Rapport in Organizations: Evidence from Fast Food

Achyuta Adhvaryu* Parker Howell† Anant Nyshadham‡
Jorge Tamayo§

December 4, 2024

Abstract

Common identity often provides a foundation for workplace rapport. Using personnel and productivity data from a large fast food chain in Colombia, we study whether mismatched gender identity across managers and workers affects the team's ability to deal with demand shocks. Leveraging the staggered expansion of a leading food delivery platform across the country, we show that stores in which managers and workers share predominantly the same gender: 1) have better communication and rapport between managers and workers; 2) have more broadly-skilled workers (who are proficient in multiple job positions); 3) exhibit the largest impacts on observed worker reallocation following the delivery platform implementation; and consequently, 4) realize more than three times the sales gains. Stores with female managers and predominantly male workers suffer less from gender mismatch, consistent with female managers being generally more aware of and responsive to workers' scheduling constraints.

Keywords: managers, productivity, workplace relationships, rapport, gender, people management, labor allocation, staffing, scheduling, quick-service restaurants, Colombia
JEL: J16, J24, M54

*University of California San Diego, NBER, J-PAL, BREAD, Good Business Lab; aadhvaryu@ucsd.edu. Adhvaryu serves on the board of directors and is an executive of Good Business Lab, a non-profit research organization which conducts research related to the topic of this article and provided support for this research. He does not receive monetary compensation for this role.

†University of Michigan; parhow@umich.edu

‡University of Michigan, NBER, J-PAL, BREAD, Good Business Lab; nyshadha@umich.edu. Nyshadham serves on the board of directors and is an executive of Good Business Lab, a non-profit research organization which conducts research related to the topic of this article and provided support for this research. He does not receive monetary compensation for this role.

§Harvard University, Harvard Business School, Digital Reskilling Lab; jtamayo@hbs.edu

1 Introduction

Rapport among workers shapes workplace performance (Ashraf and Bandiera, 2018). For example, productivity is affected by knowledge flows between co-workers (Atkin et al., 2023; Sandvik et al., 2020), as well as by the degree of interpersonal interaction with peers in nearby firms (Atkin et al., 2022). However, establishing these ties is not trivial (Gibbons and Henderson, 2012). Managers' ability to establish and maintain relationships with their workers contributes to the strong impact managers have on workplace outcomes like worker retention and productivity (Adhvaryu et al., 2023; Hoffman and Tadelis, 2021).[1]

Workplace rapport between supervisors and subordinates can impact knowledge transfers and the advancement of workers, and frictions in establishing this rapport such as language barriers and gender mismatch can hinder the development of these relationships (Cullen and Perez-Truglia, 2023; Guillouet et al., 2021). In particular, the degree to which co-workers identify with each other and the ease with which they communicate and cooperate can impact relationships both among peers as well as across the hierarchy (Akerlof and Kranton, 2000, 2005). For example, co-workers' ethnic identities can affect their willingness to cooperate with each other and, in turn, their joint productivity (Hjort, 2014). Identity-based distance – including age, gender, education, and experience – can also impede the degree to which fellow managers of parallel teams cooperate to help insure each other against shocks (Adhvaryu et al., 2021a).

Though gender is perhaps the most often studied dimension of identity among workers (Albanesi et al., 2023; Goldin and Katz, 2018; Goldin and Mitchell, 2017; Olivetti and Petrongolo, 2016, 2017), there is still much to learn about how shared or mismatched gender identity between managers and their workers might affect managers' abilities to elicit cooperation.[2] Using rich personnel records and granular productivity data from the universe of fast food restaurants of a leading chain in Colombia, we study the degree to

[1] An interesting related study of a training intervention which improves leader-subordinate relationships finds impacts on retention of the leader rather than the subordinate (Alan et al., 2023).

[2] A few recent studies have investigated related topics. Bagues et al. (2017) find that female scientific evaluators are no more favorable to female candidates for professorship positions, but that male evaluators are less favorable to female candidates when females are present on the scientific committee. Ronchi and Smith (2021) show that male managers who experience the birth of a daughter rather than a son begin to prioritize hiring women of similar quality over men, but there appears to be no performance impact of this change. Perhaps most relatedly, Buchmann et al. (2023) find that male managers are less willing to hire women for night shift jobs when they believe safe transport is not available; and Karpowitz et al. (2023) find that female majority and female leadership both improve female influence in teams. To our knowledge, ours is the first paper to study how managers' gender match with workers impacts managerial efficacy and team performance.

which mismatched gender identity between managers and workers affects managers' ability to successfully perform one of their primary responsibilities and a critical determinant of store performance: the allocation of workers to shifts to meet variable demand.

In this setting, a manager's time and attention are largely devoted to training workers in subsets of the numerous stations which need to be covered in each shift, and then allocating workers to shifts to meet variable demand. Scheduling workers is particularly complicated in the quick-service restaurant context. Demand is highly variable across hours of the day and days of the week. The product mix demanded also shifts within and across days resulting in changes in required composition of skills for different stations of the production process.

Workers must be trained and certified for a given station before they can be assigned to it, and this process involves a substantial amount of one-on-one face-time between the manager and the worker. In our sample, a given worker is trained and certified on only 4-5 of the 37 possible stations on average. As a result, managers' training and scheduling responsibilities are quite interrelated. When scheduling workers to shifts, the manager must take into account on which stations each worker is certified, how many workers are needed on each station given predicted demand for different products, and the workers' own idiosyncratic availability and/or preferences across shifts. The confluence of these constraints makes staffing an extremely complex and time-consuming process, especially in the presence of demand shocks which result in the need for changes.

The number of workers trained in each skill thus constrains the set of possible allocations of workers to shifts. Given that the training and certification process is personally costly to both workers and managers, we hypothesize that managers who have better relationships with their workers will invest more time and effort in training, and that their workers will be more willing to be trained in more stations. In line with this conjecture, in a comprehensive survey we administered of management styles and practices among the universe of current managers of all stores in Colombia, managers for whom their workers predominantly share their gender identity are 44% more likely to prefer having employees who are trained across more stations over specialized workers, and report that their workers are on average trained on 50-80% more stations, as compared to managers of stores in which predominantly male managers supervise predominantly female workers. This lower cost of training, and greater investment in training as a result, allows the (re)allocations of workers across shifts to be less constrained.

A large number of studies have documented that female workers make different labor decisions and/or have different preferences than their male counterparts in many contexts

(Goldin, 2015; Goldin and Katz, 2011; Haegele, 2021; Mas and Pallais, 2017). In particular, women often prefer to work different hours than men due to competing demands on their time at home (Ashraf et al., 2022) or safety concerns (Buchmann et al., 2023).[3] They also have higher valuations for control over which shifts they work and greater disutility from having their hours changed (Mas and Pallais, 2017); they may also have lower levels of baseline trust in others (Alesina and La Ferrara, 2002). Accordingly, we hypothesize that managers who share a common gender identity with their workers will be more likely to understand these types of preferences among their workers and, as a result, be more successful at adjusting worker shift scheduling to meet changes in demand than will gender mismatched managers. Consistent with this notion, we find that managers of gender-balanced stores are 59% more likely to report scheduling as the most important of their responsibilities for store performance. In addition, female managers may be generally more aware of or sympathetic to these constraints and preferences than are male managers. Indeed, female managers are 13% more likely to both report spending more time on scheduling than any other task and report assigning workers preferred shifts to prevent turnover.

We leverage the geographic expansion of a leading delivery platform across the country to study how well managers are able to adjust worker staffing to match resulting non-uniform increases in demand. The expansion was governed by the delivery app firm and therefore plausibly exogenous to the operations of each QSR store in our data.[4] Using event study models and two-way fixed effect (TWFE) differencing strategies, we find that stores realize a 5.62% increase in sales on average as a result of the arrival of the delivery platform to their neighborhood. We also find that stores do not on average hire more workers to meet this demand, consistent with the difficulty in hiring and training new workers for just a few hours a day and to cover a broad and variable set of stations.

Rather, stores tend to maintain a larger than necessary workforce spanning different station certifications and preferences across shifts, given the aforementioned complexity arising from the need to cover a broad set of stations at different times of the day (and frequent changes due to absenteeism and turnover). These extra workers still need to

[3]Related studies on school-aged girls have documented similar safety concerns (Fiala et al., 2022; Muralidharan and Prakash, 2017) and studies of educational outcomes of children of working mothers have demonstrated adverse impacts of reduced parental time at home (see, e.g., Løken et al. (2018)).

[4]The restaurant chain agreed to join the platform whenever it became available in each area, but the timing of that availability was governed by the spatial expansion of the delivery app, with that expansion staggered across cities and to some extent in large cities across neighborhoods with recruitment of drivers and uploading of restaurant details as the logistical constraint determining the rate and scope of incremental expansion. A lack of pre-trends in event studies supports the validity of the identification.

be given a minimum number of hours per week to satisfy labor regulations, leading to overstaffing at lean times like breakfast shifts. Consistent with this, we find that in response to the increased demand after the arrival of the delivery platform, managers harvest workers from previously overstaffed shifts like breakfast and reallocate them to peak hours (both as temporary reschedules and as more permanent changes to modal shift assignments of workers), often cutting the shifts into smaller segments. Specifically, in the 18 weeks after a restaurant partnered with the delivery app, managers rescheduled workers' shifts 21.18% more often and changed the modal shift of workers 40.29% more often as compared to stores which did not yet have access to the platform.

However, the effects on sales are highly heterogeneous. Gender-balanced stores achieved 3.5 times the sales gains of stores with predominantly male managers and female workers (7.27% vs. 2.08%). Gender-balanced stores also gain more in sales than do stores with predominantly female managers and male workers, but less dramatically so (21% larger sales increase). Patterns in survey responses indicate that gender-balanced stores have better communication and rapport between managers and workers, and that managers in these stores are more collaborative with workers in their style and practices. For example, managers of gender-balanced stores are 33% less likely to report acting without consulting workers, 80% more likely to report discussing key performance indicators (KPIs) with workers very frequently, and 52% more likely to report looking out for the welfare of their employees as compared to stores in which predominantly men manage women. Stores with predominantly women managing men also exhibit better communication and rapport than do stores in which predominantly men manage women.

Consistent with these better relationships, the greater investment in training, and the more broadly skilled workers that result, gender-balanced stores achieve 62% more permanent changes in the modal shifts of workers in response to the platform implementation; while stores in which predominantly men manage women no such impact is achieved. Similar patterns are reflected in temporary rescheduling of workers, with gender-balanced stores achieving a 2.3 times larger impact than stores in which predominantly men manage women. Taken together, we argue that the better relationships which prevail among gender-balanced stores, and to a lesser degree in women-managing-men stores, is a primary reason for the observed pattern of realized gains in sales following the introduction of the delivery platform.

1.1 Related Literature and Contributions

Our results build on the set of studies documenting the impacts of the quality of workplace interpersonal interactions and relationships on organizational performance (Ashraf and Bandiera, 2018; Atkin et al., 2023; Gibbons and Henderson, 2012; Sandvik et al., 2020). More specifically, our work relates to studies showing that (mis)matched identities among co-workers affect cooperation and, in turn, productivity (Adhvaryu et al., 2021a; Hjort, 2014). While most of these studies have focused on matching of identities among peer coworkers, we add to recent evidence from a study of mismatch between managers' attitudes and their subordinates' identities (Glover et al., 2017), a study of gendered manager-worker social interactions in the workplace (Cullen and Perez-Truglia, 2023), and a related study of language barriers across the management hierarchy (Guillouet et al., 2021).

While this prior work studies impacts on outcomes for the lower-level worker (e.g., knowledge, productivity, absenteeism, promotion), our study is the first to document effects on the manager's ability to perform their responsibilities which rely heavily on these interpersonal relationships. On a related note, another recent study documents that workers' willingness to cooperate can determine the degree to which a firm can realize value from the adoption of new technologies (Atkin et al., 2017). We show that common gender identity between managers and workers is a predictor of this willingness to cooperate and, therefore, can be an important mediator of the impact of new technology adoption.

In this sense, our findings also relate to the large and growing literature on the importance of managerial quality in determining team performance. In particular, we build on recent empirical results showing the value of personnel management skills (Adhvaryu et al., 2021b, 2022; Frederiksen et al., 2020; Hoffman and Tadelis, 2021). Recent studies have shown that the effectiveness of some practices may vary across workers and teams (Friebel et al., 2023), while others have argued that the identity of the manager may matter in addition to the practices they use (Metcalfe et al., 2023). Our results provide evidence of one way in which these mechanisms and features interact. That is, the gender identity of the manager may affect their ability to successfully execute people management tasks, but heterogeneously by the identity of the workers they manage.

Finally, our results add to the large empirical literature documenting gender differences in work-related preferences across skill-levels and incomes (see, e.g., Mas and Pallais (2017)).[5] Several studies have established the role of competing domestic demands on time

[5]Mas and Pallais (2017) find in a discrete choice experiment for call center jobs that women are more

and safety concerns in the labor decisions of women (Ashraf et al., 2022; Buchmann et al., 2023), particularly in developing country contexts. We build on these results in showing that managers who identify with the same gender as their workers are better able to navigate these preferences and constraints when scheduling workers to meet varying workloads.

2 Background and Context

2.1 Quick Service Restaurant (QSR) Industry in Latin America

Latin America and the Caribbean is the fourth-largest global region in consumer food industry sales, with a population of 646 million and sales over $243 billion (Reinhardt et al., 2021). A growing middle class, rising disposable incomes, surging urban populations, and an increasing female labor force participation rate have combined to create a thriving environment for convenience food (Smith et al., 2014).

Large quick service restaurant (QSR) chains in Latin America have also contributed to the steady incorporation of fast food into urban life. Over time, these chains have added healthier and more sustainable options in an attempt to broaden their consumer base and as the number of delivery platforms (e.g., Uber Eats, Rappi, Glovo, iFood) in Latin America quadrupled between 2014 and 2019 (Euromonitor, 2019), these QSR chains took advantage of the popularity of these platforms to add or augment their delivery offerings. This additional sales channel enabled these chains to widen their appeal in suburban areas (Reinhardt et al., 2021).

2.2 Firm Partner

Our firm partner, one of the leading QSR franchise operators in Latin America, owns and operates nearly 2000 stores across 20 countries in Latin America and the Caribbean, which employ more than 90,000 workers. Both because these stores are franchise units of a leading global brand and because all of these stores are fully owned and operated by the same firm (a few hundred more stores are sub-franchised to other operators), the operations are quite standardized across the region and any findings of analyses like those we undertake below are fairly generalizable to the broader portfolio of stores. Indeed, the

likely than men to select flexible work arrangements and have higher valuations for avoiding irregular work schedules, and Wiswall and Zafar (2018) find the same among high-ability undergraduates attending a highly selective university.

firm has expressed particular interest in our research to inform the (re)adoption of delivery platforms across their operations.

The firm operated more than 60 restaurants across 9 cities in Colombia during the span of our data. It averaged monthly sales of nearly 3 million units (items) sold through approximately 400,000 transactions–serving an average of 50,000 customers each day. Given the franchise structure, each branch's data systems are regulated by the global QSR brand which helps ensure that reporting of the performance and training indicators we rely on is standard across stores.

Equipment, suppliers, menus, pricing, and training curricula are also standardized by the global brand, but all hiring and staffing is handled by store managers. Employees at each branch receive a base salary comparable to what they could earn elsewhere within the Colombian QSR industry, but they also have the opportunity to receive short-term monetary incentives contingent upon individual and branch performance. This pay structure ensures that both managers and crew members care about the performance of the store.

2.3 Delivery Platform

The proliferation of delivery service platforms in recent years made it feasible for Colombian companies like our partner to successfully add a delivery sales channel. In fact, Colombia is a global leader for the percentage of restaurant sales driven by online channels. In 2018, the proportion of restaurant sales made online was 31% in Colombia, compared to 8% in Latin America and 11% globally (Euromonitor, 2019).

Our partner implemented online ordering via a delivery app after a third-party delivery company approached corporate leadership and brokered a deal to be the sole delivery service for all restaurant locations in Colombia as it expanded across the country. This partnership was established before the delivery platform was widely available in Colombia. Accordingly, after the partnership was established, the delivery company spread its coverage in stages across cities (and in large cities across neighborhoods) by hiring and training drivers, recruiting other restaurants, uploading their menus and prices, etc. Our partner firm's stores joined the delivery platform as soon as the service was available in their area. Accordingly, the roll-out of the app across store locations mirrored the expansion of the third-party food delivery platform across cities in Colombia. That is, store managers had no control over the timing of service availability, indicating that the roll-out was exogenous to the stores' operations.

Figure 1a shows the order in which restaurants implemented the delivery service as the app's reach expanded to their cities. All our partner's stores in areas where the app was available partnered with the third-party delivery service. Accordingly, both whether a store implemented and when it implemented are plausibly exogenous.[6] Since each store implemented the app delivery service at different times, we can measure the app's impact on different store performance indicators before and after the implementation. Figure 1b shows the month in which each store implemented the delivery service. The roll-out was spread across eight consecutive months in 2018 and 2019. The number of implementations peaked in January 2019, with the last stores implementing in July 2019.[7] Table A.1 in the Appendix presents balance checks between early and late implementations of the delivery app, and shows no significant differences in store performance, scheduling, or training measures.

2.4 Managers' Responsibilities

Each branch has a store manager and several shift managers. Some stores have station managers, assistant managers, or even second assistant managers. While it can be useful in some contexts to acknowledge differences in the duties of each type of manager, the responsibilities we study in this paper are shared across most or all levels of management. Consequently, we refer to the set of all managers in each store in our sample as a unit.

Managers are trained in most or all of a store's stations, but generally spend their time supporting crew members instead of preparing food. Keeping a restaurant running efficiently requires significant preparation. During slower hours, managers spend considerable amounts of time scheduling and rescheduling workers as well as preparing and implementing trainings. Appropriate scheduling is crucial to a restaurant's success since too few workers results in lost revenue and frustration (for both customers and crew), while too many workers results in inflated labor costs.

Scheduling workers is particularly complicated in the quick service restaurant context due to highly variable demand across hours of the day and days of the week. Composition of products demanded also shifts within and across days, resulting in changes to the required

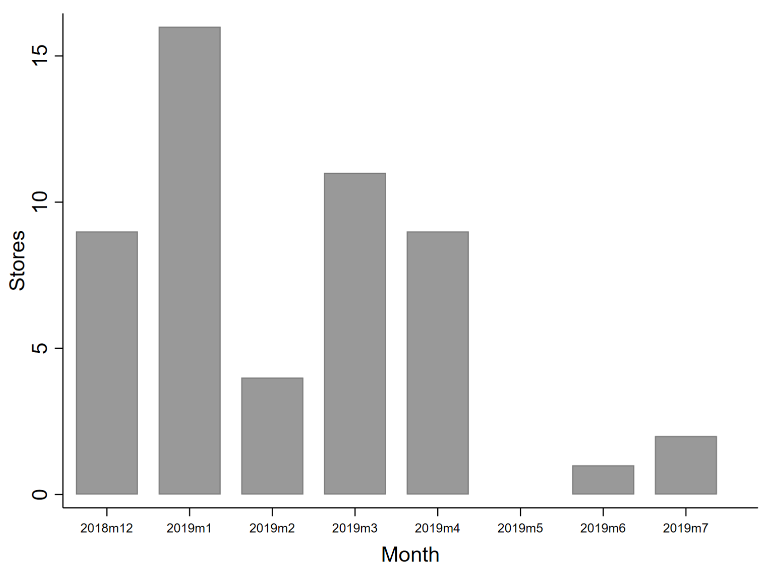
[6] Our main results use only not yet treated stores as controls, but we also explore using never treated stores for robustness.

[7] We note that our partner did have a delivery sales channel before partnering with a third-party delivery service. Despite early investment to develop an in-house delivery application, our partner's efforts were unsuccessful, and delivery sales did not comprise a significant proportion of sales until the third-party delivery app was implemented. Almost immediately after implementation, more than 80% of delivery sales came from the third-party app. The data we use for our analysis uses all store sales both in-store and delivery, with delivery including those via the third-party app as well as the partner firm's own service.

Figure 1: Delivery App Implementation

(a) Panel A: App Implementation by City in Colombia

(b) Panel B: Implementation of The Delivery Service by Month



Notes: Panel A of Figure 1 shows the month in which the first store(s) within each major Colombian city implemented the third-party delivery app. Panel B shows the month in which each store in our sample implemented the delivery service. Table A.1 shows there is no difference in observables between stores that implement the delivery app early (before March 2019) or later (after March 2019).

composition of skills for different stations of the production process. However, workers must be trained and certified for a given station before they can be assigned to it. The process of training and certifying a worker on a particular station is personally costly to both the manager and the worker, involving a substantial amount of one-on-one face-time between the two. Accordingly, on average in our sample, a given worker is only trained and certified on roughly 4-5 of the 37 stations.

Note that managers' training and scheduling responsibilities are connected. When scheduling workers to shifts, the manager must take into account on which stations each worker is certified to work, how many workers are needed on each station given predicted demand for different products, and the workers' own idiosyncratic availabilities or preferences across shifts. The confluence of these constraints makes staffing an extremely complex and time-consuming process, especially in the presence of shocks which result in the need for changes. Of course, how many workers are trained in each skill will impact how constrained are the possible allocations of workers to shifts, and therefore, how difficult is the task of scheduling. Accordingly, each manager will choose how much time and effort to devote to training her workers in different stations, trading off the gains in scheduling flexibility with the personal costs of conducting the training.

2.5 Colombian Labor Regulations

Labor laws governing the minimum and maximum number of hours an employee can work greatly affect the way managers in our sample schedule employees. Work completed in excess of 48 hours per week[8] is considered "extra time," and employers must pay 125% of the regular wage, or 175% of the regular wage if the extra hour is worked after 9 p.m., on Sunday, or another holy day.[9] Moreover, labor laws cap overtime at a maximum of 12 extra hours per week or two hours per day, meaning that even if a manager is willing to pay overtime to cover labor shortages, they can cover a maximum of two extra hours per worker per day.

Perhaps, the most significant constraints on a manager's ability to optimize scheduling are the minimum shift length and minimum hours per day requirements. No shift may be shorter than 5 hours and employees cannot work multiple shifts in a day, and full-time

[8]In 2021 the maximum number of regular hours an employee could work per week was updated to 42, but during the period of analysis, 48 hours per week was considered the maximum number of regular hours an employee could work per week.

[9]Extra hours on holidays not considered holy days are paid at 125% of the regular wage.

workers must be given at least 4 hours per day.[10] As a result, the laws governing maximum shift length restrict a manager's ability to ask employees to stay late or come early to cover variable demand, since such a request either requires higher wages or is illegal (if it gives an employee more than 2 extra hours in a day or 12 extra hours in a week). However, managers' inability to schedule shifts shorter than five hours and requirement to give each full-time worker at least four hours a day, taken together, mean that adequately staffing a store during peak hours frequently results in chronic overstaffing during nonpeak hours.

3 Data

3.1 Performance Data

The dataset we use to study performance contains granular transaction-level data. It covers 88 million transactions for 62 restaurants in Colombia from July 2018 to October 2019.[11] Each transaction includes details such as the point of sale in which the ticket (order) was processed (e.g., in-store or delivery), all items on each ticket, and their prices. This allows us to use sales, units sold, and number of tickets as our primary measures of performance. Table 1 presents summary statistics for each of these measures; monthly averages at the store-shift level. During the pre-implementation period of analysis, stores sold an average of 18,706 units on 7,367 tickets for a total sales of $29,650 USD per monthly period. Throughout the analysis, we report sales results in thousands of USD, cumulative over monthly periods. However, the granular nature of the raw data also enables us to construct versions of these and other outcomes that leverage hourly and shift-level variation.

3.2 Personnel Data

We also have access to granular personnel data at the individual level for all restaurants that implemented the delivery app. The personnel data includes demographic characteristics (age, store, position, and gender) and date of hiring and termination for each worker and is available for all stores and employees from January 2019 to October 2019. Table 1 shows that approximately 65% of employees within a store are crew members. The other group of

[10]Substantial recruitment and training costs make maintaining a roster of a large number of part-time workers also inefficient.

[11]Our main results below exclude the never-treated stores (13 out of 76), as well as one additional store which had incomplete data over the observation window; so our main sample covers 62 stores.

Table 1: Pre-Implementation Summary Statistics for Performance and Personnel Measures

	Mean	Stand. dev.
Performance Measures		
Sales (thousands, USD)	29.65	29.44
Units sold	18,706	19,368
Number of tickets	7,367	7,889
Personnel and Scheduling		
Number of workers per store	18.86	8.74
Number of crew members per store	12.29	7.58
Number of workers working per hour	8.63	5.33
Number of hours per worker per week	28.74	6.37
Length of the shift	6.51	1.31
Number of shifts	19.71	6.10
Hiring	0.08	0.27
Turnover	0.014	0.12
Absenteeism	0.36	0.48
Reschedules	8.64	7.02
Gender balance		
Share of female managers	51.6%	0.26
Share of female crew members	54.1%	0.10
Training		
New skills	1.92	2.86
Refresher	0.73	1.69

Notes: Statistics presented in Table 1 for performance measures are monthly averages at the store-shift level from the pre-implementation period for the 62 stores that implemented the delivery app. Personnel and Scheduling, and Training results are monthly averages at the store-employee level from the pre-implementation period for the 62 stores that implemented the delivery app. We use sales data from July 2018 to October 2019 and training and personnel data from January 2019 to October 2019. Sales is presented in thousands of US dollars. Units sold is the number of items sold. Number of tickets is the number of separate orders fulfilled. Absenteeism refers to an employee failing to show up for a scheduled shift. Reschedule refers to any changes logged to an employee's schedule after the initial schedule has been set. New skills training refers to all post-onboarding trainings that teach a skill for which an employee has not yet been trained. Refresher trainings include all subsequent trainings after a new skill training.

store employees are managers of some sort (store manager, shift manager, assistant manager, station manager, or equipment manager). On average during the period of our analysis, stores hired 1.5 and lost 0.26 employees each month.

3.2.1 Manager-Worker Gender Balance

Table 1 also shows that the share of crew members in a store that are female is 54% on average and the share of managers who are female is 52%. The gender identity of both managers and workers in each store feature heavily in our hypotheses and empirical analyses below. We take the gender of all managers and workers in the store from the personnel records described above and construct measures of gender "balance" at the store level. We differentiate between imbalanced stores in which the share of female managers exceeds the share of female crew by at least 25% (denoted as "Female-heavy Management" or "FHM" in graphs and tables) and stores in which the share of male managers exceeds the share of male crew by at least 25% (denoted as "Male-heavy Management" or "MHM" in graphs and tables).[12] For example, a store with 70% female managers but 20% female crew would be a female-heavy management (FHM) store while a store with 20% female managers and 70% female crew would be a male-heavy management (MHM) store. We assign stores to one of these categories using the store's personnel records from before the store gained access to the delivery platform. We note that the manager-worker gender balance of the store is quite stable over the sample period (i.e., the within-store correlation in gender-balance over time is 0.89).[13]

3.2.2 Training

The training data is available during the same period as the personnel data (January 2019 to October 2019). The dataset includes every training for every worker and specifies when the training occurred. This lets us see the days and times of day when managers choose to

[12] Justification for the 25% threshold and for the symmetric functional form in gender-balance between managers and crew is provided in B.1.1. We replicate the main findings in this section using several different gender-balance thresholds which we present in B.1.2. We also demonstrate in Figure B.2 that treatment effects are quite similar for balanced stores in which female managers and crew dominate, in which male managers and crew dominate, and in which both managers and crew are split equally across male and female, justifying the symmetrical gender-balance functional form we choose in our preferred specifications.

[13] In Figure B.3, we demonstrate that the main pattern of heterogeneity in treatment effects presented below is preserved if we use variation in gender-balance within stores over time for identification. Our preferred results utilize gender-balance in the pre-implementation period given the persistence over time and the potential for changes over the app implementation timeline in gender-balance to be endogenous.

conduct these trainings. Because crew members cannot work at a station for which they are not trained, we are also able to see the specific stations at which a crew member is certified to work at any moment.[14] The number of certifications an employee has or does not have greatly affects a manager's ability to schedule that employee, since every shift should have employees certified to work each station. Some stations require more training than others but all trainings are time-consuming, so managers need to be judicious when deciding when to train employees on new stations versus when to schedule already certified employees to cover demand (see Appendix Table A.2 which presents the average weekly certifications by station). Table 1 provides summary statistics for the number of new skill and refresher trainings conducted in each of the stores in our sample. During the period of analysis, the average employee completes 1.92 new skill trainings and 0.73 refresher trainings per month.

3.2.3 Scheduling

The scheduling data is available at individual-shift level from June 2018 to October 2019. This dataset includes information on the shifts worked by every worker in each of the restaurants our partner operated in Colombia during this time. Importantly, we can see the exact time every shift was scheduled to start, whether the manager changed the start or end time of a shift, whether a worker showed up for their shift, as well as how many employees are working at any point throughout the day. Table 1 provides summary statistics on the number of shifts and hours worked, and number of reschedules at each store during a monthly period. The average store had 19 workers who worked approximately 6.51 hours per shift for an average of 28.74 hours a week. The employees in our sample averaged 8.64 reschedules during a monthly period. This regular adjustment is in part in response to 6.8 workers being absent, 0.26 workers quitting, and 1.5 new workers being hired every month.[15]

Effective scheduling of crew members is an important feature of managerial quality, a primary responsibility of the managers in our sample, and one of the most important mechanisms managers used to respond to the implementation of the delivery app (as we document below). Managers prepare weekly schedules for all crew members but can ask to reschedule shifts based on unexpected changes in demand or availability of other crew

[14] These are complete records for any employee who joined after the start of our observation window, but are truncated for pre-existing workers. To analyze a more complete record of average certifications of all store employees we use a survey measure obtained from the store managers.

[15] These figures are computed using values of Table 1. Number of workers absent is computed as 18.86*0.36. Number of workers quitting is computed as 18.86*0.014. Number of hired workers is computed as 18.86*0.08.

members. Despite having some flexibility to change weekly schedules, a manager faces several constraints.

First, Colombian labor laws (see Section 2.5) prevent managers from keeping just a small set of workers and overworking them. These same laws make it illegal for managers to ask employees to work shifts shorter than five hours or to send employees home early when sales are slow. While these laws protect employers and workers, they also contribute to frustrations for both managers and crew members. Managers may have difficulty deciding how many employees to schedule at a given hour to meet demand, and crew members may be asked to change their schedules on short notice if managers have predicted demand and/or worker absenteeism incorrectly.

Second, the firm's own procedures prohibit managers from scheduling employees to work stations on which they are not yet certified. On average, crew members are certified to work only 4–5 stations even though stores has as many as 37 different stations. This means that even supposing that each worker could handle several of a store's stations at one time, covering 25 stations would still require at least 5 workers if each worker were certified in exactly 5 mutually exclusive stations.

Third, managers are dealing with real people who have variable schedules and idiosyncratic preferences that affect the hours they are willing and able to work. Younger employees may be enrolled in university courses during the day, parents may need to take children to or from school, and women may prefer to be home before dark due to safety concerns. Figure 2a, for example, shows that the share of female employees starting work at a given hour drops markedly after the sun goes down.

Finally, non-uniform fluctuations in demand for specific products require some stations to be doubled or tripled up at some times of the day. While managers and employees across many sectors face these issues, they are particularly intense in the QSR industry where demand is so high during certain hours of the day (afternoon and dinner time) and so low at other times (breakfast or graveyard hours). Accordingly, we see employee schedules start and end at different times throughout the day, leaving a variable number of employees working during any one hour of the day. Taken together, the manager must then both maintain and dynamically allocate – subject to workers' idiosyncratic personal schedules and preferences and in the presence of worker absenteeism and turnover – enough workers certified on each station to meet not only modal demand across stations but also peak demand for some stations at some times.

3.3 Survey of Managerial Style and Practices

We surveyed the universe of current managers of all stores in Colombia in March 2023 on their management styles and practices, including how they approach their responsibilities and decisions regarding training and scheduling. Given the amount of time which has elapsed between the roll-out of the delivery platform we study and the fielding of this survey, some managers have left the firm in the interim and others have been promoted to higher levels of management. We keep only the survey responses from the managers who match to our study sample. This matched sample includes 117 managers across 58 of the 62 stores in the analysis sample. The 4 unmatched stores are no longer managed by the same managers who oversaw the implementation of the delivery platform.[16] Table A.3 in the Appendix presents summary statistics of survey measures across stores by the gender of the manager.

We hypothesize that stores that are more gender-balanced will exhibit better communication and rapport between managers and workers and more collaborative or inclusive management practices. Accordingly, we asked managers about different ways in which they interact with workers, for example: act without consulting their workers; treat employees as their equals and look out for the welfare of employees; whether they involve all team members in goal-setting and if they achieved the goals they set; and how often they discuss key performance indicators (KPIs) with their workers.

Given that the training and certification process is personally costly to both workers and managers and relies heavily on these worker-manager relationships, we hypothesize that managers who share a common gender identity with their workers will invest more time and effort in training and that their workers will be more willing to be trained in more stations. As a result, we would expect that workers in these stores would be more easily reallocated across stations and shifts.

Accordingly, we ask managers several questions regarding their training preferences and practices, for example: whether they prefer that workers are trained broadly or are specialized in particular stations; and how many stations on which the average worker

[16]The survey has a response rate of 86% among the targeted managers (senior managers working in 2023 in the company). The response rate by manager type is 91% for store managers, 80.25% for shift managers, and 94.3% for assistant managers. We note that the gender balance of the stores at the time of the delivery platform roll-out also may not correspond perfectly to that during the time of the survey. Although the manager-worker gender balance of the store is quite stable, the within-store correlation in gender balance over time is 0.89. Nevertheless, since any mismatch should simply attenuate any relationships between survey measures of management practices and the gender composition of the stores, we still analyze these data and discuss patterns here and below.

in their team is trained.[17] Figure 3a plots some preliminary comparisons showing that managers of gender-balanced stores are 44% more likely to prefer having workers who they can shift across stations.

Figure 2b documents that the ratio of employees to sales is significantly higher during breakfast hours than any other time of the day, suggesting that stores are chronically overstaffed in the morning relative to other times of the day. This pattern suggests that managers may be able to respond to the delivery app implementation by harvesting workers from less busy hours and scheduling them to work during hours in which the new delivery demand is concentrated.[18] This reallocation of workers may be reflected in last-minute reschedules or in permanent changes to the modal shift to which a worker is assigned, so we study both in our analysis.

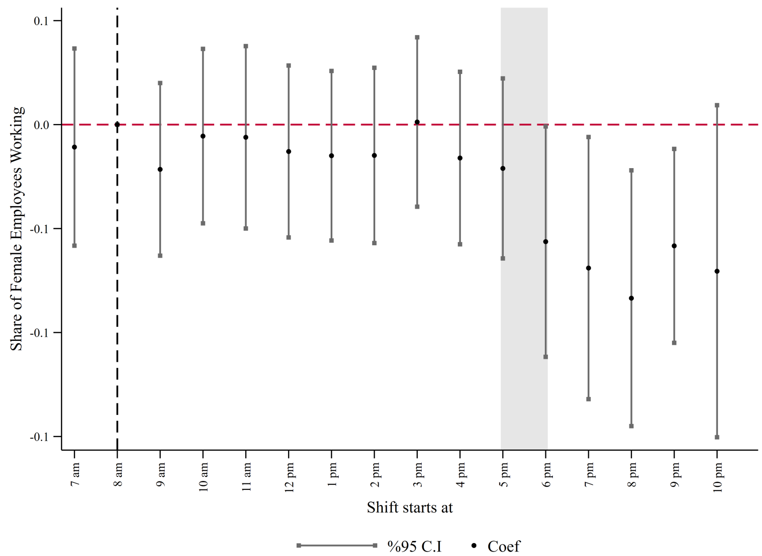
Finally, previous studies have documented that women often prefer to work different hours than men due to competing demands on their time at home (Ashraf et al., 2022) or safety concerns (Jaitman, 2020; Kondylis et al., 2020; Martinez et al., 2018); and that they have higher valuations for control over which shifts they work and greater disutility from having their hours changed (Mas and Pallais, 2017). We hypothesize that managers who share a common gender identity with their workers will be more likely to understand these types of preferences among their workers and, as a result, be more successful at adjusting worker shift scheduling to meet changes in demand than will gender mismatched managers. Accordingly, we ask managers about scheduling, for example: how important is scheduling for store performance. Figure 3b plots some preliminary comparisons showing that managers of gender-balanced stores are 59% more likely to list scheduling as the most important of their responsibilities for store performance than are managers of MHM stores. Figure A.2 in the Appendix shows that managers who rank scheduling as their most important task in the survey utilize worker shift reschedules and modal shift changes per more intensively as recorded in the administrative records data. This pattern is stronger among managers in FHM and Balanced stores relative to those in MHM stores, and helps to validate the informativeness and accuracy of the survey measures.

[17] We collected this to complement the administrative training records we have, since any trainings that workers complete prior to the start of our observation window will not be counted toward any cumulative training certifications measure we calculate using these records.

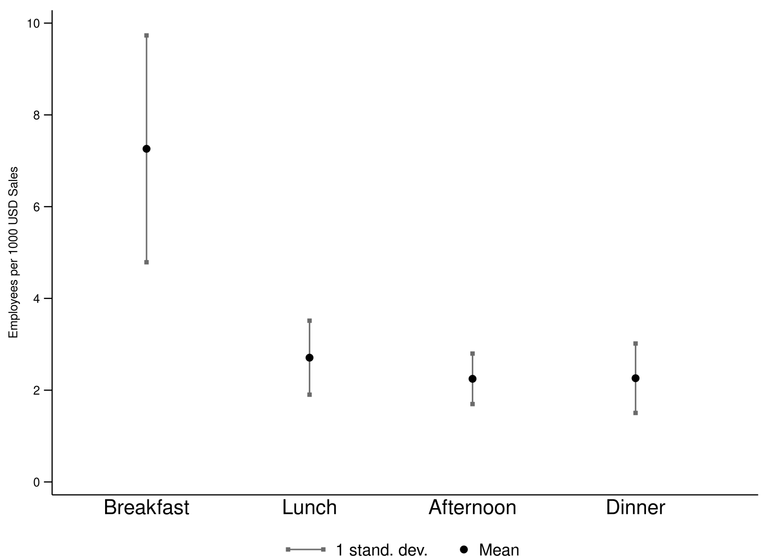
[18] Figure A.1 in the Appendix confirms that impacts on sales are not uniform throughout the day.

Figure 2: Schedule Management

(a) Share of Female Employees Working by Hour



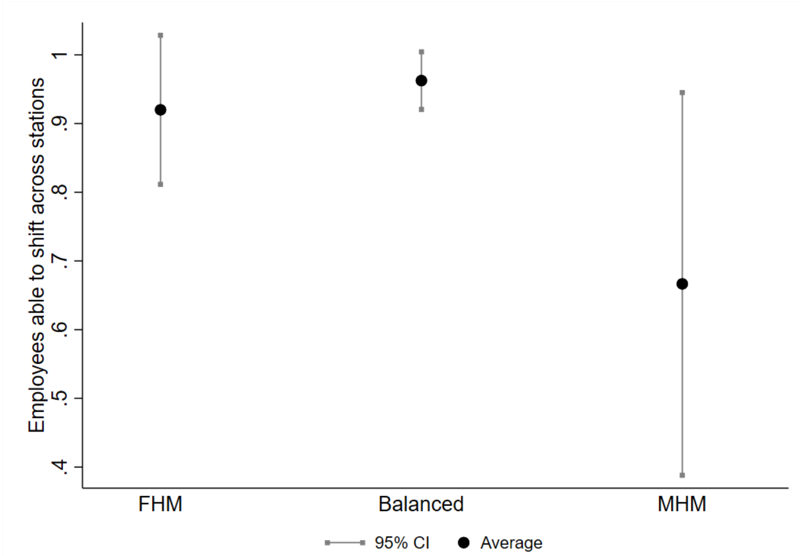
(b) Employees per $1000 of Sales by Timeband



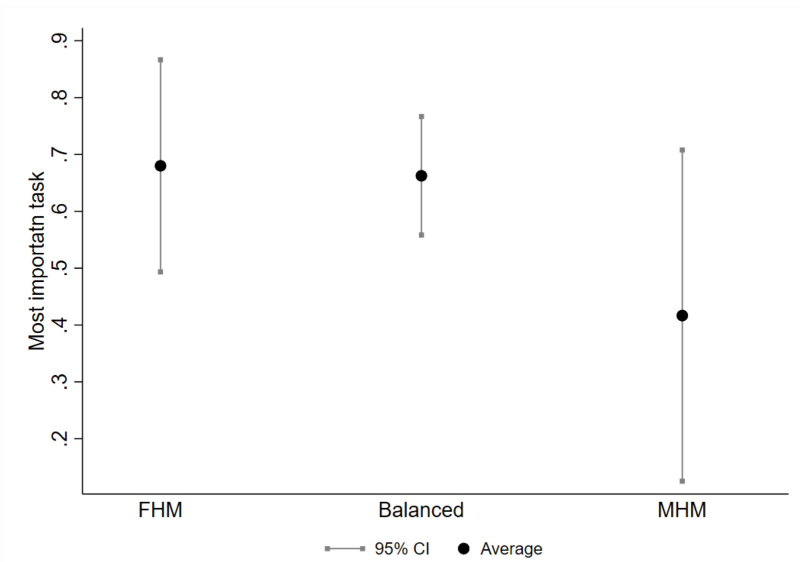
Notes: Figure 2a presents the difference of the percent share of female employees whose shifts start during each hour of the day respect the percent share of female employees at 8 am. We present results from 7 am to 10 pm because these are the hours during which a majority of stores are open, but we note that some stores are open during late-night/early-morning hours. The sample consists of pre-implementation data for the 62 stores that implemented the app between July 2018 and October 2019. The vertical gray bar corresponds to the time that the sun sets in Bogotá. The earliest sunset during the year occurs at 5:39 pm and the latest at 6:10 pm. Figure 2b shows the ratio of employees to $1000 USD of sales. The start and end times for each timeband were set by the partner firm, so we simply adopt their nomenclature. Breakfast includes the hours from 7:00 am to 10:59 am, lunch from 11:00 am to 1:59 pm, afternoon from 2:00 pm to 5:59 pm, and dinner from 6:00 pm until 11:59 pm.

Figure 3: Survey Questions

(a) Better to Have Workers Who Can Shift Across Stations



(b) Scheduling Most Important Task for Store Performance



Notes: Figure 3a shows the share of managers who report that it is better to have employees able to shift across stations by store gender balance i.e., female-heavy management (FHM), gender-balanced, or male-heavy management (MHM). The difference between MHM and gender-balanced stores is significant at 95% (p-value 0.026). Figure 3b shows the share of managers that list scheduling shifts as the most important task by gender balance of the store. The difference between MHM and gender-balanced stores is significant at 90% (p-value 0.097).

Note that in these preliminary comparisons, managers of FHM stores look more like gender-balanced stores than MHM stores. This likely reflects that female managers may be in general more aware of or sympathetic to scheduling constraints and preferences than are male managers. Consistent with this notion, we find that female managers are 13% more likely to both report spending more time on scheduling than any other task and report assigning workers preferred shifts to prevent turnover. Taken together, these patterns in survey responses validate the idea that both the gender of the manager and the gender match between managers and workers might determine the manager's approach to and success in reallocating workers across shifts in response to the introduction of the delivery platform. Accordingly, as discussed above, we continue to separate the gender imbalanced stores into FHM and MHM to allow for detecting both of these patterns.

4 Empirical Analysis

First, we aim to compare different store performance indicators before and after the introduction of the online food ordering and delivery platform across restaurants. We leverage the expansion of the food delivery platform to almost all the cities where our partner firm restaurants are located during 2019. Thus, our identification strategy relies on spatiotemporal variation in platform roll-out at the municipality level, which is plausibly exogenous to the decision-making and performance of stores and the larger firm. We present evidence in support of this below.

4.1 Baseline Specification

We estimate an event study model (Freyaldenhoven et al., 2021; Roth, 2019), which allows us to test for differential pre-trends across stores before the implementation, and to trace out the dynamic impacts over the post-introduction period. We estimate the following baseline specification:

$$Y_{s,t} = \alpha_0 + \sum_{\underline{C} \leq k \leq \overline{C}, k \neq -1} D_{st}^k \delta_k + \Phi_s + \theta_t + \varepsilon_{s,t} \tag{1}$$

where $Y_{s,t}$ is the performance measure of store-shift s at monthly period t, D_{st}^k is a relative time-shift indicator for whether the food delivery platform had introduced the app in (monthly) period $t-k$ where store s is located, defined as $D_{st}^k = 1[t = \tau_s + k]$ for $k \in (\underline{C}, \overline{C})$,

$D_{st}^{\underline{C}} = 1[t \leq \tau_s + \underline{C}]$, and $D_{st}^{\overline{C}} = 1[t \geq \tau_s + \overline{C}]$, where $1[\cdot]$ is the indicator function, k indexes the set of time indicator variables, and τ_s is the first monthly period when store-time band s get orders through the app delivery platform. The parameter δ_k measures the impact of the introduction of the platform before, during and after the event. We normalize $\delta_{-1} = 0$ and set $\underline{C} = -6$ and $\overline{C} = 6$.

We control for store-shift fixed effects Φ_s and time (monthly period) fixed effects θ_t, and cluster standard errors at the store-shift level for inference. Our parameters of interest are δ_k for $k \in [\underline{C}, \overline{C}]$. Our main specification excludes the never-treated stores (13 out of 76).

4.2 Identification

The panel data structure allows us to interpret the statistical significance of these coefficients as evidence of the causal relationship between the introduction of the food delivery platform and performance measures, provided that the roll-out of the platform varies over time and is uncorrelated with store-specific shocks that affect the performance measures (Blundell and Dias, 2009). Specifically, identification relies on two features of the roll-out of the food delivery platform. First, the introduction of the platform was unanticipated by the managers of the stores. Second, the introduction of the platform in a city is uncorrelated with manager- and worker-specific characteristics. Both assumptions are validated by the anecdotal evidence: a) the expansion of the food delivery platform across the country was executed by the platform, not our partner firm, and b) due to technological and operational limitations, the expansion of the app implementation was performed city by city. To the best of our knowledge the delivery platform did not announce much in advance which cities would be added to the platform nor when they would be added. Accordingly, managers of our partner firm could not have anticipated the app's roll-out. The lack of pre-trends in our analysis validates these assertions.

We showed in Figure 1a in Section 2, that there are no systematic geographic patterns in the order in which restaurants implemented the delivery service as the app's reach expanded to other cities. Moreover, Appendix Table A.1 presents the mean and standard deviation of different outcomes studied in the paper for the pre-implementation period for early (before March 2019) and late (after March 2019) implementing stores. The table shows no systematic pattern of significant differences between early and late implementing stores in productivity and local labor market measures, among many others, confirming the exogeneity of the roll-out of the platform. Given this context behind the app roll-out across Colombia and the demonstrated statistical balance between early and late implementing stores, we believe that stores yet to implement food delivery platform form a credible counterfactual for stores that implement the technology earlier, especially after accounting for time-invariant differences between store-shift and monthly period fixed effects.

4.3 Effects on main performance measures

We begin by implementing the baseline event-study specification (1). Panels (a) to (c) of Figure 4 present the event-study coefficients for sales, units sold, and number of tickets. The figures show no evidence of selection into the app implementation based on past store performance indicators; i.e., we do not see any evidence of divergent pre-trends in any of these outcomes. Rather, it is only after stores implement the new technology that we observe a positive effect on sales, units sold, and the number of tickets. In panels (a) to (c) of Figure 4 we observe some effects beginning during the same monthly period of the event and persisting even 6 monthly periods after the implementation of the food delivery platform.

Because the event studies show sharp onsets and persistent effects in the post period for each performance measure, we also estimate a difference-in-difference regression to recover a single average treatment effect estimate. We replace the full set of time-to-treatment and time-since-treatment indicators with a post treatment dummy (indicating before and after the app implementation). We control for store-shift and monthly period fixed effects, and cluster the standard errors at the store-shift level. Table 2 shows that the implementation of the food delivery platform increased overall sales at the store-shift-level by 5.62%, or $1,542 USD per monthly period from the pre-implementation mean; the total number of units sold by 7.20%; and the number of tickets by 6.66%. Note that the similar magnitude of effects across all of these performance measures indicate that the transactions via the delivery platform did not reflect a quantity or value of items that was systematically different from the average transactions prior to the introduction of the delivery platform.[19]

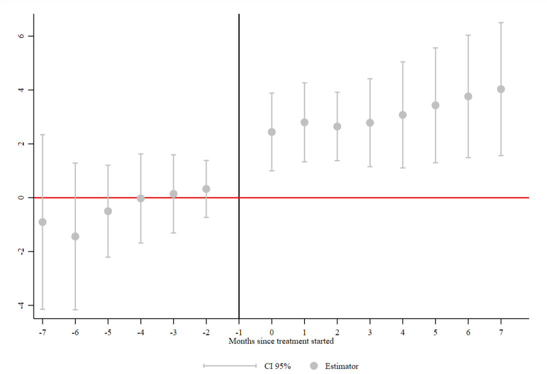
4.4 Managers' Response

We now explore the different ways that managers responded to the changes in demand due to the app's implementation. The franchising model of our implementing partner ensures

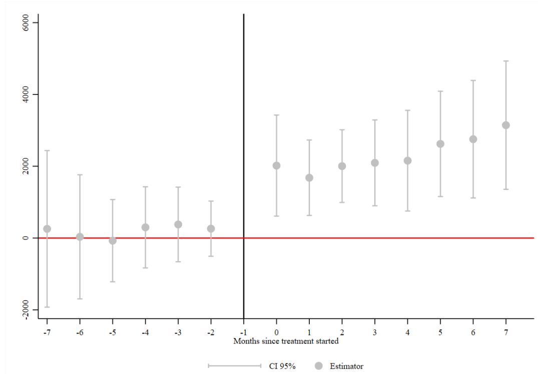
[19] We confirm this in additional results on items/ticket and avg. ticket value, but do not present these null results for the sake of brevity.

Figure 4: Effects of delivery app implementation on performance measures

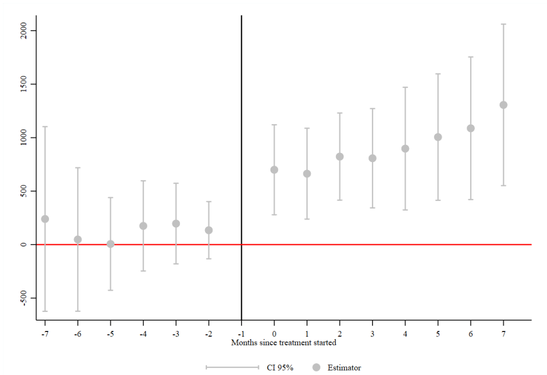
(a) Event study estimates on sales



(b) Event study estimates on units sold



(c) Event study estimates on number of tickets



Notes: Panels (a), (b) and (c) of Figure 4 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on the key performance measures. Treatment is defined as having implemented the food delivery app. Control stores are those not yet treated. The sample consists of store-shift-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Panel (a) shows the impact on sales, Panel (b) shows the impact on units sold and Panel (c) shows the impact on number of tickets (orders). The vertical line represents the time of the treatment.

that store and shift managers had a great deal of autonomy in the way they responded to the implementation of the delivery app. For example, in response to increased demand at different times of the day, managers were free to choose whether they would hire new staff, train the existing workers in more skills to cover more stations, refresh the training of existing workers to increase their productivity, or reallocate existing workers from non-peak to peak times. This variability in manager's responses is an important determinant of how well (or poorly) a given branch was able to take advantage of access to a new sales channel.

The delivery service's varied impacts on sales during different hours of the day makes a manager's response especially important. Figure A.3 in the Appendix shows mean sales

Table 2: Effects of delivery app implementation on performance measures

	(1) Total sales	(2) Units sold	(3) No. Tickets
Post app implementation	1.542***	1,245***	458.2***
	(0.571)	(463.6)	(172.0)
Mean of Dep. Var.	27.419	17,297.636	6,880.046
Relative effect	5.62%	7.20%	6.66%
Observations		4885	
Number of stores		62	

Notes: Table 2 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy. Treatment is defined as having implemented the food delivery app. The sample consists of store-shift-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Column 1 presents the estimations for total sales, Column 2 presents the estimations for units sold and Column 3 presents the estimations for the number of tickets. Numbers in parentheses are clustered standard errors at the store-shift level. * significant 10%, ** significant 5%, *** significant 1%

by time of day; while Figure A.1 shows treatment effects on sales by times of the day. We see that demand for delivery follows demand via existing points of sale. That is, the implementation of the delivery service amplified demand during peak hours, making the busiest hours even busier than before. However, the effects are concentrated over only a few hours of the day. Effects are largest in the hours of 5-7pm (around 20,000 USD in each of those hours over the bi-weekly period, an order of magnitude larger than the average treatment effect of 2.848). Effects are roughly half this peak magnitude in the hours of 2-4pm and 8pm, and essentially null for other times of the day.

This non-uniformity over the hours of the day creates an interesting dilemma: because managers cannot schedule workers for just a few hours at a time, scheduling more workers during peak hours frequently means having too many workers during non-peak hours. Thus, even if the delivery service greatly boosted sales, a corresponding increase in costs due to over-hiring and/or poor scheduling could offset this increase, resulting in small or even negative impacts on a store's profitability.[20] Moreover, if the impact of higher-stress

[20]We do not have data on profits of the stores, but can analyze inputs like labor hours in comparison to sales, as well as effects on waste. We find no effects on hours per worker nor number of workers in Table 3 and no evidence of effects on waste in Table A.4 in the Appendix, so we believe profits scale proportionally with sales. The firm anecdotally confirms that since inventory costs scale proportionally with sales and input waste is quite small compared to other costs this interpretation that labor costs would be the main reason that effects on profits would differ from those on sales.

peak hours or the unpredictability due to increased rescheduling impacts turnover and/or absenteeism, the way a manager responds becomes even more important.

We first consider the impact of the delivery app implementation on the quantity of hired labor. Column 1 of Table 3 explores effects on new hiring and shows no statistically significant impact. Though the coefficient indicates a fairly large 6.33% increase from the pre-implementation mean, this reflects more the small mean of hiring at the store-monthly level than a large coefficient. Column 2 shows that managers also did not respond by extracting more hours of work from the existing employees. That is, effects on hours per worker per week are small in magnitude (-2.43% of the mean) and insignificant.

Next, given the quantity of workers and hours available did not change, we ask whether increased demand during some hours necessitated that managers adjust employee schedules. That is, did managers reallocate workers across hours more than they had before the app was implemented. Columns 3 to 5 of Table 3 confirm that employee schedules were indeed adjusted. Column 3 shows that on average, managers added 3.083 shifts per employee in a monthly period from a mean of 19.71 shifts per employee before the app implementation. Specifically, we see in Figure A.4 that shifts beginning during the busiest hours (late afternoon and dinner time) were shortened to allow for more workers to be staffed during these peak times without having to use more worker hours overall.

Columns 4 and 5 show the magnitude of these adjustments to employee schedules. On average, managers made 1.8 more reschedules per employee-monthly period after implementing the app (from a pre-implementation mean of 8.6). By definition, a reschedule occurs only if an employee's shift is set and then changed, making this more of a "last-minute" mechanism that managers use to meet unexpected demand. Given the demand shock from joining the delivery platform was permanent and the effects on sales were persistent as shown in Figure 4, we would expect that the permanent shift schedule of workers may also change to adjust to the new demand pattern and staffing requirements. Because reschedules exclude anticipated schedule changes, we also consider changes to an employee's modal shift.

We measure an employee's modal shift by taking the mode of that employee's start time over each monthly period. If an employee's modal start time differs from that employee's modal start time in the month before, we deem this a modal shift change in the current month.[21] Column 5 of Table 3 shows a large (40.5%) and significant increase in the share of

[21] While columns 1-4 of Table 3 present results at the monthly level, we measure modal shift changes at the store-employee level for the entire pre- and post-implementation adding the number of modal shift changes at

Table 3: Effects of delivery app implementation on employees

	(1) Hiring	(2) Hours worked	(3) Number of shifts	(4) Reschedules	(5) Num. of Modal Shift Changes
Post app implementation	0.005	-0.698	3.083***	1.830**	1.145***
	(0.00973)	(0.491)	(0.672)	(0.868)	(0.317)
Observations	24,989	24,815	25,522	24,989	2,718
Mean of Dep. Var.	0.079	28.741	19.710	8.637	2.828
Relative effect	6.33%	-2.43%	15.64%	21.18%	40.49%
Stores			62		

Notes: Table 3 show the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy. For the modal shift estimation we shows the results of a simple post-treatment dummy regressed against the average number of modal shift changes at each store. Treatment is defined as having implemented the food delivery platform. The sample consists of store-employee-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Hiring is a dummy of being hired that month in that store. Hours worked is the number of hours worked per employee per month. Number of shifts is the number of shifts per worker per monthly period. Reschedules is the number of reschedules per employee per month. Number of Modal Shift Changes is the number of modal shift changes pre(post)-implementation per employee. Numbers in parentheses are clustered standard errors at the store-employee level. * significant 10%, ** significant 5%, *** significant 1%

employees whose shifts changed as a result of the app implementation.

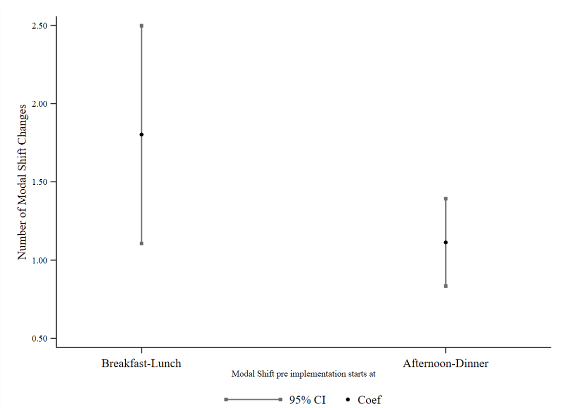
Figure 5a plots the treatment effect on the number of modal shift changes in response to the platform adoption among those whose modal shift *before* the app implementation was during the lean hours of Breakfast-Lunch, as compared to those whose pre-implementation modal shift was during the peak hours of Afternoon-Dinner. Figure 5b does the same exercise but using the timing of the worker's model shift *after* the app implementation. Taken together, the patterns across Figures 5a and 5b show that managers were predominantly shifting workers from lean Breakfast-Lunch shifts toward peak Afternoon-Dinner shifts. Figure 5d plots the analogous treatment effects on reschedules split by the timing of the worker's modal shift *after* the app implementation. Here we see again that workers are rescheduled to work the Afternoon-Dinner shifts and actually less likely to be rescheduled to the Breakfast-Lunch shifts.

Remember, our hypothesis is that managers who share the same gender as their workers will have better rapport and be better able to elicit cooperation from their workers regarding the changing of shifts. If so, we would expect that realized modal shift changes and

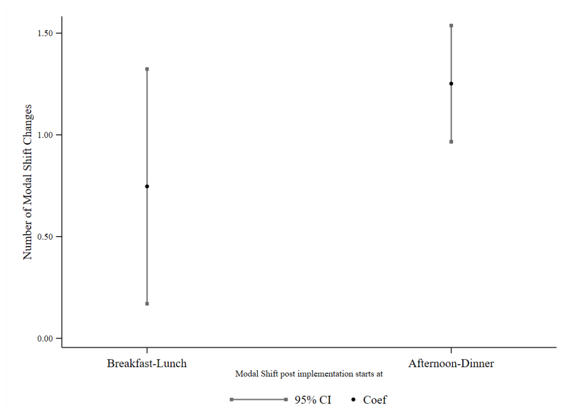
store-employee level. This change also results in a simpler regression specification in which we regress the number of modal shift changes at each store on a simple treatment dummy, defined as having implemented the food delivery platform.

Figure 5: Effect of delivery app implementation by timeband

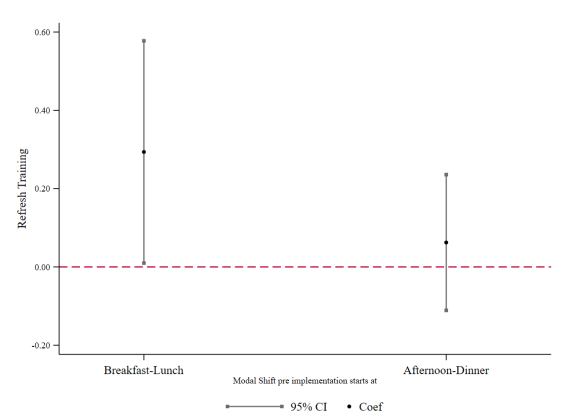
(a) Modal Shift Changes by Pre-App Timing



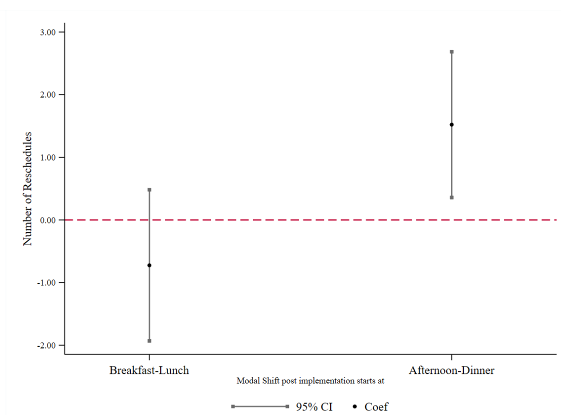
(b) Modal Shift Changes by Post-App Timing



(c) Refresh training by Pre-App Timing



(d) Re-schedules by Post-App Timing



Notes: Figure 5 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 when the modal shift of the employee falls during the Breakfast-Lunch or Afternoon-Dinner timeband. We define Breakfast-Lunch timeband from 7am to 2pm and Afternoon-Dinner timeband from 3pm to 10pm. We also include a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high-productivity store as any store with above-median productivity during the pre-implementation period. The difference between the effect at Breakfast-Lunch and Afternoon-Dinner is significant at 95% for Figure 5a (p-value: 0.047), significant at 90% for Figures 5b (0.086) and 5c (p-value: 0.090), and significant at 99% for Figure 5d (p-value: 0.0001). The sample consists of store-employee-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store-employee level.

reschedules are concentrated among workers who are the same gender as the majority of the managers in the store. In Figure A.5 in the Appendix, we plot treatment effects on modal shift changes and reschedules split by whether workers identify with the same or different gender as a majority of the managers. Indeed, the treatment effects on both reschedules and modal shift changes are significantly larger among same-gender workers as compared to those among different-gender workers.

One final lever which managers could have used to meet the new demand is training. That is, managers could have mostly left workers in their usual shifts but trained some workers to be able to contribute in part to bottleneck stations during peak hours in addition to their usual station assignments. Alternatively, managers could have left workers in their usual shifts *and stations* and used refresher training to try to improve their productivity so as to prevent or relieve bottlenecks created by increased demand. Indeed, Adhvaryu et al. (2022) find in a similar setting that managers respond to information they receive from a new performance monitoring technology primarily by employing both of these training-related levers. Table A.5 presents estimates of effects on training split into new skills and refresher training. We do not find any evidence that managers are training workers in new skills, but we see that they do increase refresher training by 54.4%.

However, when we split these effects by the modal shift of the worker being trained *before* the app implementation, we find in Figure 5c that these refresher trainings are not being allocated to workers who were already working the peak shifts, but rather to workers who were reallocated toward the peak shifts and workers who are left behind in the shifts from which managers are harvesting workers to reallocate. That is, these refresher trainings are not being used to avoid the need for more workers at peak times, but more to enable the reallocation of more workers to peak shifts. Taken together, the results in Table 3 and the patterns depicted in Figures 5a through 5d indicate that the reallocation of workers was the primary mechanism that managers used to achieve the sales gains in the face of the demand shock, rather than increasing the number of workers or hours per worker or using training to increase workers' breadth of skills or the productivity of workers already working peak shifts.

4.4.1 Turnover and Absenteeism

One might imagine that this adjustment of workers' schedules may not be well-received by all workers. In this sense, we may expect effects on turnover and absenteeism to reflect this tension. However, Appendix Table A.6 shows no statistically significant effects on turnover or absenteeism on average. Absenteeism reflects a small rise of 4.79% from the pre period mean of 0.36. Turnover reflects a small rise of 7.14% compared to the pre period mean of 0.014.

4.5 Management Styles and Practices by Gender Balance

Having documented that managers responded to the delivery app implementation primarily by reallocating workers across shifts to meet the increased demand at particular times of the day. To motivate the hypothesis that managers of gender-balanced stores may be more able to accomplish these worker reallocations, we now explore to what extent the shared gender identity between managers and workers enables better communication and rapport, and if these more cooperative relationships are reflected in managerial practices.

We begin in Panel A of Table 4 by studying survey measures reflecting the rapport between managers and lower-level workers. From left to right, the columns of Panel A show that managers of gender-balanced stores, as compared to managers of MHM stores, are 21% less likely to report keeping to themselves; 33% less likely to report acting without consulting workers; 52% more likely to report looking out for the welfare of their employees; and 38pp more likely to report treating workers as their equals as compared to the MHM store mean of 8%. Managers of FHM stores also differ in their reporting of these behaviors in the same direction as do managers of gender-balanced stores, but the magnitudes of the differences are not as large in general.

In Panel B, we study specific practices with respect to performance metrics and goal-setting. Compared to managers of MHM stores, managers of gender-balanced stores are 80% more likely to report discussing KPIs with their workers very frequently; 42% more likely to set short-term goals and 29% more likely to include their workers in the process; and 74% more likely to achieve these goals. Once again, FHM stores look more like gender-balanced stores in terms of these practices than like MHM stores.

Finally, in Panel C we study training related measures. Remember that workers who are more broadly trained are more easily reallocated, but investing in training workers in additional skills is costly for both managers and workers, particularly if manager-worker rapport is poor. We find that managers of gender-balanced stores are 44% more likely to prefer having employees who are trained across more stations over specialized workers, as compared to managers of MHM stores. Consistent with this, workers in gender-balanced stores are certified on 50-80% more stations than are workers in MHM stores. This pattern is reflected in both the manager reported average number of station certifications per worker in Column 2 as well as stock of certifications calculated from training records for all stations in Column 3 and for kitchen production stations in Column 4, validating the survey data.[22]

[22] As noted earlier, stocks calculated from the training records will be attenuated because they do not reflect any trainings before the start of our observation window. They will be accurate for short tenured workers

Table 4: Relationships and Management Practices by Gender Balance of Stores

	(1)	(2)	(3)	(4)
Panel A: Rapport	I have kept to myself	Acted w/out consulting workers	Looked out for welfare of employees	Treated employees as equals
Balance	-0.213***	-0.271**	0.262*	0.379***
	(0.046)	(0.123)	(0.154)	(0.099)
FHM Store	-0.120*	-0.273*	0.060	0.117
	(0.066)	(0.148)	(0.177)	(0.114)
Mean for MHM Stores	1.000	0.833	0.500	0.0833
Observations	117	117	117	117
Stores	58	58	58	58
Panel B: Practices	How frequently discuss KPIs with your employees?	Set short-term goals	Team goal-setting	Achieved short-term goals
Balance	0.267*	0.283**	0.196	0.308**
	(0.149)	(0.140)	(0.143)	(0.153)
FHM Store	0.187	0.333**	0.293**	0.223
	(0.171)	(0.138)	(0.143)	(0.174)
Mean for MHM Stores	0.333	0.667	0.667	0.417
Observations	117	117	117	117
Stores	58	58	58	58
Panel C: Training	Better to have employees who can shift	Avg number of stations certifications	Stock of Certifications/ Employees	Stock of Production Certifications/ Employees
Balanced Store	0.296**	2.973***	0.216*	0.0716**
	(0.140)	(0.915)	(0.120)	(0.0350)
FHM Store	0.253*	1.113	-0.00911	0.0246
	(0.148)	(0.946)	(0.0657)	(0.0221)
Mean for MHM Stores	0.667	3.727	0.433	0.0902
Observations	117	117	1,198	1,198
Stores	58	58	62	62

Notes: Table 4 shows the result of managers' responses to a survey regressed with the gender balance of the stores in which those managers worked (i.e., gender-balanced, MHM, or FHM stores). The final two columns of Panel C show the average number of certifications per employee in the period prior to the app implementation regressed with the gender balance of each store. Standard errors are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%.

4.6 Role of Store Gender Balance

Given the differences in survey measures of rapport and practices discussed above, we investigate whether managers of gender-balanced stores were more able to accomplish the potentially disruptive changes to worker staffing required to meet the non-uniform increases in demand via the delivery platform. We hypothesize that female workers may react differently to managers' attempts to adjust schedules to meet the new demand patterns than do male workers, and managers of the same gender as their workers may be more aware of and better able to navigate the preferences of their workers regarding schedule changes. Accordingly, we now explore whether the effects of the demand shock on both the movement of workers across shifts and the resulting sales are heterogeneous across gender "balanced" and "imbalanced" (FHM or MHM) stores.

We use a modified version of Equation (1) to account for heterogeneity in the gender balance of stores. The modified specification replaces the full set of time-to-treatment indicators with a simple post-treatment dummy which we then interact with a dummy equal to 1 if the store is gender balanced and a second dummy variable equal to 1 if the store is FHM (leaving MHM stores as the excluded category). Finally, in order to account for the possibility that gender balance may correlate with other dimensions of baseline productivity or skill of the store or manager, we add a control for the pre-app productivity of stores. Specifically, we add a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. For this control we follow Adhvaryu et al. (2022) and use the ratio of units sold to mean number of employees per store as a proxy for store productivity. We define a "high productivity" store as any store with above-median productivity.[23]

We use this specification to estimate heterogeneous effects on all the outcomes for which we studied average treatment effects above, but we focus on interpreting effects on sales, modal shift changes, and reschedules as the outcomes of primary interest. Figure 6a plots the treatment effect on sales by gender balance of the store. The increase in sales for MHM stores after the app implementation was only 2.1% from the baseline mean, though still significant at the 10% level; while gender-balanced and FHM stores exhibited increases

but underestimates for long tenured workers. The measure in Column 4 will be even smaller because it is restricted to kitchen production stations. For robustness, we present results for the analogous survey measure in Column 2 of Panel C.

[23] We also include results for all outcomes of interest without controlling for pre-app productivity in Appendix Table B.4.

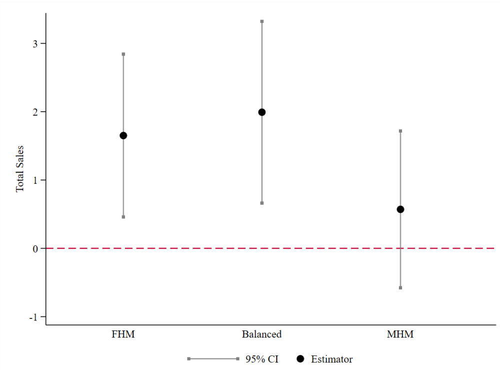
in sales 3.5 and 2.9 times the size, respectively. Table A.7 in the Appendix presents full estimation results corresponding to the effects plotted in Figure 6a and shows that the differences in these treatment effect coefficients are significant at conventional levels.

We next investigate whether MHM stores derived so little benefit from the delivery app as compared to balanced and FHM stores at least in part because they were less able to achieve the reallocation of workers across shifts to meet demand. We do so by similarly estimating treatment effect heterogeneity by gender balance of the store. Figure 6b shows that in gender-balanced stores and those in which women predominantly manage men, employee permanent schedules were changed dramatically in response to the platform (i.e., the treatment effect on modal shift changes among FHM and Balanced stores represents a roughly 60% increase from the pre-implementation mean); but no such changes were achieved in MHM stores. Figure 6c depicts analogous results for reschedules and once again shows that FHM and Balanced stores exhibited treatment effects on reschedules 2 to 2.5 times the size of those among MHM stores, respectively. These treatment effects patterns mimic notably the sales effects by gender balance of the stores presented in Figure 6a.

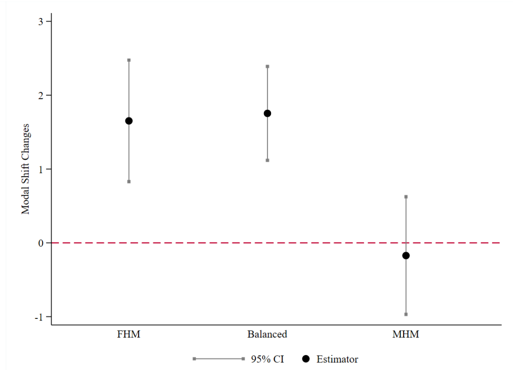
Note that this analysis implicitly assumes that the gender-balance of the store is static over time. Indeed, the within-store correlation in gender-balance over time is .89, reflecting the general validity of this assumption. Nevertheless, we demonstrate that the main pattern of heterogeneity in treatment effects presented here is preserved if we use variation in gender-balance within stores over time for identification. These results are presented in Table B.3 and Figure B.3 in the Appendix.

Figure 6: Effect of App Implementation in Balanced and Imbalanced Stores

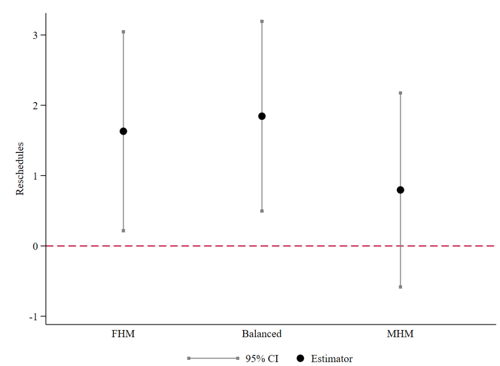
(a) Total Sales



(b) Modal Shift Changes



(c) Reschedules



Notes: Figure shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 when the gender of a store's managing team differs from the gender of the crew by more than 25% and 0 otherwise. We replace the gender-imbalance dummy with one dummy equal to 1 when the share of female managers exceeds the share of female crew members by more than 25% (i.e., female-heavy management or FHM), and another dummy equal to 1 when the difference between the share female managers and the share of female crew is between -0.25 and 0.25 percentage points (i.e., balanced stores) and 0 otherwise. Treatment is defined as having implemented the food delivery platform. We also include a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-shift(employee)-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. In Figure 6a total sales are measured for each monthly period and each store-shift. The difference between FHM and MHM is significant at the 90% level, and the difference between Balanced and MHM stores is significant at 95%. The difference between FHM and Balanced stores is not significant at conventional levels. For Figure 6b we aggregate the data at the store-employee-(before and after) level. The difference between FHM and MHM is significant at 99% level and the difference between Balanced and MHM stores is significant at the 99% level. The difference between FHM and Balanced stores is not significant at conventional levels. For Figure 6c we aggregate the data at the store-employee-monthly level. The difference between FHM and MHM stores is significant at 90%, and the difference between Balance and MHM is significant at the 95% level. The difference between FHM and Balanced stores is not significant at conventional levels. Table A.7 presents the estimated coefficients used to construct the figures and the test of the difference between coefficients. Standard errors are clustered at the store-shift(employee) level.

4.7 Robustness

4.7.1 Never-treated Stores

We replicate the baseline event study analysis using a sample that includes stores which implemented the app between December 2018 and October 2019, and stores that never implemented the app during this period. Identifying the event study coefficients hinges on the assumption that stores yet to implement the food delivery platform and stores that started to implement the app in earlier months, form a credible counterfactual for stores that start implementing the app, after accounting for time-invariant differences between stores and biweekly period shocks. Finally, we cluster standard errors at the store level for inference.

Results are similar when we estimate the regressions including never-treated stores. Figure C.1 shows the event-study coefficients for sales, units sold, and number of tickets for the whole store. We compare the patterns for this sample and outcomes with those from the baseline event-study sample (Figure 4). The results display the same takeaways as those from the baseline sample. We see an increase in store performance after the app implementation, although the magnitudes are smaller than the values reported in Table 2. The relative effect on sales is 2.76%, on units sold is 4.29%, and on tickets is 2.00%, all still statistically significant at conventional levels.

4.7.2 Heterogeneous Effects

As our treatment timing is staggered, the recent literature in econometrics has raised concerns about the possibility of negative weights in multiple-period difference-in-difference estimators since there may exist heterogeneity in treatment effects within-unit over time or between groups of units treated at different times (Athey and Imbens, 2021; De Chaisemartin and d'Haultfoeuille, 2020). The aforementioned issue may contaminate leads and lags in event studies where all treated observations are pooled together across groups (Sun and Abraham, 2020). Goodman-Bacon (2021) shows that difference-in-differences models of the form in (1) yield a weighted average of all possible permutations of pairwise difference-in-differences estimators, where, in our case, a pair is a cohort of observations treated at time t paired with a cohort of observations treated at time $t' > t$.

We address these issues in three ways. First, we estimate the cohort-specific average treatment effect suggested by Sun and Abraham (2020), which translates Callaway and Sant'Anna (2020) group-time average treatment effect from calendar time into relative periods, allowing to compare cohorts while holding their exposure to the treatment constant. Second, we estimate the group-time average treatment effect, where a group is defined by the time period when stores are first treated. The key assumption in our main sample excluding "never-treated" stores is the conditional parallel trends between stores treated in period g and groups that are "not-yet-treated" by time t. The results of these robustness checks are presented in Figures C.2 and C.3 in the Appendix and yield remarkably similar results to those from our baseline event-study specification and sample. Both methods show increases in sales, units sold, and tickets of even larger magnitude than those estimated in Table 2. The relative effect using Sun and Abraham (2020) on sales is 7.55%, on units sold is 11.06%, and on tickets is 11.85%. The relative effect using Callaway and Sant'Anna (2020) on sales is 7.99%, on units sold is 11.31%, and on tickets is 11.77%.

5 Discussion

Managers' abilities to establish and maintain strong communication and rapport with their workers contribute to the dramatic impact managers have on workplace outcomes such as worker retention and productivity. These interpersonal relationships, in turn, are affected by the degree to which managers and workers can identify with each other. Using rich personnel records and granular productivity data from the universe of fast food restaurants of a leading chain in Colombia, we study the degree to which mismatched gender identity between managers and workers affects the managers' ability to successfully allocate workers to shifts to meet demand.

We leverage the geographic expansion of a leading delivery platform across the country to study how well managers are able to adjust worker staffing to match resulting non-uniform increases in demand. We find that stores realize large and persistent gains in sales on average as a result of the arrival of the delivery platform to their neighborhood, but that store managers do not hire more workers to meet this demand. Rather, we find that in response to the increased demand after the arrival of the delivery platform, managers harvest workers from overstaffed shifts like breakfast and lunch and reallocate them to peak hours (both as temporary reschedules and as more permanent changes to modal shift assignments of workers), often cutting the shifts into smaller segments.

Using a survey of the universe of current managers, we show that stores in which managers and workers share predominantly the same gender identity have better communication and rapport between managers and workers. As a result, these same stores have more broadly skilled workers who are more easily reallocated across shifts. That is, the significant investment in face to face time between the managers and workers required in the training and certification process means that stores with better manager-worker rapport engage in more training. Since workers in these stores are trained and certified for a broader array of stations, their reallocation across shifts becomes more feasible.

We find that gender-balanced stores exhibit the largest impacts on observed worker reallocation following the delivery platform implementation and, in turn, the largest sales gains. Stores in which women predominantly manage men achieve schedule adjustments and sales gains closer to gender-balanced stores. We believe this reflects that female managers may have a greater awareness of and empathy toward the scheduling constraints and preferences of their workers, as women are more likely to have to navigate competing demands on their time and safety considerations when setting work schedules. Stores in which men predominantly manage women achieved a third to nearly one and a half times less worker reallocations and as a result realized only a third to a half the sales gain from the introduction of the delivery platforms.

5.1 Alternative Interpretations

We interpret the full pattern of results to be consistent with the notion that gender matched manager-worker pairs exhibit better rapport than do mismatched pairs such that the training and scheduling responsibilities of the manager – which depend so heavily on this rapport – are accomplished more successfully in gender matched stores. Also, the results support the notion that female managers may be more sensitive to scheduling preferences of workers, given their own constraints and considerations, as compared to male managers. We present survey evidence which supports this interpretation; however, of course, there are potential alternative interpretations.

First, we might be concerned that the gender balance of the store correlates (perhaps even incidentally) with other determinants of productivity. If so, the heterogeneity patterns we observe in the results would not only reflect the role of gender balance but these other factors as well. To address this concern, we follow the approach used in Adhvaryu et al. (2022), by controlling for a measure of baseline productivity interacted with the platform introduction dummy in all of our main results above. In the appendix we report estimates from specifications omitting this control (see Table B.4). The results are qualitatively similar across both specifications, indicating this may not have been a major concern, but our main results account for the possibility nonetheless.

In addition, we use the richness of the survey to investigate explicitly if any other style or practice measure correlates with gender balance. To do so, we employ a LASSO regression analysis in which we regress the gender-balanced dummy at the store level on the full set of survey variables (evaluated at the mean across multiple managers in the store where necessary), excluding only variables related to the training and scheduling practices and the underlying variables related to rapport and communication highlighted in Table 4. Of the more than 90 variables included in the LASSO regression, only one ("The job is quite simple and repetitive") was selected; the correlation is negative and significant at the 5 percent level, but the R-squared is only .094. We interpret this as further evidence that the gender balance measure is not picking up incidental correlations with other managerial styles and practices which may themselves affect productivity.

Second, we may be concerned that managers purposefully select the gender of their workers according to how well they believe they will be able to communicate and elicit cooperation from them. Though this is of course a very reasonable concern, we are told anecdotally by the firm that attracting and keeping workers for jobs like these is so difficult that managers must essentially hire every willing and capable worker who applies, and cannot pass over some workers on the basis of their gender. We also note that if indeed managers could choose their preferred gender of worker, MHM stores should be very few or nonexistent in our sample, as the predominantly male managers of these stores should prefer to hire men. Finally, we check explicitly if the arrival of a new male or female manager changes the gender composition of new workers hired in the future. We also repeat this exercise limiting attention to those arrivals which are "pivotal" in that they switch the store from majority male managers to majority female or vice versa. We find small and insignificant effects across all of these analyses (see Table B.5), consistent with the notion that managers have little scope for preferentially selecting the gender of their workforce.

Along similar lines, we may be concerned that the introduction of the delivery platform may affect hiring patterns or gender composition of workers as a response to changes in demand. Accordingly, we fix the gender composition of each store in our sample at its pre-adoption value for all of the analyses undertaken above. Nevertheless, we also check explicitly for whether the share of female managers or workers, gender balance of the store, or hiring of managers responded significantly to the introduction of the delivery platform. We once again find no evidence of significant effects across all of these analyses (see Table B.6).

5.2 Policy Implications

Our findings emphasize an important driver of potential gains to advancing women to managerial positions. Particularly in industries in which women dominate the frontlines – such as services, retail, food processing, and light manufacturing – having predominantly or even disproportionately male managers may significantly hinder productivity. In fact, our findings suggest that even in settings in which men dominate the front lines, female managers are not substantially less able to perform their responsibilities than are male counterparts. Taken together, interventions to increase the presence of women at managerial levels in organizations could not only improve equality, but actually yield higher productivity.

References

Adhvaryu, A., Gauthier, J.-F., Nyshadham, A., and Tamayo, J. A. (2021a). Absenteeism, productivity, and relational contracts inside the firm. Technical report, National Bureau of Economic Research.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2021b). Management and shocks to worker productivity. *Journal of Political Economy*, 130(1).

Adhvaryu, A., Murathanoglu, E., and Nyshadham, A. (2023). On the allocation and impacts of managerial training. Technical report, National Bureau of Economic Research.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2022). An anatomy of performance monitoring.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, 115(3):715–753.

Akerlof, G. A. and Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic perspectives*, 19(1):9–32.

Alan, S., Corekcioglu, G., and Sutter, M. (2023). Improving workplace climate in large corporations: A clustered randomized intervention. *The Quarterly Journal of Economics*, 138(1):151–203.

Albanesi, S., Olivetti, C., and Petrongolo, B. (2023). Families, labor markets, and policy. In *Handbook of the Economics of the Family*, volume 1, pages 255–326. Elsevier.

Alesina, A. and La Ferrara, E. (2002). Relational contracts and organizational capabilities. *Journal of Public Economics*, 85(2):207–234.

Ashraf, N. and Bandiera (2018). Social incentives in organizations. *Annual Review of Economics*, 10:439–463.

Ashraf, N., Bandiera, O., Minni, V., and Quintas-Martínez, V. (2022). Gender roles and the misallocation of labour across countries. *Working Paper*.

Athey, S. and Imbens, G. W. (2021). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*.

Atkin, D., Chaudhry, A., Chaudry, S., Khandelwal, A. K., and Verhoogen, E. (2017). Organizational barriers to technology adoption: Evidence from soccer-ball producers in pakistan. *The Quarterly Journal of Economics*, 132(3):1101–1164.

Atkin, D., Chen, M. K., and Popov, A. (2022). The returns to face-to-face interactions: Knowledge spillovers in silicon valley. Technical report, National Bureau of Economic Research.

Atkin, D., Schoar, A., and Shinde, S. (2023). Working from home, worker sorting and development. Technical report, National Bureau of Economic Research.

Bagues, M., Sylos-Labini, M., and Zinovyeva, N. (2017). Does the gender composition of scientific committees matter? *American Economic Review*, 107(4):1207–1238.

Blundell, R. and Dias, M. C. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3):565–640.

Buchmann, N., Meyer, C., and Sullivan, C. D. (2023). Paternalistic discrimination. Technical report, Working Paper.

Callaway, B. and Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Cullen, Z. and Perez-Truglia, R. (2023). The old boys' club: Schmoozing and the gender gap. *American Economic Review*, 113(7):1703–1740.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Euromonitor, I. (2019). Food service delivery in latin america: Local heroes, super apps and the search for growth. *London: Euromonitor International*.

Fiala, N., Garcia-Hernandez, A., Narula, K., and Prakash, N. (2022). Wheels of change: Transforming girls' lives with bicycles. *IZA Discussion Paper*.

Frederiksen, A., Kahn, L. B., and Lange, F. (2020). Supervisors and performance management systems. *Journal of Political Economy*, 128(6):2123–2187.

Freyaldenhoven, S., Hansen, C., Pérez, J. P., and Shapiro, J. M. (2021). Visualization, identification, and estimation in the linear panel event-study design. Technical report, National Bureau of Economic Research.

Friebel, G., Heinz, M., Hoffman, M., Kretschmer, T., and Zubanov, N. (2023). This really kneaded? identifying and eliminating potentially harmful monitoring practices. Technical report.

Gibbons, R. and Henderson, R. (2012). Relational contracts and organizational capabilities. *Organization Science*, 23(5):1350–1364.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics*, 132(3):1219—-1260.

Goldin, C. (2015). Hours flexibility and the gender gap in pay. Technical report, Center for American Progress.

Goldin, C. and Katz, L. (2011). The cost of workplace flexibility for high-powered professionals. 638(1):45–67.

Goldin, C. and Katz, L. F. (2018). *Women working longer: Increased employment at older ages*. University of Chicago Press.

Goldin, C. and Mitchell, J. (2017). The new life cycle of women's employment: Disappearing humps, sagging middles, expanding tops. *Journal of Economic Perspectives*, 31(1):161–182.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Guillouet, L., Khandelwal, A., Macchiavello, R., and Teachout, M. (2021). Language barriers in multinationals and knowledge transfers. Technical report, National Bureau of Economic Research.

Haegele, I. (2021). The broken rung: Gender and the leadership gap. Technical report, Working Paper.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4):1899—-1946.

Hoffman, M. and Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy*, 129(1).

Jaitman, L. (2020). Public transport from a gender perspective: Insecurity and victimization in latin america. the case of lima and asuncion metropolitan areas. *Journal of Economics, Race, and Policy*, 3(1):24–40.

Karpowitz, C. F., O'Connell, S. D., Preece, J., and Stoddard, O. (2023). Strength in numbers? gender composition, leadership, and women's influence in teams. *Journal of Political Economy*, conditionally accepted.

Kondylis, F., Legovini, A., Vyborny, K., Zwager, A., and Andrade, L. (2020). Demand for safe spaces avoiding harassment and stigma. *World Bank Policy Research Working Paper No. 9269*.

Løken, K. V., Lommerud, K. E., and Reiso, K. H. (2018). Single mothers and their children: Evaluating a work-encouraging welfare reform. *Journal of Public Economics*, 167:1–20.

Martinez, D., Mitnik, O. A., Salgado, E., Scholl, L., and Yañez-Pagans, P. (2018). Connecting to economic opportunity: The role of public transport in promoting women's employment in lima. *IZA Discussion Paper*.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–3759.

Metcalfe, R. D., Sollaci, A. B., and Syverson, C. (2023). Managers and productivity in retail. Technical report.

Muralidharan, K. and Prakash, N. (2017). Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics*, 9(3):321–350.

Olivetti, C. and Petrongolo, B. (2016). The evolution of gender gaps in industrialized countries. *Annual review of Economics*, 8:405–434.

Olivetti, C. and Petrongolo, B. (2017). The economic consequences of family policies: lessons from a century of legislation in high-income countries. *Journal of Economic Perspectives*, 31(1):205–230.

Reinhardt, F., Alvares, J., Buitrago, J., and Cal, M. (2021). Arcos Dorados: a QSR recovery plan. *Harvard Business School Case 721-023*.

Ronchi, M. and Smith, N. (2021). Daddy's girl: Daughters, managerial decisions, and gender inequality. Technical report, Working paper, Bocconi University.

Roth, J. (2019). Pre-test with caution: Event-study estimates after testing for parallel trends. *Department of Economics, Harvard University, Unpublished manuscript*.

Sandvik, J. J., Saouma, R. E., Seegert, N. T., and Stanton, C. T. (2020). Workplace knowledge flows. *The Quarterly Journal of Economics*, 135(3):1635–1680.

Smith, L. P., Ng, S. W., and Popkin, B. M. (2014). Resistant to the recession: low-income adults' maintenance of cooking and away-from-home eating behaviors during times of economic turbulence. *American journal of public health*, 104(5):840–846.

Sun, L. and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Wiswall, M. and Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

ONLINE APPENDIX

A Additional Statistics

Table A.1: Balance Table between Early and Late App Implementations

	Early implementers		Late implementers		Difference
	Mean	S.D.	Mean	S.D.	P-value
Store level					
Sales per employee	0.410	0.141	0.392	0.123	0.597
Units per employee	246.094	94.233	244.800	65.304	0.951
Tickets per employee	95.448	43.146	97.154	27.948	0.856
Reschedules per employee	0.424	0.136	0.405	0.097	0.523
Turnover per employee	0.006	0.004	0.011	0.034	0.330
Hiring per employee	0.008	0.005	0.037	0.163	0.313
New skills per employee	0.093	0.038	0.251	0.640	0.168
Follow up per employee	0.047	0.027	0.079	0.097	0.074
Unemployment rate (munic)	11.583	2.314	12.104	1.872	0.341
Avg income (munic)	1.324	0.216	1.316	0.206	0.873
Number of Stores	29		34		
Manager level					
Ages	22.187	2.288	22.273	2.264	0.749
Tenure (years)	2.067	1.551	2.279	1.838	0.287
Males	0.467	0.501	0.519	0.501	0.372
Number of Manager	135		154		
Number of Store Managers	29		34		
Number of Shift Managers	98		107		
Number of Second Assistant	8		13		

Notes: Table A.1 shows the descriptive statistics of the period before implementation for early (before March 2019) and late (after March 2019) implementers of the delivery app. In the last column of the table, we show the p-value associated with the difference in those characteristics between both types of stores. Unemployment rate and Avg income (in millions) are at the values of the municipality where the store is located.

Table A.2: Number of approved training programs per station

Station	Mean	Std. Dev.
Host	1.48	1.03
Kitchen opening	1.45	1.25
Service opening	1.56	1.26
Service opening and closure	1.51	1.25
Ice cream machine	3.32	4.89
Drive-thru	1.77	1.67
Drinks	1.75	1.62
Cash register	1.77	1.55
Birthday party celebration	1.54	1.12
Kitchen closure	1.46	1.20
Assembly	1.62	1.37
Breakfast assembly	1.51	1.39
Trainer	1.73	1.27
Fryer	1.50	1.29
Hospitality	2.16	2.26
Hot cakes	1.53	1.37
Bun initiator	1.52	1.15
Cleaning	2.32	2.39
Lobby, parking and playground	1.66	1.43
Quality service	1.66	1.23
Coffee	1.49	1.00
Delivery	1.54	1.03
Fries and hash brown	1.68	1.43
Grid	1.52	1.14
Breakfast / Muffin grid	1.56	1.48
Fried chicken	1.46	1.12
Desserts	2.06	1.74
Vegetables	1.56	1.43
Fried products	1.50	1.22
Food and personal security	2.40	2.54
Teamwork	2.47	2.42
Seller	1.89	1.79

Notes: Number of certifications per week per store. The sample consists of data for the 62 stores that implemented the app between July 2018 and October 2019.

Table A.3: Summary Statistics from Survey of Managers

	Female		Male		Difference Female-Male	
	Mean	S.D.	Mean	S.D.	Coef.	S.E.
I have kept to myself	0.891	0.315	0.774	0.422	0.117*	(0.0683)
I have acted without consulting my employees	0.600	0.494	0.581	0.497	0.0194	(0.0918)
In the last 3 months, I set short-term goals e.g., daily and across days	0.927	0.262	0.935	0.248	-0.00821	(0.0473)
In the last 3 months, on average, I achieved some or most of the short-term goals	0.927	0.262	0.952	0.216	-0.0243	(0.0448)
I involve all my team members in the goal-setting process	0.818	0.389	0.903	0.298	-0.0850	(0.0647)
Discuss KPIs with the employees in your store/shift. (how frequently)	0.491	0.505	0.613	0.491	-0.122	(0.0923)
Scheduling shifts is most important task	0.691	0.466	0.597	0.495	0.0941	(0.0889)
Shift scheduling is more time consuming than most tasks	0.855	0.356	0.758	0.432	0.0965	(0.0729)
Prefer to train more employees on critical stations	0.891	0.315	0.887	0.319	0.00381	(0.0587)
Better to have employees able to shift across stations than a few specialized	0.927	0.262	0.919	0.275	0.00792	(0.0496)
I have looked out for the personal welfare of my employees	0.945	0.229	0.952	0.216	-0.00616	(0.0413)
I have treated all my employees as my equals	0.400	0.494	0.339	0.477	0.0613	(0.0901)
Preferred shifts to prevent turnover	0,091	0,290	0,081	0,275	0.0103	(0.0524)
How many stations is an employee trained on for on average?	6.509	5.754	5.574	4.068	0.935	(0.934)

Notes: Table A.3 shows the result of managers' responses to a survey implemented in Colombia to capture management styles and practices. * significant 10%, ** significant 5%, *** significant 1%.

Table A.4: Effect of App Implementation on Waste measures

	(1) All	(2) Employee food	(3) Complete waste	(4) Incomplete waste	(5) Promo
Post Treatment	-0.121	-0.090	-0.025	-0.017	0.010
	(0.135)	(0.079)	(0.0400)	(0.054)	(0.047)
Observations	671	671	671	671	671
Mean of Dv	4.444	2.905	0.550	0.531	0.458
Stores			62		

Notes: Table A.4 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy. Treatment is defined as having implemented the food delivery platform. We also add a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high productivity store as any store with above-median productivity during the pre-implementation period. Column 1 presents the estimations for the whole sample. Column 2 presents the estimations for employee consumption proportion – food consumed by the employees as part of their benefits during the shift. Column 3 presents the estimations for complete waste proportion, which occurs when an item is wasted in its entirely. Column 4 shows the estimations for incomplete waste proportion, which occurs when only an item component is wasted. Column 5 presents the estimations for the promo proportion in the form of coupons or free products. The sample consists of store-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

Table A.5: Effects of delivery app implementation on training

	(1) New Skills	(2) Refresher
Post app implementation	0.337	0.395***
	(0.236)	(0.142)
Observations	27,811	27,811
Mean of Dep. Var.	1.919	0.726
Relative effect	17.56%	54.41%
Stores	62	

Notes: Table A.5 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy. Treatment is defined as having implemented the food delivery platform. The sample consists of store-employee-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Column 1 presents the estimations for number of new skill training per employee and Column 2 presents the estimations for number of refresher, or follow-up training. New skills training refers to all post-onboarding trainings that teach a skill for which an employee has not yet been trained. Refresher trainings include all subsequent trainings after a new skill training. Standard errors are clustered at the store-employee level. * significant 10%, ** significant 5%, *** significant 1%

Table A.6: Effect of delivery app implementation on turnover and absenteeism

	(1) Turnover	(2) Absenteeism
Post app implementation	0.000849	0.0173
	(0.00394)	(0.0123)
Observations	25,141	24,690
Mean of Dep. Var.	0.014	0.355
Relative Effect	7.14%	4.79%
Stores	62	

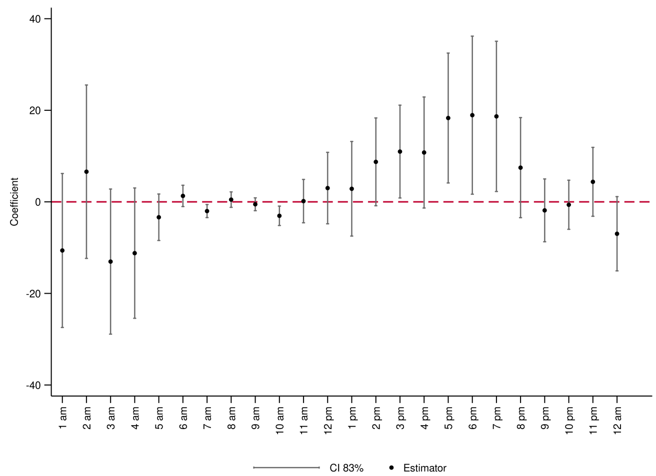
Notes: Table A.6 show the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy. Treatment is defined as having implemented the food delivery platform. The sample consists of store-employee-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Turnover is a dummy 1 if employee left the firm and 0 otherwise. Absenteeism is a dummy of 1 if employee was unexcused absent, 0 otherwise, at the store-employee-monthly level. Numbers in parentheses are clustered standard errors at the store-employee level. * significant 10%, ** significant 5%, *** significant 1%

Table A.7: Effect of delivery app implementation by Gender-Balanced, FHM, and MHM Stores

	Total Sales	Units Sold	Tickets	Hiring	Hours per worker/week	Number of shifts	Reschedules	Num. of Modal shift changes
(a) Post Treatment	0.571	518.3	120.4	0.00483	-0.358	2.922***	0.796	-0.172
	(0.586)	(427.4)	(191.0)	(0.0145)	(0.528)	(0.670)	(0.704)	(0.406)
(b) Post x Balanced store	1.422**	1,026**	453.3**	0.00303	-0.511	0.724**	1.049**	1.926***
	(0.564)	(445.7)	(180.5)	(0.0134)	(0.383)	(0.357)	(0.426)	(0.335)
(c) Post x FHM store	1.080*	873.9**	445.7***	-0.00789	-0.976**	0.534	0.835*	1.825***
	(0.553)	(391.1)	(167.8)	(0.0151)	(0.422)	(0.390)	(0.491)	(0.407)
(a+b) TE for Balanced stores	1.993***	1544.289***	573.751***	0.008	-0.868	3.646***	1.845***	1.754***
	(0.678)	(460.086)	(129.025)	(0.011)	(0.534)	(0.732)	(0.688)	(0.324)
(a+c) TE for FHM stores	1.651***	1392.224***	566.115***	-0.003	-1.334**	3.456***	1.631**	1.652***
	(0.608)	(327.739)	(202.107)	(0.013)	(0.558)	(0.731)	(0.722)	(0.420)
(b-c) TE for FHM -Balanced stores	-0.341	-152.065	-7.636	-0.011	-0.466*	-0.190	-0.214	-0.101
	(0.454)	(571.81)	(172.836)	(0.012)	(0.282)	(0.267)	(0.336)	(.305)
Observations	4,885	4,885	4,885	24,989	24,815	25,522	24,989	2,718
Mean of Dv	27.419	17297.636	6880.046	0.0786	28.741	19.710	8.637	2.828
Stores					62			

Notes: Table A.7 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 interacted with one dummy equal to 1 when the share of women on a store's managing team exceeds the share of female crew members by more than 25% (i.e., female-heavy management or FHM), and another dummy equal to 1 when the difference between the share female managers and the share of female crew is between -0.25 and 0.25 percentage points (i.e., balanced stores) and 0 otherwise. Treatment is defined as having implemented the food delivery platform. We also add a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-shift(employee)-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

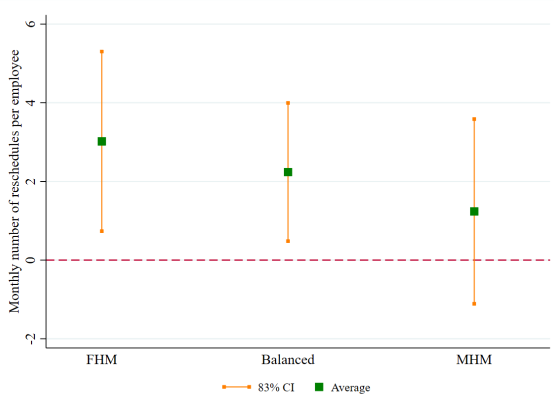
Figure A.1: Effects of delivery app implementation on sales by hour of the day



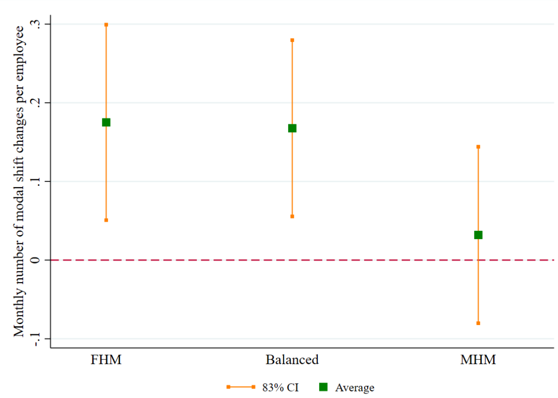
Notes: Figure shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with dummies of each hour of the day. We also include a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level.

Figure A.2: Validating survey with administrative data

(a) Reschedules



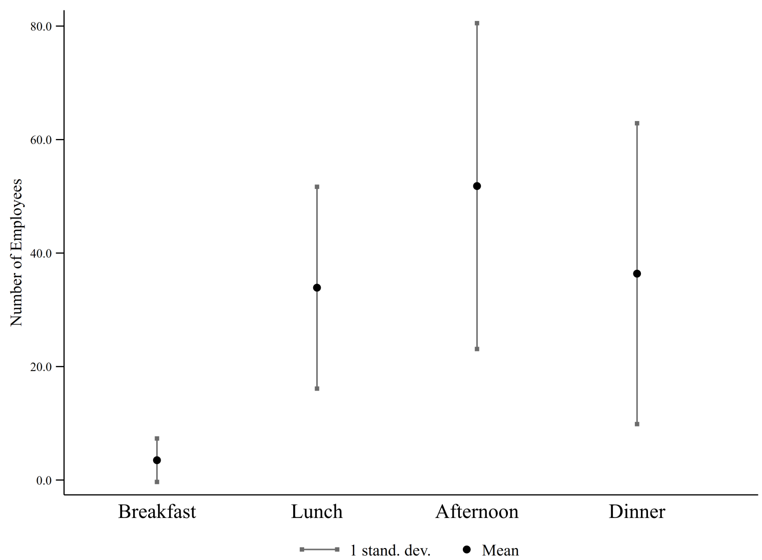
(b) Modal Shift Changes



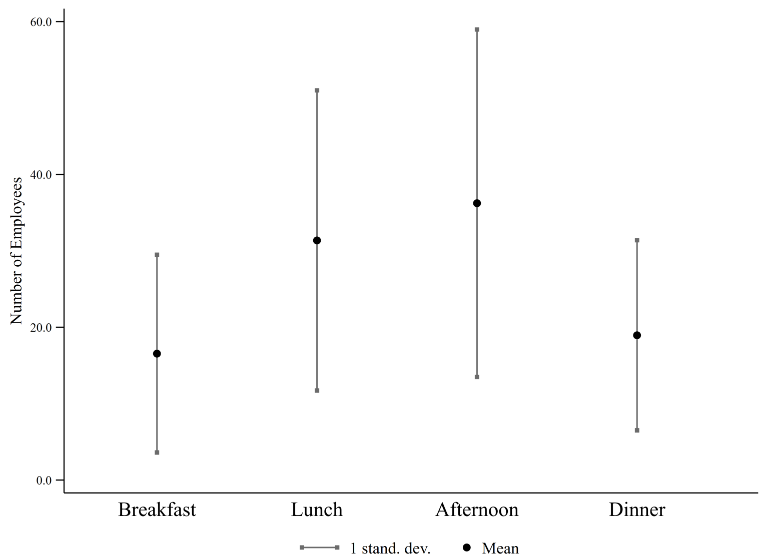
Notes: Panel a shows the difference in average reschedules per employee between managers who report scheduling as the most important task vs those who don't among the different types of store gender balance. Panel b shows the difference in average modal shift changes per employee between managers who report scheduling as the most important task vs those who don't among the different types of store gender balance. We use robust standard errors.

Figure A.3: Sales and Employees by Timeband

(a) Sales by Timeband

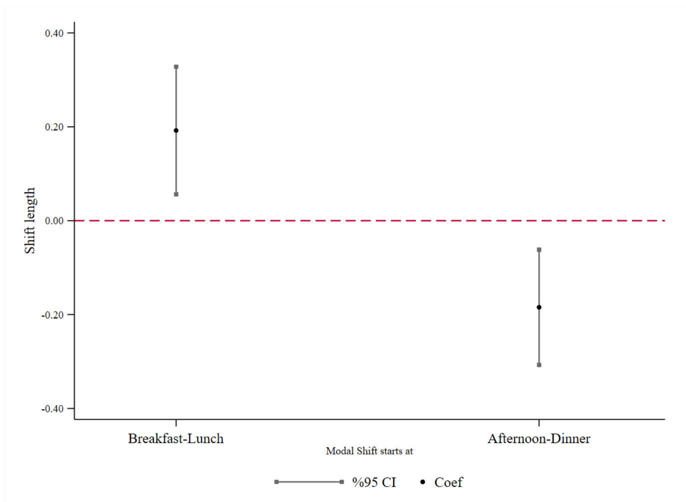


(b) Employees by Timeband



Notes: Panels (a) and (b) of Figure A.3 show the average monthly sales (in thousands of USD) and number of employees working during each of the four periods of the day, respectively. The sample consists of pre-implementation store-monthly panel data for the 62 stores that implemented the app between July 2018 and October 2019. The start and end times for each timeband were set by the partner firm, so we adopt their nomenclature. Breakfast includes the hours from 7:00 am to 10:59 am, lunch from 11:00 am to 1:59 pm, afternoon from 2:00 pm to 5:59 pm, and dinner from 6:00 pm until 11:59 pm.

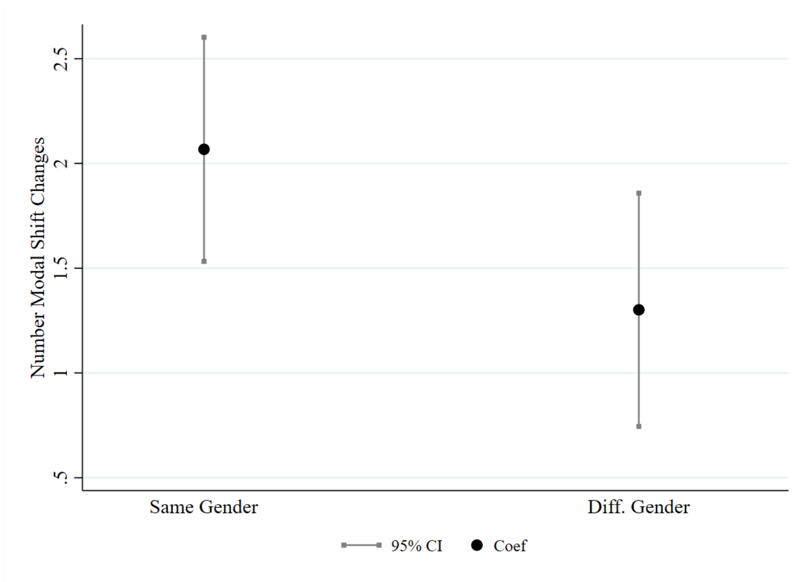
Figure A.4: Effect of App Implementation on Shift Length by Timeband



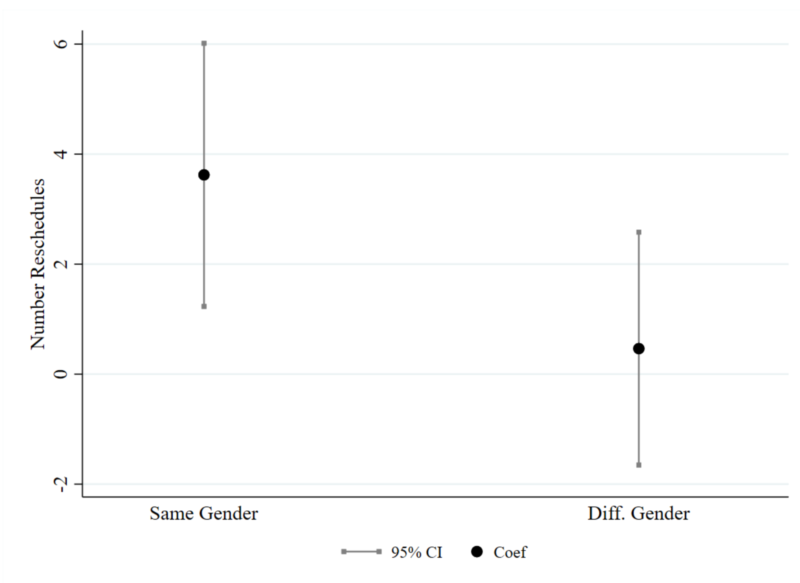
Notes: Figure shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 when the modal shift of the employee falls during the Breakfast-Lunch or Afternoon-Dinner timeband. We define Breakfast-Lunch timeband from 7am to 2pm and Afternoon-Dinner timeband from 3pm to 10pm. We also include a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-biweekly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level.

Figure A.5: Effect of App Implementation on Same and Different Gender outcomes

(a) Number of Modal Shift Changes



(b) Number of Re-schedules



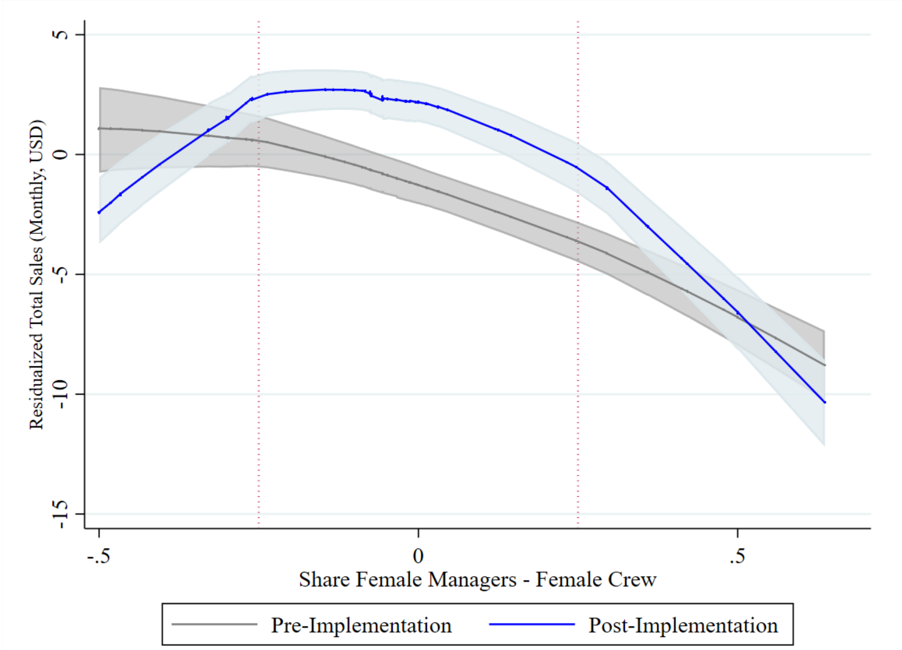
Notes: Figure shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy interacted with a dummy 1 if female (male) employees work in a store where most of the managers (above 50%) are female (males) 0 otherwise. We also include a dummy for high store productivity pre-implementation and interact this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-employee-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Difference between coefficients is significant at 90% for Number of Reschedules (p-value 0.052) and Number of Modal Shift Changes (p-value 0.052). Standard errors are clustered at the store-employee level.

B Additional Evidence

B.1 Gender Balance definition

B.1.1 Locally Weighted Smoothing of Gender-balance Shares

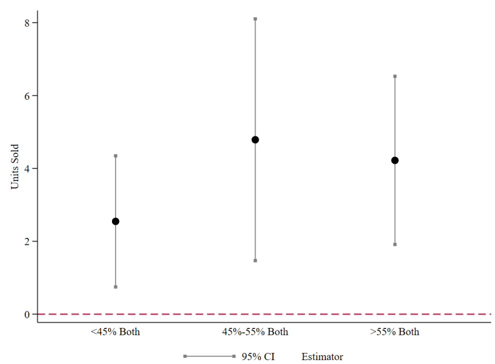
Figure B.1: LOWESS regression for Total Sales in Balanced and Imbalanced Stores



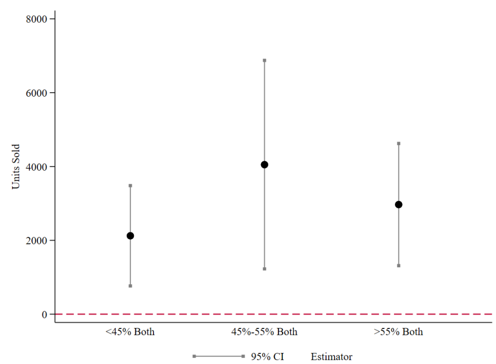
Notes: Figure B.1 shows the estimation of a locally weighted smoothing regression of effect of gender balance (i.e., the share of female managers minus the share of female crew members) within a store on average total sales in a monthly period in thousands of USD. Total sales are the residuals after controlling by Monthly FE and Shift FE. The red vertical lines are drawn at positive and negative 0.25, the thresholds chosen in the main paper to classify a store a gender-balanced or gender-imbalanced.

Figure B.2: Heterogeneous effects on balanced stores

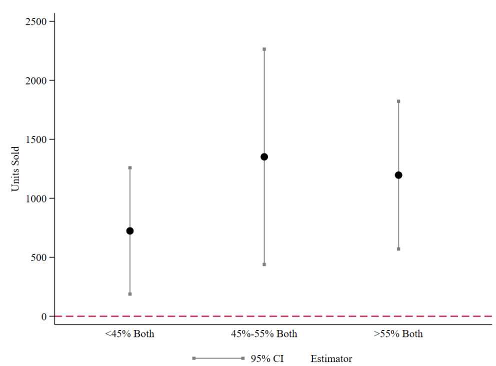
(a) Total Sales



(b) Units Sold



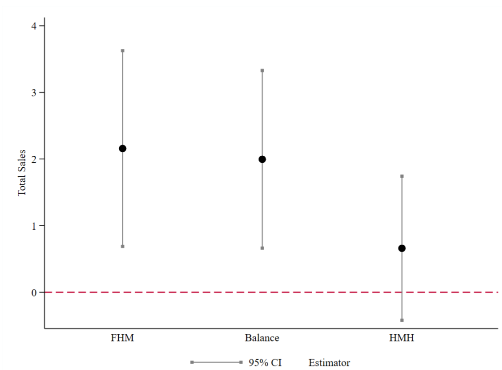
(c) Tickets



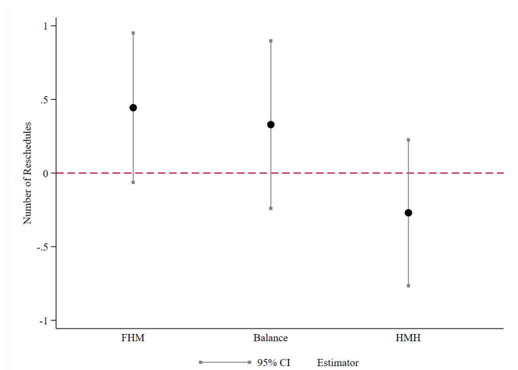
Notes: Table B.3 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 when the share of female managers and crew members is lower than 45%, a dummy equal to 1 when the share of female managers and crew members is larger than 55%, and a dummy equal to 1 when the share of female managers and crew members is between than 45% and 55%. Treatment is defined as having implemented the food delivery platform. We also add a dummy for high store productivity pre-implementation and interacted with this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high-productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-shift-monthly panel data for Balanced stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level.

Figure B.3: Robustness of heterogeneous results using floating Gender Store Balance

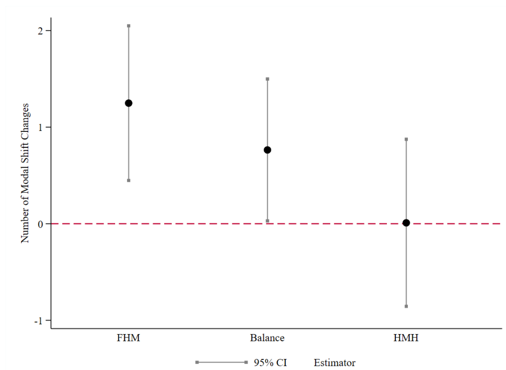
(a) Total Sales



(b) Reschedules



(c) Modal Shift Changes



Notes: Table B.3 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 interacted with one dummy equal to 1 when the share of women on a store's managing team exceeds the share of female crew members by more than 25% (i.e., female-heavy management or FHM), and another dummy equal to 1 when the difference between the share female managers and the share of female crew is between -0.25 and 0.25 percentage points (i.e., balanced stores) and 0 otherwise. Both dummies are defined using time-variant shares of female managers and crew. Treatment is defined as having implemented the food delivery platform. We also add a dummy for high store productivity pre-implementation and interacted with this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high-productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-shift-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level.

Table B.1: Balance Table between gender store balance classifications

	Female-Heavy Management		Male-Heavy Management		Balanced stores		FHN -Balanced	MHM -Balanced	FHM-MHM
	Mean	SD	Mean	SD	Mean	SD	P-value	P-value	P-value
Store level									
Sales per employee	0.411	0.121	0.382	0.093	0.402	0.144	0.816	0.593	0.511
Units per employee	251.230	59.504	222.769	46.305	249.268	91.363	0.930	0.211	0.211
Tickets per employee	97.679	27.372	89.472	24.346	97.680	40.482	1.000	0.423	0.456
Reschedules per employee	0.437	0.136	0.441	0.119	0.400	0.109	0.378	0.340	0.949
Turnover per employee	0.005	0.004	0.006	0.004	0.011	0.031	0.282	0.338	0.754
Hiring per employee	0.007	0.004	0.008	0.005	0.033	0.150	0.303	0.316	0.737
New skills per employee	0.090	0.030	0.126	0.060	0.219	0.592	0.183	0.341	0.108
Follow up per employee	0.042	0.019	0.062	0.031	0.072	0.091	0.062	0.595	0.091
Age balance	1.482	1.771	1.980	1.166	1.775	1.601	0.593	0.667	0.426
Location balance	0.810	0.266	0.831	0.123	0.875	0.132	0.392	0.355	0.799
Unemployment rate (munic)	11.544	2.353	12.385	2.380	11.824	1.954	0.705	0.505	0.409
Avg income (munic)	1.288	0.228	1.371	0.195	1.317	0.209	0.685	0.459	0.357
Number of stores	13		10		39				
Managers level									
Ages	22.235	2.446	22.250	1.934	22.196	2.356	0.918	0.888	0.976
Tenure (years)	2.269	1.784	2.156	1.725	2.056	1.685	0.444	0.762	0.774
Number of Managers	60		35		194				
Number of Store Managers	13		10		39				
Number of Shift Managers	38		22		145				
Number of Second Assistant	9		3		10				
Workers level									
Ages	20.940	2.251	20.686	2.131	20.781	2.961	0.294	0.595	0.182
Tenure (years)	0.935	1.473	0.887	1.200	0.806	1.477	0.156	0.401	0.678
Number of Workers	367		204		953				

Notes: Table B.1 shows the descriptive statistics for Female-heavy management stores, Balanced stores, and Male-heavy management stores of the period before implementation. In the last three columns of the table, we show the p-value associated with the difference in those characteristics between all the types of stores. Unemployment rate and Avg income (in millions) are at the values of the municipality where the store is located.

Table B.2: Time variation of Gender Store Balance Classification

	Interclass Correlation	Corr between pre-implementation period and last value in data
Gender Store Balance classification	0.89	0.88

Notes: Table B.2 shows how much time variation of Gender Store Balance Classification we can find in data. To that end, we compute the Inter Class Correlation and the correlation between the pre-implementation period and the last value in data of the Gender Store Balance classification.

Table B.3: Robustness of heterogeneous results using floating Gender Store Balance

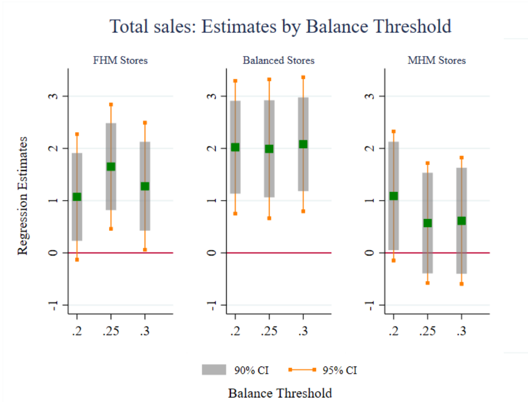
	Total Sales	Units Sold	No. Tickets	Reschedule	Modal Shift Changes
Post	0.660	788.3**	422.7**	-0.270	0.00957
	(0.552)	(358.6)	(172.7)	(0.253)	(0.441)
Post*Balanced	1.336	802.3	6.063	0.599*	0.755**
	(0.814)	(541.3)	(214.9)	(0.329)	(0.298)
Post*FHM	1.498**	666.3	9.180	0.714**	1.239***
	(0.725)	(445.8)	(203.3)	(0.328)	(0.305)
Balanced	0.425	197.5	184.9	1.994***	-0.426
	(0.834)	(472.2)	(220.0)	(0.485)	(1.017)
FHM	-0.187	92.08	203.3	1.223**	-1.458
	(1.006)	(542.1)	(243.1)	(0.531)	(1.110)
Observations	4,681	4,681	4,681	24,988	2,718

Notes: Table B.3 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 interacted with one dummy equal to 1 when the share of women on a store's managing team exceeds the share of female crew members by more than 25% (i.e., female-heavy management or FHM), and another dummy equal to 1 when the difference between the share female managers and the share of female crew is between -0.25 and 0.25 percentage points (i.e., balanced stores) and 0 otherwise. Both dummies are defined using time-variant shares of female managers and crew. Treatment is defined as having implemented the food delivery platform. We also add a dummy for high store productivity pre-implementation and interacted with this dummy with our post-treatment dummy. We use the ratio of units sold to mean number of employees per store as a proxy for store productivity and define a high-productivity store as any store with above-median productivity during the pre-implementation period. The sample consists of store-shift(employee)-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store-shift(employee) level. * significant 10%, ** significant 5%, *** significant 1%.

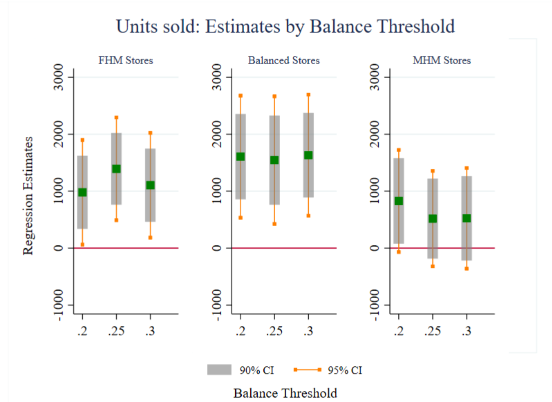
B.1.2 Robustness Across Gender-Balance Thresholds

Figure B.4: Robustness Across Gender-Balance Thresholds for performance outcomes

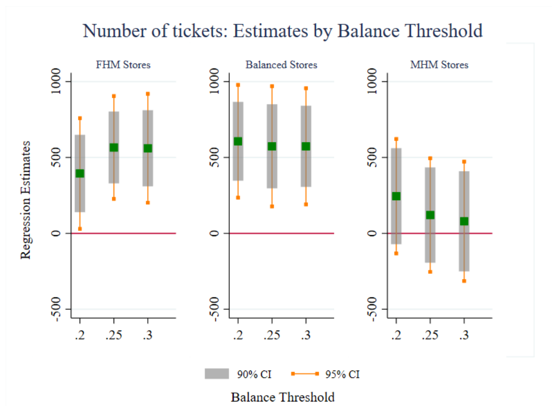
(a) Total Sales



(b) Total Units



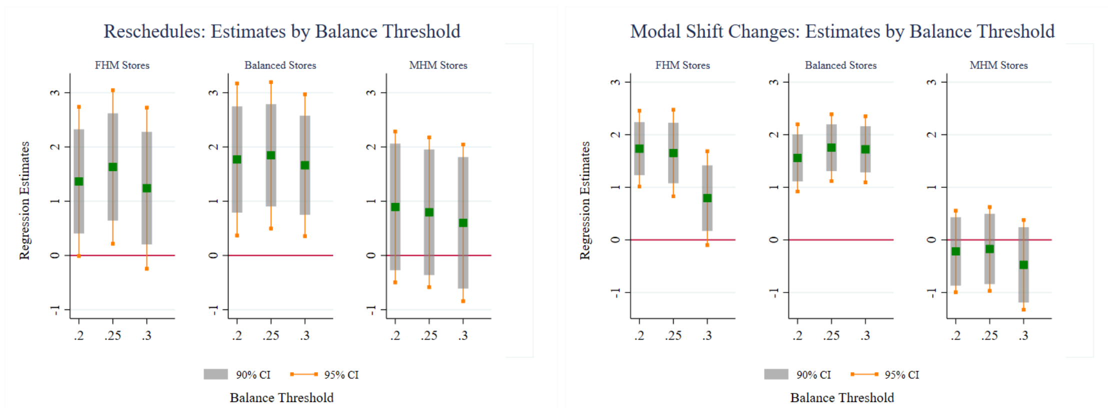
(c) Num. Tickets



Notes: Figure shows how the estimated effects change for Female-heavy Management, Gender-Balanced, and Male-heavy Management Stores by different gender-balance thresholds.

Figure B.5: Robustness Across Gender-Balance Thresholds for Reschedules and Modal Shifts Changes

(a) Re-schedules (b) Modal Shift Changes



Notes: Figure shows how the estimated effects change for Female-heavy Management, Gender-Balanced, and Male-heavy Management Stores by different gender-balance thresholds.

B.2 Tables

Table B.4: Effect of delivery app implementation by Gender-Balanced, FHM, and MHM Stores without controlling for productivity

	Total Sales	Units Sold	Tickets	Hiring	Hours per worker/week	Number of shifts	Reschedules	Modal Shift Changes
Post Treatment	0.571	518.3	120.4	0.00539	0.768**	1.200***	-0.465	-0.211
	(0.586)	(427.4)	(191.0)	(0.0133)	(0.359)	(0.367)	(0.400)	(0.316)
Post x Balanced Store	1.422**	1,026**	453.3**	0.00264	-0.330	0.453	0.849**	1.922***
	(0.564)	(445.7)	(180.5)	(0.0124)	(0.376)	(0.340)	(0.422)	(0.336)
Post x FHM Store	1.080*	873.9**	445.7***	-0.00812	-0.751*	0.273	0.505	1.818***
	(0.553)	(391.1)	(167.8)	(0.0149)	(0.413)	(0.376)	(0.495)	(0.403)
TE for Balanced stores	1.993***	1544.289***	573.751***	0.008	0.438*	1.653***	.384*	1.711***
	(0.678)	(571.810)	(129.025)	(0.011)	(0.259)	(0.317)	(0.206)	(0.206)
TE for FHM Stores	1.651***	1392.224***	566.115***	-0.003	0.017	1.474***	0.04	1.607***
	(0.608)	(460.086)	(202.107)	(0.012)	(0.291)	(0.326)	(0.307)	(0.278)
TE for FHM-Balanced Stores	-0.341	-152.065	-7.636	-0.011	-0.422	-0.179	-0.344	-0.104
	(0.454)	(327.739)	(172.836)	(0.011)	(0.281)	(0.268)	(0.339)	(0.301)
Observations	4,885	4,885	4,885	24,989	24,815	25,522	24,989	2,718
Mean of Dv	27.419	17297.636	6880.046	0.079	28.741	19.710	8.637	2.828
Stores					62			

Notes: Table B.4 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy and the interaction of the post-treatment dummy with another dummy equal to 1 interacted with one dummy equal to 1 when the share of women on a store's managing team exceeds the share of female crew members by more than 25% (i.e., female-heavy management or FHM), and another dummy equal to 1 when the share of males on a store's managing team exceeds the share of male crew members by more than 25% (i.e., male-heavy management or MHM) and 0 otherwise. Treatment is defined as having implemented the food delivery platform. The sample consists of store-shift(employee)-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store-shift(employee) level. * significant 10%, ** significant 5%, *** significant 1%

Table B.5: Effect of a new manager on hiring of female workers

	New Managers			New "pivotal" Managers		
	Share Female's hiring	Share Net Female's hiring	Share Female's turnover	Share Female's hiring	Share Net Female's hiring	Share Female's turnover
			A: New Female Manager			
Post New Manager	-0.0419	-0.0639	-0.0312	-0.0168	-0.0584	0.00382
	(0.0517)	(0.0582)	(0.0538)	(0.0510)	(0.0771)	(0.0476)
			B: New Male Manager			
Post New Manager	-0.00568	-0.00629	0.00164	0.0331	-0.0187	-0.0274
	(0.0344)	(0.0512)	(0.0509)	(0.0458)	(0.0404)	(0.0376)
Observations	1,080	1,080	1,080	1,080	1,080	1,080
Stores			62			

Notes: Table B.5 shows the effect of the arrival of a new manager on the hiring of female workers. We define a pivotal manager as the manager that changes the gender-balanced composition after arriving. The sample consists of store-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

Table B.6: Effect of delivery app implementation on gender composition

	(1) Share of female managers	(2) Share of female workers	(3) Gender-Balanced stores	(4) Hiring of managers
Post-Implementation	0.0106	-0.0523	-0.0114	-0.0854
	(0.0949)	(0.119)	(0.166)	(0.0622)
Observations	1,051	1,051	1,051	1,051
Stores		62		

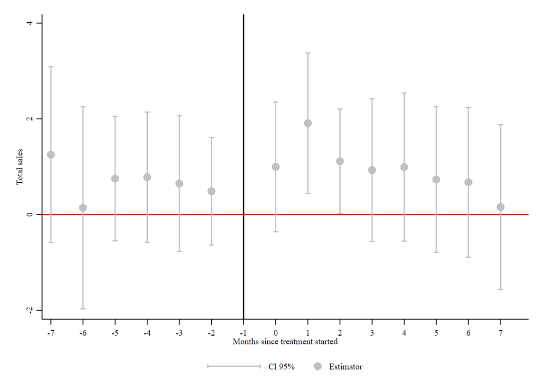
Notes: Table B.6 shows the results for the estimation of equation (1), replacing the full set of time-to-treatment indicators with a simple post-treatment dummy. The sample consists of store-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Standard errors are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

C TWFE Robustness Results

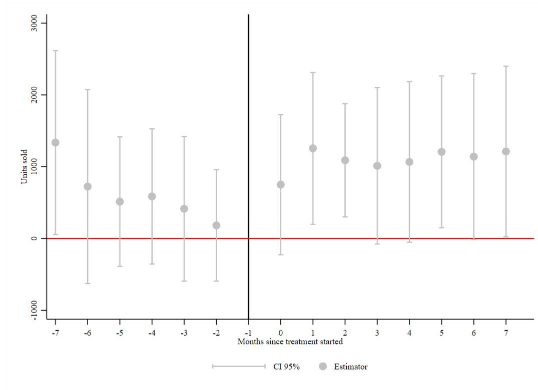
C.1 Never-Treated Stores

Figure C.1: Effects of delivery app implementation on performance measures for Never-Treated Stores

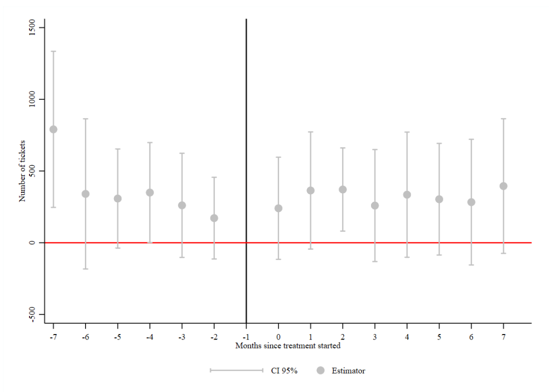
(a) Total Sales (USD)



(b) Units sold



(c) Number of tickets

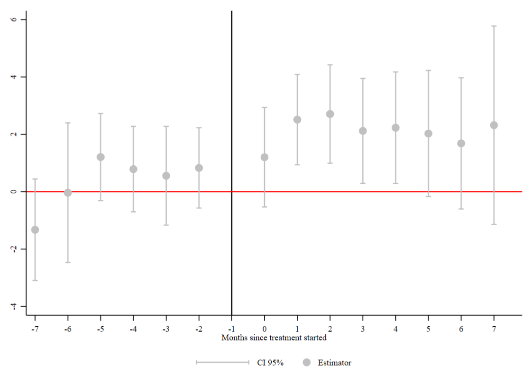


Notes: Panels (a) to (c) of Figure C.1 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on the key performance measures. Treatment is defined as having implemented the food delivery platform. Control stores are those not yet treated. The sample consists of store-shift-monthly panel data for the 62 stores implementing the app and 13 stores that never implemented the app between July 2018 to October 2019. Panel (a) shows the impact on sales, Panel (b) on units sold, and Panel (c) on the number of tickets (orders). The vertical line represents the time of the treatment.

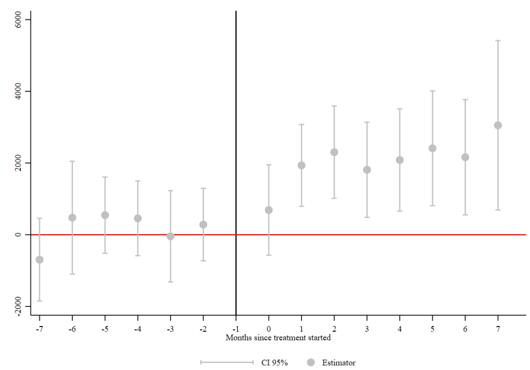
C.2 Sun and Abraham (2020)

Figure C.2: Effects of delivery app implementation on performance measures

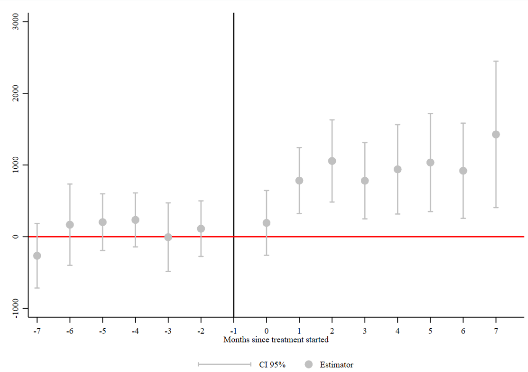
(a) Sales



(b) Units sold



(c) Number of tickets

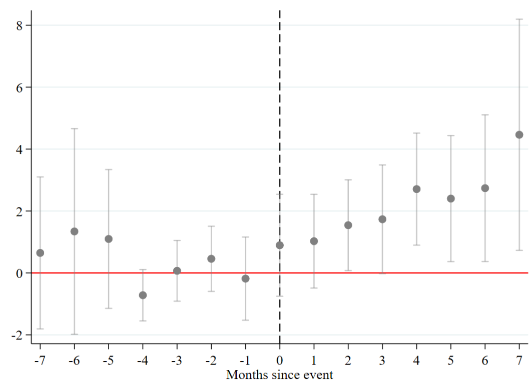


Notes: Panels (a) to (c) of Figure C.2 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on the key performance measures in logs, after correcting for staggered treatment timing per Sun and Abraham (2020). Treatment is defined as having implemented the food delivery platform. Control stores are those not yet treated. The sample consists of store-shift-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Panel (a) shows the impact on sales, Panel (b) on units sold, and Panel (c) on the number of tickets (orders). The vertical line represents the time of the treatment.

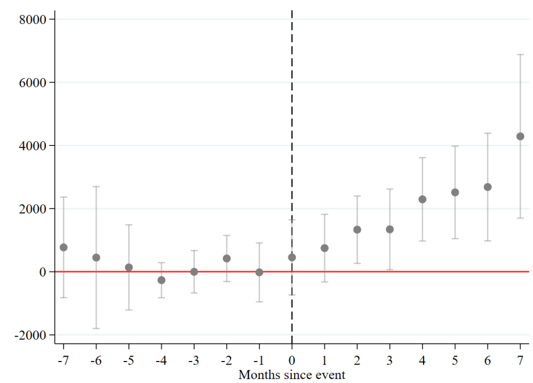
C.3 Callaway and Sant'Anna (2020)

Figure C.3: Effects of delivery app implementation on performance measures

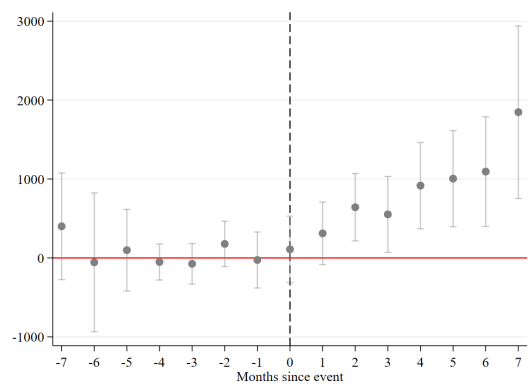
(a) Sales



(b) Units sold



(c) Number of tickets



Notes: Panels (a) to (c) of Figure C.3 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on the key performance measures in logs, after correcting for staggered treatment timing per Callaway and Sant'Anna (2020). Treatment is defined as having implemented the food delivery platform. Control stores are those not yet treated. The sample consists of store-shift-monthly panel data for the 62 stores that implemented the app between July 2018 to October 2019. Panel (a) shows the impact on sales, Panel (b) on units sold, and Panel (c) on the number of tickets (orders). The vertical line represents the time of the treatment.