INSEAD

The Business School for the World®

# Working Paper Series

2015/88/TOM

# When You Work with a Super Man, Will You Also Fly? An Empirical Study of the Impact of Coworkers on Performance

Tom Fangyun Tan
Southern Methodist University, ttan@cox.smu.edu

Serguei Netessine
INSEAD, Serguei.netessine@insead.edu

We examine a large operational data set in a casual restaurant setting to study how coworkers' sales ability (measured as servers' sales premium) affects workers' performance in terms of service speed and service quality. We find that servers react non-linearly to their coworkers' ability. In particular, when coworkers' overall sales ability is low, increasing this ability may trigger servers to redouble both upselling and cross-selling efforts at the expense of slower service speed. When overall coworkers' ability is high (approximately one standard deviation $1.02 above the sample mean of $0.02), however, further increasing their ability may stimulate servers to accelerate their service at the expense of reduced sales efforts. Our empirical findings imply that managers should mix servers having heterogeneous ability levels during the same shift. Through a counterfactual analysis, we find that considering the inverted-U-shaped peer effects to optimize current servers' schedules without changing their capacity may increase total sales by 2.7%.

When You Work with a Super Man, Will You Also Fly? An Empirical Study of the Impact of Coworkers on Performance

Tom Fangyun Tan

*Cox Business School, Southern Methodist University, Dallas, Texas, U.S.A.*

*ttan@cox.smu.edu*

Serguei Netessine

*INSEAD, Singapore*

*serguei.netessine@insead.edu*

**Abstract**

We examine a large operational data set in a casual restaurant setting to study how coworkers' sales ability (measured as servers' sales premium) affects workers' performance in terms of service speed and service quality. We find that servers react non-linearly to their coworkers' ability. In particular, when coworkers' overall sales ability is low, increasing this ability may trigger servers to redouble both upselling and cross-selling efforts at the expense of slower service speed. When overall coworkers' ability is high (approximately one standard deviation $1.02 above the sample mean of $0.02), however, further increasing their ability may stimulate servers to accelerate their service at the expense of reduced sales efforts. Our empirical findings imply that managers should mix servers having heterogeneous ability levels during the same shift. Through a counterfactual analysis, we find that considering the inverted-U-shaped peer effects to optimize current servers' schedules without changing their capacity may increase total sales by 2.7%.

# 1 Introduction

Although service sectors play a more and more important role in the global economy, they generally suffer from low labor productivity. An OECD study shows that the labor productivity in wholesale and retail trade, hotels and restaurants across OECD countries is at a level typically about three fourths that of industry sectors, such as manufacturing (Freeman, 2008), thus creating opportunities for productivity improvement. One such opportunity lies in effective team management, not only because in service sectors employees often work in interdependent teams/groups[1] (Cohen and Bailey, 1997), but also because good teams often

[1]In this paper, we refer to a group of employees simultaneously performing similar tasks together as either a team or a group. Admittedly, the terms 'teams' and 'working groups' sometimes have different definitions in the management literature (e.g., Katzen-

create value more than the sum of their individual parts. Effective teams tend to promote knowledge sharing, repeated collaborations and creativity (e.g., Argote and Epple, 1990; Uzzi and Spiro, 2005; Guimera et al., 2005), and reinforce individual goals and accountability (Katzenbach and Smith, 1993).

For a long time operations management literature modeled service employees as having homogeneous (but perhaps random) capacity. Of course, a labor force is typically heterogeneous on the skill/ability dimension, and this consideration has been recently incorporated into labor decisions in settings such as skill-based routing, staffing and scheduling, as well as structural flexibility decisions (see Section 2 for references). Still, most of these studies use analytical models that assume heterogeneous workers work independently from each other, and little empirical research has been done to understand how to exploit the spillover/peer effects – that is, how team members with heterogeneous abilities affect each other. Our paper aims to empirically examine the spillover and its implications for operational performance through better scheduling, while demonstrating the value of the data analytic approach to labor management. One specific decision we analyze is whether or not to polarize team ability during the same shift and how to schedule the best or worst workers.

Following prior research, we define the spillover between coworkers' ability and the performance of others as peer effects (Chan et al., 2014a). Research has proposed various mechanisms, including free riding, competition, monitoring and social pressure. Empirical studies on peer effects in the workplace have typically found linear peer effects, either positive or negative. In this paper, we propose to examine a potential non-linear effect of peers' ability and performance, which may reconcile the seemingly conflicting linear effects found in previous research. We analyze a large operational data set to study how coworkers' ability affects workers' performance in terms of both service speed and quality. We study the setting of a full-service casual restaurant chain because 1) servers exhibit wide ability heterogeneity; 2) servers have significant influence on both service time and sales; and 3) multiple servers are scheduled to work during the same shift. Specifically, we collect detailed transaction-level data from the restaurant chain's point-of-sales

bach and Smith, 1993), wherein a group is characterized by individual accountability and a team is characterized by both individual and mutual accountabilities. We do not make such a distinction and use the terms interchangeably.

system, which contains approximately 226,350 check-level observations for three restaurants from January 2011 to June 2012. Using an instrumental variable approach, we demonstrate the inverted-U relationship between performance (sales) of the focal employee and coworkers' sales ability and further show how labor mix can be leveraged to optimize the employee schedule. Particularly, we show that mixing servers having various ability levels is associated with improved financial performance. We also propose a heuristics to more efficiently schedule the servers without adding the capacity of additional high-ability workers. Lastly, we conduct a counterfactual analysis of the impact on sales when managers consider the inverted-U peer effects on scheduling: a 2.7% sales increase at no additional cost.

## 2 Related Literature

Our research contributes mainly to two streams of literature, optimal scheduling/rostering decisions and peer effect studies.

How to schedule workers to meet the stochastic demand is a classic problem in services, and we refer our readers to Gans et al. (2003), Akşin et al. (2007) and Van den Bergh et al. (2013) for their excellent literature reviews. Although classical models here tended to be constrained by assumptions that workers are all the same and independent from each other for modeling tractability reasons, recent studies in operations management have developed new approaches to managing a group of heterogeneous workers in various areas, including skill-based routing (e.g., Wallace and Whitt, 2005; Ata and Van Mieghem, 2009; Mandelbaum et al., 2012; Mehrotra et al., 2012; Ward and Armony, 2013); hiring and retention (e.g., Arlotto et al., 2013); structural flexibility (e.g., Hopp et al., 2004; Iravani et al., 2007; Narayanan et al., 2009; Kesavan et al., 2014) and incentive design (e..g, Siemsen et al., 2007; Roels and Su, 2013). Similarly, researchers have also started to incorporate workers' heterogeneity in scheduling/rostering decisions. Scheduling is different from staffing, which determines the number of workers for a shift. For example, Tan and Netessine (2014b) find an inverted-U-shaped relationship between workload and performance. Accordingly, if the current workload is less than the optimum, they suggest that reducing the number of workers per hour may not only reduce labor costs but also achieve a sales lift. Unlike that study, this paper focuses on scheduling, which determines the team composition during a particular shift. Cezik and L'Ecuyer (2008) and Bhulai et al. (2008) devise efficient techniques for scheduling call center agents having different skills and labor costs to handle calls requiring varying skills in order to minimize costs. Bard and Wan (2008) consider workers having non-symmetric movement restrictions between work sites to find the best mix of employees to satisfy demand at minimum cost.

Furthermore, although considerable research has been devoted to making optimal scheduling decisions analytically, less attention has been paid to explicitly evaluating the impact of scheduling on workers' performance. As an exception, in studying retail labor mix decisions, Kesavan et al. (2014) find that increasing temporary labor mix from zero to its optimal value increases sales by 6.78%, and that increasing part-time labor mix from zero to its optimal value increases sales by 15.04%. However, the focus of their paper is on labor force flexibility, while our paper is about peer effects within a team. In addition, Akşin et al. (2015) find that the new recruits at the London Ambulance Service who have worked with more different partners in the past tend to perform more efficiently during patient pick-up and handover processes than those who have worked with the same partners, because they have benefited from group learning. Thus, their paper primarily studies the diversity of partnership experience, which is again different from the skill heterogeneity and peer effects examined in this paper.

The examination of peer effects has recently attracted attention in the labor economics literature. In a lab setting, Falk and Ichino (2006) discover that peer effects improve workers' envelope-stuffing productivity as a consequence of peer pressure. In an academic performance setting where students are encouraged to learn from each other, Carrell et al. (2009) find that the ability of a cohort at the U.S. Military Academy, measured in terms of average SAT verbal score, should have a positive effect on the academic performance of every member of the cohort. In practice, however, after implementing an intervention based on theoretical prescription, Carrell et al. (2013) observe that peer effects turn out to have a negative impact on low-performance students because these students tend to gather without interacting with high-performance students. More relevant to our setting is the research on peer effects in the workplace. Mas and Moretti (2009) study a supermarket register checkout setting, where workers are paid a fixed hourly rate. They find evidence of positive productivity spillovers from highly productive workers because of social pressure. Similarly, Schultz et al. (2010) show that workers on a production line adjust their speed toward the average speed of their coworkers since their work stations are interdependent. In a setting without externalities, where workers are paid a piece rate, Bandiera et al. (2010) still find that a fruit picker's productivity increases when he/she works with more capable friends since workers having social preferences desire to socialize with their friends. Chan et al. (2014a) analyze the sales performance of the salespeople at cosmetics counters and argue that the incentive scheme determines the direction of peer effects. According to their findings, while team-based commissions produce positive peer effects because workers may help each other, individual-based commissions create negative peer effects because strong salespeople may gain customers from lower ability coworkers. Even in a knowledge-based workplace, Staats et al. (2015) find that cardiologists are more likely to choose the same treatment procedures as more experienced colleagues do because the experienced surgeons may exert group pressure.

Our study contributes to this stream of literature in three ways. First, most peer effect studies tend to focus on linear peer effects. Our study, however, focuses on non-linear peer effects, which may reconcile previous conflicting findings. Second, although considerable research has been devoted to assessing peer effects, much less attention has been paid to examining the value of incorporating peer effects into labor decisions. For example, in Mas and Moretti (2009), managers are not responsible for assigning individual workers to particular shifts. In contrast, our setting allows us to explicitly stress the value of accounting for peer effects to make optimal scheduling decisions (Campbell and Frei, 2011). Third, previous work has analyzed peer effects on either service speed or service quality; our study shows how peer effects affect workers in terms of both service time and quality. In particular, we analyze how workers make speed/quality trade-off decisions (Hopp et al., 2007; Debo et al., 2008; Anand et al., 2011; Kostami and Rajagopalan, 2013; Alizamir et al., 2013; Zhan and Ward, 2013; Tan and Netessine, 2014b; Batt and Terwiesch, 2015).

## 3 Hypotheses Development

With one in three Americans having worked in the restaurant industry at some point of their life (Mill, 2006), restaurant servers come from different backgrounds, and have a wide heterogeneity in ability. In this paper, we focus on servers' sales skill levels because 1) sales have a direct impact on both restaurants' and servers' income; 2) any increase in sales is particularly significant in the casual dining industry, where the profit margin is only 3% to 9%; 3) a well-executed sales job will substantially enhance customers' dining experience. A high-ability server tends to have a pleasant personality and favorable attitude. He/she has thorough knowledge of both the food and the wine, thus having table-side confidence to successfully conduct suggestive sales. A high-ability server is also able to read diners and anticipate their needs. For example, he/she rarely leaves diners' glasses empty, maximizing the sales opportunities of beverages and wines. By contrast, a low-ability server may just mention the cheap $9.99 special once diners are seated, failing to sell more expensive items.

These servers having heterogeneous sales skills usually work with a group of other servers during a shift. Although servers are sometimes considered as "independent business people" (Walker, 2007) because the majority of their income comes as tips from the tables that they serve, they are also trained to collectively contribute to the whole restaurant. In other words, team interaction can influence an individual's service performance. Effective server teamwork is important for restaurant operations since it improves customer satisfaction, encouraging repeat visits and therefore increasing long-term financial performance; it promotes the sense of achievement, equity and camaraderie, keeping servers motivated and satisfied and thus reducing turnover, which is usually costly; it facilitates learning in the workplace, strengthening collaboration. Because of the importance of teamwork, servers are often referred to as "team members" in the restaurants.

Desirable as high ability is, it does not always precisely translate into higher performance, which also requires servers' contextual efforts. Similar to many other multitasking agents (e.g., D. and R., 2007), servers in a casual restaurant expend two important types of effort to serve seated diners, sales and speed. Sales efforts consist of both upselling more expensive items and cross-selling additional items, which both lead to higher sales per check. In addition, speed efforts may include carrying multiple items from the kitchen to save trips and time (Tan and Netessine, 2014b). Although such efforts are not directly observable, they may be inferred from observable performance metrics (i.e., the sales and the meal duration of each meal), for which we develop hypotheses about the peer effects in this subsection. We categorize the theories into three main types of effects: *positive spillover* effects, *anti-productive emotion* effects and *external capacity* effects.

**Positive Spillover Effects** Positive spillovers are defined as the phenomenon that high-ability workers improve the performance of their coworkers (Mas and Moretti, 2009). Behind this phenomenon there are at least three theories from economics and social psychology. First, due to social pressure, a worker may experience disutility if he/she works less hard than the high-ability coworkers for fear of sanctions or shame by those coworkers (Mas and Moretti, 2009). To reduce this disutility, that worker may expend more effort in both service quality and speed to catch up with those higher-ability coworkers. In addition, social pressure may help mitigate the free-riding problem, especially when employees work as a team (Kandel and Lazear, 1992). Second, knowledge spillover implies that information about how to do a job well may transfer from one worker to the next (Argote and Ingram, 2000; Moretti, 2004a,b; Chan et al., 2014b). This advantageous knowledge is usually possessed by the high-ability workers, who may choose to share it with lower-ability coworkers for prosocial reasons (Itoh, 1991, 1993; Siemsen et al., 2007), or it can be learned by low-ability workers through observation (Song et al., 2015). Third, social comparison suggests that people may compare themselves to others for self-evaluation (Festinger, 1954) and show conformism to their peers (Fehr and Schmidt, 1999; Charness and Rabin, 2002). If people work with high-ability coworkers, they may feel inferior and exhibit behind-averse behavior by working harder to minimize the disparity (Roels and Su, 2013). Social comparison may also trigger competition or contagious enthusiasm for the lagging workers to work harder (Lazaer and Rosen, 1981; Bandiera et al., 2010). Positive spillovers could happen to restaurant servers who work as a group during the same shift. Although they may be considered independent workers because they earn their tips only from the tables that they serve individually, waiters are trained to collectively contribute to the whole restaurant (Walker, 2007). If a server particularly lags, other servers may report him/her to management or ostracize him/her socially, creating social pressure for that server to expend more effort. In addition, servers may learn from the higher-performing coworkers either by watching or by exchanging ideas during the service meetings before the shift. Furthermore, servers compare their tips at the end of the shift, which should motivate behind-averse servers to improve their performance.

**Anti-productive Emotion Effects** Overly capable coworkers may provoke anti-productive emotions among other workers. When coworkers are excessively capable, they may pose a threat that hinders workers from reaching their goals, which may further reduce their motivation and commitment (e.g., O'Connor et al., 1984; Barankay, 2012). Furthermore, comparison with highly capable coworkers may create negative feelings about oneself (Buunk et al., 1990), which may cause lower motivation and reduce effort. Instead of working hard, these demoralized workers may try to sabotage the high-ability coworkers (Lazear, 1989; Chen, 2003), negatively impacting their performance. When restaurant servers work with other highly capable coworkers, they may similarly feel the aforementioned anti-productive emotions. They may become demoralized because the highly capable coworkers will hinder them from reaching their desired goals, which may include getting assigned to favorable table sections or winning the implicit sales contest during the shift. Workers may also feel disappointed about themselves when benchmarking against their extremely capable coworkers. Consequently, they may give up devoting more effort to service.

**External Capacity Effects** In a service network, the throughput and quality of each stage is dependent on the load of other stages, particularly the "bottleneck" stages that have the lowest capacity (e.g., Goldratt et al., 2004; Ata and Van Mieghem, 2009). For the throughput, an overloaded bottleneck service stage will create extra idle time for the downstream stages, thus prolonging the total service time (e.g., Suresh and Whitt, 1990). In terms of quality, a congested service stage may cause more errors and require rework because each of the service units may receive less attention from the service provider (e.g., KC, 2013). Restaurants operate in such a service network because servers' performance is dependent on the capacity of other stages, such as the kitchen or bar. High-ability servers tend to sell more items, such as cocktails or desserts. These additional items create more work for the kitchen and the bar, which may further extend the duration of a meal because a late order simply needs to wait for earlier orders to be processed. Furthermore, the extra items sold by the high-ability servers may create confusion and errors, which may further prolong the meal duration and discourage diners from ordering more items.

**Prediction Regarding Overall Impact on Sales** When coworkers' overall sales ability is low, positive spillovers may dominate the anti-productive emotions and external capacity effects because the low-ability coworkers may not yet sell enough additional items to overload bottlenecks such as the kitchen or bar. That is to say: positive spillovers will trigger servers to expend more sales effort and improve sales performance. In addition, servers may devote more effort to sales instead of speed because a sales-maximizing (i.e., tips-maximizing) server should exploit the sales opportunities when common resources, such as kitchen and bar, are not yet overly utilized by his/her low ability coworkers. However, when the overall coworkers' sales ability is high, a further increase in their ability may create anti-productive emotions and overload other restaurant functions, thus impairing service quality (sales). For example, servers may have to repeatedly apologize to diners about the delays due to the congested kitchen or bar, which may lower diners' appetite to consume more items because of the fear of a long wait, leading to lost sales opportunities. Furthermore, when the coworkers are highly capable, the positive spillover may be diminished because it is constrained by servers' physical and cognitive capacity. Therefore, we propose:

HYPOTHESIS 1 (H1): *As coworkers' sales ability increases, the sales performance of the focal employee will first increase and then decrease: that is, there is an inverted-U-shaped relationship between coworkers' ability and sales.*

**Prediction Regarding Overall Impact on Meal Duration** When coworkers' overall sales ability is low, in response to the positive spillover effects from higher-ability coworkers , servers may spend more time and effort selling more items, which accordingly take additional time to consume and prolong meal duration.

Moreover, although the positive spillover may stimulate servers to increase their effort level, this additional effort is likely to be largely sales related instead of speed related because of the aforementioned unique sales opportunities of working mostly with low-ability coworkers and less utilized common resources. Besides, the external capacity effects may further contribute to lengthening meal duration. When coworkers' overall sales ability is high, however, working with ever more capable coworkers may induce anti-productive emotions and may trigger servers to rush diners by selling fewer items, which should decrease meal duration. In addition, when coworkers are highly capable of generating sales from their tables and they consume a lot of common resources (e.g., having bartenders make special drinks) to achieve their high sales, other servers may have a strong incentive to speed up. Turning a table more quickly allows servers to be more frequently available to be assigned a new party because a host considers which servers currently have the most empty tables when seating incoming diners (Walker, 2007). Admittedly, the external capacity effect alone would suggest that the meal duration may keep lengthening as coworkers become capable of more and more sales; nevertheless, this effect may be reduced by the increased service rate because of servers' promptness and rushing of diners. Hence, we propose:

HYPOTHESIS 2 (H2): *As coworkers' sales ability increases, meal duration for the focal employee will first increase and then decrease: that is, there is an inverted-U-shaped relationship between coworkers' ability and meal duration.*

**Prediction Regarding the Effect of Coworker Proximity on Peer Effects** Positive spillover and anti-productive emotion effects require that the focal worker can observe his/her coworkers and be observed by them. If the worker can neither observe nor be observed by other coworkers, he/she will hardly feel the social pressure from the high-ability coworkers because the fear of sanctions or shame do not reach him/her. Similarly, both social comparison and anti-productive emotions are not likely to materialize if the worker simply does not know the ability level of his/her coworkers. Research further suggests that proximity facilitates communication, coordination, mutual support, effort and cohesion in the team (Hoegl and Proserpio, 2004). Examining U.S.-based pharmaceutical plants over a 13-year period, Gray et al. (2015)

find that physical proximity through geographical colocation between manufacturing and R&D activities improves conformance quality. In addition, in a study of supermarket cashiers, Mas and Moretti (2009) find that only those cashiers who can directly observe fast coworkers in front of them may experience an improvement in their productivity because of social pressure.

On the other hand, knowledge spillover and external capacity effects do not necessarily rely on the observability condition because high-ability workers can still share their knowledge at work and external capacity effects do not change workers' intrinsic capacity or capability. In the restaurant setting, servers are typically assigned to different table sections, some of which are closer to or farther away from each other. Although servers who are in close proximity to each other are more likely to observe each other, they are still aware of the entire team of the shift because they meet them during the pre-shift meetings and near the kitchen and the common area. Hence, we do not think that the coworkers whose table sections are far away exert no peer effects at all on the focal worker. Instead, we hypothesize that distance weakens the peer effects:

HYPOTHESIS 3 (H3): *Coworkers whose work sections are in proximity to the focal employee have stronger peer effects than those farther way.*

**Prediction Regarding the Effect of Team Heterogeneity on Storewide Sales** Our last prediction concerns the "best" team composition. For the sake of argument, suppose each store has four high-ability servers and four low-ability servers to allocate to two shifts, with each server working only one shift. Similar to Chan et al. (2014a), we consider two team composition schemes. In the heterogeneous composition scheme, two high-ability and two low-ability servers work during the same shift, while under the homogeneous scheme, either four high-ability or four low-ability servers staff one shift. Hypothesis 1, which postulates an inverted-U-shaped relationship between coworkers' overall sales ability and the sales of the focal worker, implies that worker heterogeneity should increase individual workers' sales because forming homogeneous groups of all high-ability or low-ability workers instead of mixing them will cause the coworkers' overall sales ability to be either too high or too low, corresponding to the two lower ends of the inverted-U-shaped curve. For the same reason, H2 implicitly suggests that worker heterogeneity may induce the longest meal duration, which may increase the opportunity cost of the store capacity, especially during peak hours. Some previous studies report related findings. For example, Chan et al. (2014a) find that in a cosmetics store both high-ability and low-ability salespeople generate higher total sales under heterogeneous staffing while using team-based compensation, but they also suggest that heterogeneity will hurt firms with individual-based compensation. Admittedly, in our setting, servers are primarily rewarded for their own performance because they earn tips from the tables that they are assigned to. Nevertheless, they are also trained to help each other as a team because of the aforementioned importance of teamwork in a restaurant. Furthermore, even adjusting for the average team ability, Hamilton et al. (2003) find that more heterogeneous teams in terms of ability were more productive in manufacturing garments because of mutual learning and intrateam bargaining. Similarly, Shafer et al. (2001) show that worker heterogeneity in terms of learning rates produces higher output among a group of workers operating independently of one another. For these reasons, we posit:

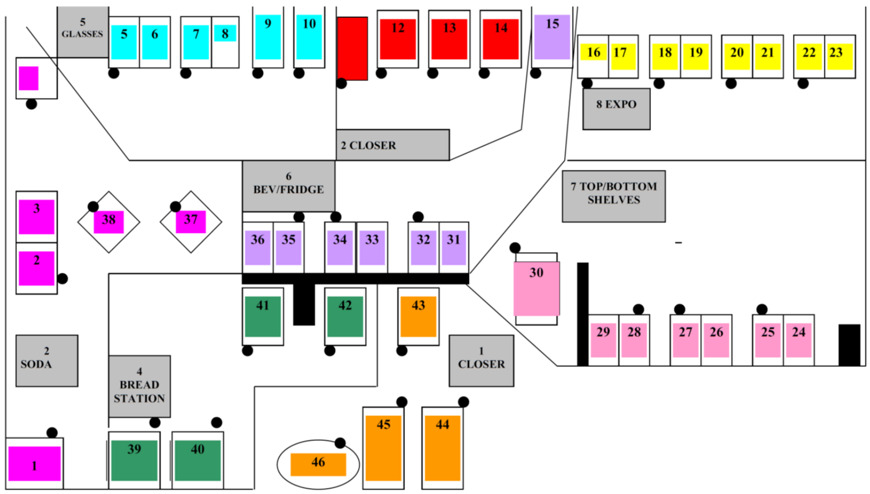
HYPOTHESIS 4 (H4): *Worker heterogeneity in a team increases storewide sales performance.*

# 4 Empirical Setting, Variables and Descriptive Statistics

## 4.1 Empirical Setting

Similar to other family-style restaurants such as TGI Friday's and Applebee's, our empirical setting is a restaurant chain that offers casual American-style dining with table service, and is set in the Boston suburbs. In this setting, servers are responsible for waiting on assigned tables, from which they earn gratuities in addition to a restaurant minimum hourly wage. Servers are typically scheduled to work on shifts of varying lengths (e.g., six hours). The hours of the restaurants are from 11:30 am to 10:00 pm from Monday to Thursday, and from 11:30 am to 11:00 pm from Friday to Sunday. From three restaurants of this chain, we collected 18 months of the point-of-sales (POS) data from January 2011 to June 2012, including detailed information about servers, sales, party size, and service start and end time for each check. The floor plan of one restaurant is shown in Figure 1, where tables are numbered and sections are color coded. As indicated by the black dots, each server is typically assigned up to four parties in a section. We were able to obtain the floor plan of only this particular restaurant because we worked with this restaurant to implement a new scheduling system (more details on this system are provided in Subsection 5.3).

Figure 1: Typical Floor Plan




Since our empirical analysis focuses on the main dining room, where peer observation and interaction are most likely to happen, we exclude the bar and take-out services. Further, we drop the transactions which include the day's top and bottom 7.5% of checks to reduce the influence of outliers (e.g., very large parties and private events). Our final data set is comprised of approximately 226,350 check-level observations. This setting has a number of advantages for examining peer effects on individual work performance. First, restaurant servers have heterogeneous abilities, creating a unique challenge and opportunity. Second, we know who is working at any moment in time, so that we can identify the working group and the peers to further assess the implications for labor decisions.

## 4.2 Variables and Summary Statistics

Our main analysis is conducted at the individual check level (similar to Tan and Netessine, 2014b) because this granularity of analysis contains more information about servers' behavior (sales and speed) in handling each dining party. Hence, we provide check-level variable definitions in this subsection. We use sales and meal duration to measure servers' performance because they are evaluated by the management and capture the dining aspects over which servers can exert leverage.

*Sales$_i$*. Our first dependent variable measures the sales (in dollars) of check *i,* which is exclusively assigned to one server in our focal restaurant. We consider it to reflect service quality, which is concerned less with service accessibility as often assumed in the service operations literature than with a standard for service contents.

*MealDuration$_i$*. A measure of service time of check *i,* which is inferred from the difference between check opening and closing times recorded in our POS data. This inferred duration is a common meal duration measure in restaurant operations (e.g., Kimes, 2004) because 1) no technology is available to systematically track the whole process from diners' entering the restaurant to their leaving; and 2) this meal duration measure directly captures the server's interaction with the diners and omits the host's interaction with them before the server opens the check.

We transform *Sales* and *MealDuration* into their natural logarithms to linearize the exponential forms of sales and meal duration models (Kleinbaum et al., 2007). Log transformation also increases the normality of the errors, which ensures that our hypothesis test statistics follow *t*-distribution. Similar to previous literature (e.g., Mas and Moretti, 2009; Chan et al., 2014a), we construct the key independent variable *CoworkerAbility$_i$* in two steps. First, we employ a fixed-effect model specified in Subsection 5.1 to estimate the sales premium or intrinsic sales skill of each server, $\theta_j$ ($\theta_j$ can be either positive or negative). We elect to evaluate a server by his/her sales ability because sales performance is typically the most critical performance metric tracked by restaurant management. Second, we take the average of the intrinsic sales abilities of the coworkers working during the same hour as worker *j*, who opens the focal check *i* being analyzed, to form a peer effect variable, *CoworkerAbility$_i$*. In other words, $CoworkerAbility_i = \bar{\theta}_{-j} = 1/n \sum_{k \neq j} \theta_k$. For example, suppose check $i$ is handled by server $j$. When the check is opened, server $j$ has four coworkers, whose intrinsic sales abilities are \$1, \$2, -\$2, and \$4, respectively. Our peer effect measure *CoworkerAbility$_i$* is (\$1

+ \$2 - \$2 + \$4)/4 = \$1.25. Alternatively, servers can also be classified by their speed ability, but sales ability is arguably more important than speed ability in our setting, where the restaurants under study rarely reach their capacity constraints. Nevertheless, we check the peer effects in terms of intrinsic speed ability and find that they do not significantly affect the sales performance of the focal server. Hence, we focus on servers' sales ability in this paper.

Servers' sales abilities may fluctuate over time for reasons such as learning (e.g., Argote and Epple, 1990; Lapré et al., 2000); forgetting (e.g., Shafer et al., 2001); and task variation (e.g, Wiersma, 2007; Staats and Gino, 2012). Since the focus of our study is peer effects instead of learning, forgetting or task variation effects, we do not separately identify those factors of intrinsic abilities. Our approach is to control for all these effects through estimating servers' intrinsic sales ability every month ( $\theta_{jm}$), which reflects both a relatively stable underlying ability within a month and a variable ability over a longer time period.

Note that we measure the effect of coworkers' intrinsic sales ability on the focal worker's contemporaneous performance. An alternative and equally interesting independent variable would be coworkers' contemporaneous performance[2] (i.e., performance during the same shift rather than some average performance). In this study, we use the dependence of the peer effect on servers' intrinsic sales ability because 1) servers talk to each other and often compare their tips at the end of the shift, so they may have no knowledge about who are high-performing servers currently but they may know who is high performing on average; 2) waiters may be so busy waiting their own tables that they have little time to observe coworkers' contemporaneous performance; 3) workers' intrinsic ability should generally correlate with contemporaneous performance; 4) performance shocks that affect all servers at a particular moment can cause a spurious relationship between focal server and coworker performance. Nevertheless, both intrinsic ability and contemporaneous performance may simultaneously affect servers' performance, and our independent variable *CoworkerAbility* should possibly be interpreted as a combination of a true effect of intrinsic sales ability and a true effect of contemporaneous effort. Most important, only intrinsic sales ability can be used in proactively scheduling

[2]Similar to Mas and Moretti (2009), we use 'effort' and 'performance' interchangeably in this paper.

servers because it can be calculated in advance.

Averaging coworkers' sales ability reflects the absolute effect of the peers' sales abilities, which is consistent with prior work by Mas and Moretti (2009) and Carrell et al. (2009). These absolute peer effects are comparable across the three restaurants because the three stores belong to the same chain offering standardized menus. We believe that the absolute measure is more appropriate than the relative measure (e.g., Chan et al., 2014a[3]) for the following reasons: First, servers may not have an accurate evaluation of their own sales ability for such reasons as superiority bias (Hoorens, 1993), causing the relative peer effect measure to be inaccurate. Second, the relative measure cannot distinguish the group having high-ability coworkers (e.g., $\hat{\theta}_{-j} = \$5$) and a high-ability focal server (e.g., $\theta_j = \$5$) from another group having low-ability coworkers (e.g., $\hat{\theta}_{-k} = -\$5$) and a low-ability focal server (e.g., $\theta_k = -\$5$) because the relative measure may yield the same quantitative results ($\hat{\theta}_{-j} - \theta_j = \hat{\theta}_{-k} - \theta_k = 0$). Third, the absolute measure is arguably more intuitive for managers to implement in the scheduling process than the relative measure because the absolute measure only needs to consider the average team ability instead of additionally considering the focal server's. We do use the relative measure in the robustness check section and the results are largely similar.

We control for several variables that, according to previous literature, could affect servers' performance. Variable $OwnAbility_i$ is $\theta_j$ estimated from Model 1, that is, the intrinsic sales ability of server $j$ responsible for check $i$. Variable $PartySize_i$ controls for the number of diners in a particular party $i$, which should affect sales and meal duration. Variable $AbilityStDev_i$ is the standard deviation of the sales abilities of all the servers working when check $i$ is opened. We use this variable to adjust for ability dispersion/heterogeneity, which is known to affect worker performance in Chan et al. (2014a). We do not use the coefficient of variation to measure the dispersion because our intrinsic sales ability has both negative and positive values. In addition, we control for a one-hour lagged effect of $CoworkerAbility_i$ calling it $LagCoworkerAbility_i$, because the peer effect may propagate over time (Mas and Moretti, 2009; Carrell et al., 2009). Furthermore, following Tan and Netessine (2014b), who find a non-linear relationship between servers' workload and

[3]The relative measure is probably more appropriate in the setting of Chan et al. (2014a) because they compare cross-counter peer effects.

their performance, we control for the individual workload $AvgTables_i$ and its quadratic form. Similar to Tan and Netessine (2014b), variable $AvgTables_i$ is the average number of tables (parties) that a server handles simultaneously with the focal check $i$ being analyzed. For instance, suppose check $i$ lasts 50 minutes, when it shares its server with another table (party) for 10 minutes. The workload measure $AvgTables_i$ is (50 min + 10 min)/(50 min) = 1.2 tables. Calculated in the same way we calculated the individual workload, variable $StoreTables_i$ is the average number of tables occupied during check $i$, which is used to control for the demand in the kitchen and the bar (a similar variable is used in Tan and Netessine (2014b) as a control variable for meal duration). Finally, we include additional fixed effects of the time/date/location of check $i$ to control for temporal and spatial factors, such as demand. In particular, we include a categorical variable $Hour_i$, time when check $i$ was opened, to control for systematic intra-day difference in demand. We include another categorical control, $DayWeek_i$, indicating the day of the week because weekends are usually busier than weekdays. In addition, in order to adjust for seasonality and economic trends, we use a categorical control variable, $YearWeek_i$, which starts at one from the first week of January 2011 and ends at 79 in the last week of June 2012. Finally, we include a categorical variable $Store_i$ for each store $i$ to control for time-invariant aspects of store fixed effects (e.g., location, traffic).

### 4.3 Descriptive Statistics

Table 1 shows the descriptive statistics of the check-level variables. Each check has an average sales total of \$45 and an average meal duration of approximately 51 minutes. There are on average 2.51 diners in each check, which translates to \$45/2.51 $\approx$ \$18 per diner. Furthermore, there is a considerable heterogeneity in coworkers' sales ability and focal servers' intrinsic sales ability. For example, the coworkers' sales ability ranges from -\$4.64 to \$8.09, with the focal servers' intrinsic sales ability ranging from -\$13.59 to \$15.8. Each server on average handles 2.32 tables simultaneously, and the entire store has on average 16.36 tables occupied.

Figure 2a shows the histogram of the sales ability distribution, illustrating a wide variation in servers'

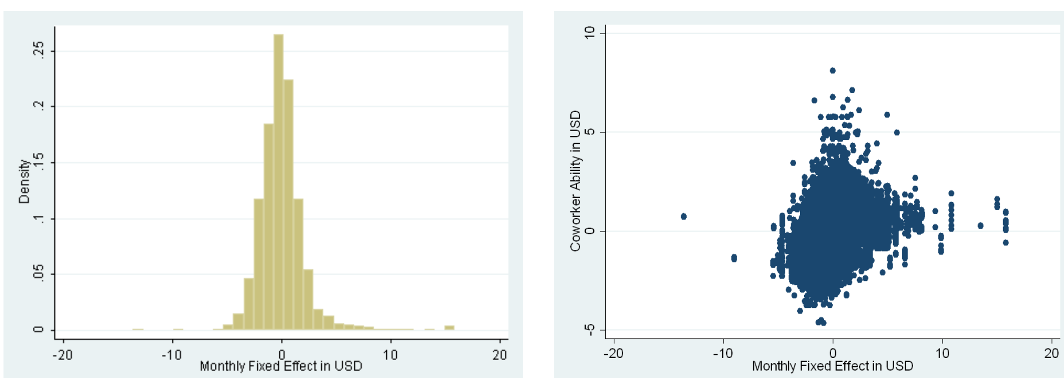
Table 1: Summary Statistics of Check-Level Variables

| | *Sales* | *MealDuration* | *Coworker Ability* | *OwnAbility* | *PartySize* | *Ability StDev* | *LagCoworker Ability* | *AvgTables* | *StoreTables* |
|---|---|---|---|---|---|---|---|---|---|
| N | 220,923 | 220,923 | 220,923 | 220,923 | 220,923 | 220,923 | 206,257 | 220,923 | 220,923 |
| Mean | 45.01 | 50.80 | 0.02 | 0.03 | 2.51 | 1.13 | 0.03 | 2.32 | 16.36 |
| Stdev | 22.87 | 19.21 | 1.02 | 1.50 | 1.04 | 0.51 | 1.03 | 0.80 | 6.48 |
| Min | 5.11 | 20 | -4.64 | -13.59 | 1.00 | 0.00* | -3.51 | 1 | 1 |
| P5 | 18.27 | 28 | -1.73 | -2.32 | 1.00 | 0.47 | -1.74 | 1 | 5.04 |
| P25 | 28.96 | 37 | -0.93 | -0.95 | 2.00 | 0.81 | -0.95 | 1.75 | 11.55 |
| P50 | 39.67 | 47 | 0.31 | 0.01 | 2.00 | 1.06 | 0.32 | 2.27 | 16.94 |
| P75 | 55.16 | 59 | 0.76 | 0.94 | 3.00 | 1.38 | 0.76 | 2.79 | 21.34 |
| P95 | 91.22 | 89 | 1.31 | 2.42 | 5.00 | 2.01 | 1.31 | 3.72 | 26 |
| Max | 149.97 | 139 | 8.09 | 15.80 | 8.00 | 6.02 | 6.59 | 10.58 | 37.32 |

*The exact value is 0.00039.

intrinsic sales abilities. Figure 2b further displays the scatter plot of servers' own intrinsic sales ability (*OwnAbility*) and their coworkers' average sales ability (*CoworkerAbility*). As can be seen, there seems to be a moderately positive relationship between *OwnAbility* and *CoworkerAbility*.

Figure 2




(a) Histogram of Intrinsic Sales Ability (b) Scatter Plot of Own and Coworker Sales Ability

Table 2 shows the correlations of the check-level variables. As expected, log(*Sales*) is positively associated with log(*MealDuration*) (correlation = 0.391) and *PartySize* (correlation = 0.599). Similarly, log(*MealDuration*) is positively correlated with *PartySize* (correlation = 0.1811). It is noteworthy that the correlations between the two dependent variables and *CoworkerAbility* are low (-0.0046 and 0.0225, respectively). These low correlations do not necessarily indicate that peer effect has no relationship with the focal server's performance because their relationship may be non-linear. In addition, the correlations between other predictors are generally quite low, except for the correlation between *LagCoworkerAbility* and *CoworkerAbility,* which is quite high (0.8897), but we compute the variance inflation factors to find them to be below 10, indicating that we are not likely to have multicollinearity problems.

Table 2: Correlation Matrix of Check-Level Variables

| | log (*Sales*) | log(*Meal Duration*) | *Coworker Ability* | *OwnAbility* | *PartySize* | *AbilityStDev* | *LagCoworker Ability* | *AvgTables* |
|---|---|---|---|---|---|---|---|---|
| log (*Sales*) | 1.0000 | | | | | | | |
| log(*MealDuration*) | 0.3907* | 1.0000 | | | | | | |
| *CoworkerAbility* | -0.0046* | 0.0225* | 1.0000 | | | | | |
| *OwnAbility* | 0.0269* | -0.0035 | 0.4961* | 1.0000 | | | | |
| *PartySize* | 0.5985* | 0.1811* | -0.0719* | -0.0990* | 1.0000 | | | |
| *AbilityStDev* | 0.0531* | 0.0453* | 0.2957* | 0.1887* | 0.0058* | 1.0000 | | |
| *LagCoworkerAbility* | 0.0106* | 0.0303* | 0.8897* | 0.6044* | -0.0758* | 0.2647* | | |
| *AvgTables* | -0.0100* | 0.1043* | 0.1220* | 0.0939* | -0.0418* | 0.0365* | 0.1187* | |
| *StoreTables* | 0.1733* | 0.1337* | 0.1689* | 0.1078* | 0.0843* | 0.1290* | 0.1731* | 0.3440* |

*Significant at the 0.05 level.

# 5 Empirical Analysis Strategy and Results

## 5.1 Intrinsic Sales Ability

First we employ a fixed-effect model to estimate the focal server's intrinsic sales ability, on which we construct our independent variable. We first divide our data into 18 months (from January 2011 to June 2012). Then following Chan et al. (2014a), who measure a focal salesperson's permanent sales productivity (i.e., dollar sales per hour), we specify the following fixed-effect model to estimate the intrinsic sales ability of server $j$ at hour level $t$ and run this fixed-effect model for each month $m$, separately:

$$\sum_{i \in jtm} \frac{Sales_i}{PartySize_i} = \beta_0 + \theta_{jm} + \beta_1 tm + \beta_2 Controls_i + \varepsilon_{jtm} \quad \forall m \in 1, \cdots, 18. \tag{1}$$

In the specification, $\sum_{i \in jtm} \frac{Sales_i}{PartySize_i}$ is measured by averaging the sales of all the checks that are opened during hour $t$ and handled by server $j$ over all the diners who contribute to these sales (the server*'s* average per person dollar sales (PPA) in each hour). PPA is a financial measure often used by restaurant management

(Mill, 2006). We calculate PPA in dollars instead of log-transforming it for interpretation purposes. We estimate the intrinsic sales ability at the hour level because 1) servers typically work hourly shifts; 2) hourly aggregation instead of daily aggregation ensures both an adequate sample size in each hour for each server and enough observations over time for statistically significant estimates. In addition, $Controls_i$ include $DayWeek_i$, $Hour_i$, $YearWeek_i$ and $Store_i$ to adjust for the time, date and location factors, which is equivalent to a store fixed-effect model. We include store-specific time-invariant factors to control for unobserved heterogeneity among stores, such as the income level of the neighborhood and other time-invariant omitted variables.

## 5.2 Performance Analysis

Unlike previous studies that analyze the data either at hourly or other aggregate level (e.g., Mas and Moretti, 2009; Chan et al., 2014a), we conduct our main analysis at the check level (similar to Tan and Netessine, 2014b) because granular-level analysis tends to be more informative than aggregate-level analysis and we have sufficient data. In particular, we employ the following ordinary least squares (OLS) regression models:

$$\begin{aligned}
\log(Sales_i) &= \alpha_0 + \alpha_1 CoworkerAbility_i + \alpha_2 CoworkerAbility_i^2 + \alpha_3 OwnAbility_i + \\
&\quad \alpha_4 PartySize_i + \alpha_5 AbilityStDev_i + \alpha_6 LagTeamAbility_i + \\
&\quad \alpha_7 AvgTables_i + \alpha_8 AvgTables_i^2 + \alpha_9 Controls_i + \varepsilon_i \qquad (2)
\end{aligned}$$

$$\begin{aligned}
\log(MealDuration_i) &= \beta_0 + \beta_1 CoworkerAbility_i + \beta_2 CoworkerAbility_i^2 + \beta_3 OwnAbility_i + \\
&\quad \beta_4 PartySize_i + \beta_5 AbilityStDev_i + \beta_6 LagTeamAbility_i + \\
&\quad \beta_7 AvgTables_i + \beta_8 AvgTables_i^2 + \beta_9 StoreTables_i + \beta_{10} Controls_i + \xi_i \qquad (3)
\end{aligned}$$

The independent variables *CoworkerAbility* and $CoworkerAbility^2$ are centered around their means for interpretation purposes. The coefficient of $CoworkerAbility_i^2$ (e.g., $\alpha_2$ in the sales model) will be negative if there is an inverted-U-shaped relationship between peer effect and sales. In addition, the critical point of the performance measure is expected to be at $-\alpha_1/(2\alpha_2)$. $Controls_i$ represents the same control variables as in Model 1. We also calculate heteroscedasticity-consistent standard errors so as to allow the fitting of our model to contain potential heteroscedastic residuals.

Although useful as a preliminary estimator, these regression models may not address potential omitted variable bias towards both sales and meal duration estimations. The potential omitted variables should affect both the performance measures and the team composition decisions. In other words, those omitted variables, such as consumers' price sensitivity or their intrinsic level of hunger, will not bias our estimation because this type of consumption-behavior-related factor is likely to be uncorrelated with team composition. Rather, we highlight two types of significant omitted variables that are related to team composition scheduling. In the sales model, one major omitted variable is managers' unobserved demand forecast, which should be positively correlated with sales. In addition, it should affect team ability, but the direction of the correlation may be ambiguous ex ante. On one hand, managers may be inclined to schedule servers having high sales ability to work during high-sales days either to match the demand or to reward high-performing servers. On the other hand, managers may wish to schedule servers having high sales abilities to work during low-sales days because the extra sales improvement for high-performing servers may be more significant during low-sales shifts than during high-sales shifts. Hence, the direction of the omitted variable bias in the OLS sales model is inconclusive a priori.

Furthermore, although the meal duration model controls for the kitchen and the bar workload via *StoreTables*, it may still suffer from omitted variable bias because we do not observe the workload in the kitchen and the bar, which should be positively correlated with team ability because high-performing servers tend to sell more items, which can increase the load of the kitchen or bar. However, the correlation between kitchen workload and meal duration may be equivocal ex ante. On one hand, kitchen workload may be positively correlated with meal duration because high workload may cause a longer wait for food/drink preparation, assuming that the processing rate is fixed. On the other hand, kitchen workload may be negatively correlated with meal duration because service speed in the kitchen may increase (see evidence and explanations of this effect in other contexts such as Schultz et al., 1999; KC and Terwiesch, 2009). Moreover, workload may even have an inverted-U-shaped relationship with meal duration (Tan and Netessine, 2014a,b). For these reasons, similar to the sales model, the direction of the omitted variable bias is not clear a priori. Although the direction of these biases cannot be determined ex ante, these omitted variables may still cause inaccurate estimations, so we performed additional Hausman endogeneity tests and rejected the null hypotheses that those peer effect measures were exogenous. In order to alleviate these biases, we resort to an instrumental variable two-stage-least-square (2SLS) approach (Angrist and Krueger, 1994) in the next subsection.

### 5.3 Instrumental Variable 2SLS Estimation

We rely on an instrumental variable 2SLS approach to address the endogeneity issues because it can provide consistent estimates of the dependent variables using a large sample (Angrist and Krueger, 1994). For an instrumental variable to be valid, it should satisfy both relevance and exclusion restriction assumptions (Wooldridge, 2002), which means that it should be uncorrelated with the error (i.e., exclusion restriction) and correlated with the endogenous regressor (i.e., relevance). We introduce two types of instruments, which should satisfy these two conditions. First, we observe an exogenous shock to the scheduling decision during our study period. In the middle of 2011, one of the three restaurants switched to a computer-based scheduling system insead of relying on managers' discretionary decisions. The new computer-based system does not explicitly advise which servers should be scheduled for each shift; however, it analyzes 13-week historical sales data to forecast the demand for servers for the next week. The new computer-based staffing level forecast is likely to be different from a manager's forecast because it analyzes more data and tends to be more consistent than a manager typically can. Because of the adjusted staffing level, the team composition may mechanistically change accordingly. For example, the team composition of an eight-server team would be by definition different from a nine-server team. Hence, the new scheduling system should affect the average team ability and coworker ability, thus meeting the relevance condition. Furthermore, we expect that the implementation of the new system should affect sales and meal duration only through team-composition decisions because diners do not observe the implementation of this labor scheduling system. Therefore, the implementation of the system should also meet the exclusion restriction condition. To operationalize the instrument, we create a dummy variable, *System*, which equals one for all the checks affected by the new scheduling system, and zero for all other observations.

We supplement our analysis using another type of instrumental variable, the lagged values of the endogenous independent variables (e.g., Bloom and Van Reenen, 2007; Siebert and Zubanov, 2010), In particular, we first construct the average team ability during the same hour $t$ as check $i$ opened, $TeamAbility_{it} = 1/n \sum_{j \in t} OwnAbility_j$, where $j$ is a server who worked during hour $t$ in the same restaurant. Then *LWTeamAbility* and $LWTeamAbility^2$ are computed to represent the *TeamAbility* and $TeamAbility^2$ of the same restaurant during the same hour of the previous week to be used as instruments for the current week[4]. As an illustration, suppose check $i$ was opened at 7:30 pm on 8/8/2011 at restaurant $k$. Its instrument is *TeamAbility* of the 7:00 pm slot on 8/1/2010 at restaurant $k$. We also mean-center these instruments for interpretation purposes. We elect to use the one-week lag because the focal restaurants schedule workers one week in advance and these weekly schedules tend to be quite stable (the correlation between *LWTeamAbility* and *TeamAbility* is about 0.8). For this reason, the weekly lagged variables should correlate with the current team composition and thus *CoworkerAbility*, thus satisfying the relevance condition. Since the scheduling decisions from a week ago should not determine the unobserved factors for sales and meal duration during the current week, these lagged instrumental variables should also satisfy the exclusion restriction condition. It is true that the lagged team composition may not be ideal in the event of common demand shocks that are correlated over time. We adjust for these common demand shocks, which are basically trends (Villas-Boas and Winer, 1999), in our models with the categorical control variable *YearWeek*, thus lessening this concern. Additional relevant statistics and further discussion to show the validity of these instruments are provided after the main results are presented in Subsection 5.4.

[4]Note that we are unable to construct a lagged *CoworkerAbility* variable because a server does not always work the same shift every week.

## 5.4 Results

Table 3 shows the results of our check-level analysis of the impact of coworkers' ability on servers' performance. In the sales models (columns 1 and 2), the coefficients of *CoworkerAbility*$^2$ are consistently significant and negative in both OLS and 2SLS estimations (-0.0033 and -0.0102, respectively), providing support for H1, which states that peer effect has an inverted-U-shaped relationship with sales. In addition, although the linear term *CoworkerAbility* is statistically insignificant in the OLS model, it becomes significant and positive (0.0208) after we correct the potential endogeneity bias by 2SLS estimation. Interpreting the coefficients from the 2SLS, we find that the critical average coworker ability is about $0.0208/(2 \times 0.0102) \approx \$1$, which is about one standard deviation (\$1.021) above the sample mean (\$0.02). We further calculate that changing the current average coworkers' ability to the optimal value would have generated ($0.0208 \times 1 - 0.0102 \times 1^2 \approx 1\%$) sales lift per check for the focal server on average, holding party size and other factors constant. As a robustness check, we test the model specification including only the linear term *CoworkerAbility* with control variables. Its coefficient turns out to be insignificant in both OLS and 2SLS models, further suggesting a plausible non-linear relationship between coworkers' ability and servers' sales performance.

For the control variables in the sales models, as expected, *OwnAbility* is positively associated with sales. In particular, its coefficient in column 2 is 0.0310, so increasing a server's intrinsic sales ability by \$1 may increase his/her sales by approximately 3.1%. In addition, *PartySize* is significant and positive across all models because a larger party size should be positively associated with higher sales per check. Control variables *AbilityStDev* and *LagCoworkerAbility*, however, are insignificant. The other two control variables of check-level workload, *AvgTables* and *AvgTables*$^2$, are significantly positive and negative, respectively, which is consistent with Tan and Netessine (2014b), who find an inverted-U-shaped relationship between workload and servers' sales performance.

An inverted-U-shaped relationship is also observed in the meal duration models (columns 3 and 4). The coefficients of *CoworkerAbility*$^2$ are consistently negative (-0.0035 from OLS estimation and -0.01 from

2SLS estimation), supporting H2, which states that peer effect has an inverted-U-shaped relationship with meal duration. In addition, the coefficients of *CoworkerAbility* are statistically undifferentiated from zero in both OLS and 2SLS models, implying that the longest meal duration seems to happen right at the sample mean ($0.02). As we do with the sales model, we also re-estimate the model including only the linear term *CoworkerAbility* with control variables and find that its coefficient is insignificant, which suggests that a non-linear relationship is more likely to reflect the true peer effects than a linear relationship. For the control variables in the meal duration models, it is noteworthy that the signs of *OwnAbility* estimates are consistently negative in all models, which seems to suggest that high-sales-ability servers may not only generate more sales, but also work faster. In addition, *AbilityStDev* estimates have significantly positive signs across all models, which implies that team sales ability heterogeneity is associated with long meal duration. For the individual workload controls, similar to the results in the sales model, the signs of *AvgTables* and $AvgTables^2$ are significantly positive and negative, respectively, which is also congruent with the finding of the inverted-U-shaped relationship between workload and meal duration in Tan and Netessine (2014b). Overall, these findings indicate strong support for H1 and H2.

The 2SLS estimation results rely on the validity and the asymptotic consistency of instrumental variable estimators. Hence, we now check both the relevance condition and the exclusion restriction condition. Our combined instrumental variables are not "weak", and they should satisfy the relevance condition because when both endogenous variables are regressed, the coefficients of *System* are both positive, which implies that the implementation of the new scheduling system may have increased the average team ability[5]. The coefficients of the one-week lagged instrumental variables are also statistically significant with expected signs. Finally, the $F$-statistics for the joint significance of the first-stage estimations are both over 1,000, which is higher than 10, the suggested rule of thumb for weak instruments (Staiger and Stock, 1997). Just as important as the relevance condition, the exclusion restriction condition should be satisfied for our instrumental variables. First, we conduct Sargan tests of over-identifying restrictions to test the exclusion restriction con-

[5]The first stage results are available upon request.

Table 3: Check-Level Peer Effect *CoworkerAbility* on log(*Sales*) and log(*MealDuration*)

| | (1) Sales Estimated by OLS | (2) Sales Estimated by 2SLS | (3) Meal Duration Estimated by OLS | (4) Meal Duration Estimated by 2SLS |
|---|---|---|---|---|
| *CoworkerAbility* | 0.0029 | 0.0208* | -0.0016 | -0.0085 |
| | (0.0022) | (0.0097) | (0.0019) | (0.0085) |
| *CoworkerAbility*$^2$ | -0.0033*** | -0.0102** | -0.0035*** | -0.0100** |
| | (0.0008) | (0.0036) | (0.0006) | (0.0033) |
| *OwnAbility* | 0.0299*** | 0.0310*** | -0.0045*** | -0.0055*** |
| | (0.0008) | (0.0010) | (0.0007) | (0.0009) |
| *PartySize* | 0.2751*** | 0.2751*** | 0.0579*** | 0.0580*** |
| | (0.0008) | (0.0008) | (0.0007) | (0.0007) |
| *AbilityStDev* | -0.0010 | -0.0010 | 0.0051** | 0.0092* |
| | (0.0020) | (0.0034) | (0.0019) | (0.0047) |
| *LagCoworkerAbility* | -0.0008 | -0.0097 | 0.0051** | 0.0092* |
| | (0.0021) | (0.0054) | (0.0019) | (0.0047) |
| *AvgTables* | 0.0690*** | 0.0692*** | 0.1496*** | 0.1500*** |
| | (0.0044) | (0.0045) | (0.0057) | (0.0036) |
| *AvgTables*$^2$ | -0.0137*** | -0.0138*** | -0.0197*** | -0.0196*** |
| | (0.0008) | (0.0008) | (0.0011) | (0.0007) |
| *StoreTables* | | | -0.0015*** | -0.0017*** |
| | | | (0.0002) | (0.0002) |
| *Controls* | Yes | Yes | Yes | Yes |
| Observations | 206,257 | 201,567 | 206,257 | 201,567 |
| Prob>Chi-sq | <.001 | <.001 | <.001 | <.001 |

1. Standard errors are shown in parentheses. 2. *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

dition (Kennedy, 2003). The $p$-values of the Sargan tests are over 0.1 for both models, which suggests that we fail to reject the null hypothesis that the error terms of the structural models are uncorrelated with the instrumental variables. Further, the implementation of the new scheduling system should affect restaurant performance only through labor decisions because it should not affect demand factors.

## 5.5 Mechanisms of Servers' Performance Variation

Having just argued that peer effects may exhibit inverted-U-shaped relationships with both sales and meal duration, we want now to understand the mechanisms of such performance impacts. Although our observational data do not allow us to examine all possible mechanisms, we examine two that may complement existing peer effect studies.

### 5.5.1 Servers' Sales Actions

What actions do servers take in response to their peers' ability? We analyze the number of items sold, which can help us understand the two different actions that servers generally take to influence sales, cross-selling or upselling. Cross-selling is selling more items, such as desserts or wines, which will take additional time for servers to sell and for diners to consume. Upselling means selling more expensive items, such as steaks or seafood instead of chicken. Further, there are two interwoven factors that may influence servers' speed and meal duration. First, servers often face a speed/quality trade-off: achieving high sales/quality takes time (e.g., influencing diners to purchase more items takes additional time and effort). Second, servers may simply change their promptness to affect the meal duration without affecting the number of items sold. To delineate these interwoven factors, we first analyze the effect of the coworker ability on the number of items sold during a check, which is a reasonable proxy for servers' cross-selling efforts. In particular, we use the same 2SLS strategy and the same set of instruments employed in Subsection 5.3 to estimate the effect of *CoworkerAbility* on a new dependent variable, log($Items_i$), which is the logarithm of the number of items sold during check *i*. Then, we control for the impact of log(*Items*) on meal duration to isolate the cross-selling effort. The additional impact of coworker ability on meal duration should then be due to servers' adjustment of promptness. Finally, we control for log(*Items*) in the sales model to examine the impact of coworker ability on upselling effort.

Table 4 shows the results of the number-of-items-sold analysis. In the log(*Items*) model (column 1), the coefficient of $Coworker^2$ is negative and significant (-0.0111), while the coefficient of *CoworkerAbility* is effectively zero, which suggests that coworker ability has an inverted-U-shaped relationship with servers' cross-selling efforts, with an inflection point near the sample mean ($0.02). In other words, as coworker ability increases, servers first sell more items, but then sell fewer items as coworker ability continues increasing. In the log(*MealDuration*) model conditioned on the number of items sold (column 2), the coefficient of $CoworkerAbility^2$ is significant and negative (-0.008), while the coefficient of *CoworkerAbility* is statistically zero, which implies that servers may decelerate as coworker ability increases up to the the sample mean, and yet accelerate after coworker ability surpasses this inflection point. Finally, in the log(*Sales*) model conditioned on the number of items sold (column 3), we find that the coefficient of $CoworkerAbility^2$ is still significantly negative (-0.0068), while the coefficient of *CoworkerAbility* is significantly positive (0.0197), not only hinting that coworker ability has an inverted-U-shaped relationship with servers' upselling behavior, but also suggesting that the inflection point is about (0.0197/(2$\times$0.0068) $\approx$ 1.4) dollars above the sample mean. Putting the three models together, Table 4 suggests that when overall coworker ability is below the sample mean, increasing coworker ability may trigger servers to expend more upselling and cross-selling efforts, while reducing their service speed. As coworker ability reaches the sample mean, increasing coworker ability may start to stimulate servers to work more promptly and reduce their cross-selling effort, which further shortens meal duration (the coefficient of log(*Items*) is positive in the log(*MealDuration*) model). Although cross-selling effort is decreased, servers continue to redouble their upselling effort until the coworker ability reaches \$1.4 above the sample mean, at which point upselling effort also starts to fall.

Together with the results shown in Table 3, these results provide insights into the decomposition of the sales effects into servers' cross-selling and upselling activities. From Table 3, we calculate that the optimal coworker ability is about \$1 above the sample mean. Given this optimal coworker ability, the average sales may increase by 1.3% because of upselling ($0.0197\times 1 - 0.0068 \times 1^2 \approx 1.3\%$). However, this optimal coworker ability may reduce servers' cross-selling effort, which may cause 1.11% fewer sold items (column 1 of Table 4). We compute that these 1.11% fewer sold items may cause further sales reduction by $1.11\%\times 0.3097 \approx 0.3\%$, using the coefficient of log(*Items*) in column 3 of Table 4. In total, Table 4 suggests that the effect of optimal coworker ability should increase sales on average by ($1.3\% - 0.3\% = 1\%$), which is consistent with the estimation in Table 3.

### 5.5.2 Table Proximity

As shown in Figure 1, we have information on the assigned tables of each server in one of the three restaurants in our study and the locations of these tables. We define an observable server if he/she is assigned to a table that is adjacent to any of the tables where the focal server works. Then we reconstruct the peer effect variables *AdjacentCoworkerAbility* and *SeparateCoworkerAbility* using the intrinsic abilities (*OwnAbility*) of only those adjacent servers and separate servers, respectively, $AdjacentCoworkerAbility_i = \bar{\theta}_{-j}^{adj} = 1/n \sum_{k \neq j} \theta_k^{adj}$, where server *j* handles check *i* and server *k* is observable to server *j*; and $SeparateCoworkerAbility_i = \bar{\theta}_{-j}^{sep} = 1/n \sum_{k \neq j} \theta_l^{sep}$, where server *l* is unobservable to server *j*. Mean-centering and replacing these new peer effect variables and their quadratic terms for *CoworkerAbility* and *CoworkerAbility*$^2$ in models 2 and 3, respectively, we use OLS to test if the peer effect changes because of the servers' observability.

Table 4: Mechanisms of Servers' Performance Variation

| | (a) Number of Sold Items Analysis | | | (b) Peer Effects of Observable and Unobservable Servers | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | log(*Items*) | log(*MealDuration*) | log(*Sales*) | log(*Sales*) | log(*MealDuration*) | log(*Sales*) | log(*MealDuration*) |
| *CoworkerAbility* | 0.0035 | -0.0093 | 0.0197* | | | | |
| | (0.0106) | (0.0082) | (0.0089) | | | | |
| *CoworkerAbility*$^2$ | -0.0111** | -0.0080* | -0.0068* | | | | |
| | (0.0040) | (0.0032) | (0.0034) | | | | |
| *AdjacentCoworkerAbility* | | | | -0.0232*** | -0.0129** | | |
| | | | | (0.0056) | (0.0049) | | |
| *AdjacentCoworkerAbility*$^2$ | | | | -0.0102*** | -0.0034* | | |
| | | | | (0.0020) | (0.0017) | | |
| *SeparateCoworkerAbility* | | | | | | -0.0029 | 0.0017 |
| | | | | | | (0.0032) | (0.0026) |
| *SeparateCoworkerAbility*$^2$ | | | | | | -0.0002 | -0.0016 |
| | | | | | | (0.0010) | (0.0008) |
| *OwnAbility* | 0.0194*** | -0.0095*** | 0.0250*** | 0.0304*** | -0.0090*** | 0.0276*** | -0.0074*** |
| | (0.0011) | (0.0009) | (0.0010) | (0.0015) | (0.0012) | (0.0019) | (0.0015) |
| *PartySize* | 0.2822*** | 0.0006 | 0.1877*** | 0.2786*** | 0.0659*** | 0.2796*** | 0.0642*** |
| | (0.0009) | (0.0009) | (0.0009) | (0.0015) | (0.0012) | (0.0018) | (0.0015) |
| *AbilityStDev* | 0.0075* | 0.0170*** | -0.0033 | -0.0016 | 0.0024 | -0.0088 | -0.0034 |
| | (0.0038) | (0.0030) | (0.0032) | (0.0044) | (0.0037) | (0.0056) | (0.0047) |
| *LagCoworkerAbility* | -0.0029 | 0.0098* | -0.0088 | -0.0047 | 0.0048 | 0.0013 | -0.0018 |
| | (0.0058) | (0.0045) | (0.0049) | (0.0039) | (0.0033) | (0.0048) | (0.0040) |
| *log(Items)* | | 0.2037*** | 0.3097*** | | | | |
| | | (0.0017) | (0.0019) | | | | |
| *AvgTables* | 0.0656*** | 0.1381*** | 0.0489*** | 0.0828*** | 0.1823*** | 0.0959*** | 0.2114*** |
| | (0.0044) | (0.0035) | (0.0037) | (0.0085) | (0.0096) | (0.0116) | (0.0142) |
| *AvgTables*$^2$ | -0.0144*** | -0.0169*** | -0.0093*** | -0.0180*** | -0.0291*** | -0.0205*** | -0.0340*** |
| | (0.0008) | (0.0006) | (0.0007) | (0.0017) | (0.0020) | (0.0024) | (0.0031) |
| *StoreTables* | | -0.0020*** | | | -0.0003 | | -0.0005 |
| | | (0.0002) | | | (0.0003) | | (0.0004) |
| Observations | 201,567 | 201,567 | 201,567 | 59,399 | 59,399 | 36,780 | 36,780 |
| Prob>Chi-sq | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 |

1. Controls always included. 2. Standard errors are shown in parentheses. 3. *$p \leq .05$, **$p \leq .01$, ***$p \leq 0.001$.

Table 4 shows the results of the peer effects of observable servers and unobservable ones. In columns 4 and 5, where coworkers are observable, the coefficients of the quadratic terms of coworkers' ability are significant and negative in estimating both log(*Sales*) and log(*MealDuration*) (-0.0102 and -0.0034, respectively), supporting the inverted-U-shaped peer effects. However, the coefficients of the coworkers' ability are all insignificant in columns 6 and 7, which seems to suggest that the peer effect is effectively zero if these coworkers are unobservable, which supports H3. Simply put, out of sight, out of mind. Previous work by Mas and Moretti (2009) finds similar evidence that only those cashiers who can directly observe fast coworkers in front of them may gain an improvement in their productivity. Although Table 4 provides some evidence that the peer effect is only present if the coworkers are observable at work, it does not preclude that servers working at table sections far apart may affect each other in common areas, such as the kitchen, the drink stations and the registers. Rather, the goal of this analysis is to suggest that servers working in close proximity may have a more significant peer effect, consistent with H3.

### 5.6 Hour-Level Analysis

We further conduct an hour-level analysis 1) to provide a robustness check because previous studies on peer effects analyze this level of aggregation (Mas and Moretti, 2009; Chan et al., 2014a); and 2) to offer practical implications for optimal storewide scheduling because restaurants tend to schedule servers on an hourly basis. We define the hour-level dependent variables in terms of hourly average sales per check and hourly average meal duration, $HRAvgSales_{tk} = \frac{\sum_{i \in tk} Sales_i}{HRChecks_{tk}}$ (mean = \$42.49, stdev = \$11.08) and $HRAvgMealDuration_{tk} = \frac{\sum_{i \in tk} MealDuration_i}{HRChecks_{tk}}$ (mean = 48.77 minutes, stdev = 9.12 minutes), respectively, where $i$ is a check starting in hour $t$ at restaurant $k$. In addition, $HRChecks_{tk}$ is the total number of checks starting in hour $t$ at restaurant $k$ (mean = 13.07, stdev = 8.82). The independent variable $TeamAbility_{tk}$ is defined as the average team ability during hour $t$ at restaurant $k$ (equal to $TeamAbility_i$, as defined in Subsection 5.3, and with a mean of \$-0.03 and an stdev of \$1.09). It is then centered around its mean for interpretation purposes. Unlike *CoworkerAbility*, which measures the average ability of the coworkers of a focal server, *TeamAbility* assesses the average ability of all the servers. Accordingly, we exclude *OwnAbility* in the hour-level model. If we still observe an inverted-U relationship between *TeamAbility* and the performance measures, we will find additional evidence for the inverted-U relationship between peer effects and servers' performance. Moreover, the quadratic specification of *TeamAbility* will imply an optimal average team ability for the store. For other controls, we change the check-level party size control to

$AvgPartySize_{tk} = \frac{\sum_{i \in tk} PartySize_i}{HRChecks_{tk}}$, which is the average party size of the checks during hour $t$ at restaurant $k$ (mean = 2.43 diners, stdev = 0.47 diners). $TeamStDev_i$ is measured every hour, so we simply change its subscript to $tk$ in the hourly model. We also change the one-hour lagged peer effect into *LagTeamAbility,* which is the one-hour lagged variable of *TeamAbility.* We further control for $HRChecks_{tk}$, the store traffic, and divide it by the number of servers who processed at least one check in the same hour to create $HRTableLoad_{tk}$ and its quadratic term and to adjust for the average individual workload (mean number of servers per hour = 6.26, stdev = 3.41 servers; mean of $HRTableLoad_{tk}$ = 1.97 tables per server, stdev = 0.69 tables). Finally, we use the same set of time/date/location variables as in models 2 and 3.

We specify our model as follows:

$$\begin{aligned}\log(Performance_{tk}) \quad = \quad & \alpha_0 + \alpha_1 TeamAbility_{tk} + \alpha_2 TeamAbility_{tk}^2 + \alpha_3 AvgPartySize_{tk} + \\ & \alpha_4 AbilityStDev_{tk} + \alpha_5 LagTeamAbility_{tk} + \alpha_7 HRChecks_{tk} + \\ & \alpha_8 HRTableLoad_{tk} + \alpha_9 HRTableLoad_{tk}^2 + \alpha_{10} Controls_{tk} + \varepsilon_{tk}., \end{aligned} \tag{4}$$

where $Performance_{tk}$ is $Sales_{tk}$ and $MealDuration_{tk},$ our two performance measures, respectively. We estimate these models by 2SLS using the same instruments as in the check-level analysis. Table 5 shows the hourly analysis results. In the log(*HRAvgSales*) model, the coefficient of $TeamAbility^2$ is significant and negative (-0.0156), while the coefficient of *TeamAbility* is significant and positive (0.0552). These results suggest that peer effect is likely to have an inverted-U-shaped relationship with sales per check, and the optimal team ability to maximize sales is greater than the sample mean, consistent with our check-level results. Interpreting these estimated coefficients, we find that the optimal *TeamAbility* for the entire store is about \$1.76 above the sample mean (-\$0.03). In addition, the optimal *TeamAbility* would have increased *HRAvgSales* by $(0.0552 \times 1.76 - 0.0156 \times 1.76^2) \approx 4\%$. Note that the 4% sales lift from optimal team ability composition is the total effect of increasing the average team ability, which includes the direct effect of using higher-ability servers, and the indirect effect via peer effects. This total effect is quantitatively congruent with our check-level estimation, where we find that optimally increasing every server's sales ability by one dollar will be associated with a 3% direct sales lift and a 1% indirect sales lift via peer effects. Hence, the hour-level analysis results are both qualitatively and quantitatively consistent with our check-level analysis results. In the log(*HRAvgMealDuration*) model, the coefficient of *TeamAbility*$^2$ is both significant and negative (-0.0105), while the linear term is statistically insignificant at the 0.05 level. The meal duration results are qualitatively congruent with our main results – team ability initially concavely increases until it reaches the sample mean and then concavely decreases the average meal duration.

The inverted-U-shaped relationship between hourly average team ability and average sales lends support for H4, which states that heterogeneous team ability levels may increase storewide sales performance because scheduling all high-ability or low-ability servers together (i.e., polarizing the team ability) causes the average team ability to be either too high or too low. In addition, whereas the coefficient of *AbilityStDev* is insignificant in the check-level sales model (Table 3), the coefficient of *AbilityStDev* is significant and positive in the hour-level analysis (0.0101), which seems to further corroborate H4. Moreover, in Model 4, we control for the average team ability with *TeamAbility*. Hence, our results also provide evidence that worker ability may increase storewide sales performance even conditioned on the mean, which is consistent with previous work (Shafer et al., 2001; Hamilton et al., 2003).

There are two caveats of our test of H4. First, following the same logic of the non-linear sales effects, the inverted-U-shaped relationship between average team ability and meal duration also implies that team heterogeneity may increase average meal duration, which may inflate the opportunity costs of store capacity especially during peak hours. However, our focal restaurants are generally not very highly utilized (the mean of *HRTableLoad* is 1.97 parties per server, with a standard deviation of 0.69, while the capacity is typically four parties). In addition, our models control for the store traffic with *HRChecks*. Second, the effect of team ability heterogeneity is known to depend on the compensation system. According to Chan et al. (2014a), heterogeneity in team ability may improve team performance under team-based incentives, while it may inhibit team performance under individual-based incentives. Although we agree that compensation systems should affect the effect of ability heterogeneity, we find counter-evidence to the effect of the heterogeneity in worker ability under individual-based incentives probably because Chan et al. (2014a) study linear peer effects and our setting is not a pure individual-based compensation system.

Table 5: Impacts of Hour-Level Team Ability on log(*HRAvgSales*) and log(*HRAvgMealDuration*)

| | log(*HRAvgSales*) | log(*HRAvgMealDuration*) |
|---|---|---|
| *TeamAbility* | 0.0552*** | 0.0019 |
| | (0.0147) | (0.0141) |
| *TeamAbility*$^2$ | -0.0156** | -0.0105* |
| | (0.0048) | (0.0046) |
| *AvgPartySize* | 0.3012*** | 0.0759*** |
| | (0.0036) | (0.0035) |
| *AbilityStDev* | 0.0101** | 0.0160*** |
| | (0.0038) | (0.0037) |
| *LagTeamAbility* | -0.0084 | 0.0054 |
| | (0.0073) | (0.0070) |
| *HRChecks* | 0.0032*** | 0.0020*** |
| | (0.0003) | (0.0003) |
| *HRTableLoad* | 0.0679*** | 0.0618*** |
| | (0.0100) | (0.0096) |
| *HRTableLoad*$^2$ | -0.0142*** | -0.0100*** |
| | (0.0020) | (0.0019) |
| *Controls* | Yes | Yes |
| Observations | 14,880 | 14,880 |
| Prob>Chi-Sq | <.001 | <.001 |

1. Standard errors are shown in parentheses. 2. *$p\leq .05$, **$p\leq .01$, ***$p\leq .001$.

# 6 Robustness Checks

## 6.1 Familiarity Weighted Peer Effect

Similar to airline crew and other fluid teams (Huckman et al., 2009), the team composition of every shift at restaurants changes quite substantially over time for reasons such as waiter time preferences and staff turnover. Consequently, during any shift, servers have familiarity of varying degrees with their coworkers. Familiarity may have implications for worker performance: for example, Huckman et al. (2009) find that team familiarity has a significant positive effect on performance. Just as important, Mas and Moretti (2009) report that peer effect is insignificant if the coworkers have low schedule overlap with the focal worker. We therefore examine whether or not the inverted-U-shaped peer effect is robust, depending on the degree of familiarity. Specifically, suppose servers $j$ and $k$ worked together in the same hour when check $i$ started. We define $F_{jki}$ as the number of overlapped hours that servers $j$ and $k$ had by the hour when check $i$ started, and then we weight *CoworkerAbility* by the familiarity. In other words, $FWCoworkerAbility_i = \bar{\theta}_{-j}^{Familiarity} = 1/n\sum_{k\neq j}(\theta_k \cdot F_{jki})$, where $\theta_j$ is the intrinsic sales ability of server $j$. Although $F_{jki}$ is left censored due to data limitation, the potential bias should be small because 1) the restaurant industry has one of the highest staff turnover rates; and 2) the effect of familiarity is likely to decrease over time (Huckman et al., 2009). We then replace this weighted peer effect measure and its quadratic term for *CoworkerAbility* and *CoworkerAbility*$^2$ in models 2 and 3 and employ 2SLS with the same instruments to test the robustness of the non-linear peer effects weighted by familiarity. We find that the peer effects weighted by familiarity have qualitatively the same inverted-U-shaped relationships with both sales and meal duration[6].

## 6.2 Alternative Inverted-U Testing: Spline Regressions

Our study is one of the first that examines the inverted-U relationship between peer effects and performance in the workplace. To identify this non-linear relationship in our main analysis, we utilize a commonly used criterion, the significance of the quadratic term of *CoworkerAbility*. Nevertheless, the quadratic specification may suffer from two issues. First, it may create an extreme point even though the true relationship is concave and monotone (Lind and Mehlum, 2010). Second, the quadratic criterion may be limited to the "non-local" assumption, which implies that the fitted dependent variables, $\log(\widehat{Sales})$ and $\log(\widehat{MealDuration})$, at a given $CoworkerAbility = CoworkerAbility_0$ depend heavily on *CoworkerAbility* values far from $CoworkerAbility_0$. The first issue should not apply to our analysis because our extreme points are within one standard deviation from the sample mean in the sales model and at the sample mean in the meal duration model. In order to address the second issue, we follow the literature (e.g., Kesavan et al., 2014) to apply spline regressions, choosing one knot that splits *CoworkerAbility* into two equal-sized groups. Then we estimate two spline regressions to fit piecewise linear functions of *CoworkerAbility1* (the lower 50%) and *CoworkerAbility2* (the

[6]The regression results are available from the authors upon request.

higher 50%)[7]. In both models, the coefficients of *CoworkerAbility1* are significant and positive, suggesting that as average coworker ability rises, both average sales and meal duration first increase. However, the coefficients of *CoworkerAbility2* are both significant and negative, implying that as average coworker ability rises further, both sales and meal duration then drop.

### 6.3 Alternative Measures of Peer Effects

In the main analysis, we measure average coworker ability in dollar values. We explore two alternative measures of peer effects. First, we measure average coworker ability in terms of their ability rankings. In particular, we rank the servers in each restaurant by their *OwnAbility* every month from low to high (a higher ranking value means a higher ability) and then assign to each server his/her relative ranking by percentile, that is, ranking/(the number of servers during that month in the restaurant). We then create an alternative independent variable, $CoworkerRank_i$, which is the average percentile ranking of the coworkers who work during the same hour with the focal server responsible for check *i*. We use these rankings as an alternative measure for ability because some restaurants use them to intuitively evaluate servers (Tan and Netessine, 2014a), and because literature on workplace tournament also uses the rankings to assess workers (Blanes i Vidal and Nossol, 2011). Second, we analyze how coworkers' ability relative to the focal server affects his/her performance. In other words, we create another alternative independent variable, *RelativeAbility* = *CoworkerAbility* - *OwnAbility*, which is also used in previous literature (e.g., Schultz et al., 2010; Chan et al., 2014a). We then mean-center these alternative independent variables and employ the same 2SLS models using the same instrumental variables as in Subsection 5.3. The quadratic terms of both alternative measures are consistently significant and negative in all four models, supporting our main finding that peer effects may have an inverted-U relationship with both sales and meal duration.

[7] We also divide the sample into three equal-sized groups and use OLS estimations. The results are still qualitatively similar.

# 7 Managerial Implications and Conclusions

## 7.1 Managerial Implications

The U.S. restaurant industry employs about 13 million (Mill, 2006) workers, the majority of whom are servers, and suffers from the lowest labor productivity in the service sectors (Freeman, 2008). How to manage these servers with diversified sales skills to optimize financial performance and to improve productivity has become an ever more pressing challenge for restaurant managers facing increasing pressures in a highly competitive industry. Our study provides three main managerial insights into how to manage this ability heterogeneity through optimal scheduling and team composition decisions.

First, in the hour-level analysis (Table 5), we find an inverted-U relationship between average team ability and storewide sales, which suggests that the average sales of the entire store will drop if the average team ability is either too high or too low. This drop in sales results from individual servers' decreasing sales performance when their coworkers working during the same hour have too strong or too weak sales ability (Table 3). This finding implies that scheduling all the "superstars" (high-performing servers) together during the same shift or all the "underdogs" (low-performing servers) together is suboptimal not only for individual servers' sales but also for the entire store's sales performance because polarizing the team ability causes it to be either too high or too low (the two lower ends of the inverted-U-shaped curve). Rather, managers should mix servers with heterogeneous ability levels so as to average out the team ability, ideally to achieve the optimal ability point (in our focal restaurants, the optimal point is about \$1.76 or 1.6 standard deviations above the sample mean). Moreover, since peer effects are stronger in close proximity, when the average team ability sometimes happens to be too low, managers may consider placing the few high performers in more visible sections in order to maximize their positive spillovers.

Second, using our 2SLS check-level estimates, we calculate that having an optimal team ability (\$1.76 above the sample mean) in our focal restaurants may increase sales by approximately 4%, which includes 3% of direct effect from increasing average sales ability and 1% of indirect peer effects. This delineation of the total effect of optimal scheduling implies that the value of having higher-ability servers at work can not only generate more sales from their own tables (direct effect) but also significantly improve the performance of their coworkers (indirect peer effects). Managers may therefore need to reconsider their compensation schemes to retain and reward higher-performing servers. Just as important, how to schedule the "superstars" and the "underdogs" in the optimal working group needs careful consideration. For example, in our focal restaurants where the current average team ability is below the optimum, hiring better servers and scheduling them to work more often can certainly boost the current average team ability. However, to do so can be costly. To make it feasible to increase the team ability without adding the capacity of higher-ability servers, we propose a heuristic to schedule the "superstars" to work during those shifts that have fewer servers scheduled. To put it mathematically, suppose Team 1 has $n_1$ servers, while Team 2 has $n_2$ servers, and $n_1 > n_2$. Without loss of generality, we assume the average team skill of Team 1 is the same as the average team skill of Team 2. In other words, $\bar{x}_1 = \bar{x}_2 = \bar{x}$. Now suppose in both teams we exchange a low-performing server with a high-performing server, whose skill is $x_k > \bar{x}$. Then the new average skill of Team 1 is $\bar{x}'_1 = \frac{1}{n_1}[(n_1 - 1)\bar{x} + x_k]$, while the average skill of Team 2 is $\bar{x}'_2 = \frac{1}{n_2}[(n_2 - 1)\bar{x} + x_k]$. The difference between the two teams, $\bar{x}'_2 - \bar{x}'_1 = (\frac{1}{n_2} - \frac{1}{n_1})\bar{x} - (\frac{1}{n_1} - \frac{1}{n_2})x_k > 0$. Hence, scheduling a "superstar" to work in a small team increases average team ability more than scheduling him in a large team. Equivalently, we suggest scheduling the "underdogs" to work during those shifts that have many servers because their negative contribution of low abilities is minimized across other servers.

Third, we conduct a counterfactual analysis and estimate that considering the inverted-U-shaped peer effects to optimize current servers' schedules without changing their capacity may increase sales by 2.7%. For illustrative purposes, we make simplifications in the scheduling model. We first assume there are two shifts in a day, (i.e., 14 shifts in a week). We further assume there are 14 types of servers having sales ability equally distanced in the empirical distribution of intrinsic sales ability (Subsection 5.1) from the bottom 1% (-\$4) to the top 1% (\$5.53). Then we round servers' original sales ability to these 14 new ability levels. Using the new sales ability levels, we re-estimate the peer effects in the check-level sales model (Model 2). The new coefficients turn out to be very close to the original ones (e.g., the coefficient of the linear term is equal to 0.021, while the coefficient of the quadratic term equals -0.01). After that, we compute both the average staffing requirements for each of the 14 shifts (in number of servers per shift) across weeks and stores and the capacity of each of the 14 servers (in terms of shifts). The lunch shifts require on average eight servers from Monday to Wednesday and nine servers from Thursday to Sunday, whereas dinner shifts require on average 10 servers on Mondays and Tuesdays; 11 servers on Wednesdays, Thursdays and Sundays; and 12 servers on Fridays and Saturdays. As to the server capacity, since the distribution of intrinsic sales ability seems to be bell curved as shown in Figure 2a, the medium-ability servers unsurprisingly have the highest capacity (27 shifts), while both the highest-ability and the lowest-ability servers have the lowest capacity (0.54 and 1.35 shifts, respectively). Finally, we run a non-linear optimization problem considering peer effects to maximize the average sales per check, subject to staffing requirements and capacity constraints. The non-negative decision variables are the scheduling decisions for each type of server during each shift. These decision variables are assumed to be continuous for computation purposes. We then include the potential peer effects calibrated by the aforementioned re-estimated sales model of the 14 types of servers to compute the counterfactual average sales for each server during each shift. In the end, the average sales per check achieved with optimal staffing turns out to be approximately 2.7% above the current sample mean ($42.49). The sales lift of this counterfactual analysis is less than the aforementioned 4% sales lift including both direct and indirect effects because this optimization precludes increasing the capacity of high-performing servers. It is also greater than the 1% pure peer effect because this optimization schedules servers more efficiently. For example, it may schedule high-performing servers to smaller shifts. In addition, the optimal schedule has a twice as high average standard deviation of different types of scheduled servers per shift as the current schedule (1.5 versus 0.72), which further provides evidence that heterogeneity of team ability may increase sales performance. Admittedly, in practice, managers face more complex scheduling constraints (e.g., certain servers cannot work together or some servers have specific availability). Nevertheless, the point of this counterfactual analysis is to highlight the value of empirical research in improving labor decisions

– using peer effects to optimize schedules without changing worker capacity can achieve a significant sales lift.

## 7.2 Conclusion

This empirical study makes contributions to two streams of literature, optimal scheduling/rostering decisions and peer effects studies. First, by studying the spillover effects of heterogeneous ability among coworkers, we identify a formerly overlooked assumption for analytical scheduling models that consider workers' heterogeneity (e.g., Cezik and L'Ecuyer, 2008; Bhulai et al., 2008; Bard and Wan, 2008). Future research on scheduling/rostering and other operational labor decisions may include such peer effects when making assumptions about servers' performance. Second, our finding about the non-linear peer effects may reconcile the seemingly conflicting linear peer effects found in earlier studies – the direction of peer effects may depend on the general level of coworkers' ability. Third, we provide empirical evidence for the analytical research on the speed/quality trade-off decisions, which is a topic of growing interest in the service operations area (e.g., Hopp et al., 2007; Debo et al., 2008; Anand et al., 2011; Zhan and Ward, 2013; Tan and Netessine, 2014b; Batt and Terwiesch, 2015).

Our empirical findings also have implications for scheduling decisions. First, the inverted-U-shaped peer effects imply that managers should mix servers with heterogeneous ability levels during the same shift because polarizing team ability is suboptimal. Second, we provide a heuristic to schedule high-ability servers to work during smaller shifts to increase average team ability because the optimal team ability is above the sample mean in our focal restaurants. Third, through a counterfactual analysis, we find that considering the inverted-U-shaped peer effects to optimize current servers' schedules without changing their capacity may increase sales by 2.7%, which highlights the value of empirical research for labor decisions (e.g., Campbell and Frei, 2011; Freeman et al., 2015)

It should be noted that this study has several important limitations. First, although our data set is more granular than previous studies on peer effects, we do not observe the social relationships among the coworkers, which have been found in prior work to moderate these peer effects (Mas and Moretti, 2009; Bandiera et al., 2009; Chan et al., 2014a). We also lack other service-quality data, such as complete tips data and customer-satisfaction survey data (However, as a robustness check, we examine the tips paid through credit cards, the only tips data available to us, and find the the tips/sales ratio is quite stable probably because of the strong social norm of tipping in the United States). Furthermore, since our servers are a subset of casual dining servers, one may raise concerns about the generalizability of our findings. We argue that the server body under study as a whole is drawn from the same pool as other casual dining restaurants throughout the United States, probably reducing this generalizability concern. Nevertheless, the servers may operate under different incentive schemes in other countries/cultures, which still require additional study. Finally, while we were primarily interested in examining the impact of peer effects on contemporaneous performance in sales and meal duration, other research may be fruitful in examining how peer effects affect other dependent variables, such as job satisfaction, retention and long-term performance.

## References


Akşin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665–688.

Akşin, Z., S. Deo, J. Jonasson, K. Ramdas. 2015. Learning from many: Partner diversity and team familiarty in fluid teams. London Business School Working Paper.

Alizamir, S., F. de Véricourt, P. Sun. 2013. Diagnostic accuracy under congestion. *Management Science* **59**(1) 157–171.

Anand, K. S., M. F. Paç, S. Veeraraghavan. 2011. Quality-speed conundrum: Tradeoffs in customer-intensive services. *Management Science* **57**(1) 40–56.

Angrist, J., A. B. Krueger. 1994. Why do World War II veterans earn more than nonveterans? *Journal of Labor Economics* **12**(1) 74–97.

Argote, L., D. Epple. 1990. Learning curves in manufacturing. *Science* **247**(4945) 920–924.

Argote, L., P. Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes* **82**(1) 150–169.

Arlotto, A., S. E. Chick, N. Gans. 2013. Optimal hiring and retention policies for heterogeneous workers who learn. *Management Science* **60**(1) 110–129.

Ata, B., J. A. Van Mieghem. 2009. The value of partial resource pooling: Should a service network be integrated or product-focused? *Management Science* **55**(1) 115–131.

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

Bandiera, O., I. Barankay, I. Rasul. 2010. Social incentives in the workplace. *Review of Economic Studies* **77**(2) 417–458.

Barankay, I. 2012. Rank incentives: Evidence from a randomized workplace experiment. Wharton Working Paper.

Bard, J. F., L. Wan. 2008. Workforce design with movement restrictions between workstation groups. *Manufacturing & Service Operations Management* **10**(1) 24–42.

Batt, R. J., C. Terwiesch. 2015. How a service process adapts to load: An econometric analysis of patient treatment in the emergency department. University of Wisconsin Working Paper.

Bhulai, S., G. Koole, A. Pot. 2008. Simple methods for shift scheduling in multiskill call centers. *Manufacturing & Service Operations Management* **10**(3) 411–420.

Blanes i Vidal, J., M. Nossol. 2011. Tournaments without prizes: Evidence from personnel records. *Management Science* **57**(10) 1721–1736.

Bloom, N., J. Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics* **122**(4) 1351–1408.

Buunk, Bram P, Rebecca L Collins, Shelley E Taylor, Nico W VanYperen, Gayle A Dakof. 1990. The affective consequences of social comparison: either direction has its ups and downs. *Journal of personality and social psychology* **59**(6) 1238.

Campbell, D., F. Frei. 2011. Market heterogeneity and local capacity decisions in services. *Manufacturing & Service Operations Management* **13**(1) 2–19.

Carrell, S. E., R. L. Fullerton, J. E. West. 2009. Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics* **27**(3) 439–464.

Carrell, S. E., B. I. Sacerdote, J. E. West. 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica* **81**(3) 855–882.

Cezik, M. T., P. L'Ecuyer. 2008. Staffing multiskill call centers via linear programming and simulation. *Management Science* **54**(2) 310–323.

Chan, T. Y., J. Li, L. Pierce. 2014a. Compensation and peer effects in competing sales teams. *Management Science* **60**(8) 1965–1984.

Chan, T. Y., J. Li, L. Pierce. 2014b. Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science* **33**(4) 463–484.

Charness, G., M. Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* **117**(3) 817–869.

Chen, K. P. 2003. Sabotage in promotion tournaments. *Journal of Law, Economics, and Organization* **19**(1) 119–140.

Cohen, S. G., D. E. Bailey. 1997. What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management* **23**(3) 239–290.

D., Nicole, Ananth R. 2007. Store manager incentive design and retail performance: An exploratory investigation. *Manufacturing Service Operations Management* **9**(4) 518–534.

Debo, L. G., L. B. Toktay, L. N. Van Wassenhove. 2008. Queuing for expert services. *Management Science* **54**(8) 1497–1512.

Falk, A., A. Ichino. 2006. Clean evidence on peer effects. *Journal of Labor Economics* **24**(1) 39–57.

Fehr, E., K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* **114**(3) 817–868.

Festinger, L. 1954. A theory of social comparison processes. *Human relations* **7**(2) 117–140.

Freeman, M., N. Savva, S. Scholtes. 2015. Gatekeepers at work: An empirical analysis of a maternity unit. London Business School Working Paper.

Freeman, Rebecca. 2008. Labour productivity indicators: Comparison of two oecd databases, productivity differentials and the Balassa-Samuelson effect. *OECD Statistics Directorate* .

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* **5**(2) 79–141.

Goldratt, E. M., J. Cox, D. Whitford. 2004. *The goal: A process of ongoing improvement*, vol. 3. Great Barrington, MA: North River Press.

Gray, J. V., E. Siemsen, G. Vasudeva. 2015. Colocation still matters: Conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. *Management Science* **61**(11) 2760–2781.

Guimera, R., B. Uzzi, J. Spiro, L. A. N. Amaral. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**(5722) 697–702.

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

Hoegl, M., L. Proserpio. 2004. Team member proximity and teamwork in innovative projects. *Research Policy* **33**(8) 1153–1165.

Hoorens, V. 1993. Self-enhancement and superiority biases in social comparison. *European Review of Social Psychology* **4**(1) 113–139.

Hopp, W. J., S. M. R. Iravani, G. Y. Yuen. 2007. Operations systems with discretionary task completion. *Management Science* **53**(1) 61–77.

Hopp, W.J., E. Tekin, M. P. Van Oyen. 2004. Benefits of skill chaining in serial production lines with cross-trained workers. *Management Science* **50**(1) 83–98.

Huckman, R. S., B. R. Staats, D. M. Upton. 2009. Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Science* **55**(1) 85–100.

Iravani, S. M. R., B. Kolfal, M. P. Van Oyen. 2007. Call-center labor cross-training: It's a small world after all. *Management Science* **53**(7) 1102–1112.

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

Itoh, H. 1993. Coalitions, incentives, and risk sharing. *Journal of Economic Theory* **60**(2) 410–427.

Kandel, E., E. P. Lazear. 1992. Peer pressure and partnerships. *Journal of Political Economy* **100**(4) 801–817.

Katzenbach, J. R., D. K. Smith. 1993. *The discipline of teams*. Watertown, MA: Harvard Business Press.

KC, D. S. 2013. Does multitasking improve performance? Evidence from the emergency department. *Manufacturing & Service Operations Management* **16**(2) 168–183.

KC, D. S., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486–1498.

Kennedy, P. 2003. *A guide to econometrics*. Cambridge, MA: The MIT Press.

Kesavan, Saravanan, Bradley R Staats, Wendell Gilland. 2014. Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science* **60**(8) 1884–1906.

Kimes, S. E. 2004. Restaurant revenue management: Implementation at Chevys Arrowhead. *Cornell Hotel and Restaurant Administration Quarterly* **45**(1) 52–67.

Kleinbaum, D. G., L. L. Kupper, K. E. Muller. 2007. *Applied regression analysis and other multivariable methods*. Duxbury Press.

Kostami, V., S. Rajagopalan. 2013. Speed-quality trade-offs in a dynamic model. *Manufacturing & Service Operations Management* **16**(1) 104–118.

Lapré, M. A., A. S. Mukherjee, L. N. Van Wassenhove. 2000. Behind the learning curve: Linking learning activities to waste reduction. *Management Science* **46**(5) 597–611.

Lazaer, E. P., S. Rosen. 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* **89**(5) 841–864.

Lazear, E. P. 1989. Pay equality and industrial politics. *Journal of Political Economy* **97**(3) 561–580.

Lind, J. T., H. Mehlum. 2010. With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics* **72**(1) 109–118.

Mandelbaum, Avishai, Petar Momcilovic, Yulia Tseytlin. 2012. On fair routing from emergency departments to hospital wards: QED queues with heterogeneous servers. *Management Science* **58**(7) 1273–1291.

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

Mehrotra, V., K. Ross, G. Ryder, Y. P. Zhou. 2012. Routing to manage resolution and waiting time in call centers with heterogeneous servers. *Manufacturing & Service Operations Management* **14**(1) 66–81.

Mill, R. C. 2006. *Restaurant management: Customers, operations, and employees*. Upper Saddle River, NJ: Prentice Hall.

Moretti, Enrico. 2004a. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics* **121**(1) 175–212.

Moretti, Enrico. 2004b. Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review* **94**(3) 656–690.

Narayanan, S., S. Balasubramanian, J. M. Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Science* **55**(11) 1861–1876.

O'Connor, E. J., L. H. Peters, A. Pooyan, J. Weekley, B. Frank, B. Erenkrantz. 1984. Situational constraint effects on performance, affective reactions, and turnover: A field replication and extension. *Journal of Applied Psychology* **69**(4) 663–672.

Roels, G., X. Su. 2013. Optimal design of social comparison effects: Setting reference groups and reference points. *Management Science* **60**(3) 606–627.

Schultz, K. L., D. C. Juran, J. W. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Management Science* **45**(12) 1664–1678.

Schultz, K. L., T. Schoenherr, D. Nembhard. 2010. An example and a proposal concerning the correlation of worker processing times in parallel tasks. *Management Science* **56**(1) 176–191.

Shafer, S. M., D. A. Nembhard, M. V. Uzumeri. 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science* 1639–1653.

Siebert, W. S., N. Zubanov. 2010. Management economics in a large retail company. *Management Science* **56**(8) 1398–1414.

Siemsen, E., S. Balasubramanian, A. V. Roth. 2007. Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Science* **53**(10) 1533–1550.

Song, H., A. Tucker, K. Murrel, D. Vinson. 2015. Learning from the best: The effects of public relative performance feedback on variability and productivity. Harvard Business School Working Paper.

Staats, B., D. S. KC, F. Gino. 2015. Blinded by experience: Prior experience, negative news and belief updating. University of North Carolina Working Paper.

Staats, B. R., F. Gino. 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science* **58**(6) 1141–1159.

Staiger, D., J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**(3) 557–586.

Suresh, S, W Whitt. 1990. The heavy-traffic bottleneck phenomenon in open queueing networks. *Operations Research Letters* **9**(6) 355–362.

Tan, T. F., S. Netessine. 2014a. The implications of worker behavior for staffing decisions: Empirical evidence and best practices. *Cornell Hospitality Quarterly* **55**(3) 277–286.

Tan, T. F., S. Netessine. 2014b. When does the devil make work? an empirical study of the impact of workload on worker productivity. *Management Science* .

Uzzi, B., J. Spiro. 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology* **111**(2) 447–504.

Van den Bergh, J., J. Beliën, P. De Bruecker, E. Demeulemeester, L. De Boeck. 2013. Personnel scheduling: A literature review. *European Journal of Operational Research* **226**(3) 367–385.

Villas-Boas, J. M., R. S. Winer. 1999. Endogeneity in brand choice models. *Management Science* **45**(10) 1324–1338.

Walker, J. R. 2007. *The Restaurant, Study Guide: From Concept to Operation*. New York City, NY: John Wiley & Sons.

Wallace, R. B., W. Whitt. 2005. A staffing algorithm for call centers with skill-based routing. *Manufacturing and Service Operations Management* **7**(4) 276–294.

Ward, A. R., M. Armony. 2013. Blind fair routing in large-scale service systems with heterogeneous customers and servers. *Operations Research* **61**(1) 228–243.

Wiersma, E. 2007. Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Science* **53**(12) 1903–1915.

Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. The MIT press.

Zhan, D., A. R. Ward. 2013. Threshold routing to trade off waiting and call resolution in call centers. *Manufacturing & Service Operations Management* **16**(2) 220–237.