16 seats. Section assignments are based on expected demand, the number and quality of workers, and managerial preferences. With more workers on days with high customer volume, the average size and quality of a section is lower than on days with low customer volume.[6]

[3]Formally, consider an outcome q_b measured at the bill level: $q_b = \mathbf{X}_b + u_b$, where $\mathbf{X}_b$ contains observed factors affecting the outcome and u_b measures unobserved bill-varying factors, such as consumer preferences, that influence q_b. When averaged over the course of a shift, the outcome becomes $\bar{q} = \bar{\mathbf{X}} + \bar{u}$. A rules-based process implies errors are conditionally independent. If errors have mean zero it follows that $\bar{u} = 0$.

[4]An alternative to the design studied in this paper is to increase base wages and have workers transfer more tip earnings to the firm. Since in the past workers objected to such changes by reducing productivity, this design was not implemented.

[5]While workers have minimal discretion over shift length, trades to prolong shifts are possible.

[6]'Quality' on the type of seats in each section. Sections with booth seats have the best quality, sections with benches are second, and sections with chairs have the lowest quality.

Two rules govern the matching process for consumers and workers on days with high customer volume. First, when the restaurant is below capacity, consumers are matched to workers based on the start time of the worker. Second, when the restaurant is at capacity, consumers are quoted an expected wait time and then matched with the next available worker. In both cases neither consumers nor workers have control over the identity of the other party in the match.

Experimental Research Design

The experiment exogenously changes the incentives of workers in November and January of the 2009-2010 season. Before the change, at the start of the 2009-2010 high season and for the entire 2008-2009 season, workers' earnings came from tips and a fixed hourly wage. During the experimental period workers earned money through tips, hourly wages, and a simple linear performance bonus for customer volume.

Workers received a bonus if customer volume, adjusted for shift length and section size, exceed an exogenously determined performance standard.[7] Bonuses were proportional to the distance between actual performance and the performance standard. In consultation with the CEO proportions were chosen so that workers who exceeding the performance standard by a standard deviation earned between $20 and $30, or more than 10% of average daily earnings.[8] Proportions and performance standards were the same for all workers.

The use of a performance standard is motivated by how tasks are allocated between worker and firm. While the firm is responsible for inducing consumer participation, through advertising and marketing, workers are responsible for convincing consumers to purchase more items or to substitute towards more expensive items (*i.e.* up-selling). Since the firm effectively guarantees a minimum level of customer volume for each worker a performance standard was used to ensure workers were not rewarded for maintaining the *status quo*.

In early October of 2009, the CEO and I informed workers about a research project being conducted at the restaurant. We informed workers that the general objective of the research was to study how to reduce wait times for consumers. During this month, I conducted interviews with each worker.[9] We did not inform workers about the performance bonuses for customer volume until November of 2009.

Upon workers' arrival for a previously scheduled shift (on a busy day), I informed workers that they would be paid performance bonuses for customer volume. Since adjustments to performance

[7] More specifically, to calculate the standard I: first, computed long run averages for customer volume, hours worked, and section size on high demand days; second, divided service volume by hours worked to reduce the gains from gaming the system (by trading end times); third, divided customer volume per hour to ensure equitable earnings across workers. The second step yields values of 2.11 consumers per hour for Fridays and 2.72 consumers per hour for Saturdays. The third step yields .4 for Fridays and .41 for Saturdays.

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

[9] All but one worker agreed to participate in the interviews.

outcomes and standards are somewhat involved, I asked workers to explain how the bonuses worked in the context of several hypothetical examples. Workers were paid privately when the shift was completed.[10]

Each worker's experience followed a similar pattern on subsequent treated days. To minimize the influence of sorting on the empirical results the length of the treatment period was not revealed to workers.

The workplace has natural features that help minimize biases from Hawthorne and experimenter-demand effects. One important feature is that, having known about the trade off from higher customer service, the CEO has used other instruments, such as contests and non-monetary incentives, to direct the attention of workers to customer volume on busy days. In light of these instruments, the treatment was introduced to workers as a 'contest' that pays performance bonuses for customer volume.

Overall the experiment combines elements of a cluster design with elements of a randomized block design [List, Sadoff, and Wagner, 2010]. While the unit of analysis is at the worker level, the unit of randomization is calendar date level. A key feature of the data is that the *same* worker is observed in more than one cluster. An alternative strategy where workers are randomized within shifts was not used because the CEO had equity concerns about within shift earnings differences and because complementarities in production might have led to cross contamination between treatment and control groups.

3 Theoretical Framework

In this section I develop a model of individual productivity when workers are rewarded by consumers, through tips, and insured by the firm, through a fixed hourly wage. I use the model to illustrate how and why the interests of workers might conflict with those of the firm. The model is also used to generate predictions about behavior under the experimental treatment.

Setup

Workers are hired to serve customers and to up-sell menu items on behalf of the firm. Individual productivity is measured by daily sales

$$Y = N(h) \times \sum_i q_i(e_i, \epsilon_i) \tag{3.1}$$

[10] In contrast with the natural field experiment considered in [Bandiera, Barankay, and Rasul, 2009], this experiment is not ideal because: as per ethical review, I had to identify myself to workers as a researcher from the University of Toronto and inform them of their participation in a research study; as per the CEO, I had to administer the experiment.

where $N(h)$ is the number of consumer arrivals for employees who work h hours and q_i is category-specific sales per customer. q_i depends on e_i, the effort given to category-specific up-selling, and on a random variable, ϵ_i, measuring category-specific tastes for the 'average' consumer.[11]

The primary earnings source for workers is daily tip earnings, which are proportional to daily sales,

$$DTE = t \times Y. \tag{3.2}$$

Tip percentages, $t = t(\mathbf{e}, \boldsymbol{\tau})$, depend on the effort spent on each category and a random variable, $\boldsymbol{\tau}$, representing consumer preferences over effort and the overall experience at the firm. Hourly earnings then consist of hourly tips earnings and hourly wages,

$$HE = \frac{DTE + wh}{h} = tN(\sum_i q_i) + w. \tag{3.3}$$

Workers are assumed to have quasi-linear preferences over hourly earnings and hourly effort, $U(HE, \mathbf{e}) = HE - c(\mathbf{e})$, where $c_{e_i}, c_{e_ie_i} > 0$ and cross partial derivatives $c_{e_ie_j} \geq 0$ measure the degree of substitutability across tasks.

Timing

The timing of the model is consistent with the actual timing of transactions between consumers and workers:

- Workers are randomly matched with consumers.
- The random vector $(\boldsymbol{\epsilon}, \boldsymbol{\tau})$ is realized, observed by workers, but not by the firm.
- Workers use expectations about $N(h)$ to decide on an effort allocation.
- Consumers use $t(\mathbf{e}, \boldsymbol{\tau})$ to pay workers.

Misaligned Interests

I use a simplified model with two choice variables to illustrate the conflict and to explore the impact of the experimental treatment on behavior. In the model workers can allocate effort to entree (q_1) and post-entree (q_2) sales per customer. The problem for the worker is to

[11] Menu prices are excluded from the analysis because there were no price changes during the principal periods under study. However the model allows one to consider the influence of prices or, more specifically, the influence of complementarities in demand. This is done by letting $q_i = q_i(e_i, \mathbf{p}, \epsilon_i)$, where $\mathbf{p}$ is a vector of category-specific prices. Menu prices are especially important when designing optimal employment contracts [Slade, 1996].

$$\max_{e_1,e_2} E_N[HE] - c(\mathbf{e}),$$

where $E_N[HE] = t\lambda[q_1 + q_2] + w$, $t_{e_i} > 0, t_{e_i e_i} \leq 0$, and e_1 and e_2 have complementary effects on tips, $t_{e_1 e_2} \geq 0$. I assume effort directed at post-entree sales has a smaller impact on overall sales per customer than effort directed at entree sales, $\frac{\partial q_1}{\partial e_1} > \frac{\partial q_2}{\partial e_2} \geq 0$. I also assume $N(h)$ is generated by a Poisson Process where the arrival rate is decreasing in the effort allocated to post-entree sales, $\lambda_{e_2} \leq 0$,[12] at a diminishing rate, $\lambda_{e_2 e_2} \leq 0$.

These assumptions highlight the source of conflict between worker and firm. When workers sell more post-entree items they earn more tips, with small gains in sales, at the expense of longer bill durations. While gaining \$7-8 from the sales increase, the firm loses the additional revenue that comes from serving the next customer. Or, to be more precise, this behavior results in longer quoted wait times and induces customers to visit their next best alternative. Over the course of a shift, as more customers decide to visit their next best alternative, this behavior results in large profit losses for the firm.

The first order conditions for this problem imply optimal effort is characterized by

$$\frac{t_{e_1} + \frac{t}{S}[\lambda \frac{\partial q_1}{\partial e_1}]}{t_{e_2} + \frac{t}{S}[\lambda_{e_2}(\sum_i q_i) + \lambda \frac{\partial q_2}{\partial e_2}]} = \frac{c_{e_1}}{c_{e_2}}, \tag{3.4}$$

where $\frac{t}{S}$ measures the weight workers place on earnings from sales relative to earnings from tips.

Equation (3.4) is more informative when compared with the efficient effort allocation,[13] which is characterized by

$$\frac{t_{e_1} + \frac{(1+t)}{S}[\lambda \frac{\partial q_1}{\partial e_1}]}{t_{e_2} + \frac{(1+t)}{S}[\lambda_{e_2}(\sum_i q_i) + \lambda \frac{\partial q_2}{\partial e_2}]} = \frac{c_{e_1}}{c_{e_2}}. \tag{3.5}$$

Relative to Equation (3.4) more weight is placed on earnings from sales than on earnings from tips. Since under the assumptions of the model, $\lambda \partial q_1/\partial e_1 < \lambda_{e_2}(\sum_i q_i) + \lambda \partial q_2/\partial e_2$, too much attention is directed to post-entree sales. Interestingly, Relations (3.4) and (3.5) coincide when waiters are

[12] These diseconomies of scale arise because during busy periods workers spend less time with each customer and/or provide lower quality service to each customer [Rosen, 1981].

[13] The level of effort that maximizes total surplus satisfies

$$\max_{e_1,e_2} [1 + t(\mathbf{e}, \tau)]\lambda(e_2)[q_1(e_1, \epsilon_1) + q_2(e_2, \epsilon_2)] - c(\mathbf{e}).$$

simply order takers and effort has no influence on customer arrivals, *i.e* when $\lambda_{e_2} = \frac{\partial q_1}{\partial e_1} = \frac{\partial q_2}{\partial e_2} = 0$, or when effort has no influence on tips, *i.e* when $t_{e_1} = t_{e_2} = 0$.

Behavior Under the Treatment

I now show how a simple bonus scheme can generate more favorable outcomes for the firm. A major benefit of the scheme is that it is a low cost alternative to increasing capacity (adding seats) in order to deal with excess demand. While an increase in capacity might improve profits in periods with high customer volume, it also means more unused capacity in periods with low customer volume.

When offered simple linear bonuses for customer volume the problem for workers is to

$$\max_{e_1, e_2} E_N[HE + \alpha(N - T)I(N \geq T)] - c(\mathbf{e}),$$

where I is the indicator function, T is the performance standard, and α governs the magnitude of the bonus based on the expected distance to T. Note that the expectation E_N is decreasing in e_2.[14] The first order conditions now imply

$$\frac{t_{e_1} + \frac{t}{S}[\lambda \frac{\partial q_1}{\partial e_1}]}{t_{e_2} + \frac{t}{S}[\lambda_{e_2}(\sum_i q_i) + \lambda \frac{\partial q_2}{\partial e_2}] + \frac{\alpha}{S}\frac{\partial E_N}{\partial e_2}} = \frac{c_{e_1}}{c_{e_2}}. \tag{3.6}$$

Since $\frac{\partial E_N}{\partial e_2} < 0$ the expression shows the bonus scheme creates strong incentives for workers to direct more attention away from e_2 than before (*c.f.* (Equation 3.4)). Note that as the performance standard becomes very large Equations (3.4) and (3.6) coincide because $\frac{\partial E_N}{\partial e_2}$ approaches 0.

Comparative Statics

To properly study the incentive effects of the experimental bonus one must consider both the influence of the piece rate and the influence of the performance standard. In this section only

[14] To see this consider the expression

$$E_N[(N - T)I(N \geq T)] = \sum_{j=T+1}^{\infty} (j - T)e^{-\lambda(e_2)} \frac{\lambda(e_2)^j}{j!}.$$

Differentiating yields

$$\frac{\partial E_N[(N - T)I(N \geq T)]}{\partial e_2} = e^{-\lambda(e_2)} \lambda_{e_2} \{\sum_{j=T}^{\infty} \frac{\lambda(e_2)^j}{j!}\} < 0.$$

the influence of the piece rate is considered. This is done to simplify the analysis and because a second treatment was implemented later in the same season to generate independent variation in performance standards. Since the piece rate was the same for both treatments the second treatment allows me to identify the incentive effects of the piece rate.

As a consequence, I base predictions about outcomes, when workers are paid bonuses for customer volume, on behavioral responses to α. In terms of the attention given to post-entree sales the model predicts an unambiguous reduction in e_1. Less attention is directed to post-entree sales under the experimental treatment. Or, more formally, $\frac{\partial e_2}{\partial \alpha} < 0$ when the second order conditions are satisfied (because $\frac{\partial e_2}{\partial \alpha}$ is proportional to $(\frac{\partial^2 E_N[HE]}{\partial e_1 e_1} - c_{e_1 e_1})$).

The model also predicts an ambiguous effect on the attention paid to entree sales (since $\frac{\partial e_1}{\partial \alpha}$ is proportional to $-(\frac{\partial^2 E_N[HE]}{\partial e_1 e_2} - c_{e_1 e_2})$). The ambiguity occurs because there are costs to shifting attention from post-entree to entree sales and because post-entree sales has an ambiguous effect on the marginal benefits of entree sales (*i.e.* the sign for $\frac{\partial^2 E_N[HE]}{\partial e_1 e_2}$ is indeterminate *ex ante*). When workers have large substitution costs or the marginal benefits are decreasing in post-entree sales, more attention is devoted to entree sales. On the other hand, when workers have small substitution costs and the marginal benefits are decreasing in entree sales, workers devote less attention to entree sales.

Given the behavioral predictions, the main consequences for observed input choices are: first, more customers are served under the experimental treatment because $\frac{\partial \lambda}{\partial \alpha} = \frac{\partial \lambda}{\partial e_2}\frac{\partial e_2}{\partial \alpha} > 0$. Second, workers sell fewer post-entree items, $\frac{\partial q_2}{\partial \alpha} = \frac{\partial q_2}{\partial e_2}\frac{\partial e_2}{\partial \alpha} < 0$. Third, the effect for per customer sales of other items $\frac{\partial q_1}{\partial \alpha} = \frac{\partial q_1}{\partial e_1}\frac{\partial e_1}{\partial \alpha}$ is ambiguous.

The main predictions for individual productivity and earnings are: first, since $\frac{\partial t}{\partial \alpha} = t_{e_1}\frac{\partial e_1}{\partial \alpha} + t_{e_2}\frac{\partial e_2}{\partial \alpha}$ the treatment has an ambiguous influence on tip percentages. However, there is no effect $\frac{\partial t}{\partial \alpha} = 0$ when effort is not a factor used to determine tip percentages. Second, workers are more productive under the experimental treatment $\frac{\partial Y}{\partial \alpha} = \frac{\partial \lambda}{\partial \alpha} q + \lambda(\frac{\partial q_1}{\partial \alpha} + \frac{\partial q_2}{\partial \alpha}) > 0$ when daily sales is more responsive to changes in the arrival rate than to changes in sales per customer. Third, when paid bonuses for customer volume, workers earn more if $\frac{\partial t}{\partial \alpha} = 0$.

An important feature of the model is that it allows arrival rates to vary within shifts. This consideration is important because it maps more closely with actual ongoings at the firm, where busy days consist of periods of both high and low demand. In high demand periods the marginal benefit to up-selling entrees increases (through $\lambda\frac{\partial q_1}{\partial e_1}$), while post-entree sales have an ambiguous effect on these benefits because of the increase in $\lambda\frac{\partial q_2}{\partial e_2}$ and the decrease in λ_{e_2}.[15]

One caveat of the model is that it assumes workers fully observe $(\boldsymbol{\epsilon}, \boldsymbol{\tau})$ before deciding on effort allocations [Baker, 1992]. In reality decisions are based on signals about $(\boldsymbol{\epsilon}, \boldsymbol{\tau})$. However this

[15] The model also allows consumer preferences to vary within shifts. However managers and workers have both emphasized that there is relatively less variation with shifts on high demand days. This is because 'diners', consumers who typically require more attention and more knowledgeable servers, visit the firm on low demand days, spend more on menu items, and are better tippers than their counterparts who visit on high demand days.

assumption allows me to capture the essence of the informational problem at the firm. Workers have better information because of their interactions with consumers and because of the high costs of monitoring each interaction. The noise from each interaction allows workers to direct more attention to appeasing consumers than is desired by the firm.

Worker Heterogeneity

To explore how behavioral responses vary across workers I first assume that the direct cost to generating effort is larger for less able workers, $\frac{\partial c_{e_i e_i}}{\partial \theta} < 0$, where θ represents the ability of the worker. The assumption implies less attention is directed to post-entree sales per customer by workers with lower ability. Or, more formally, that for each worker θ if $\theta' > \theta$ then

$$\frac{\partial e_2}{\partial \alpha}(\theta) < \frac{\partial e_2}{\partial \alpha}(\theta') < 0. \tag{3.7}$$

The prediction is obtained because small shifts in attention away from post-entree sales result in relatively large cost reductions for low-ability workers. The differences in cost reductions imply that low-ability workers have stronger incentives to reduce post-entree sales when offered performance bonuses for customer volume.

When I assume that more able workers also have lower substitution costs, $\frac{\partial c_{e_1 e_2}}{\partial \theta} < 0$, the model predicts that performance bonuses have a smaller impact on these workers relative to their less able counterparts. Formally, when $\theta' > \theta$, I obtain

$$\frac{\partial e_1}{\partial \alpha}(\theta') < \frac{\partial e_1}{\partial \alpha}(\theta). \tag{3.8}$$

The prediction in (3.8) is obtained because reductions in attention directed at post-entree sales has a smaller influence on the marginal cost of entree sales. When the response $\frac{\partial e_1}{\partial \alpha}(\theta)$ is non-negative (for every ability type), this means that performance bonuses have a lesser influence on the attention directed to entree sales by high-ability workers.

In addition to having analogous consequences for *observed* entree and post-entree sales per customer, the relations in (3.8) and (3.7) have other implications for outcomes used in the econometric analysis. First, Relation (3.7) implies that, unless arrival rates are more responsive to the effort e_2 of high-ability workers, workers with lower ability will have higher arrival rates. Second, in cases where arrival rates are highly responsive to e_2 we should see improvements in overall productivity (daily sales). Third, if there is a worker who earns smaller tip percentages when paid performance bonuses then workers with lesser ability will also earn smaller tip percentages. Moreover, for these workers, the percentages decrease with ability.

4 Data and Context

Scale Economies and Multitasking

In this workplace labor is largely a fixed production cost for the firm.[16] The firm can lower average costs simply by increasing output (as measured in dollars). In the short run, where capacity constraints bind, a good alternative to increasing capacity is to provide workers with incentives for customer volume.

Evidence of this mechanism is provided in Figures 1 and 2, where I use data from the 2006-2007 season to plot average sales and average labor costs in the short run. The figures show that on days with high customer volume small changes in average sales are accompanied by significant reductions in average labor costs. Assuming constant input costs, this implies short run profits are higher on busy days if workers focus more on customer volume. In the absence of monitoring, which has large costs on these days (volume can exceed 800 customers), a conflict of interest can result in large profit losses for the firm.

Data

This study uses data from a large-scale restaurant in the Greater Toronto Area to examine the role of incentive problems related to multitasking in the workplace. For the primary analysis I collected 6 months of transaction level data for October, November, and January 2008-2009 and October, November, and January 2009-2010. The analysis uses days with high customer volume (Fridays and Saturdays), where a significant number of customer arrivals are not served by the firm. Workers were offered performance bonuses for customer volume on these days in November 2009 and January 2010. Overall there are 40 workers, 52 days, and 937 worker-day observations.[17]

Two additional samples are used to supplement the primary analysis. I collected information from February-May of 2008-2009 and 2009-2010 to study the longer term consequences of the initial treatment and to separately identify the incentive effects of piece rates from incentive effects of performance standards. To control for aggregate factors that might confound year-over-year comparisons, such as the recent economic turmoil, I collected information from a second comparable restaurant during the 2009-2010 season. The first sample consists of 42 workers, 104 days, and 1845 observations, while the second consists of 64 workers, 54 days, and 1644 observations. Robustness checks that use both samples are found in the appendix.

Figures 3-7 and the top panel of Table 1 summarize daily individual-level information on input

[16] In reality labor costs become variable following peak demand periods. After the peak period labor decisions are based on the number of seated consumers and on expectations about the number of late arrivals. As a share of total labor costs the variable component is often quite small.

[17] I ignored data from December because customer arrivals are more evenly spread across days of the week and subsequently because there are few days where large numbers of customer arrivals are not served by the firm.

choices, productivity, and tip percentages for the period under study. The raw data helps resolve some of the ambiguity present in the model. An examination of the raw statistics reveals an increase of 2.71+.56=3.27 in the number of customers served, a decrease of -($-.42-$1.08)=$1.50 in the sales to each customer, and an overall improvement of $91.45-$7.42= $84.03 for individual productivity. This evidence suggests input choices which favor customer volume have a stronger impact on individual productivity. Since tip percentages are apparently governed by factors other than effort (.16-.16=0), the evidence also supports the notion that workers earned more under the experimental treatment.[18][19]

More detailed summary evidence on input choices is presented in Figures 8-11. This data reveals reductions in daily post-entree (by (47.22-43.84)-(45.56-39.30)=$2.88) and alcohol sales (-(165.95-169.02)-(163.43-159.42))=$7.08), and improvements in daily entree ((804.24-723.25)-(736.17-749.85) =$94.67) and appetizer/salad sales ((100.54-92.88)-(97.31-90.90)=$1.25). The raw result for post-entree sales is consistent with behavior predicted by the model. Specifically, that workers will reduce post-entree sales to increase arrival rates.

The bottom panel of Table 1 summarizes daily firm-level information on the number of customer arrivals not served. While the raw evidence weakly supports claims about short run profits, -(28.63-31.11) =2.49 fewer customer arrivals left without receiving service from the firm, the conditional evidence (in the next section) shows the treatment had a strong impact in this dimension. A similar case is made for the share of arrivals not served by the firm.[20]

5 Empirical Results

Identification and Econometric Framework

The period under study permits an identification strategy based primarily on within worker differences-in-differences across months and years. Since the firm experienced abnormally low turnover during this period, more than 75 percent of workers appear in both periods, I can compare changes in outcomes for the *same* worker across years to obtain estimates of the treatment effect.[21]

To obtain treatment effects for individual level outcomes I estimate variants of the regression

[18] The incongruence between tip percentages described in Figure 6 and in Table 1 is because the figure uses weekend averages, while the table uses daily averages.

[19] To compute tip percentages I use all credit and debit card transactions over the sample period (75.6% of all bills). The daily tip percentage is just the average for each day, taken over the number of credit and debit card transactions.

[20] The number of arrivals not served is measured with error. Some customers, having seen the length of the queue upon arrival, decide not to visit the firm. Since the firm does not track these customers (because of large costs) the measure used in this study likely underestimates the true number of arrivals not served.

[21] For robustness purposes I consider a second identification strategy in the appendix of this paper that uses within-worker comparisons across comparable restaurants on the *same* day. The estimates from each strategy are strikingly similar.

	Period					
	Oct 08	Nov 08/Jan 09	Change 1	Oct 09	Nov 09/Jan 10	Change 2
Sales per Customer	37.04 (4.36)	38.12 (4.42)	1.08	38.76 (4.63)	38.34 (4.60)	-.42
Customer Volume	28.42 (8.74)	27.86 (8.04)	-.56	26.75 (8.57)	29.46 (8.87)	2.71
Sales	1043.50 (320.22)	1050.92 (290.99)	7.42	1026.64 (314.66)	1118.09 (322.84)	91.45
Tip Percentage (After Tax Sales)	12.53 (2.52)	12.69 (2.52)	.16	12.84 (2.37)	13.00 (2.70)	.16
Performance Bonus	-	-	-	-	11.75 (19.43)	-
Share Receiving Bonus	-	-	-	-	.51 (.50)	-
Observations	135	284		190	328	
Customer Arrivals not Served	32.7 (35.58)	63.81 (49.61)	31.11	23.5 (25.56)	52.13 (37.79)	28.63
Share not Served	4.00 (4.00)	7.94 (4.85)	3.94	3.00 (2.87)	6.31 (4.06)	3.31
Days	10	16		10	16	

Table 1: **Descriptive Statistics.**

model

$$y_{id} = \alpha_i + \beta_1 I_{Nov/Jan} + \beta_2 I_{09-10} + \beta_{DID} I_{Nov/Jan} I_{09-10} + \mathbf{X}_{id}\boldsymbol{\beta} + \epsilon_{id}. \tag{5.1}$$

$\mathbf{X}_{id}$ controls for time-varying factors common to all workers, such as day (Friday or Saturday) effects, calendar week effects, and customer arrival rates, as well as for time-varying factors specific to each individual, such as section characteristics (the number of booth seats, bench seats, and chair seats), days in sample, and average days in sample for members of the peer group.[22] Regressions do not control for calendar date fixed effects because the treatment is not randomized within shifts.

ϵ_{id} represents other time-varying individual specific factors that might influence performance. Apart from statistical error, the main factors driving variability in ϵ_{id} are: first, time variation in the preferences of the average consumer matched to each worker. Such variation poses a threat to the identification strategy if, for example, recent economic shocks induced a shift in consumer types who visit on high demand days or in how consumers who visit spend their money.

Second, time variation in the behavior of support workers and managers at the firm. Changes in support worker or manager behavior related to the treatment incentive scheme, either directly or indirectly, could pose problems for identification. Behavioral changes from other agents at the firm are a threat if, for example, they respond differently to increased customer volume during the

[22]Peer days' in sample proxies for the helping effort available to each worker. In theory this variable has an ambiguous impact on individual productivity. While workers who work more frequently are more able to help others, they are also more likely to spend time socializing.

experimental period.

Experimental Results

The first result shows the experimental treatment had the expected effect on input choices.

RESULT 1. While workers serve more customers in response to performance bonuses for customer volume, it comes at the expense of sales per customer. Consonant with the idea that 'you get what you pay for' [Gibbons, 1998], when rewards are mostly based on customer service, as with the original contract, customer volume gets neglected.

Columns 1-4 of Table 4 show workers served 2.01-2.99 more customers during the treated period, where estimates are statistically significant at a $p < .01$ level (against a two-sided alternative). Columns 5-8 show a statistically significant reduction, at the $p < .05$ level, in sales per customer of $1.51-$1.78. In columns 9-12 I show that, relative to customer volume, sales to each customer fell by 10-14% ($p < .01$). Overall, the average worker sold ≈ 2.49*40=$94.62 more, because of volume improvements, and sold $\approx 1.73 * 28 \approx$ $48.44 less, because of reductions in the sales to each customer. Based on this conservative difference, the average worker produced $94.62 - $48.44 = $46.18 more when paid bonuses for customer volume.

Result 1 reflects the fact that, under the original employment contract where the firm relies on buyer monitoring [Jacob and Page, 1980] and buyer rewards to motivate performance, workers direct more attention sales to per customer than is otherwise desired. As noted in Gibbons [1998], this behavior is consistent with the idea that 'you get what you pay for'. When only some aspects of a job are rewarded then, unsurprisingly, other aspects are neglected.

Estimates from columns 9-12 of Table 5 are revealing about how workers reallocate effort to improve customer volume. The evidence shows a per customer reduction in pre-entree sales of $.39 ($p < .05$), in alcohol sales of $.96 ($p < .05$), and in post-entree sales of $.25 ($p < .05$). The reduction in post-entree sales to each customer is consistent with predictions of the model. Under the experimental treatment workers benefit from selling fewer post-entree items because of the reduction in bill durations. The other results are consistent with ambiguous predictions from the model.[23]

To better grasp the mechanisms underlying the volume improvements I study treatment effects for hours worked and customer volume per hour (columns 1-4 of Table 5). Column 2 shows the average employee worked 16.2 (.27*60) more minutes ($p < .05$) during the treatment period. This means, in addition to experimental earnings and improved tip earnings, the average worker earned $2.23 ($8.25*.27) more in hourly wages. It also means the labor bill for waiters rose by $44.6 during

[23]Intuitively, if alcohol consumption induces longer stays by customers and has a minimal impact on tips then workers have an incentive to reduce alcohol sales. On the other hand, an immediate explanation for the reduction in pre-entree sales is not obvious. It may, in fact, reflect a perverse behavioral response to the treatment.

the experiment. Estimates in columns 1 shows that, even with the increase in hours worked, the average worker served .28 ($p < .10$) more customers per hour.

The evidence on customer volume per hour, which suggests that workers actually moved faster, is supported by the within-shift descriptive evidence in Figures 12. In these figures restaurant sales are calculated at 15 minute intervals, averaged over the number of workers present during the interval, and then averaged over the number of shifts in the (control or treatment) period. The evidence in the bottom right panel indicates there was an increase in sales during peak and post-peak periods under the experimental treatment. When considered in tandem, evidence from the figures and the estimated reduction in sales to each customer suggest this is because of an increase in hourly customer volume.[24]

When one considers the implications for individual productivity the data reveals the following result:

RESULT 2. Offering performance bonuses for customer volume yields large gains in daily sales. These gains suggest that a discrepancy between the tasks covered by employment contracts and the tasks carried out by the worker can significantly distort productivity in the workplace.

Evidence for Result 2 is provided in columns 1-4 of Table 6. Estimates of the treatment effect range from \$49 to \$92, where the results in columns 2-4 are significant at between the 1 and 5 percent levels. The most robust estimate (column 4) implies a \$66 gain in daily sales for the average worker. With an average of 20 workers per shift, the overall gain in daily sales for the firm is $\$66 \times 20 \approx \$1320$.[25] When combined with Result 1, Result 2 supports the prediction that the benefits to improved customer volume, in terms of daily sales, outweigh the costs of reductions in sales to each customer.

The next result investigates relates tip percentages to worker behavior:

RESULT 3. Changes in worker behavior during the treatment period did not influence tip percentages. This suggests that the behavior induced by performance bonuses for customer volume did not adversely affect the overall experience of the consumer.

As shown in columns 5-8 of Table 6 the influence on tip percentages is small and statistically insignificant.[26] At a minimum, Result 3 strengthens the link between multitask incentive problems

[24]Managers rationalized the increase in sales following the peak period as a spillover effect from the behavior of workers during the peak period. Improvements in hourly customer volume during the busiest periods leads to a reduction in wait times quoted to customers and enhances the relative attractiveness of waiting for a table to become available after the busy period.

[25]Patterns in columns 1-5 allow for inferences about the nature of attrition over the sample period. Specifically, they suggest there was an improvement in the labor pool over this period.

[26]One concern is that the estimates for tip percentages (and tip earnings) are an artifact of how they are calculated. If the number of credit and debit card transactions is small then the true tip percentage is measured with considerable noise. As a precaution I ran regressions with tip percentages (and tip earnings) as the dependent variable for different

in this workplace and the classic problem, as originally exposited in [Holmstrom and Milgrom, 1991]. Basically, this is because Result 3 implies the original contract pays piece rates that are not influenced by the behavior of workers.

Result 3 has three main explanations. First, consumers do not use tip percentages to reward effort allocations (or, in the context of the model, $t_{e_1} = t_{e_2} = 0$). Second, new consumers visit the restaurant on high demand days, have no basis for comparison, and simply behave in accordance with conventions on tipping. Third, the treatment did not significantly alter the perceived experience of repeat consumers. Irrespective of the correct interpretation, the findings suggest consumers were not made worse off from the behavior induced by the treatment.[27]

As is the case with individual productivity and tip percentages model predictions about tip earnings are ambiguous. Further exploration of the data yields the following result.

RESULT 4. Workers who direct more attention to consumer volume under the experimental contract earn more money from tips and hourly wages.

Evidence for this result are found in columns 9-12 of Table 6. The estimate in column 12, despite being marginally insignificant at the $p < .1$, suggests that there was a $12 improvement in tip earnings.[28] The estimate is consistent with the evidence used to support Results 2 and 3, which imply that tip earnings improved by .125∗66≈ $8.25. When combined with the change in hourly wages and expected experimental earnings, the average worker earned at least $8.25+2.23+.51*11.75 = $16.47 more per shift.

While Results 3 and 4 do not suggest a decline in outcomes for consumers and workers, Result 2 does not fully reveal the influence on outcomes for the firm. The following result considers the implications for the profitability of the firm:

RESULT 5. The firm is more profitable in the short run when workers are paid bonuses for customer volume in addition to their tips and hourly wages. This, in combination with the previous results, is in line with early predictions from contract theory, that using multiple instruments to deal with incentive problems in multitasking workplaces can improve outcomes for workers and the firm.

Based on the most conservative treatment effect for daily sales ($66), the experimental treatment

sample sizes, according to the number of credit and debit card transactions. The results were qualitatively similar.

[27]A fourth explanation for Result 3 is that if smaller bills are associated with higher percentages reductions in average service quality has two opposing effects. Tip percentages are higher because bill sizes are smaller and are lower because of perceived reductions in service quality. In this case Result 2 reflects the fact that the *net* effect is zero. To explore this possibility I plotted tip percentages against bill size and found a slight inverse relationship for very large parties (with 10 or more consumers). I then eliminated large parties from the sample. This reduced the number of bills by less than 5%.

[28]One explanation for the imprecision of these estimates is that tip percentages are measured with a considerable amount of noise. This is because the number of bills used to calculate tip percentages are often too small to accurately calculate the 'true' mean tip percentage.

resulted in a 18% increase in profits per day. This effect consists of a change in daily revenue $60× 20 workers = $1200, the average daily cost of providing the incentive ≈ $.5*11.75*20 = $117.50, an increase in average daily labor costs $8.25 per hour × .27 hours × 20 = $44.60, and an increase in (food) input costs. Profits plus input costs (net revenue) are then given by $1320-117.50-44.60 ≈ $1158. Using conservative values for the input cost per additional dollar of net revenue (δ) yields the %18 ($\delta \times 1158$/(average daily profits) × 100) increase in daily profits.[29]

Results 3 through 5 are consonant with predictions made in [Holmstrom, 1979], [Holmstrom, 1982], [Holmstrom and Milgrom, 1994], and [Gibbons, 1998]. Specifically, as noted in [Holmstrom, 1979], that the firm can base employment contracts on additional information about worker performance and improve the welfare of workers and the firm.[30]

I use alternative data, at the firm-level, to investigate the factors driving improvements in short-run profitability. Specifically, I use firm-level information to estimate the specification:

$$q_d = \beta_1 I_{Nov/Jan} + \beta_2 I_{09-10} + \beta_{DID} I_{Nov/Jan} \times I_{09-10} + \mathbf{X}_d \boldsymbol{\beta} + \epsilon_d. \tag{5.2}$$

In equation (5.2), q_d equals either the number or share of customer arrivals not served by the firm. $\mathbf{X}_d$ controls for time-varying factors influencing unmet demand, such as the number of arrivals, the day of the week (Friday or Saturday), and the weather, while ϵ_d represents unobserved time-varying factors influencing unmet demand. The identifying assumption for this specification is that unobserved changes in unmet demand from October 2008 to November 2008/ January 2009 are the same on average as the unobserved changes from from October 2009 to November 2009/ January 2010.

Estimates of equation (5.2) are presented in Table 7.[31] Column 1 shows 15.50 fewer customer arrivals ($p < .1$) are not served by the firm in the treatment period. Column 2 supports this result, it shows a 2 percentage point reduction ($p < .1$) in the share of arrivals not served.[32]

Multiple Incentive Instruments and Heterogeneous Responses

If these labor contracts are so profitable why don't firms use them more often? One reason that came from discussions with the CEO is that the benefits to augmented labor contracts are sensitive to the composition of workers employed by the firm. In other firms, where workers might have low

[29]This information (δ and average daily profits) is confidential.

[30]Precise verification of this claim requires a structural econometric model. This is left for future research.

[31]To ensure the accuracy of inferences based on OLS estimates of equation (5.2) I test for $AR(1)$ serial correlation in the errors. The t-statistic from the baseline test for $AR(1)$ serial correlation is equal to -.25 for the regression in column 1 and -.37 for the regression in column 2. Both statistics lead to (strongly) not rejecting the null of no serial correlation in the error terms.

[32]Since the number of arrivals not served is underreported the estimates likely understate the true influence of the experimental treatment.

ability or inadequate training, these contracts might not have a significant impact on profits.

In this and the next section, I use 2038 observations from control-period days (with high and low customer volume) to investigate how treatment responses differ from worker to worker and to draw inferences about how the benefits from the performance bonuses vary across firms with different labor pools. As a first step I obtain individual estimates of average productivity and of average inputs used in the production process. The estimates are based on the following specification:

$$y_{id} = \theta_i + \gamma_d + \mathbf{X}_{id}\boldsymbol{\beta} + \epsilon_{id} \tag{5.3}$$

γ_d is a calendar date fixed effect, and $\mathbf{X}_{id}$ includes days in sample, the square of days in sample, and controls for quality of the section assigned to the waiter (the number of booth seats, bench seats, chair seats). The random variable ϵ_{id} measures the transitory component of individual productivity (or input choices).

When the dependent variable is an output, θ_i measures the average permanent productivity of the worker. Conversely, when the dependent variable measures an input in the production process, such as with customer volume or the sales to each customer, θ_i represents the average permanent input of the worker in that dimension.

Figure 14 provides a graphical view of various estimates of θ_i. The x-axis represents the average permanent effort dedicated to customer volume (per hour worked) and the y-axis represents the average permanent effort dedicated to sales per customer. The number at each coordinate pair is the average permanent productivity of the worker, measured using sales per hour worked as the outcome, of the worker. This figure suggests the behavior of high-ability workers, whose average permanent productivity ranges from \$132 to \$144 per hour, under the original contract was more closely aligned with the interests of the firm than the behavior of low-ability workers. Specifically, under the original contract high-ability workers directed relatively more attention to customer volume.

The empirical analysis exploits the fact that workers naturally fall into three to four categories, based on the 'level sets' in Figure 14. In Figure 15 I plot the productivity distribution and illustrate the precise criteria used to allocate workers into groups: low ability, workers whose averages sales per hour is more than one standard deviation below the mean sales per hour (across all workers); average ability, workers whose average sales per hour is within one standard deviation of the mean; high ability, workers whose averages sales per hour is more than one standard deviation above the mean. In all, there are 6 low-ability workers (16.2%, with 121 observations in treatment-control period), 24 average-ability workers (64.8%, with 639 observations), and 7 high-ability workers (18.9%, with 172 observations). This characterization yields the following result.

RESULT 6. Performance bonuses for customer volume has the largest impact on the productivity of high-ability workers. The benefits from augmenting labor contracts to deal with multi-task agency problems largely depend on the response of high-ability workers.

Evidence for the first part of Result 6 is presented in column 8 (row 3) of Table 8. When paid bonuses for customer volume high-ability workers sold \$90.60 ($p < .1$) more per day. Estimates from column 1 (rows 1 and 2) of Table 8, on the other hand, show for average- and low-ability workers the treatment had no statistically discernible influence on daily sales.

While the treatment only had a significant impact on the productivity of high-ability workers, it had a strong impact on the effort allocations of *all* workers. Column (3) of Table 8 shows that relative to customer volume the treatment induced a 15% ($p < .1$) reduction in sales to each customer for low-ability workers, a 10% ($p < .1$) reduction for average-ability workers, and a 17% ($p < .01$) reduction for high-ability workers.

When one considers the level effects on customer volume, the data is consistent with the idea that arrival rates are more responsive to the effort allocations of high-ability workers. This claim is supported by the estimated productivity differences across workers (as per the model) and by the estimates in Column (1) of Table 8, which show customer volume improved by 3.14 ($p < .01$) customers for high-ability workers and by 2.24 ($p < .05$) for average-ability workers. While the estimate for low-ability workers is larger than for workers with average ability, it is (marginally) statistically insignificant at the 10% level.

In contrast, the level effects on sales per customer are strongest for low-ability workers. Column (2) shows the performance bonuses induced a reduction in sales per customer of \$2.62 ($p < .05$) and \$2.15 ($p < .05$) for low- and high-ability workers, respectively. The treatment did not have statistically significant impact on sales per customer for workers with average ability.

The estimates in the middle panel of Table 8 provide information about the mechanisms underlying differences in behavior at this workplace. It shows the treatment induced larger behavioral changes for low and average-ability workers than for high-ability workers: low-ability workers reduced per customer sales of pre-entree and alcohol items by \$.47 ($p < .1$) and \$1.37 ($p < .01$), respectively; average-ability workers reduced per customer sales of pre-entree, alcohol, and post-entree items by \$.44 ($p < .1$), \$.90 ($p < .01$), and \$.35 ($p < .01$); high-ability workers reduced per customer sales of alcohol items by \$.90 ($p < .1$). This evidence is suggestive about what separates good from bad workers in multitasking workplaces:

RESULT 7. For the most able workers there are large productivity gains from small changes in up-selling behavior. Most of the improvement occurs because these workers move faster.

This evidence is also somewhat consistent with model predictions about heterogeneous responses to performance bonuses for customer volume. The model predicts that, first, the reductions in sales of items which typically lower arrival rates are largest for less able workers. Evidence for this prediction is presented in columns (4) and (6), which shows that the reductions in sales of alcohol and pre-entree items (per customer) are largest for low-ability workers. In contrast with predictions from the model, the evidence in column (7) shows that only average-ability workers reduced their sales of post-entree items. One explanation for this phenomenon might be that low- and high-ability workers did not direct their attention towards selling these items on busy days. This explanation is consistent with the idea that high-ability workers exploited the trade off under the original contract.

A second prediction about heterogeneity in responses to performance bonuses is that, in terms of sales of entree items, the response for more able workers is weaker than for less able workers. Although not statistically significant, the evidence in column (5) is consistent with this prediction: high-ability workers reduce entree sales by \$.38; average-ability workers increase entree sales by \$.13; low-ability workers increase entree sales by \$.19.

A natural next step is to determine if the patterns, particularly for low-ability workers, are congruent with a reduction in the overall service quality delivered to consumers. The data on tip percentages reveals that this is, in fact, the case.

RESULT 8. When the firm uses multiple instruments to deal with multi-task agency problems, the cost includes large reductions in the overall service quality delivered to customers of low-ability workers.

Initial evidence for Result 8 is presented in column (9) of Table 8. While the estimate is marginally significant at the 10% level, it suggests that, when paid performance bonuses, low-ability workers experience an 8% reduction in tip percentages.

A closer examination of the data (Table 2) reveals, consonant with model predictions, a stronger effect (both in terms of magnitude and statistical significance) as one moves further out in the left tail of the ability distribution. Columns (2) and (3) respectively show a 9% ($p < .1$) and 11% ($p < .05$) reduction in tip percentages for workers who are 1.1 and 1.2 standard deviations below the mean ability level.

Short Run Profits

I complement the empirical analysis with information on input costs and itemized profit margins to evaluate the short run relationship between the profits from augmented labor contracts and the composition of workers employed by the firm. This exercise is based on the expected change in profits per worker

$$E[\Delta\Pi] = E[\Delta\Pi|Low]Pr(Low) + E[\Delta\Pi|Average]Pr(Average) + E[\Delta\Pi|High]Pr(High), \quad (5.4)$$

	Dependent Variable = ln(Tip Percentages)		
	Distance to Mean Ability		
	1 Standard Deviation	1.1 Standard Deviations	1.2 Standard Deviations
	(1)	(2)	(3)
Nov/Jan × 2009-2010 ×			
Low Ability	-.08·	-.09*	-.11**
	(.05)	(.05)	(.05))
Average Ability	-.002	.01	.01
	(.04)	(.04)	(.04)
High Ability	.02	-04	-.04
	(.05)	(.06)	(.06)
R^2	.05	.05	.05
Observations	926	926	926
Workers	38	38	38
Days	52	52	52

Table 2: **Reductions in the Perceived Quality of Service.** Regressions use daily data at the individual level and the same restaurant last year as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and · for estimates marginally significant at the 10 percent level. All regressions control for worker fixed effects, calendar week fixed effects, the day of the week (Friday or Saturday), section characteristics, customer arrivals, and days in sample. Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for the individual and average days in sample for the peer group.

where treatment effects for individual productivity (Table 8, Column 8) are used to calculate the difference in sales, hours worked (Table 8, Column 11) and the minimum wage in Ontario ($8.25/hour) are used to calculate differences in the labor bill, and experimental earnings are used to calculate incentive costs. To compute the expected percentage gain in daily profits I subtract other input costs from equation (5.4), multiply by the average number of workers, and divide by average daily profits.

Calculations are presented in Table 3. Column 4 (6) evaluates the gains at the lower (upper) confidence bound of the treatment effect for daily sales. Column (5) sets the statistically insignificant coefficients to zero. While profits are generally smaller when there are fewer high-ability workers, column (5) shows that profits become negative when they comprise less than 15% of the workforce. Overall, the calculations suggest that augmented labor contracts have larger benefits in workplaces where high-ability workers are more common.

Long Run Profits

A major concern with this study, especially when one considers the behavior of low-ability workers, involves the long run consequences of augmented contracts that deal with multi-task agency problems. To explore this concern I use transaction level information from the control and treatment periods on the number of visits by each consumer type, where types are defined by

	Shares, by Ability			Change in Profits per Worker (%)		
	Low	Average	High	(4)	(5)	(6)
	0.16	0.33	0.50	-6.14	12.09	50.86
	0.18	0.36	0.45	-7.04	10.45	50.37
	0.20	0.40	0.40	-7.94	8.82	49.89
	0.21	0.43	0.35	-8.84	7.18	49.40
	0.23	0.46	0.30	-9.74	5.55	48.91
	0.25	0.49	0.25	-10.64	3.91	48.42
	0.26	0.53	0.20	-11.54	2.27	47.93
	0.28	0.56	0.15	-12.44	0.64	47.44
	0.30	0.59	0.10	-13.34	-1.00	46.96
	0.31	0.63	0.05	-14.23	-2.63	46.47
	0.33	0.66	0.00	-15.13	-4.27	45.98
Performance Bonus	8.51	11.69	14.19			
	(16.34)	(19.37)	(21.43)			
Share Receiving Bonus	.44	.50	.59			
	(.50)	(.50)	(.49)			

Table 3: **Multiple Incentive Instruments and Short Run Profits.** Estimates in columns (4)-(6) are based on treatment effects for daily sales from Table 8. Column (4) uses the lower confidence bound for each worker type. Column (6) uses the upper bound. \$90.60 for average-ability workers and \$0 for high and low-ability workers.

consumers who paid using credit cards *and* by the first four and last two digits of the credit card number. The information is summarized in Figure 13. The raw evidence in this figure suggests that, since repeat visits are not common on days with high customer volume, the adverse behavior of low-ability workers under the experimental treatment has minimal long term implications for the profitability of the firm.[33]

6 Conclusion

This paper uses field evidence from a large-scale restaurant to investigate distortions generated by simple contracts that reward workers who operate on more than margin, in terms of the impact on outcomes for consumers, workers, and the short-run performance of the firm. This paper also investigates why firms might not find it profitable to expand the scope of existing contracts to deal with incentive problems in multitask jobs. The data shows that distorted incentives under the original simple contract generated large losses in short-run profits, earnings for workers, but had a negligible impact on tip percentages. The data also reveals that augmented labor contracts are profitable when at least 15% of the workforce has high ability.

A natural extension considers the interplay of incentive pay offered under the experimental treatment with the risk preferences of workers. Such an exercise would provide a basis for quanti-

[33]The evidence in Figure 13 is also consistent with observations made by managers and workers at the firm. 'Diners', who are most likely to be affected by the change in behavior, visit the restaurant more regularly but do so on days with low customer volume.

fying inefficiencies associated with the experimental contract relative to the optimal contract with incomplete information (See *e.g.* [Ferrall and Shearer, 1999]).

Fundamentally, as is the case with many incentive problems, the conflict between worker and firm under the original contract is rooted in a trade off between short and long term rewards. Workers who appease consumers earn more in the near term, but might earn less over the long term if the firm bases future decisions (about *e.g.* shift and section allocations) on current performance.[34] This study suggests that existing long run incentives are not strong enough to achieve the desired allocation. This is a common problem in more sophisticated jobs. Executives who are rewarded based on accounting profits, for example, have an incentive to sacrifice future for current profits through reductions in expenses such as those used to finance R & D [Murphy, 1999].

This paper also has implications for compensation policies used by governments. Specifically, results from this paper suggest the application of minimum wage laws to pay for performance jobs is not prudent. While internal compensation policies are important instruments for dealing with incentive problems, minimum wage laws that regulate hourly wages constrain firms to provide more insurance than is otherwise optimal. Alternative minimum wage policies, which allow for the use of incentive pay in lieu of hourly wages, could improve outcomes for consumers, workers, and the firm.

[34]For more detailed discussions on the role of current performance for future rewards within firms see [Kahn and Huberman, 1988] and [Prendergast, 1993].

References

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.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, July 2009.

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.

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.

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.

Robert Gibbons. Incentives in organizations. *Journal of Economic Perspectives*, 12(4):115–132, Fall 1998.

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.

Glenn W. Harrison and John List. Field experiments. *Journal of Economic Literature*, XLII: 1009–1055, December 2004.

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 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.

Bengt Holmstrom and Paul Milgrom. The firm as an incentive system. *The American Economic Review*, 84(4):972–991, September 1994.

Nancy L. Jacob and Alfred N. Page. Production, information costs, and economic organization: The buyer monitoring case. *American Economic Review*, 70(3):476–478, June 1980.

Charles Kahn and Gur Huberman. Two-sided uncertainty and 'up-or-out' contracts. *Journal of Labor Economics*, 6(4):423–444, October 1988.

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

John A. List, Sally Sadoff, and Mathis Wagner. So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design. *Experimental Economics*, 2010. Forthcoming.

Alexandre Mas and Enrico Moretti. Peers at work. *American Economic Review*, 99(1):112–145, March 2009.

Kevin J. Murphy. Executive compensation. *Handbook of Labor Economics*, 3(2):2485–2563, 1999.

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.

Canice Prendergast. The role of promotion in inducing specific human capital acquisition. *Quarterly Journal of Economics*, 108(2):523–534, May 1993.

Canice Prendergast. The provision of incentives in firms. *Journal of Economic Literature*, 37(1): 7–63, March 1999.

Ram T.S Ramakrishnan and Anjan V. Thakor. Competition and cooperation in agency. *Journal of Law, Economics, and Organization*, 7(2):248–283, 1991.

Sherwin Rosen. The economics of superstars. *American Economic Review*, 71(5):845–858, December 1981.

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

Margaret E. Slade. Multitask agency and contract choice: An empirical exploration. *International Economic Review*, 37(2):465–486, May 1996.

Joseph E. Stiglitz. Symposium on organizations and economics. *Journal of Economics Perspectives*, 1(2):15–24, Spring 1991.

A Robustness Checks

In this section I present evidence against alternative explanations for the main (average) results found in this paper. Specifically, using two years of data from the same restaurant and data from a comparable restaurant during the 2009-2010 I rule out explanations based on changes in consumer behavior (due to the most recent economic crisis), changes in the behavior of other agents (managers and support staff) at the firm, and temporary changes in worker behavior. Evidence from this section reinforces the main conclusions drawn in the paper.

Incentive Effects of Piece Rates or Performance Standards

To explore the relative importance of piece rates a second treatment was conducted on days with high customer volume in May 2009. In the second treatment each worker is presented with a common piece rate and an individual-specific performance standard. Performance standards are calculated using the section assignment of the workers. At the start of each shift I observed the section assignment of each worker. Using the long run customer volume for each section I calculated targets for each worker.

(REMAINDER TO BE WRITTEN)

A.1 Economic Shocks

Estimates of the impact of the experimental incentive scheme are prone to (upward) bias if recent trends in the overall economy induced an artificial reduction in the difference in outcomes from October 2008 to November/January 2009 relative to the difference for October 2009 to November/January 2010. Such reductions can be attributed to: first, a change in the type of consumers that visit the restaurant. The difference in outcomes for the 2008-2009 season are artificially inflated if, for example, more consumers visited the firm in the October 2008 relative to October 2009. Second, a change in behavior for consumers that continued to visit the firm in the presence of an economic crisis. In this case the difference is inflated if each consumer spent more in October 2008.

The top panel of Table 9 provides information about changes in consumer behavior at the treated restaurant across the various periods under study. Evidence from the top left panel of this table suggests an increase in consumer spending did occur in October of 2008. The proportion of bills (using a single payment method) paid by credit card was .61 in October 2008 and .54 in November 2009/January 2010. In comparison the proportion was .52 in October 2009 and .55 in November 2009/January 2010. The corresponding differences-in-differences is equal to .10. Similar evidence from the top right panel suggests, that while there are large differences in the proportion of bills paid by credit, there are relatively small differences in the proportion of bills using more than one payment method.

To assess this threat to internal validity I compare outcomes for workers at the treated franchise with outcomes for workers at another franchise in the same corporation during the 2009-2010 season.[35] More specifically the identification strategy used in this appendix is to compare *within* worker differences *across* restaurants on the *same* day. Overall this strategy relies on 1533 observations, 50 days, and 64 workers to estimate the impact of the treatment incentive scheme on worker behavior.

Evidence for the comparability of the control restaurant is presented in the bottom panel of Table 9 and Figure 16. The bottom panel of Table 9 shows changes in consumer spending in the control restaurant in October 2009 (relative to November 2009/January 2010) are roughly similar to changes at treated restaurant in October 2009. Figure 16 shows that, while there are level differences in the number of customer arrivals, the patterns are strikingly similar for 2009-2010 season across restaurants.

Estimates of the impact of the treatment incentive scheme are based on the specification:

$$y_{ird} = \alpha_i + \beta_1 T_{1rd} + \beta_2 T_{2rd} + \gamma_d + \mathbf{X}_{ird}\boldsymbol{\beta} + \epsilon_{ird}. \tag{A.1}$$

where $\mathbf{X}_{ird}$ includes days in sample, peer days in sample, the number of consumer arrivals, arrivals squared, and controls for quality of the section assigned to the waiter (the number of booth seats, bench seats, chair seats) in restaurant r ($r \epsilon \{1, 2\}$). γ_d is a fixed effect for the calendar date. The random variable ϵ_{ird} measures the transitory component of the performance of workers at restaurant r on date d. T_{1rd} (T_{2rd}) indicates if workers in restaurant r received the first (second) treatment on date d.

Regression estimates for (A.1) are presented in Table 11. Interestingly, the estimates are similar in sign and magnitude to the results presented in Table 4. Note that estimates of the impact are measured imprecisely in columns (7) and (8). The imprecision is not surprising because, in contrast with the earlier identification strategy, the estimates are based on comparisons of workers at two different restaurants rather than on comparisons for the same worker across years.

Estimates in columns (1)-(4) of Table 12 provides information about the channels used to increase customer volume.[36] Columns (1) and (2) reveal that, when offered performance bonuses for customer volume, the average worker served 1.11 more customers *per hour* ($p < .01$) and worked .25 fewer hours ($p < .05$) than workers from the control restaurant. In contrast with the estimates in columns (4)-(6) of Table **??**, which suggest an year-over-year increase in hours worked, these

[35] A differences-in-differences identification strategy is used (rather a triple difference approach) because information from the control restaurant during the 2008-2009 season was not available.

[36] Note that results from Table 13 are each robust to each of the specifications explored in Table 11.

results suggest a reduction in hours worked for the average worker at the treated restaurant relative to workers at the control restaurant. Consonant with this evidence in columns (3) and (4) I show the average treated worker served .87 more bills ($p < .01$) and reduced average bill duration by .07 hours ($p < .01$).

In columns (4)-(12) I investigate the impact of performance bonuses on the sales of items from various categories. Columns (5)-(7) provides evidence for an increase in the daily sales of most items, including pre-entree, entree, and alcohol items. The result is not surprising because of a mechanical relationship between the sales of these items and the number of customers served. Of greater import is the result in column (8), which shows the treatment had no statistically discernible impact on daily post-entree sales. This result suggests that relative to workers in the control group the experimental contract better aligned the interests of workers with those of the firm. The treatment induced workers to trade-off the added benefits from post-entree sales for the added benefits from customer volume. The estimates in columns (9)-(12) show that workers sold fewer post-entree items to each customer (valued at \$.18, $p < .05$) and that the treatment did not effect per customer sales of other items for the average worker.

Table 13 shows that with a control restaurant the treatment incentive scheme had a similar impact on profits, consumer satisfaction, and tip earnings. Consistent with the magnitude of previous estimates, columns (1)-(4) suggests an increase in daily revenue of between \$73.56-\$92.52 ($.01 < p < .05$) while columns (5)-(8) again suggest a neglible impact on consumer satisfaction. Estimates of the impact on tip earnings (columns (9)-(12)) are similar in magnitude to previous estimates but at best are marginally statistically significant.

A.2 Manager and Coworker Behavior

Matching Consumers with Workers

A second concern involves changes in the assignment mechanism, used to match consumers with workers, in response to the treatment. My estimates are biased if hosts and/or managers, agents who observe consumer characteristics before allocating consumers to workers, match consumers with workers based on expected bill durations. To investigate this mechanism as a potential confounding factor I estimated Specification (5.1), a triple-difference specification using data from the treated restaurant over two seasons, and Specification (A.1) with average table usage:

$$y_{ird} = (\text{seats filled/table capacity})_{ird}$$

as the dependent variable. This proxy measures, in part, the assignment decisions of hosts/managers.

Regression results for this dependent variable are provided in Table 16. Estimates based on full seasons (columns (5)-(12)) show the treatment incentive scheme did not have a statistically significant impact on agents responsible for the assignment mechanism at the firm. Estimates from columns (1)-(8), while marginally significant at $p < .1$, suffer from similar consumer selection issues discussed earlier. More specifically, if the recent downturn is concurrent with a reduction in the average group size of visiting patrons then it could introduce upward bias into these estimates.

Section Assignments

A related concern is that managers can influence consumer-worker matches through section designations, how tables are divided into sections, and/or the assignment of sections to workers. To explore the impact of this concern, I used two years of control period data (leading up to the introduction of the first treatment) from high demand days to obtain long run measures of service volume for each table, computed the average service volume of the section assigned to each worker (for each shift), and estimated treatment effects for the constructed measures. Estimates are provided in Table 14. Columns (1)-(4) use binary measures of section quality (in terms of volume), indicating if the average volume of the worker's section is above mean (or median) average service volume across all tables, and probit regressions to estimate the effect of the treatment on section assignments. Columns (5)-(6) use the constructed continuous measures of section quality. All regressions suggest the treatment had a negligible effect on the section quality of workers.

A.3 Other changes in Worker Behavior

Hawthorne Effects

To address concerns about whether the results are driven by transitory responses to the treatment I estimate the baseline specification (Equation 5.1) for various time windows around the initial introduction of the treatment. Estimates are presented in Table 15. The data reveals that parameter estimates approach the values previously obtained within three weeks of the first treatment date.

B Figures

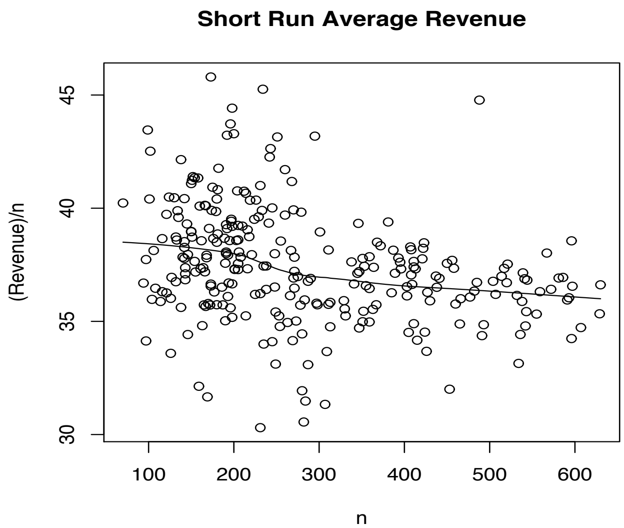

Figure 1: **Short run average revenue for the period September 01, 2006 to June 06, 2006.** The x-axis measures the number of customers served. The y-axis measures total daily revenue per customer served. Each point in the figure represents a unique calendar date. A lowess estimator is used to fit the data.

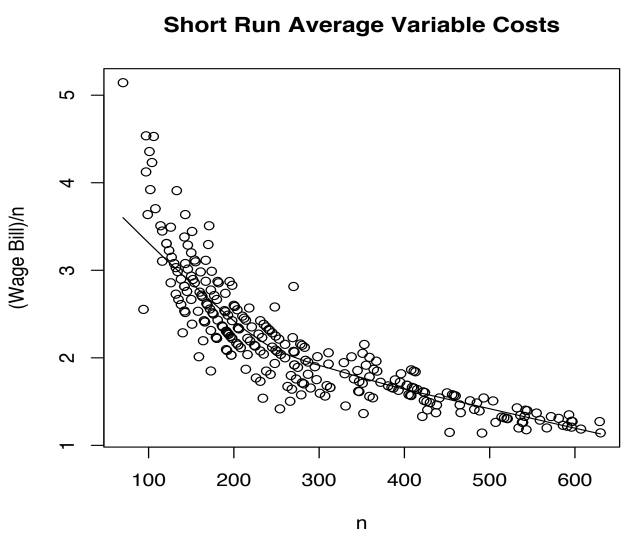

Figure 2: **Short run average cost for the period September 01, 2006 to June 06, 2006.** The x-axis measures the number of customers served. The y-axis measures total daily labour cost (at the minimum hourly wage of $8 per hour) per customer served. Each point in the figure represents a unique calendar date. A lowess estimator is used to fit the data.

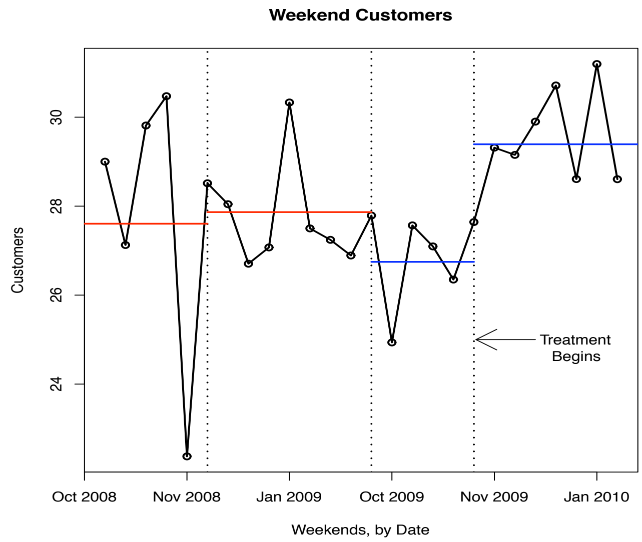

Figure 3: **Customers.**

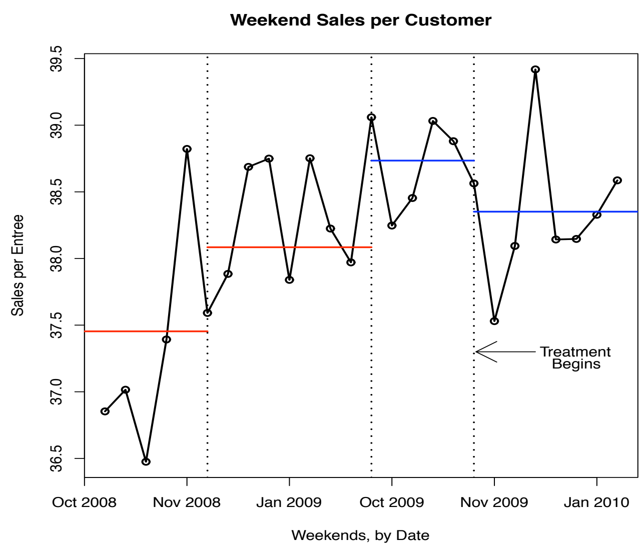

Figure 4: **Sales per Customer.**

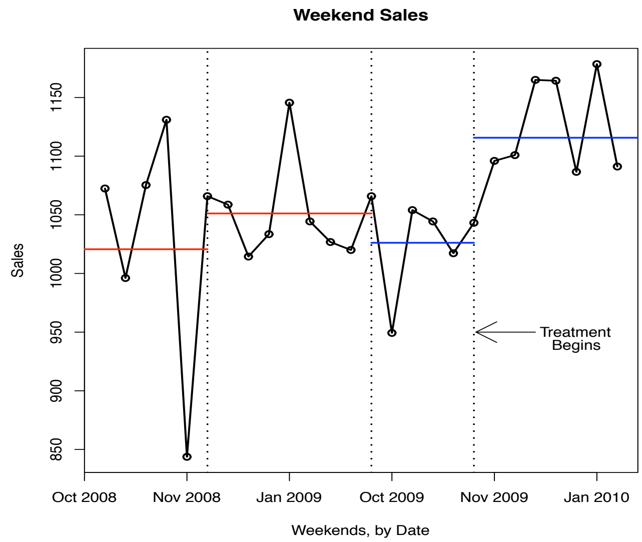

Figure 5: **Sales.**

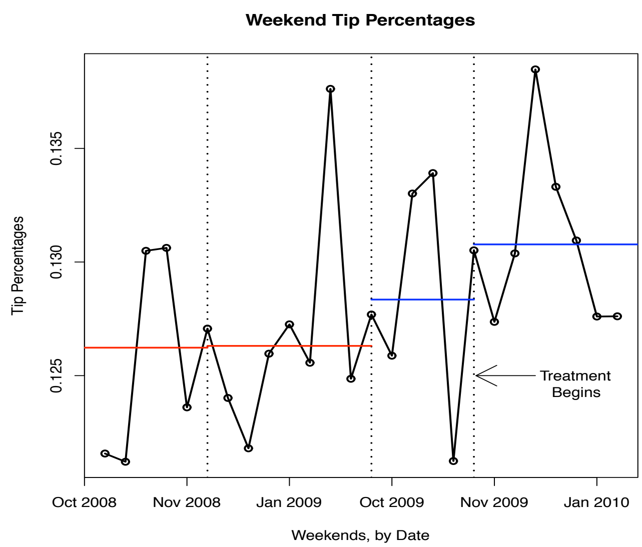

Figure 6: **Tip Percentages.**

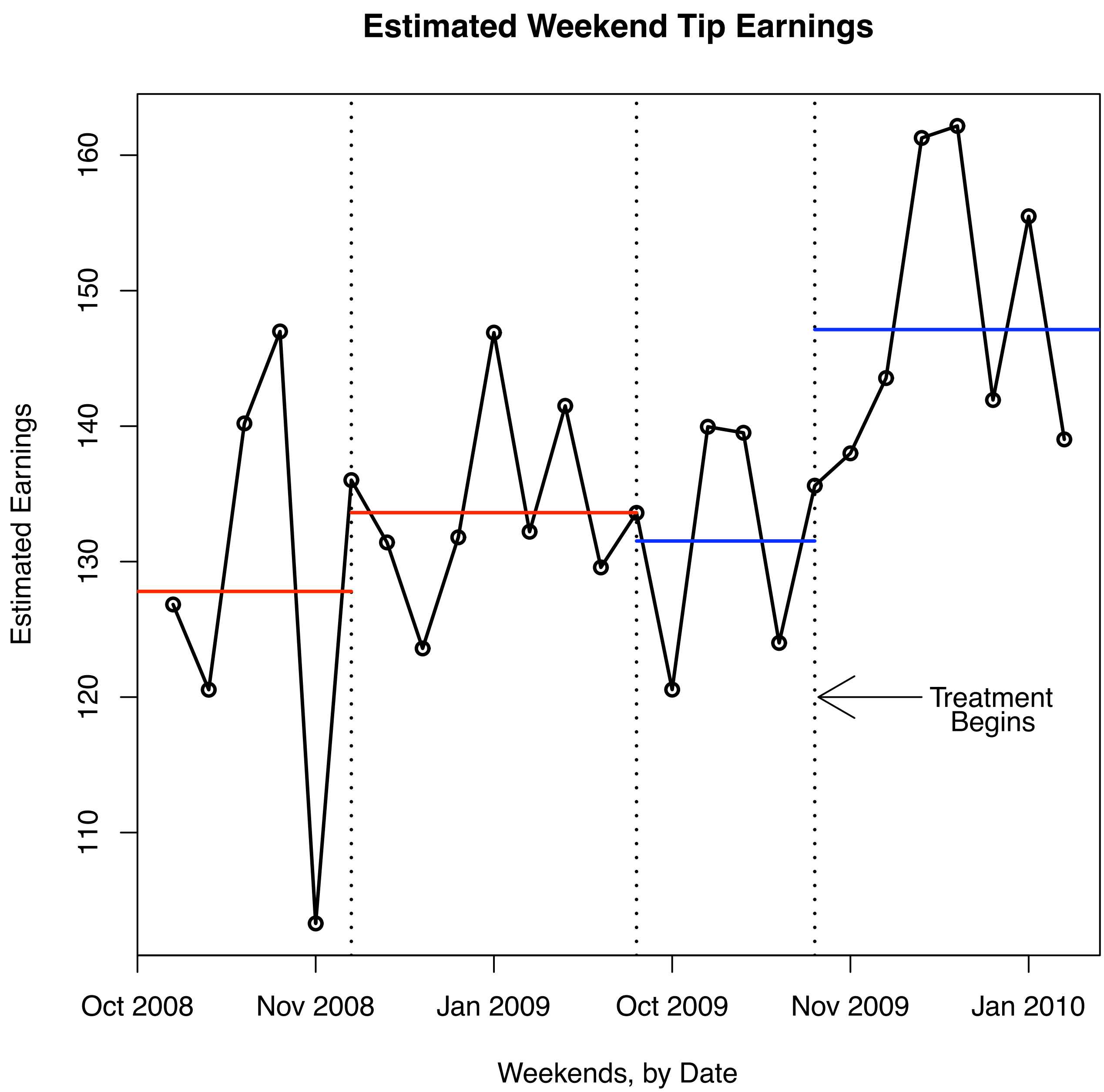

Figure 7: **Estimated Tip Earnings.**

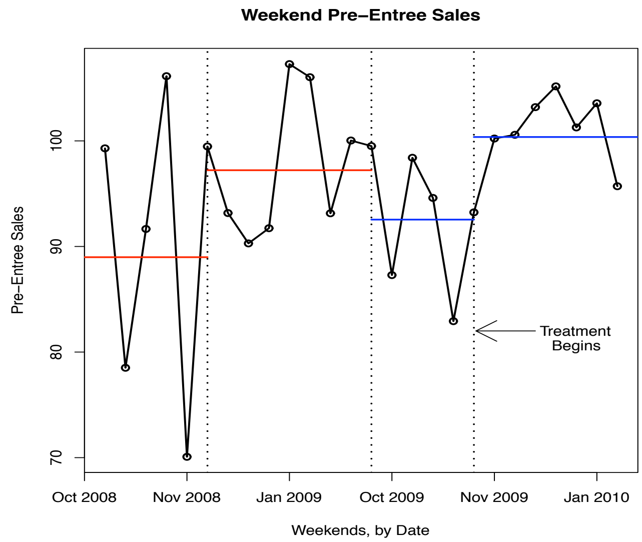

Figure 8: **Sales of Appetizers and Salads.**

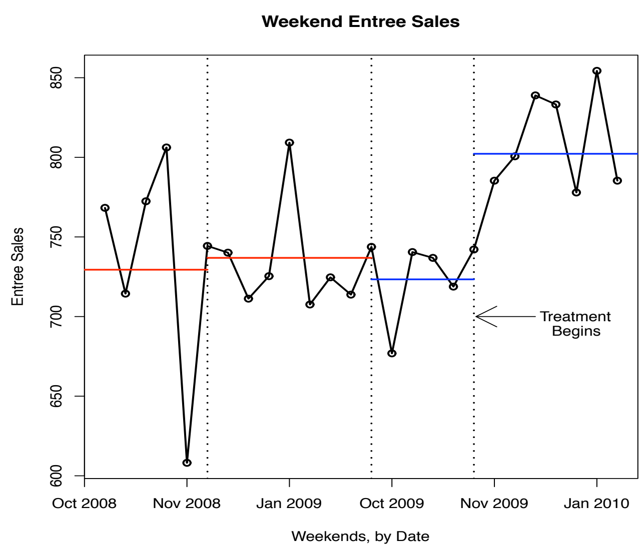

Figure 9: **Sales of Entree Items.**

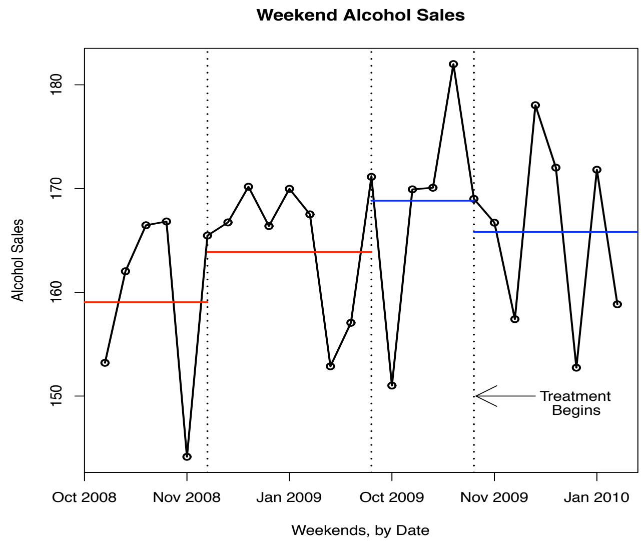

Figure 10: **Sales of Alcoholic Beverages.**

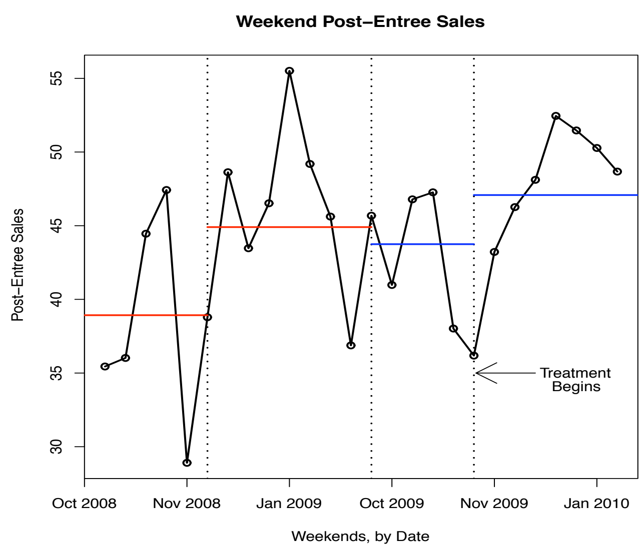

Figure 11: **Sales of Desserts, Coffee, and Tea.**

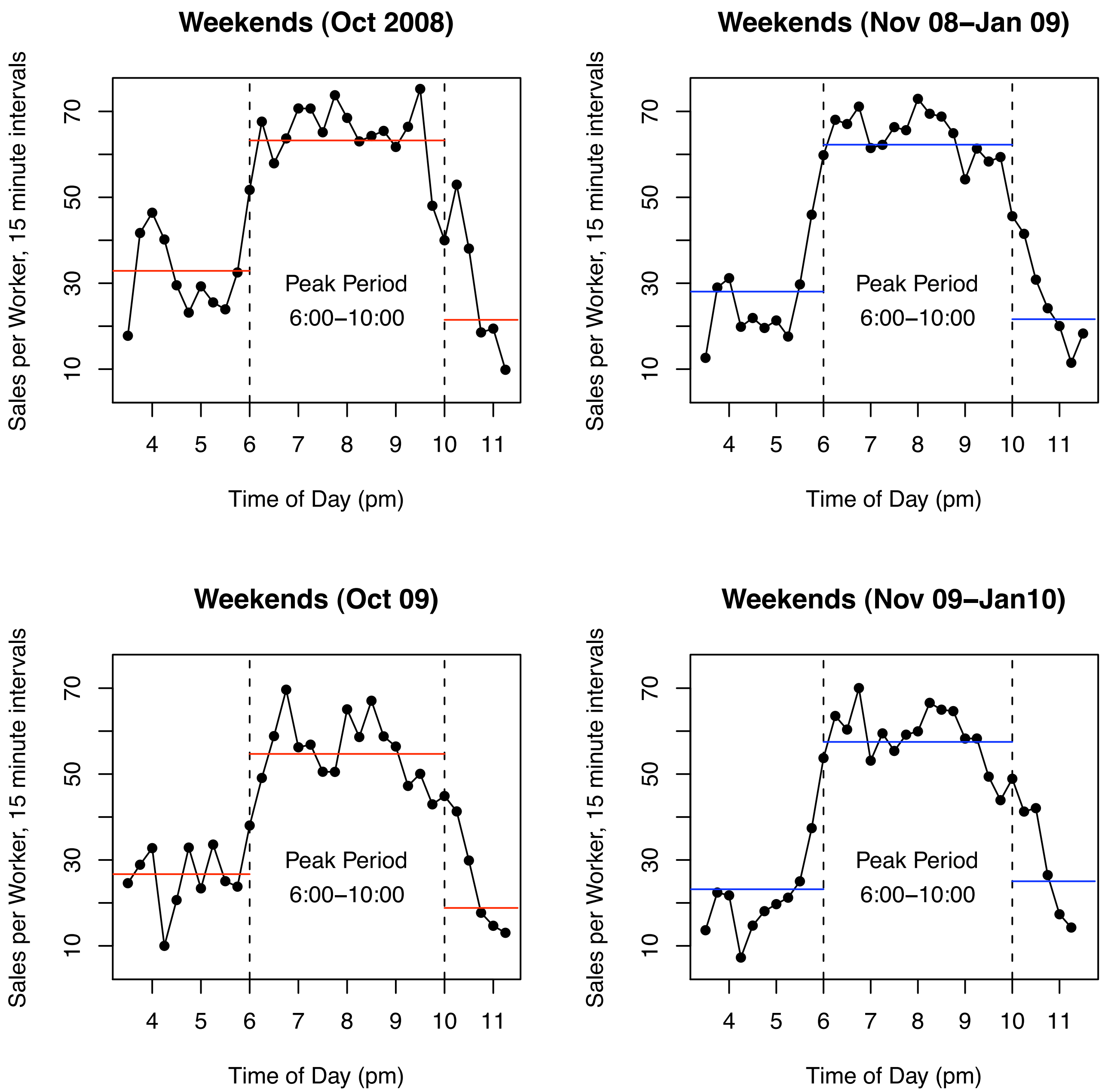

Figure 12: **Weekend Sales, by 15 minute intervals.** Restaurant sales are calculated at 15 minute intervals, averaged over the number of workers present during the interval, and then averaged over the number of shifts in the (control or treatment) period.

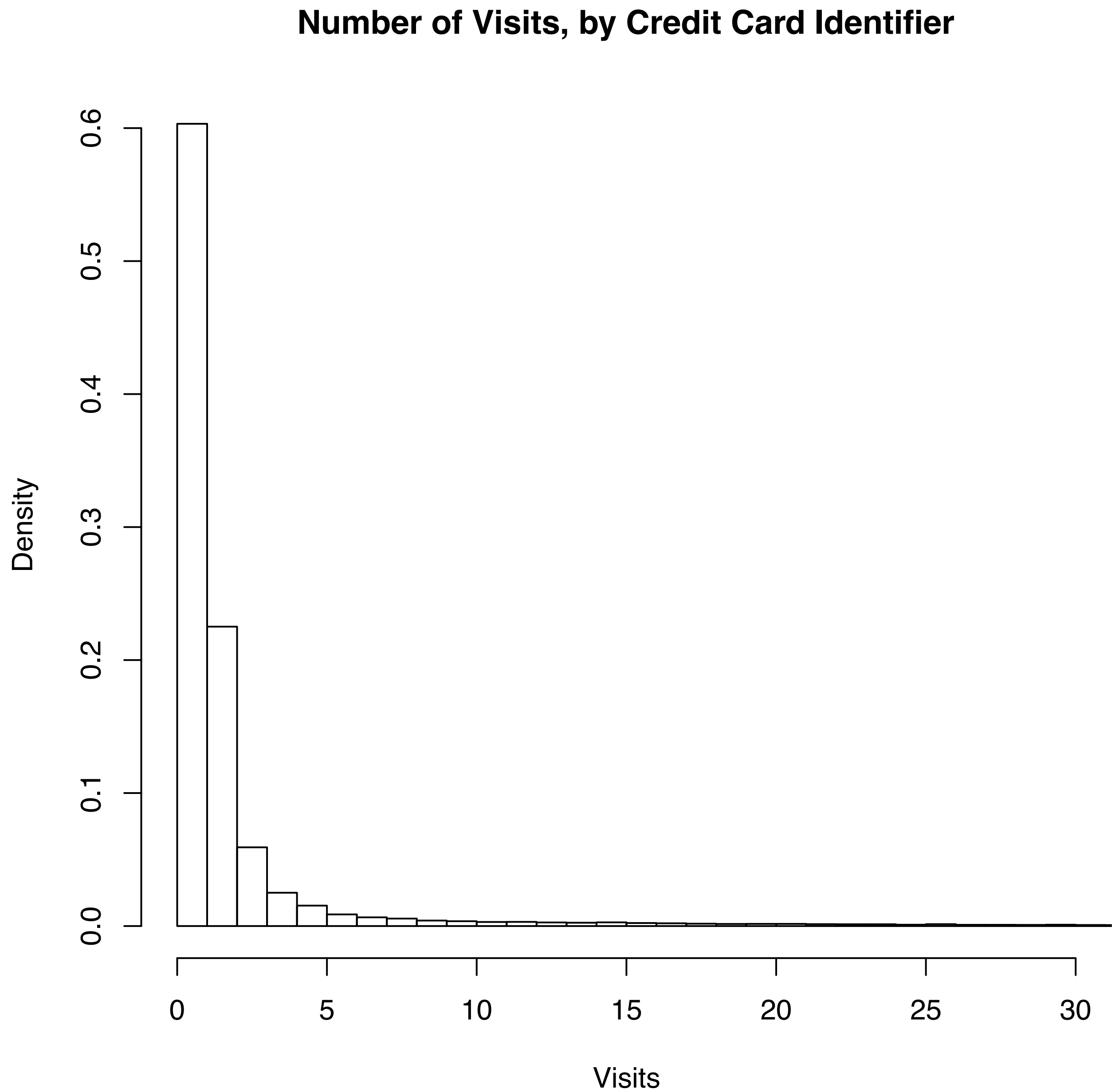

Figure 13: **Repeat Visits.**

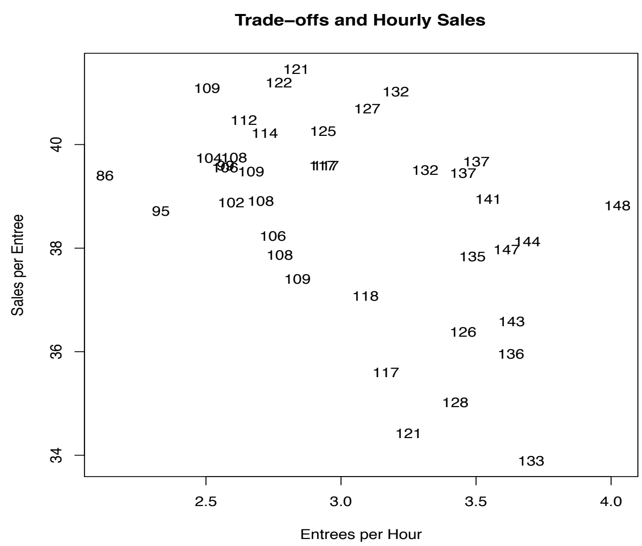

Figure 14: The x-axis represents average permanent effort dedicated to service volume (per hour worked) and the y-axis represents the average permanent effort dedicated to service quality. The number at each coordinate pair is the average permanent productivity, measured using sales per hour worked as the outcome, of the worker. 38 of 39 workers are represented in the figure. The omitted worker had a negative fixed effect for Entrees per Hour.

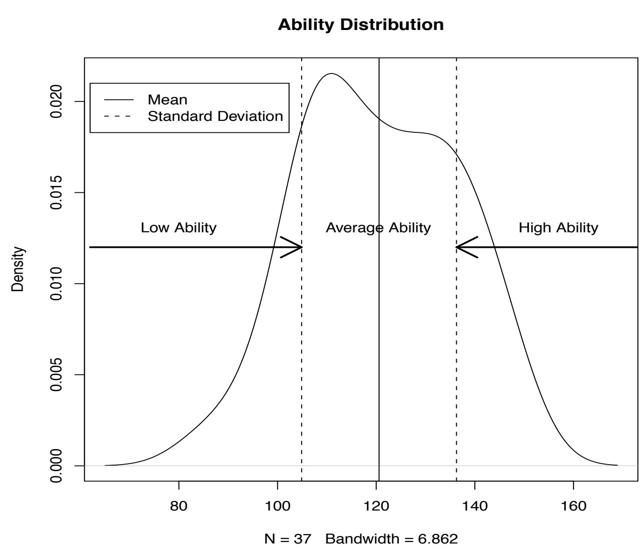

Figure 15: Ability is measured using Equation (5.3) with sales per hour as the dependent variable.

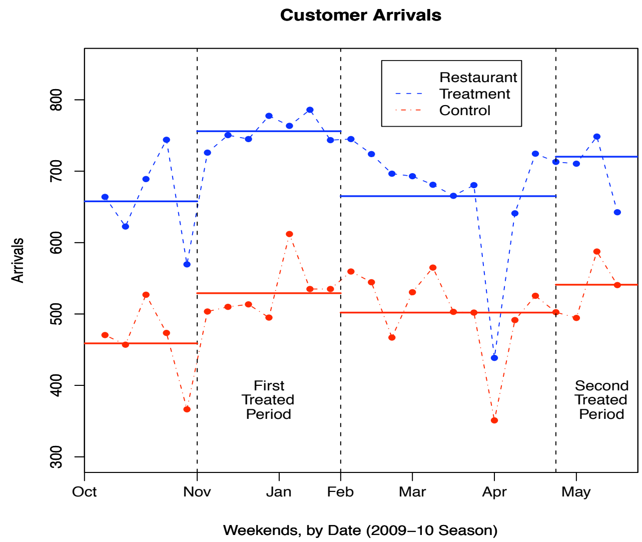

Figure 16: Demand Conditions, in terms of the number of customer arrivals, for the treated and control restaurants in the 2009-2010 season.

C Tables

	Dependent Variable											
	Number of Customers				Sales per Customer				$ln(\frac{\text{Sales per Customer}}{\text{Number of Customers}})$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(11)	(10)	(12)
Nov/Jan × 2009-2010	2.01**	2.99***	2.86***	2.49***	-1.51**	-1.54**	-1.78**	-1.73**	-.10***	-.13***	-.14***	-.13***
	(.89)	(.80)	(.77)	(.77)	(.73)	(.75)	(.83)	(.84)	(.04)	(.04)	(.04)	(.04)
2009-2010	.46	-.19	-2.35	-1.53	1.86***	1.89***	7.30***	7.20***	.03	.05	.28***	.25
	(1.12)	(.89)	(2.43)	(2.44)	(.66)	(.62)	(2.24)	(2.25)	(.05)	(.04)	(.10)	(.10)
R^2	.20	.38	.38	.43	.06	.07	.08	.08	.18	.28	.29	.33
Observations	938	938	938	938	937	937	937	937	937	937	937	937
Workers	40	40	40	40	40	40	40	40	40	40	40	40
Days	52	52	52	52	52	52	52	52	52	52	52	52
Section Characteristics	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Arrivals	N	N	N	Y	N	N	N	Y	N	N	N	Y
Days in Sample (Own and Peers')	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y

Table 4: **Substitution Effects.** Regressions use daily data at the individual level and the same restaurant last year as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions control for worker fixed effects, calendar week fixed effects, and the day of the week (Friday or Saturday). Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for the individual and average days in sample for the peer group.

	Dependent Variable											
	Entrees per Hour	Hours Worked	Bills	Bill Duration	Sales, by category				Sales per Customer, by category			
					Pre-Entree	Entree	Alcohol	Post-Entree	Pre-Entree	Entree	Alcohol	Post-Entree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Nov/Jan × 2009-2010	.59***	.27**	.08	-.07	.24	68.20***	-15.27**	-2.51	-.39**	-.05	-.96***	-.25***
	(.23)	(.13)	(.39)	(.05)	(4.72)	(18.04)	(7.08)	(2.91)	(.19)	(.30)	(.29)	(.09)
2009-2010	-.07	-.62	2.26**	.07	7.99	1.79	33.94	1.81	.83	1.90	2.10	.03
	(.75)	(.45)	(1.21)	(.12)	(11.82)	(68.55)	(19.48)	(7.32)	(.46)	(.72)	(.85)	(.25)
R^2	.12	.33	.23	.06	.18	.39	.13	.14	.05	.06	.12	.06
Observations	851	851	938	938	938	938	938	938	937	937	937	937
Workers	39	39	40	40	40	40	40	40	40	40	40	40
Days	52	52	52	52	52	52	52	52	52	52	52	52

Table 5: **Substitution Channels.** Regressions use daily data at the individual level and the same restaurant last year as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions control for worker fixed effects, calendar week fixed effects, the day of the week (Friday or Saturday), section characteristics, customer arrivals, and days in sample. Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for the individual and average days in sample for the peer group.

	Dependent Variable											
	Sales				ln(Tip Percentage)				Estimated Tip Earnings			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Nov/Jan × 2009-2010	48.63	91.59***	81.61**	66.15**	.001	.001	-.01	-.01	13.53	17.97*	13.73*	12.22
	(39.32)	(34.67)	(33.10)	(32.98)	(.04)	(.04)	(.04)	(.04)	(11.14)	(10.50)	(8.02)	(8.00)
2009-2010	69.00	41.47	45.47	79.44	.003	.003	.15	.15	5.02	2.25	36.09	38.85
	(46.21)	(34.67)	(112.17)	(112.14)	(.03)	(.03)	(.10)	(.10)	(6.74)	(6.20)	(25.80)	(25.40)
R^2	.16	.34	.34	.43	.03	.03	.03	.03	.04	.06	.06	.06
Observations	938	938	938	938	933	933	933	933	935	935	935	935
Workers	40	40	40	40	40	40	40	40	40	40	40	40
Days	52	52	52	52	52	52	52	52	52	52	52	52
Section Characteristics	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Arrivals	N	N	N	Y	N	N	N	Y	N	N	N	Y
Days in Sample (Own and Peers')	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y

Table 6: **Individual Productivity and Earnings.** Regressions use daily data at the individual level and the same restaurant last year as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions control for worker fixed effects, calendar week fixed effects, the day of the week (Friday or Saturday), section characteristics, customer arrivals, and days in sample. Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for the individual and average days in sample for the peer group.

	Dependent Variable	
	Customer Arrivals not Served	Share of Arrivals not Served
Nov/Jan × 2009-2010	-15.50*	-2.02*
	(8.27)	(1.06)
November/January	10.73	1.60
	(6.96)	(.99)
2009-2010	1.19	.15
	(5.78)	(.69)
Mean for Dependent Variable	46	5.73
R^2	.91	.87
Observations	52	52

Table 7: **Aggregate Effects.** Regressions use data aggregated at the daily level. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. Controls for the weather include mean temperature, total precipitation (in millimetres), and maximum windgust (in kilometres per hour). Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. The share of arrivals not served equals the number of arrivals not served divided by the total number of arrivals.

	Dependent Variable										
	Customers	Sales per Customer	ln(Ratio)	Sales per Customer, by category				Sales	ln(Tip Percentage)	Estimate Tip Earnings	Hours
				Pre-Entree	Entree	Alcohol	Post-Entree				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Nov/Jan × 2009-2010 ×											
Low Ability	2.28·	-2.62**	-.15*	-.47*	.19	-1.37***	-.04	33.83	-.08·	-1.76	.11
	(1.41)	(1.20)	(.08))	(.28)	(.41)	(.35)	(.13)	(41.03)	(.05)	(6.59)	(.40)
Average Ability	2.24**	-1.41	-.10*	-.44**	.13	-.90***	-.35***	62.19	-.002	7.48	.42*
	(.99)	(.95)	(.05)	(.20)	(.32)	(.29)	(.10)	(40.07)	(.04)	(7.76)	(.22)
High Ability	3.14***	-2.15**	-.17***	-.28	-.38	-.91*	-.14	90.60**	.02	12.05	.36
	(.67)	(.87)	(.04)	(.23)	(.43)	(.47)	(.16)	(35.72)	(.05)	(8.04)	(.24)
R^2	.43	.07	.33	.05	.06	.12	.39	.12	.05	.17	.20
Observations	932	931	931	931	931	931	931	932	926	927	852
Workers	38	38	38	38	38	38	38	38	38	37	37
Days	52	52	52	52	52	52	52	52	52	52	52

Table 8: **Heterogeneous Treatment Effects.** Regressions use daily data at the individual level and the same restaurant last year as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and · for estimates marginally significant at the 10 percent level. All regressions control for worker fixed effects, calendar week fixed effects, the day of the week (Friday or Saturday), section characteristics, customer arrivals, and days in sample. Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for the individual and average days in sample for the peer group.

	Bills Paid with Credit			Bills Paid with more than one payment method		
	Oct	Nov - Jan	Change	Oct	Nov - Jan	Change
Season						
2008-09	.61	.54	-.07	.18	.24	.06
	(.49)	(.50)		(.39)	(.43)	
Bills	1102	2200		1352	2901	
2009-10	.52	.55	.03	.23	.30	.07
	(.50)	(.50)		(.42)	(.46)	
Bills	1499	2468		1942	3536	
Difference-in-Differences			.10			.01
Restaurant						
Control	.63	.64	.01	.24	.34	.10
	(.48)	(.48)		(.43)	(.47)	
Bills	612	985		802	1496	
Treatment	.52	.55	.03	.23	.30	.07
	(.50)	(.50)		(.42)	(.46)	
Bills	1499	2468		1942	3536	
Difference-in-Differences			.02			-.03

Table 9: **Consumer Selection.** The outcome in the left panel is the proportion of bills paid with credit where a single payment is used. The outcome in the right panel is the proportion of bills where more than one payment method is used. The top panel summarizes this information when the same restaurant last year is used as a control group. The bottom panel summarizes this information when a comparable restaurant from the 2009-2010 season is used as a control group. The number of bills used for calculations in each quadrant are: 13861 for the top left panel, 18660 for the top right panel, 11465 for the bottom left panel, and 16112 for the bottom right panel.

	Dependent Variable									
	Customers	Sales per Customer	ln(Ratio)	Sales per Customer, by category				Sales	ln(Tip Percentage)	Estimate Tip Earnings
				Pre-Entree	Entree	Alcohol	Post-Entree			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Restaurant ×										
Piece Rate	2.25***	-.90	-.10**	-.15	-.06	.16	-.18*	81.70***	.02	8.33
(Nov 09/Jan 10 or May 10)	(.76)	(.74)	(.05)	(.21)	(.27)	(.30)	(.10)	(30.43)	(.05)	(5.86)
Performance Standard	-1.96*	.90	.15**	-.11	.20	-.15	.11	-71.39	-.06	-8.52
(May 10)	(1.18)	(.97)	(.06)	(.24)	(.41)	(.37)	(.16)	(56.38)	(.11)	(8.19)
R^2	.44	.06	.29	.09	.13	.10	.05	.43	.05	.21
Observations	1514	1514	1514	1514	1514	1514	1514	1514	1511	1503
Workers	64	64	64	64	64	64	64	64	64	64
Days	50	50	50	50	50	50	50	50	50	50

Table 10: **Incentive Effects of Piece Rates or Performance Standards.** Regressions use daily data at the individual level and a comparable restaurant as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions include controls for worker fixed effects and calendar date fixed effects. Controls for section characteristics include the number of booth seats, the number of bench seats, the number of chair seats, days in sample, peer days in sample. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for individual i and the average days in sample for the peer group in the same restaurant.

	Dependent Variable											
	Number of Entrees Sold				Sales per Entree				$\ln(\frac{\text{Sales per Entree}}{\text{Number of Entrees Sold}})$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
First Treatment	2.52***	2.71***	1.99***	2.25***	-1.16*	-1.15*	-.94	-.90	-.12**	-.12***	-.09*	-.10**
(Nov 2009/Jan 2010)	(.86)	(.65)	(.75)	(.76)	(.64)	(.64)	(.76)	(.74)	(.05)	(.04)	(.05)	(.05)
Second Treatment	-1.38	.02	.25	.28	-.21	-.19	-.24	-.01	.11**	.06	.05	.05
(May 2010)	(.99)	(.89)	(.90)	(.93)	(.90)	(.89)	(.88)	(.89)	(.05)	(.05)	(.04)	(.05)
R^2	.25	.42	.42	.44	.05	.05	.05	.06	.16	.27	.27	.29
Observations	1533	1533	1533	1514	1533	1533	1533	1514	1533	1533	1533	1514
Workers	64	64	64	64	64	64	64	64	64	64	64	64
Days	50	50	50	49	50	50	50	49	50	50	50	49
Section Characteristics	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Arrivals	N	N	N	Y	N	N	N	Y	N	N	N	Y
Days in Sample (Own and Peers')	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y

Table 11: **Robustness Checks for Substitution Effects.** Regressions use daily data at the individual level and a comparable restaurant as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions include controls for worker fixed effects and calendar date fixed effects. Controls for section characteristics include the number of booth seats, the number of bench seats, the number of chair seats, days in sample, peer days in sample. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for individual i and the average days in sample for the peer group in the same restaurant.

	Dependent Variable											
	Entrees per Hour	Hours Worked	Bills	Bill Duration	Sales, by category				Sales per Customer, by category			
					Pre-Entree	Entree	Alcohol	Post-Entree	Pre-Entree	Entree	Alcohol	Post-Entree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
First Treatment	1.20***	-.39**	.68***	-.07***	4.73	58.58***	14.73***	.20	-.15	-.06	.16	-.18**
(Nov 2009/Jan 2010)	(.24)	(.14)	(.30)	(.03)	(4.42)	(18.06)	(6.19)	(2.71)	(.21)	(.26)	(.30)	(.10)
Second Treatment	.19	-.16	-.11	.03	-5.65	15.34	-2.91	-1.65	-.26	.14	.01	-.07
(May 2010)	(.16)	(.11)	(.46)	(.05)	(5.21)	(28.19)	(7.60)	(3.54)	(.17)	(.36)	(.31)	(.12)
Mean (standard dev.) for	4.33	6.37	9.91	1.54	93.63	706.29	145.57	38.37	3.54	26.14	5.53	1.45
Dependent Variable	(1.78)	(1.19)	(3.17)	(.30)	(40.00)	(223.82)	(57.90)	(23.00)	(1.38)	(2.32)	(2.03)	(.84)
R^2	.19	.33	.24	.07	.20	.47	.19	.13	.09	.13	.10	.05
Observations	1412	1412	1514	1514	1514	1514	1514	1514	1514	1514	1514	1514
Workers	63	63	64	64	64	64	64	64	64	64	64	64
Days	49	49	49	49	49	49	49	49	49	49	49	49

Table 12: **Robustness Checks for Other Outcomes.** Regressions use daily data at the individual level and a comparable restaurant as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions include controls for worker fixed effects, calendar date fixed effects, section characteristics, customer arrivals, and days in sample. Controls for section characteristics include the number of booth seats, the number of bench seats, the number of chair seats, days in sample, peer days in sample. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for individual i and the average days in sample for the peer group in the same restaurant.

	Dependent Variable											
	Sales				ln(Tip Percentage)				Estimated Tip Earnings			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
First Treatment	87.24***	96.18***	69.31***	81.70***	.03	.03	.03	.02	9.68	10.60·	4.78	5.85
(Nov 2009/Jan 2010)	(37.63)	(30.33)	(29.47)	(30.43)	(.04)	(.04)	(.05)	(.05)	(7.38)	(6.51)	(6.59)	(6.93)
Second Treatment	-70.70	-7.48	1.92	10.31	-.02	-.03	-.03	-.04	-9.85	-2.93	-1.26	-.42
(May 2010)	(50,52)	(45.42)	(46.87)	(47.45)	(.08)	(.08)	(.08)	(.08)	(7.81)	(6.68)	(6.67)	(6.77)
R^2	.23	.41	.42	.43	.05	.05	.05	.05	.10	.19	.19	.20
Observations	1533	1533	1533	1514	1530	1530	1530	1511	1523	1523	1523	1504
Workers	64	64	64	64	63	63	64	64	64	64	64	64
Days	50	50	50	49	50	50	50	49	50	50	50	49
Section Characteristics	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Arrivals	N	N	N	Y	N	N	N	Y	N	N	N	Y
Days in Sample (Own and Peers')	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y

Table 13: **Robustness Checks for Revenue, Consumer Satisfaction, and Tip Earnings.** Regressions use daily data at the individual level and a comparable restaurant as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions include controls for worker fixed effects and calendar date fixed effects. Controls for section characteristics include the number of booth seats, the number of bench seats, the number of chair seats, days in sample, peer days in sample. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for individual i and the average days in sample for the peer group in the same restaurant.

	Dependent Variable					
	Indicator for good section based on				Average Section Quality	
	Mean	Mean	Median	Median		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-.05	-.02	-.07	-.03	-.01	.001
	(.20)	(.20)	(.20)	(.20)	(.06)	(.05)
Log Likelihood	-550.04	-594.62	-553.35	-596.85		
R^2					.04	.05
Observations	883	937	931	937	937	937
Workers	36	40	37	40	40	40
Worker Fixed Effects	Y	N	Y	N	Y	N
Worker Random Effects	N	Y	N	Y	N	Y

Table 14: **Selection into Sections.** Probit and OLS tests for section-based selection. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. Regressions use the entire sample and control for the treated period, the year, day of the week (Friday or Saturday), the number of customer arrivals, the number of customer arrivals squared, days in sample, peer days in sample, and a weekly trend. I measure average section quality by: first, computing the long run average number of entrees per hour (from the 2008-2009 data) for each table in the restaurant; second, averaging over the long run averages for the tables assigned to each worker. In columns 1 and 2 good sections have averages that exceed the mean number of entrees per hour (over all tables). In columns 3 and 4 good sections have averages that exceed the median number of entrees per hour (over all tables). The specifications differ in the number of observations because some workers always have 0's or always have 1's while in the sample.

	Window				
	1 week	2 week	3 weeks	4 weeks	5 weeks
Nov/Jan × 2009-2010	-25.20	35.20	52.32	81.71**	64.11*
	(169.66)	(56.00)	(42.07)	(33.30)	(33.47)
November/January	81.37	-16.71	-33.03	-69.56**	-48.97
	(128.44)	(48.24)	(39.06)	(30.64)	(29.83)
2009-2010	-298.71	-90.66	-64.21	-210.25**	-109.54
	(1695.22)	(317.29)	(179.53)	(89.95)	(81.74)
R^2	.49	.50	.50	.49	.48
Observations	131	274	418	561	625
Workers	35	36	39	39	39
Worker Fixed Effects	Y	Y	Y	Y	Y

Table 15: **Temporary Incentive Effects.** Regressions use daily data at the individual level and the same restaurant last year as a control group. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. All regressions control for the day of the week (Friday or Saturday) Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for the individual and average days in sample for the peer group.

	Dependent Variable = Number of Consumers Seated/Table Capacity											
	Control = Same Firm Last Year								Control = Another Firm			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
First Treatment	.09	.09	.10·	.10*	.07	.05	.06	.06	-.005	-.002	-.02	-.04
(Nov 2009/Jan 2010)	(.06)	(.06)	(.06)	(.06)	(.07)	(.06)	(.06)	(.06)	(.03)	(.03)	(.03)	(.03)
Second Treatment					-.02	-.03	-.03	-.03	-.0001	-.007	.0004	-.005
(May 2010)					(.05)	(.05)	(.05)	(.05)	(.05)	(.05)	(.05)	(.05)
R^2	.03	.07	.08	.08	.03	.08	.08	.08	.04	.07	.07	.08
Observations	938	938	938	938	1845	1845	1845	1845	1644	1644	1644	1625
Workers	40	40	40	40	42	42	42	42	64	64	64	64
Days	52	52	52	52	104	104	104	104	54	54	54	54
Controls												
Saturday	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
2009-2010	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
2009-2010×Feb-May	-	-	-	-	Y	Y	Y	Y	N	N	N	N
Section Characteristics	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Arrivals	N	N	N	Y	N	N	N	Y	N	N	N	Y
Days in Sample (Own and Peers')	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y

Table 16: **Consumer to Worker Matches.** Regressions use daily data at the individual level. Robust Standard Errors are in parentheses with *** for $p < .01$, ** for $.01 < p < .05$, * for $p < .1$, and $\cdot$ for estimates marginally significant at the 10 percent level. Columns (1)-(8) include controls for calendar week fixed effects. Columns (9)-(12) include controls for calendar date fixed effects. Controls for section characteristics include the number of booth seats, the number of bench seats, and the number of chair seats. Controls for Customer Arrivals include the total number of arrivals and the total number of arrivals squared. Controls for days in sample include days in sample for individual i and the average days in sample for the peer group in the same restaurant.