An Anatomy of Performance Monitoring*

Achyuta Adhvaryu[†], Anant Nyshadham[‡], and Jorge Tamayo[§]

June 8, 2022

Abstract

Does better monitoring improve firm performance? We study the introduction of a technology that enabled managers to track the progress of drive-thru orders in a large quick-service restaurant chain. Sales increased by 5%, but impacts diminished to half that within two months. Worker skill dynamics play an important role in explaining this pattern. Managers provided greater training inputs at key workstations, but only a subset provided "refresher" training to counteract skill depreciation. Stores in which managers utilized refresher training intensively pre-technology had more persistent gains, suggesting that managers' attention to skill dynamics is critical to the success of performance monitoring. .

Keywords: performance monitoring, management, worker skills, skill depreciation, managerial inattention, on-the-job training, productivity, information technology, multitasking, quick-service restaurants, Puerto Rico
JEL: J24, M10, M53

*We thank Estefania Saravia, Cristian Chica, Luisa Estrada, Sofia Calderon, and Mamta Pimoli for excellent research assistance, and Gabriel Serber and his team for their incredible support, guidance, and most importantly patience in teaching us about fast food operations, sharing administrative data, and helping us understand the Puerto Rican QSR context. We thank Harvard Business School for supporting travel and other research-related expenses. We thank seminar participants at UPenn, Michigan, HBS, NBER/BREAD, UWA, Georgia State, University of Toronto (Rotman), The Inter-American Development Bank, and Wilfred Laurier University for excellent comments and ideas. Thanks to Teresa Amabile, Vittorio Bassi, Jere Behrman, Lauren Bergquist, Hanming Fang, Guido Friebel, Teresa Fort, Robin Greenwood, Mitch Hoffman, Amit Khandelwal, Rem Koning, Shannon Liu, Tommaso Porzio, Ananth Raman, Raffaella Sadun, Michael Toffel, and Eric Verhoogen for very helpful conversations and insights.

[†]University of Michigan, NBER, BREAD, William Davidson Institute, Good Business Lab; adhvaryu@umich.edu

[‡]University of Michigan, NBER; BREAD, Good Business Lab; nyshadha@umich.edu

[§]Harvard University, Harvard Business School; jtamayo@hbs.edu

1 Introduction

The manager's ability to effectively monitor team performance and identify and solve problems as they arise has long been held as a key dimension of managerial quality (Adhvaryu et al., 2021a,b; Bloom and Van Reenen, 2010). Indeed, our fundamental understanding of the manager-worker relationship relies heavily on what exactly can be monitored and how well (Baker, 2002, 1992; Holmstrom and Milgrom, 1991; Prendergast, 1999). Despite a large and still growing empirical literature linking variation in managerial quality to productivity both across (Bandiera et al., 2020; Bloom et al., 2019, 2013; Bloom and Van Reenen, 2007) and within firms (Bertrand and Schoar, 2003; Frederiksen et al., 2020; Hoffman and Tadelis, 2021; Lazear et al., 2015), few studies examine how managers respond when the ability to monitor workers improves, and whether productivity increases as result.

We aim to fill this gap by leveraging the introduction of a performance monitoring technology that gives managers real-time data on the efficiency of production in a large quick-service restaurant (QSR) chain in Puerto Rico. The technology's continual feedback is meant to enable managers to better target their attention and thus alleviate productivity constraints more effectively. Specifically, we study the implementation of an information technology (IT) which provides real-time tracking of meal orders placed at restaurant drive-thru windows. The technology uses sensors in drive-thru lanes as well as data on order fulfillment, and automatically synthesizes these and reports them in real time on monitors located inside the food production and cashier areas of the restaurant. Managers can use the technology to track key performance indicators (KPIs) such as the number of cars currently queued in the drive-thru; the average time between when an order was placed and when it was completed (so-called takt time); and the average time a customer has to wait on the queue before placing an order. These KPIs help the manager identify bottlenecks in various segments of the production process and take appropriate actions to mitigate them.

We leverage the staggered roll-out of this technology across restaurants to identify its effects on productivity as well as the intensity and targeting of managerial inputs. There are several features of this context that are conducive to accomplishing our study's goals. First, almost all production processes, employee and manager functions, and capital are by design identical across the restaurants; this setup lends itself well to empirically detecting the impacts of the performance monitoring technology's introduction separate from other determinants of restaurant-level productivity. Second, while managerial time and attention allocations are in general quite difficult to capture empirically, there is one very clearly defined (and contextually important) indicator of managerial input in our data, namely, the amount of on-the-job training the manager provides to employees. This is one of the key responsibilities of managers in this context: managers spend a large fraction of their time training workers in new skills and refreshing existing skills. We are fortunate to observe these investments in our data, allowing us to measure the quantity as well as the targeting of managerial inputs across workers and stations, and how these change in response to the introduction of the monitoring technology. Finally, the time span of the roll-out and the availability of granular administrative data allows us to estimate dynamic effects beyond the short-run.

To better understand mechanisms through which monitoring technology could improve performance, we set out a simple theoretical framework in the vein of Holmstrom and Milgrom (1991). Performance monitoring technology allows the manager to better observe drive-thru productivity in real-time, which in turn affects the optimal targeting and quantity of training inputs. The framework elucidates how a clearer signal of drive-thru productivity due to the technology's introduction should cause managers to increase training investments in food production tasks in particular (which feed into both drive-thru and counter sales), thus speeding up the flow of orders and increasing drive-thru productivity. Given the trade-off between increased kitchen efficiency and decreased attention to the counter due to managerial multitasking concerns, the sign of the predicted impact on counter sales is ambiguous.

In line with these predictions, we find that the introduction of the technology indeed caused a substantial increase in managerial input – as measured by training investments – particularly for kitchen tasks. Correspondingly, overall sales increased by about 4.5 percent (or about 3400 USD per week) after the technology's introduction, an effect driven by drive-thru sales, which increased by nearly 6 percent. Counter sales increased slightly as well, though effects were substantially smaller than the impacts on drive-thru sales. Impacts came entirely through an increase in the number of orders, i.e., the number of customers served. There was very little change in the average value per order. This is congruent with the particular features of the technology we study, in that the main new information conveyed to managers via the technology was related to the speed of completing orders (and not to the average value of orders).

Given the strong positive correlation between drive-thru and counter sales in the data prior to technology implementation, and the technical complementarity between the kitchen and the two points of sale, this muted effect on counter sales might be interpreted as evidence of a multitasking trade-off. That is, we might expect that the closer monitoring of the pace of drive-thru order fulfillment yields gains in drive-thru sales at the expense of some other less closely monitored dimensions of performance. However, additional results from the study of order composition and, perhaps more convincingly, inventory waste, do not support a multitasking trade-off interpretation.[1]

Strikingly, treatment effects on sales diminish 6-8 weeks after the technology's introduction. This is true both for effects on drive-thru as well as counter sales. We uncover a key mechanism for these dynamics that is related to the depreciation of skills over time. Specifically, our results suggest that while increased training in new skills due to the technology's implementation was initially effective at raising productivity, "skill capital" depreciates over time and serves to dampen the impacts of the new technology. Indeed, in the administrative data, we can distinguish between training in new skills versus *refresher* training, which managers provide to workers who are already certified in certain skills in order to re-up their knowledge and sharpen their practice in those areas. We find that restaurants in which managers tended to invest in refresher training frequently prior to the technology implementation exhibit large treatment effects in the use of this type of training post-implementation. In contrast, restaurants in which managers do not use refreshers pre-implementation exhibit no treatment effects on this type of training. Importantly, this analysis is conditional on baseline store productivity by time interactions, which absorb any general changes in the productive value of overall managerial quality as a result of the introduction of the technology.

Accordingly, when we perform a similar analysis of impacts on sales across this refresher training distinction, we find that these restaurants in which managers did not tend to utilize refresher training prior to implementation, and accordingly did not respond to the technology implementation with greater refresher training, account for the decline in treatment effects on sales over time we see on average. On the other hand, the high refresher training restaurants retain most of the benefits of the monitoring technology on sales even after the initial months following implementation. This final set of results – the fact that not all managers optimized refresher training investments despite the very similar environments in which they managed and the identical tools available to them – echoes the literature on the importance of so-called behavioral concerns, especially managerial inattention, in the determination of managerial quality (Adhvaryu et al., 2021a,b; Bandiera et al., 2014; Frederiksen et al., 2020; Halac and Prat, 2016; Hortaçsu et al., 2017; Kahn

[1] Note that in our setting, counter sales are partially observed (just as drive-thru sales are), even if close monitoring of counter productivity is not enabled by the technology. As Hong et al. (2018) argue, a cleaner test of multitasking would involve a performance measure which is not readily observed by the principal (i.e., central corporate management, in our case), and also one which is not as positively correlated with the observed performance measure so as to improve mechanically despite a lack of effort devoted to it. We thus study waste as a measure which, if anything, *ex ante* worsens as sales increase due to a speed-precision style trade-off and find that waste did not significantly rise in response to the monitoring technology implementation. We propose that this absence of multitasking issues could reflect the specific managerial response which enabled sales to rise. That is, the increased managerial training input may have reduced the likelihood of errors which often lead to increased waste when crew members in the kitchen are working at a faster pace.

and Lange, 2014).

Our main contribution is to the understanding of the relationship between management and firm productivity (Adhvaryu et al., 2021b; Bandiera et al., 2020; Bloom et al., 2013; Bloom and Van Reenen, 2007; Friebel et al., 2022). Performance monitoring has long been held as an essential component of managerial quality, and is a mainstay tool for managers in organizations of all kinds (see, e.g., (Bloom and Van Reenen, 2010)). Yet despite this ubiquity, there is still little rigorous causal evidence on whether – and, just as importantly, why – performance monitoring affects productivity. To our knowledge, only one recent paper, by Gosnell et al. (2020), tackles this question head-on, in a study of highly skilled workers (airline pilots). Our study contributes several novel findings related to this question. First, like Gosnell et al. (2020), we find that performance monitoring can have substantial positive impacts on productivity, even in a low-tech, labor-intensive context.[2] Second, managerial investments in worker skills play a key role in mediating this productivity effect. In particular, the introduction of performance monitoring technology enabled clearer insight into bottlenecks, allowing managers to deliver more skill investments to workers in key production areas. This result shows how performance monitoring can complement on-the-job training in enhancing firm productivity by enabling better targeting of skill enhancement (Bloom et al., 2019, 2016). Third, initial impacts diminish considerably after 6 weeks on average, but managers who use refresher training intensively can achieve sustained productivity gains, suggesting that the impacts of skill depreciation are critical – yet are often overlooked by managers (Bandiera et al., 2014; Frederiksen et al., 2020; Halac and Prat, 2016; Hortaçsu et al., 2017; Kahn and Lange, 2014).

We also contribute to the literature on the impacts of information technology on organizational design and firm productivity (Acemoglu et al., 2007; Bloom et al., 2014, 2012; Bresnahan et al., 2002). The closest work to ours is that by Hubbard (2000, 2003) on trucking, Bartel et al. (2007) on valve manufacturing, and Athey and Stern (2002) on emergency health care services. This literature emphasizes the potentially central role of managerial decision-making and coordination in actualizing the value of IT. We build on these important studies by leveraging a novel measure of managerial investment in workers – skills training – along with granular data on productivity, which together allow us to peer further inside the "black box" of IT adoption to uncover precisely in which ways managers shift their everyday behaviors in response to new information from performance monitoring IT.

Finally, we contribute to the body of work testing seminal theories of multitasking

[2] This result is similar in spirit to Jackson and Schneider (2015), who study the impacts of provision and monitoring of checklists for auto mechanics, and Duflo et al. (2012), who examine the impacts of monitoring for schoolteachers.

(Baker, 1992; Holmstrom and Milgrom, 1991). Hong et al. (2018) review the state of the literature as consisting of several studies showing mixed evidence of either subtle trade-offs or none at all,[3] and then provide perhaps the strongest evidence to date of a quantity-quality trade-off in a real-world workplace.[4] They emphasize the importance of the principal's inability to observe the not-contracted-upon dimension in generating multitasking trade-offs. Chetty et al. (2014) provide relevant evidence in the context of journal refereeing and Amodio and Martinez-Carrasco (2019) study similar issues of asymmetric information regarding input quality between managers and workers in a multitasking environment using data from an egg producer in Peru.

In addition to this issue of observability, our results emphasize that the shape of the production technology turning common inputs into multiple dimensions of output can also discipline the scope for trade-offs in multitasking environments. That is, we find store managers are not trading off counter sales for drive-thru sales, both because counter sales are also just as complementary with kitchen production as are the now more closely monitored drive-thru sales and perhaps also because counter sales are still (imperfectly) monitored by the principal. But we also find that waste, which is not readily observed by the principal in our context, does not increase as might be expected given the speed-precision trade-off described to us by management and observed in the pre-implementation data. This is likely because in this context the specific managerial response induced by the monitoring technology was to increase training, which might reduce waste-inducing errors that otherwise rise with kitchen production pace. Our results on the speed-waste trade-off relate closely to the evidence in Atkin et al. (2017) showing that, when monetary incentives are linked to output quantity and not to waste, workers are likely to resist the adoption of waste-saving (but potentially speed-reducing) technology at the expense of firm profits.

2 Context

2.1 The QSR Industry in Puerto Rico

Puerto Rico is home to a robust QSR sector with a market size of approximately four billion USD in annual sales, directly employing more than 60,000 workers.[5] We partnered

[3] Hong et al. (2018) provide a comprehensive review of the empirical literature. Much of this evidence comes from education (see, e.g., Fryer et al. (2012); Glewwe et al. (2010)) and health settings (see, e.g., Feng Lu (2012); Mullen et al. (2010)).

[4] Hansman et al. (2020) provide nice complementary empirical evidence of the implications of similar trade-offs for the boundary of the firm in the Peruvian fish meal industry.

[5] Statistics reported by the Asociación de Restaurantes de Puerto Rico (ASORE), last updated in 2018.

with one of the leading QSR franchises in Latin America and the Caribbean. Over the period of analysis, the chain owned more than 60 fast food restaurants in Puerto Rico, largely concentrated in the San Juan region and in the perimeter zone of the island. On average, monthly sales across across all stores in Puerto Rico amounted to 4.1 million units (items) sold through approximately 1 million transactions over the observation period.

2.2 Production Process

In our partner's restaurants, an order may be placed at two points of sale – the counter (in-store) or the drive-thru.[6] Once an order is placed it is recorded and automatically displayed on a monitor in the kitchen.[7]

In the kitchen, orders are completed in five stations: grill, fryer, assembly (including condiments), soda fountain, and desserts (see Appendix Figure B.1). Each employee is typically assigned to a specific station within the production line for a given shift. For example, when a burger is ordered, one worker toasts the bun, places it in a box or wrapping paper, and adds condiments and vegetables. Afterwards, another worker adds the meat from the grill, after having cooked it according to strict specifications, and a third worker then closes the packaging and adds it to the bag or tray for the order along with any other items on the order ticket. The final assembled order for that ticket is then delivered at the counter or at the pick up window of the drive-thru by the worker staffed at the corresponding point of sale. In addition to these kitchen operations, other support operations occur in parallel (e.g., the cleaning of facilities, machine maintenance, security and parking management, etc.).

2.3 Store Managers

A store manager is in charge of the overall functioning of each restaurant, leading teams of approximately 35 people on average spanning multiple shifts and the above stations. Store managers oversee all operations which ensure efficient daily operations, including calibration of equipment (time and temperature), kitchen and dining area sanitation, product availability (inventories), final product quality checks, waste, and perhaps most importantly, personnel management (employee schedules, recruitment, and training). Given a manager's critical role in all operations of the restaurant, she can have a large impact on the restaurant's performance.

[6] Online orders are also now available at some of our partner's restaurants across Latin America, though this technology has not yet been introduced in Puerto Rico.

[7] The origin of the order, i.e., whether it was placed at the counter or through the drive-thru, is not specified on this screen.

In order to be a store manager, a crew member first must complete online and in-restaurant training sessions for approximately six weeks, after which they achieve certifications in all of the following areas: (1) production, where they learn how to run the kitchen, maintain the production area (food safety), and how to effectively maintain the restaurant sales and inventory records; (2) counter and drive-thru, where they learn about money management and administrative procedures like invoicing, cash register operation, and different payment methods, as well as final order delivery via each point of sale; (3) additional responsibilities, where they learn about cleanliness, hospitality and customer service, parking and playground maintenance.At this stage the candidate may take on some daily leadership responsibilities as a team leader. Over the next six months, candidates undergo further specialized training in two to three week courses on "client experience", "product specialization" and "people management." After these additional courses, the candidate is able to serve as a store manager.

All of this substantial training ensures both that managers are able to effectively oversee the many critical operations of a restaurant, but also that they are experts who can, in turn, train crew members to effectively perform operations at the station to which they are assigned. As discussed below, crew members must first be trained and certified by managers on a station before being able to work at that station. Note that the manager spends the entirety of their time floating between stations monitoring work flow, with the majority of these stations (e.g., counter, drive-thru, and kitchen operations) being generally visible from any point in the kitchen (as shown in Appendix Figure B.1). Accordingly, we are told that managers are in general able to easily detect and address major causes of delay in customer service and acute fluctuations in service pace such as those due to shirking or distraction among some workers, errors in operating machines, or machine malfunctions.

On the other hand, managers may not fully realize the impacts of frequent but small errors or inefficiencies on ultimate service delivery pace, and therefore may not act to prevent such issues if they seem small or inconsequential or are hard to detect. More subtle imbalances in productivity and slow onset or chronic reductions in pace would be difficult to observe with the naked eye without some metric or timer reflecting pace, particularly relative to some peak pace achieved in the past or a target pace. These seemingly small productivity imbalances across stations and subtle but chronic reductions in pace can accumulate to greatly impact the rate at which orders are taken, produced, and delivered (i.e., takt time). It is mainly for this reason that central corporate management decided to invest in the pace measurement and reporting tool described below.

2.4 Crew Members, Stations, and Training

Crew members execute all daily restaurant tasks, including manning points of sale and kitchen operations, as well as general restaurant support tasks like cleaning, maintenance, and dining room hospitality. Crew members can rotate across any "station" for which they have been trained and certified. There are 23 different stations within each restaurant at which crew members can be certified.

Managers are responsible for evaluating and certifying crew members at a station.[8] Crew members participate in this training and evaluation during their normal shifts. At the counter and drive-thru stations, workers learn about cash register operation and the use of different payment methods, as well as final order delivery to the customer. In the kitchen or production area stations (i.e., fryer, grill, assembly, drinks, and desserts) they learn to produce food orders of all kinds. Additionally, workers can be trained in restaurant cleaning, hospitality, customer service, parking, security, and playground maintenance. Managers may also choose to provide refresher training in any of the stations to a crew member as needed.

2.5 Organization of the Firm

The firm with which we partner is the exclusive franchisor in the region for one of the world's largest QSR brands. The firm owns and operates directly the vast majority of the thousands of stores across Latin America and the Caribbean, with the remaining stores being sub-franchised by the firm to external owners and operators. This pattern holds true in Puerto Rico as well.

All of the stores in our sample are fully owned by the firm. As discussed above, each store employs tens of workers and is managed fairly autonomously by a single store manager. That is, the store manager makes, for example, daily personnel and inventory management decisions with little oversight or need for approval from central corporate management. Of course, as a corporate-owned franchise, materials and equipment, suppliers, menu items, pricing, and training curricula are all standardized by the firm (or mostly by the global QSR brand with some adaptation to the regional context decided by the firm).

Importantly, the main data systems are also provided by the global QSR brand. Accordingly, the most carefully maintained and harmonized data are the transactions records from which total sales are calculated, as these inform the calculations of the royalty payments to the global brand. That is, central corporate management of the partner firm is most easily and frequently able to track sales of each store. Personnel records

[8] The actual training prior to evaluation and certification is either conducted by managers themselves or sometimes by trainers, who are senior crew workers.

and inventories (and accordingly waste), for example, are kept in separate and less easily referenced systems.

Accordingly, any awards, bonuses, and promotion decisions are linked to store sales, to the exclusion of other likely important metrics like waste, training investments, and worker turnover. Central corporate management was aware of the limitations of tracking only sales and explicitly introduced the performance monitoring technology we study to supplement sales with a metric more reflective of productivity. The drive-thru performance monitoring technology which was introduced represented a readily available solution on the market which was easily adopted and covered the primary point of sale for stores in Puerto Rico (i.e., drive-thru).

2.6 Drive-thru Performance Monitoring Technology

The drive-thru service was one of the brand's most significant innovations in the industry, and in Puerto Rico nearly all of our partner's restaurants provide such a service. The drive-thru contributes approximately two-thirds of the total sales of the chain's restaurants in Puerto Rico. Given the importance of the drive-thru to the company's overall sales and performance on the island, beginning in 2019 the chain implemented in staggered fashion a drive-thru performance monitoring system in 51 Puerto Rico restaurants to track productivity at the primary point of sale.[9] The technology in question was developed by a firm specialized in creating products and services for the QSR industry worldwide. It was designed to help managers and worker crews reach greater levels of productivity through a comprehensive system that allows the team to monitor performance in real time.

The system has three technological components: underground detectors, a dashboard showing the drive-thru timer, and a data storage system. Underground detectors installed at the order, cashier, presenter and put forward points (see Appendix Figure B.2) record the exact time in which a car reaches each point.[10] The dashboard (see Appendix Figure B.3) is displayed on a monitor installed in the drive-thru station, by the presenter window, and visible to the kitchen crew. It displays detection point times (i.e., time a car spends at the order, cashier and presenter windows), average times (updating after every car departure), and car counts (per hour, pace estimate based on the past 20 minutes speed of service).

Armed with this information the manager can be better informed on the causes of a bottleneck at any point of the process and address it appropriately. The issues could be

[9] Some additional restaurants implemented the technology before the start of our observation period, and 4 stores still had not implemented the technology by the end of our observation period. We do not consider these stores in the empirical analysis in this paper.

[10] Most stores in our data have a single drive-thru lane. Only three restaurants have two drive-thru lanes.

related to motivation or attention, but we are told that those types of acute issues are generally easily observed first hand by the manager while making rounds. As discussed above, corporate management invested in the technology with the aim of identifying more subtle but systematic opportunities for improving pace. Broadly speaking, the resulting adjustments could be of two broad forms: 1) making existing workers at bottleneck stations better or faster, or 2) adding workers to bottleneck stations to balance productivity with other stations achieving faster pace. We study below both the degree to which managers seem to respond via each of these two ways, as well as which portion of operations (i.e., point of sale or kitchen) seems to receive the most adjustment.

The upper management at the franchise decided the order of technology roll-out across restaurants. Appendix Figure B.4 shows a map of the restaurants, color-coded according to timing of installation. No clear geographical pattern appears in the implementation of the technology. The event studies we present later in the results also show no evidence of pre-trends across stores which had implemented compared to those which had not yet. Nevertheless we also present a host of robustness checks to bolster the claim to internal validity of our study design.

3 Data and Descriptive Statistics

We use data from three main sources for 51 restaurants in Puerto Rico which implemented the monitoring technology over the observation period. The first source is data on all the transactions of the stores that took place during our sample period. As mentioned above, these transaction data are the only records the corporate central management collects and analyzes on a regular basis. The introduction of the performance monitoring technology was motivated by an interest in supplementing sales records with a productivity measure. The second is employment and training data, and the third is waste and yield data for each store. Finally, we combined these outcome data sets with data on the timing of the monitoring technology roll-out.

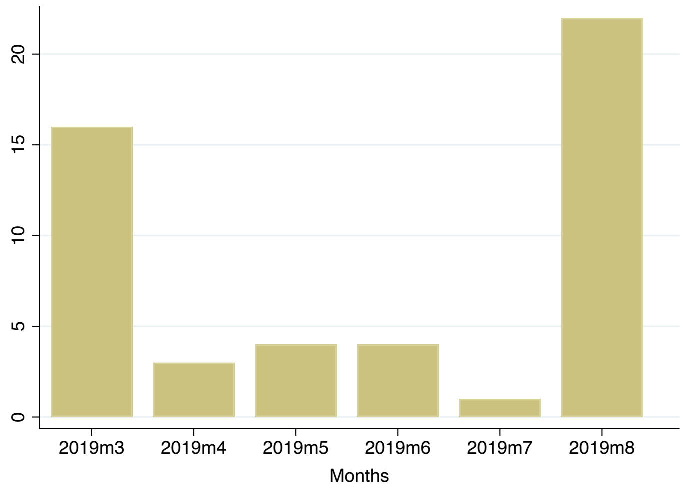
3.1 Adoption of the drive-thru monitoring technology

For the 51 stores which implemented the technology during the observation period, we know the week in which each store implemented the monitoring technology. Since in general, stores implemented the technology at different times, we can measure the technology's impact on different store performance indicators before and after the implementation.[11]

[11] Unfortunately, we do not have information collected by the monitoring technology (e.g., drive-thru average order fill-out times and car counts at each restaurant) *before* the implementation of the monitoring technology.

Figure 1 shows the staggered implementation, starting in April 2019, and intensifying in August 2019.[12]

Figure 1: Monthly implementations



Notes: Figure 1 presents the staggered implementation of the monitoring technology in Puerto Rico's stores, starting in April 2019 until September 2019. The figure shows the number of stores that implemented the technology by month.

3.2 Transaction data

We have panel data at the restaurant-day-transaction level from March 2018 to October 2019. The dataset includes 30 million transactions that took place during this period. In addition, it includes detailed information on each ticket issued and printed by each store, such as the point of sale of the order (counter or drive-thru), the items included in each ticket (e.g., two ice creams and a cheeseburger), and the value of each item as well as that of the total ticket. As mentioned above, these records are particularly rich and well-archived because they inform calculations for royalty payments to the franchisor (which designed most of the basic data systems for franchisee use).

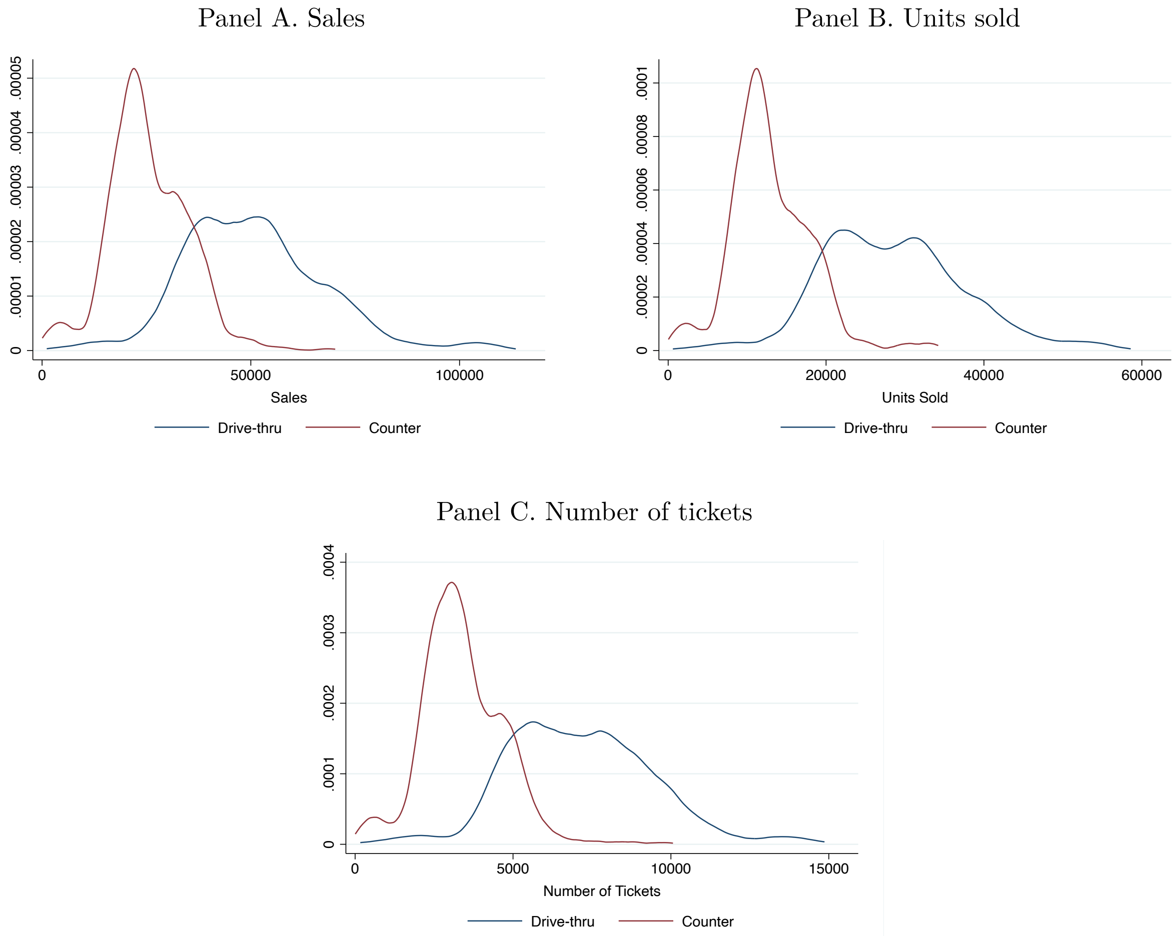
We focus on three performance measures: sales, units sold, and the number of tickets. We aggregate total sales, units sold, and the number of tickets issued for each biweekly period for each store. Figure 2 plots the distribution of the three performance measures for the two points of sales. The drive-thru represents roughly two-thirds and the counter one-third of total sales. We see that total sales, units sold, and the number of tickets

[12]Figure B.4 in the Appendix, shows the geographic implementation of the monitoring technology and confirms no systematic spatial pattern in the roll-out across stores. Given that the majority of the roll-out occurred in two of the months of the observation period (though week of implementation varied), we investigate robustness to just using stores implementing during those two waves.

have similar distributions within a point of sale (drive-thru and counter) and that the relationship between the distributions for each point of sale is similar across these measures. However, note that for all three measures there is a great deal of variation in performance across stores.

Table 1 presents summary statistics of the main variables of interest for the stores in our sample. Our sample comprises 2,074 store x biweekly observations. The average of sales across stores and biweekly periods is around $76,000 per period. The average of sales is around $50,000 at the drive-thru and $25,000 at the counter.

Figure 2: Performance measures distribution



Notes: Figure 2 plots the distribution of key performance measures divided by the two sales points (drive-thru and counter) for the stores in our sample from March 2018 to October 2019. Sales is the average of the total value of the sales, units sold is the average total number of items sold, and the number of tickets is the average number of tickets issued for each biweekly period and each store.

Table 1: Summary statistics performance measures data

	Whole store	**Drive-thru**	**Counter**
		Panel A. Sales	
Mean	76,080	50,734	25,432
Stand. Dev.	24,979	16,955	9,504
		Panel B. Units sold	
Coefficient	41,432	28,488	12,989
Stand. Err.	13,478	9,138	5,133
		Panel C. Number of tickets	
Coefficient	10,531	7,137	3,405
Stand. Err.	3,354	2,276	1,293
Observations	2,074	2,073	2,069
Stores		51	

Notes: Table 1 presents the mean and standard deviation of the main performance key measures for the stores in our sample from March 2018 to October 2019. Sales is the average of the total value of the sales, units sold is the average total number of items sold, the number of tickets is the average number of tickets issued for each biweekly period and each store. The first column provides summary statistics of the measures for both sales points (drive-thru and counter), the second for the drive-thru, and the third for the counter.

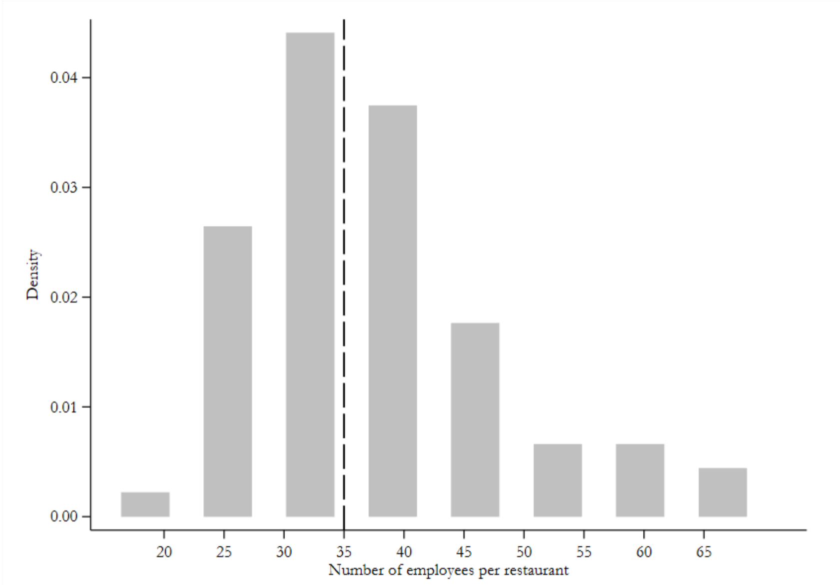
3.2.1 Order Composition

We compute two measures related to the order composition: the average value per ticket and the number of distinct items per ticket. For the average value per ticket, we compute the average value of all the tickets issued for each biweekly period for each store. Figure B.6 Panel A plots the distribution of the average value per ticket for the two sales points. Both distributions are similar, but the average value per ticket is slightly higher for the tickets issued at the counter than at the drive-thru: the average value per ticket is around $7.10 at the drive-thru and $7.50 at the counter. Figure B.6 Panel B shows the number of distinct items per ticket, and the bar represents the median equal to 3 items per ticket. We define complex orders as those for which the number of distinct items per order is higher than the median. In our sample, the share of complex orders is equal to 38%.

3.3 Employees, Training and Stations

We have access to the firm's employee data for the stores in our sample. The data provides information on nearly 3,000 workers distributed across 23 stations that compose each restaurant's production process. Figure 4 shows the distribution of employees at the store level. Note that the number of employees per store varies from 15 to 65 employees.

Figure 4: Employees distribution



Notes: Figure 4 plots the distribution of the number of employees per store, for the 51 stores considered in our sample from January 2019 to November 2019. The median is 35 employees per store.

We also have access to worker-level training data (or certifications) spanning from January 2019 to November 2019. There are three major types of training: first, there is the training of new workers in a new skill (this is done during on-boarding and can be used to measure new hires); second, there is the training of existing workers in new skills; and finally, there is refresher training. Refresher training refers to those courses which managers provide to workers who are already certified in certain skills in order to re-up their knowledge and sharpen their practice in those areas. Finally, we know the exact date of the training program (or more specifically, the evaluation following training) and if the training was approved or unapproved (i.e., whether the worker passed or failed the evaluation).

Table 2 shows that there are on average ten new training programs successfully completed per biweekly period per store, and the approval (i.e., passing) rate is around 60%. The number of successful refresher training programs completed is about 6 per biweekly period per store. Notice that the standard deviation is large for both new and refresher training, indicating substantial variation in the amount of training taking place across stores and weeks. Table B.1 presents a summary of the different types of training programs. These programs correspond to the different stations of each store.

Table 2: Summary statistics traning data

	All	**Approved**	**Unapproved**
		Panel A. New skills	
Mean	15.59	9.57	6.02
Stand. Dev.	15.63	12.83	6.72
		Panel B. Refresher	
Coefficient	10.18	6.43	3.76
Stand. Err.	11.37	9.09	4.64
Observations		1,202	
Stores		51	

Notes: Table 2 presents the mean and standard deviation of the training program certifications. Panel A provides summary statistics for new skills in a specific station where a crew member has not received training before. Panel B provides summary statistics for the refresher training, which are skills or certifications taught and approved previously. The first column provides summary statistics of all the certifications. The second column provides statistics only for the approved certifications, and the third column for unapproved certifications.

3.4 Waste data

Finally, we explore the impact of the monitoring technology on the proportion of waste in restaurants and inventory use. In our context, waste is measured in input units. There are four types of waste: (1) promotions in the form of coupons or free products offered by the store; (2) complete waste measuring when an item is wasted in its entirety (e.g., the complete burger is returned by a customer); (3) incomplete waste measuring when a component of the product is wasted (e.g., the patty is burned); and (4) employee consumption measuring products consumed by the employees as part of their benefits (e.g., during their lunchtime).

We have daily data on inventory and data for the four types of waste by input products from March 2018 to November 2019 for 362 input products used. To construct our waste measure, we calculate the proportion of usage by subtracting the final inventory from the initial inventory (which gives us the daily usage for each input) and divide the quantity wasted of each input by the total quantity of that input used to calculate percentage waste. Finally, we take the average of this percentage wasted for each biweekly period for each store. Table 3 presents the average proportion of quantity wasted for each type of waste.

We note above that these data are maintained in a separate system than transactions and are not easily or frequently referenced by central corporate management. That is, the manager of each store knows these metrics but also knows that these metrics are not generally used to assess her performance for bonuses or promotions. We are told anecdotally that waste often rises when restaurants are busier, for example during peak hours, because making more orders or working faster often means more errors are made. We find using the takt time data from the week that the technology was implemented that indeed waste rises with the speed of order fulfillment.

Table 3: Summary statistics waste data

	Mean	**Std. Dev.**
Promo	1.02	0.77
Complete waste	1.07	0.52
Incomplete waste	1.75	1.61
Employee consumption	1.51	0.75
Total waste	5.35	2.69
Observations	2,385	
Stores	51	

Notes: Table 3 presents the mean and standard deviation of waste proportions. The first row provides summary statistics for promotions—in the form of coupons or free products. The second row is for complete waste, which occurs when an item is wasted entirely. The third row is for incomplete waste, which occurs when an item of the product is wasted. The last row is for employee consumption—consumed by the employees as part of their benefits during the shift.

4 Empirical Analysis

Our aim is to compare different store performance indicators before and after the implementation of the technology. We first test if the introduction of the new technology impacted key performance measures, such as sales, units sold, and the number of tickets. We then look for evidence of multitasking trade-offs in counter sales, composition of orders, and waste. Then, we explore how managers responded to the introduction of the new technology to illuminate mechanisms by which impacts on store performance were achieved. Specifically, we estimate the impact on hiring and managerial training input. Finally, we explore the degree to which managerial investments in refresher training help to explain the dynamics of the effects on sales.

4.1 Theoretical framework

To better understand the mechanisms through which improved monitoring increases performance, we set out a basic theory based on the seminal work of Holmstrom and

Milgrom (1991). A formal mathematical model is presented in Appendix A; we provide the intuition for and prediction of this framework below.

The model has three types of agents: the principal, the store manager, and workers. More specifically, each store is equipped with a counter and a drive-thru to take customer orders, and a kitchen in charge of food preparation. Hence, the employees in the store perform in one of the three positions mentioned. We stipulate that employees can become more productive in their positions depending on the amount of training (measured in time invested) that their store manager provides.

Notice that the store generates sales at the counter and the drive-thru, although their productivity is directly linked to the kitchen's efficiency. Therefore, when the principal is designing the manager's contract, it is enough for her to decide on a set of rewards for the net production of those two stations. Naturally, the principal's problem is then how to balance those rewards to maximize the store's production, while the manager's problem is how to invest her training time optimally given this reward structure. The manager's choices related to the vector of optimal training investments is based on the compensation plan specified in her contract, as well as other factors such as the cost of training and her risk aversion.

We assume that there are unobserved and random factors in the production process that affect sales (e.g., worker's effort and motivation, or the customer's rate of arrivals). A unique equilibrium for the simultaneous problems of the principal and the manager can be attained under certain conditions laid out in Appendix A (Propositions 1 and 2). In particular, for the manager's problem, Corollary 1 illustrates that the manager's rewards are directly linked to her investment in training workers (which in turn increases sales through greater worker productivity).

We are ultimately interested in describing the behavior of optimal contracts as well as training inputs in response to the introduction of monitoring technology – i.e., a decrease in the uncertainty of drive-thru production. In Corollaries 2 and 4 we obtain that a better forecast will always enhance the training of workers positioned at the drive-thru window and in the kitchen, and also increase total sales. The sign of training responses for workers at the counter station are ambiguous; there are certain conditions that need to be satisfied for this response to be positive, which are laid out in Corollaries 2 and 3.

4.2 Baseline Specification

Our baseline specification estimates the impact of the implementation of the drive-thru monitoring technology on different restaurant performance measures using 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 estimate the dynamic consequences in the post-implementation period. We estimate the following specification:

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

where $Y_{s,t}$ is the performance measure of store s at biweekly period t, D^k_{st} is a relative time to treatment indicator for whether the store had implemented the technology in (biweekly) period $t-k$, defined as $D^k_{st} = 1[t = \tau_s + k]$ for $k \in (\underline{C}, \overline{C})$, $D^{\underline{C}}_{st} = 1[t \leq \tau_s + \underline{C}]$, and $D^{\overline{C}}_{st} = 1[t \geq \tau_s + \overline{C}]$, where $1[\cdot]$ is the indicator function, k indexes the set of time indicator variables, and τ_s is the first biweekly period when store s introduces the monitoring technology.

The parameters of interest δ_k for $k \in [\underline{C}, \overline{C}]$ measure the impact of the implementation before, during and after the event. We normalize $\delta_{-1} = 0$ and set $\underline{C} = -6$ and $\overline{C} = 6$. Finally, our specification controls for biweekly time period fixed effects θ_t and store fixed effects Φ_s. We cluster standard errors at the store level for inference.[13]

4.3 Identification

Identification relies on two features of the roll-out of the monitoring technology: a) the implementation of the technology was unanticipated by the managers of the stores, and b) the timing and sequence across stores was uncorrelated with manager- and worker-specific characteristics. These assumptions are consistent with the anecdotal evidence shared by the partner firm and would be empirically validated by a lack of pre-trends in our analysis below.

Additionally, the panel data structure allows us to interpret the statistical significance of these coefficients as evidence of the causal relationship between the monitoring technology implementation and store performance outcomes, provided that the treatment of the technology varies over time and is uncorrelated with store-specific performance shocks (Blundell and Dias, 2009). That is, shocks with onset coincident with the timing of the implementation and correlated with the performance of the stores after but not before the implementation, are a threat to our identification. A lack of meaningful effects in the pre-period once again would suggest that this type of dynamic selection is unlikely. Accordingly, provided that we find common pre-trends for all outcomes in the event study analysis below, we contend that stores yet to implement the monitoring technology form a credible counterfactual for stores which have already implemented the technology, after

[13]When a "never-treated" group of units is "too small" Callaway and Sant'Anna (2020) suggest that researchers may favor the conditional parallel trends assumption based on "not-yet-treated' groups. Accordingly, our main specification excludes the never-treated stores (4 out of 55 for which we have complete data over the observation period); thus, we compare the outcomes of stores in event year k to the outcomes of the stores that implemented the technology in the future in the year before their event.

accounting for time-invariant differences between stores and biweekly periods fixed effects. We report robustness results in section 5 to support this assertion.

4.4 Effects on main performance measures

We first implement the baseline event-study specification (1) to test if the introduction of the new technology impacted key performance measures. Panels A-C of Figure 5 illustrate the event-study coefficients for sales, units sold, and number of tickets for the whole store. Reassuringly, we find no evidence of selection into the technology implementation based on past store performance indicators (i.e., no evidence of divergent pre-trends in any of these outcomes). Instead, it is only after stores implement the new technology that we observe a positive effect on sales, units sold, and the number of tickets. Although we observe some effects the same week of the event in Panels A-C of Figure 5, the largest impacts (i.e., 10% higher sales, 9% more units sold, and 11% more tickets) occur three biweekly periods after the implementation. Finally, after three biweekly periods, the effects of the technology begin to diminish. That is, though they remain positive after the first 2 months, they are measurably reduced in magnitude and no longer statistically significant.

4.5 Multitasking

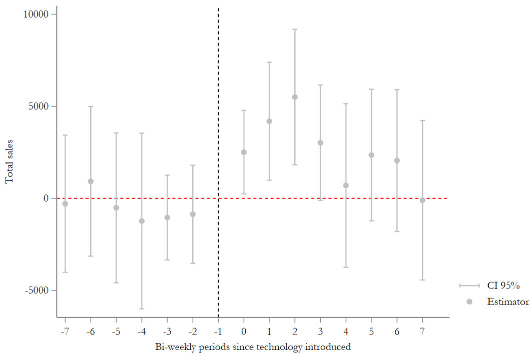
4.5.1 Counter Sales

We further study the effect of the technology implementation on the performance at the counter and the drive-thru, separately. The performance monitoring technology provides real-time tracking of fulfillment rates of orders placed at restaurant drive-thru windows, so we should expect a significant increase in sales at the drive-thru point of sale. Based on anecdotal evidence, we also expect some complementarities between the kitchen-counter and kitchen-drive-thru production relationships. These complementarities, if strong enough, should generate a strong positive correlation between counter and drive-thru order fulfillment rates and, in turn, sales. We indeed find that this is true in the pre-implementation sales data. We formalize in our model the common complementarity between each point of sale and the kitchen as the source of the strong positive correlation between performance at the two points of sale seen in the data. If these complementarities are substantial, our model predicts that sales at the counter should also increase.

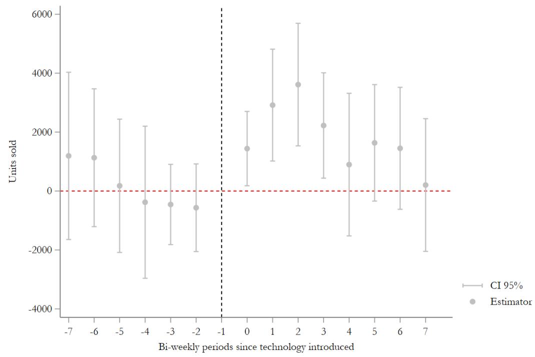
Figures C.1, C.2, and C.3 show the results for sales, units sold, and the number of tickets, respectively, for the orders placed at the counter and drive-thru. We find a similar pattern for these performance measures for the two sales points, although the effects are larger (close to 6%) for the drive-thru than the counter (close to 1.6%). Finally, in Table

Figure 5: Effects of technology implementations on performance measures

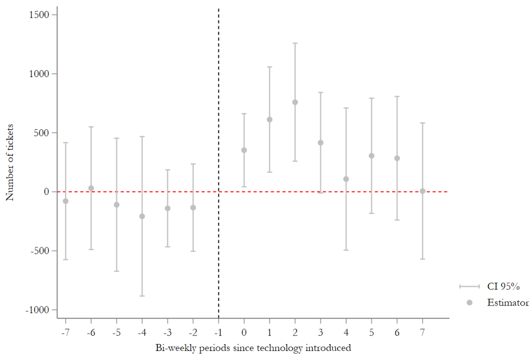
Panel A. Event study estimates on sales



Panel B. Event study estimates on units sold



Panel C. Event study estimates on number of tickets



Notes: Panels A to C of Figure 5 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 stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows the impact on sales, Panel B on units sold, and Panel C on the number of tickets. The vertical line represents the time of the treatment.

C.1 we run a difference-in-differences regression, replacing the full set of time to treatment indicators with a post-treatment dummy (before and after).[14] We again control for store and biweekly period fixed effects and cluster the standard errors at the store level. Table C.1 shows that the performance monitoring technology increased sales by 4.4% for the whole store, 5.6% at the drive-thru, and 1.8% (though not significant) at the counter.

These results are consistent with both the predictions of the model and the empirical evidence that performance is strongly positively correlated across the two points of sale. That is, the results show that improvements in drive-thru performance do not appear to come at the expense of performance at the counter as would be predicted by the standard multitasking model. That said, the attenuated magnitude of the effect at the counter might still be consistent with a multitasking trade-off and as such studying an outcome which does not have a strong positive correlation with the now more closely monitored drive-thru sales performance measure would be a clearer test of multi-tasking. We propose that waste is such an outcome and study it below.

4.5.2 Order Composition

In a similar vein, we might expect that since the order fulfillment time is being monitored but the value, complexity, or size of each order is not, workers or managers may prioritize filling small or simple orders or may even refrain from encouraging customers to add items to their order to help minimize fulfillment times. This would be another form of a multitasking trade-off in this environment.

Panel A of Figure 6 shows the results for the average value per ticket. We find a small, negative effect on the average value per ticket, which becomes statistically significant after four biweekly periods of the implementation. With a maximum magnitude of only around 7 cents we interpret these coefficients as well-estimated null results.

Next, we explore if the introduction of the monitoring technology changed the composition of the average order. Panel B of Figure 6 shows the event-study coefficients for the share complex orders for the whole store. Consistent with the average value per ticket results, we observe a small decrease in the share of complex orders. Panel C and D of Figure C.6 show that this decrease is driven by a decrease in the share of complex orders at the drive-thru point of sale. However, given the magnitudes, we again interpret these coefficients as well-estimated null results.[15]

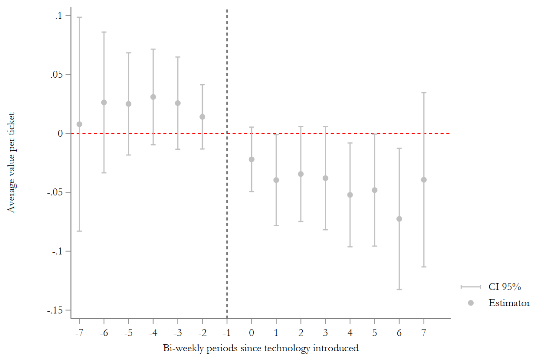
Taken together, these results do not convey a multitasking trade-off in which order

[14] Table C.2 shows the analogue on the log of performance measures.

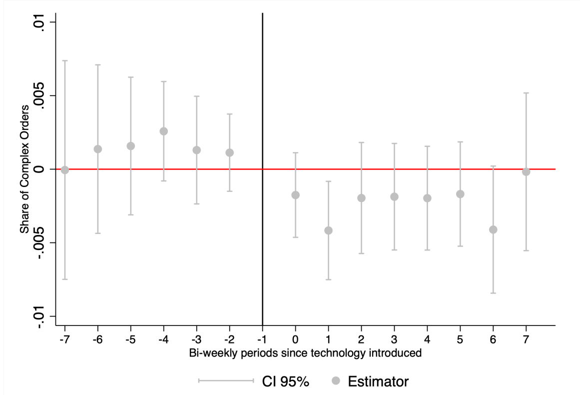
[15] In Table C.4 we run a difference-in-differences regression, replacing the full set of time to treatment indicators with a post-treatment dummy (before and after). We control for store and biweekly period fixed effects and cluster the standard errors at the store level. Table C.4 confirms the well-estimated null results.

Figure 6: Effects of technology implementations on the composition of the order

Panel A. Event study estimates on average value per ticket



Panel B. Event study estimates on share of complex orders



Notes: Panels A and B of Figure 6 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on average value per ticket and share of complex orders, respectively, where complex orders are orders with more than three distinct items (see Figure B.6b). Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. The vertical line represents the time of the treatment.

fulfillment pace is optimized at the expense of size or complexity of orders. Once again, given that overall store sales are still monitored closely it is not clear that we would expect such a trade-off. Accordingly, we next study waste as the best available test of multitasking in this setting.

4.5.3 Effects on waste

In Table 4, we test if the implementation of the performance monitoring technology caused an increase on the proportion of waste and inventory use. In Section 3.4 we mentioned that waste is measured in input units, and there are four types: promotions, complete waste, incomplete waste, and employees consumption. On average, we do not find significant increases in waste after the implementation of the monitoring technology.[16] Table 4 shows that total waste and three of its four components exhibited no significant effect, while promotions in fact decreased by roughly 13%. We interpret this as the best test of multitasking in our context and the results once again do not support a multitasking trade-off.

[16]Figure C.7 plots the event study coefficients for waste and confirms that the overall waste did not change with the technology implementation.

Table 4: Effects of technology implementations on proportion of waste

	All	**Promo**	**Complete**	**Incomplete**	**Employee food**
Coefficient	-4.80e-05	-0.131**	0.041	0.133	-0.0439
Stand. Err.	(0.245)	(0.0581)	-0.0529	(0.137)	(0.0807)
Mean of the dep. var.	5.349	1.02	1.07	1.75	1.51
Relative effect	0.00%	12.78%	3.84%	7.60%	-2.91%
Observations			2,385		

Notes: Table 4 shows the results for the estimation of equation (1), replacing the full set of time to treatment indicators with a post-treatment dummy (before and after) on waste. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Column 1 presents the estimations for the whole sample. Column 2 presents the estimations for the promo proportion in the form of coupons or free products. 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. The last column presents the estimations for employee consumption proportion—consumed by the employees as part of their benefits during the shift. Standard errors (in parenthesis) are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

4.6 Managers response

Senior executives within the firm noted that managers would likely have a dramatic impact on the gains from the monitoring technology in that they are granted a great deal of autonomy in running the daily operations of the store. That is, the performance monitoring technology was indeed seen as a tool which might help managers make operational decisions and as such the nature of their responses (or lack thereof) could largely determine the how both benefits and costs change after the implementation.

On the one hand, they posited that managers might react to the new technology by hiring new staff who can help achieve greater productivity. The firm gives individual store managers complete autonomy to react in this manner. However, if the introduction of this technology does boost store productivity, then any increase in store sales could well be partly or entirely offset by an increase in hiring and salary costs.

Alternatively, if managers rather invest effort in improving productivity of existing employees, for example by way of training them on additional operations or refreshing their training on operations they already perform but at which they could perhaps be more efficient, then the gains in sales could lead to substantial gains in profit. Indeed, the managers' allocation of their time to training and re-training employees as well as the targeting of this effort across employees is one of the key responsibilities of the store manager and one important way in which an attentive manager can strongly contribute to the performance of the store.

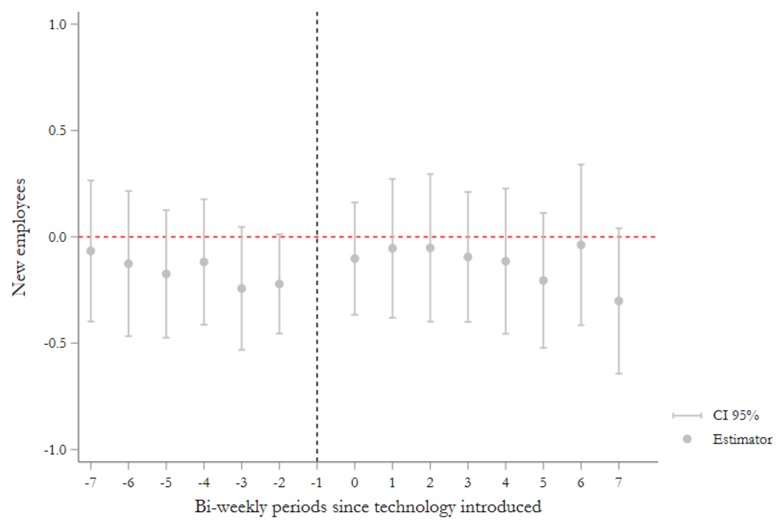
As discussed above, we have access to granular data on training at the store level and beyond. This data contains three major types of training: first, there is the training of new workers in a new skill (this is done during on-boarding and can be used to measure new hires); second, there is the training of existing workers in new skills; and finally, there is refresher training. Refresher training refers to those courses which managers provide to workers who are already certified in certain skills in order to re-up their knowledge and sharpen their practice in those areas.

Accordingly, we first explore whether managers respond to the implementations with more new employee hiring. Second, we examine how managers adjusted their training efforts on new skills for existing workers, particularly for operations associated with kitchen production. Finally, we explore the role of refresher training in sustaining the gains in sales over time.

4.6.1 Hiring

Figure 7 shows the event study estimates for new employees. Contrary to the effects on sales, units sold and tickets, we do not find effects on hiring in any of the biweekly periods following the implementation event. Thus, the increases in performance outcomes were achieved without hiring additional employees.

Figure 7: Event study estimates of implementation on new hires



Notes: Figure 7 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on new hires. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between January 2019 to November 2019. The vertical line represents the time of the treatment.

4.6.2 Training on new skills

One key prediction from the model relates to the impacts of the performance monitoring technology on the allocation of managerial effort on training. Specifically, the model states that when managers can better observe productivity at the drive-thru, this should increase the overall training they provide. We show that under fairly reasonable assumptions, the impacts on kitchen food production training should be largest.

First, we test if the monitoring technology implementation impacted the overall training provided by the managers. As we mentioned before, training programs can be classified as new or refresher. New training programs consist of new station-specific skills for which a crew member has not received training or testing before. Refreshers consist of training in skills or certifications in a station for which a given workers has already been approved to work in the past.

Table 5 shows that managers increased the overall training by 22% after stores implemented the new technology. The average effect is concentrated in the new skills: managers increase the number of crew members trained in new skills every biweekly period by 37% (around four more successfully completed worker-station training every biweekly period).

Table 5: Effects of technology implementations on training

	All	**New Skills Approved**	**New Skills Unapproved**	**Refresher Approved**	**Refresher Unapproved**
Coefficient	5.566**	3.554**	0.382	1.000	0.630
Stand. Err.	(2.237)	(1.437)	(0.797)	(0.817)	(0.438)
Mean of the dep. var.	25.78	9.57	6.02	6.43	3.76
Relative effect	21.59%	37.13%	6.35%	15.56%	16.76%
Observations			1,202		

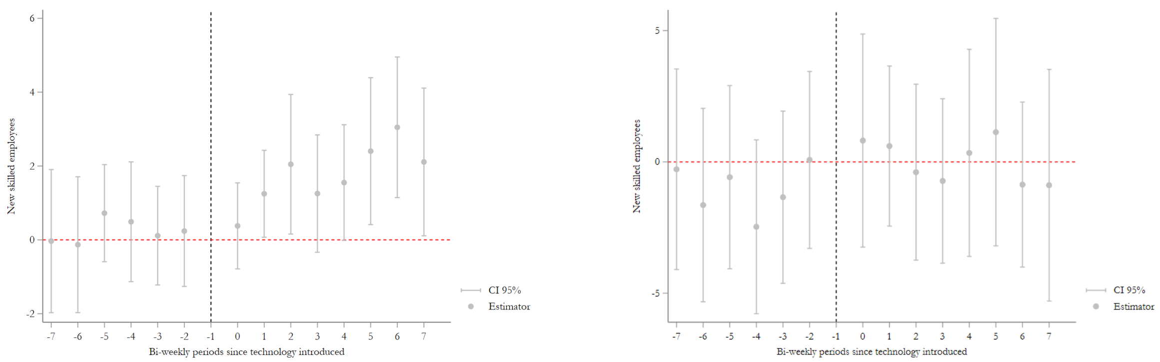
Notes: Table 5 shows results for the estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy (before and after) on training. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between January 2019 to November 2019. Column 1 presents the estimations for the whole sample. Column 2 presents the estimations only for approved new training in a specific station in which a member has not received training or testing before, while the third column presents the estimations for unapproved new training. Column 4 presents estimations for approved refresher training, which are skills taught and approved at a previous time, and the last column presents estimations for unapproved refresher training. Standard errors (in parenthesis) are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

Next, we explore the effects of the technology on each station. In line with the predictions of the model, we find that the key kitchen production stations (desserts, fried products, assembly, and grill) had the greatest allocation of training effort post-implementation.[17] In Figure 8, we present the event-study coefficients for new skill training of existing employees, grouped by kitchen and non-kitchen stations. The kitchen stations include those discussed above; while the non-kitchen group includes training in stations such as cashier and security and operations such as store opening, cleaning and maintenance, and hospitality. We observe a positive and significant effect only for training on new kitchen skills.[18]

Figure 8: Event study estimates of implementation on new skilled employees at kitchen and non-kitchen stations

Panel A. Kitchen

Panel B. Non-kitchen



Note: Figure 8 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on new skilled employees. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores implementing the new technology between January 2019 to November 2019. Panel A shows the estimations for the kitchen stations and panel B shows the estimations for the non-kitchen stations. The vertical line represents the time of the treatment.

4.6.3 Refresher training

Having shown that the increase in the store sales due to the monitoring technology implementation was seemingly driven by an increase in new skills training of existing employees in kitchen operations, we next ask: why did the effects of the monitoring technology on sales diminish after six weeks? We explore whether skill depreciation might

[17]See Figure B.7 for station-specific difference-in-difference estimates. Following the implementation of the monitoring technology, the greatest increase in crew certifications was for desserts station. The fried products, assembly, and grill stations all increased by 41.7%, 25.3%, and 25%, respectively. Figure B.8 shows that crew members working at hospitality, cleaning and maintaining, and lobby, parking, and playgrounds experience the most significant increase in new certifications.

[18]Table C.6 shows the difference-in-differences regression, replacing the full set of time to treatment indicators with a post-treatment dummy (before and after). We control for store and biweekly period fixed effects and cluster the standard errors at the store level. Table C.6 shows that the performance monitoring technology increased new kitchen training skills by 34.47%.

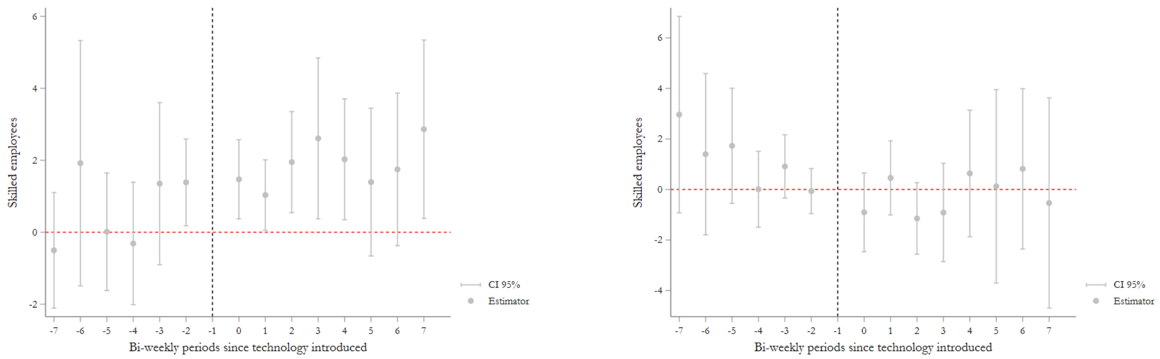
explain the dynamics we see by studying the store managers' commitment to refresher training.

First, we split our sample between stores where managers invested above and below the sample median of refresher training in the pre-implementation period.[19] Panel A of Figure 9 shows that those restaurants in which managers exhibited a commitment to refresher training pre-implementation exhibit large and persistent treatment effects in the use of this type of training in the post-implementations period as well. Conversely, we observe no impact on refresher training for managers that did not frequently invest (below the median) in refresher training pre-implementation.

Figure 9: Effects on kitchen refresher training by high and low pre-implementation refresher training

Panel A. Stores with $>$ median refresher training pre-implementation

Panel B. Stores with $<$ median refresher training pre-implementation



Notes: Figure 9 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on new skilled employees. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between January 2019 to November 2019. Panel A shows estimations for high pre-implementation training and panel B shows estimations for low pre-implementation training. The vertical line represents the time of the treatment.

Next, we explore if stores in which managers frequently invested in refresher training pre-implementation experienced a larger and/or more persistent impact of the new technology implementation on store sales than managers who did not. To make for a more parsimonious and more easily interpreted specification, we split our post-treatment dummy into two post-treatment dummies. The first one is equal to 1 for the first six weeks after the technology implementation by store s, and the second is equal to 1 after six weeks of the implementation by store s. Then we perform a similar analysis of impacts on sales,

[19]To split our sample, we computed the ratio between the number of workers trained in refresher training in a biweekly period and the average number of employees per store. Next, we split our sample between stores where managers invested above and below the sample median of refresher training in the pre-implementation period standardized by the average number of employees per stores.

unit sold, and number of tickets (equation 2) across this post-treatment distinction. To control for stores productivity pre-implementation, we use the ratio between the number of units sold and the average number of employees per store as a proxy for store productivity. We then define stores with "high productivity" as those with a proxy value above the median pre-implementation, and interact this dummy variable with our post-treatment dummies. This control helps to capture any degree to which the productive value of determinants of store performance such as general managerial quality might change after the implementation of the monitoring technology.

We estimate the following regression:

$$
\begin{aligned}
Y_{s,t} = \beta_0 + \sum_{\tau \in 1,2} \beta_1 MT_{s,t,\tau} + \beta_2 MT_{s,t,\tau} \times 1[high \;\; refresher \;\; training]_s \\
+ \beta_3 MT_{s,t,\tau} \times 1[high \;\; productivity]_s + \Phi_s + \theta_t + \epsilon_{s,t},
\end{aligned}
\tag{2}
$$

where $MT_{s,t,1}$ is a post-treatment dummy variable equal to 1 if the store s had implemented the monitoring technology in period $t <= \nu_s + 3$ (first three biweekly periods), where ν_s is the first biweekly period when store s introduces the technology. Similarly, $MT_{s,t,2}$ is a post-treatment dummy variable equal to 1 if the store s had implemented the monitoring technology in period $t > \nu_s + 3$ (after three biweekly periods). We cluster standard errors at the store level for inference.

Table 6 shows that restaurants in which managers do not utilize refresher training account for the decline over time we see on average. The high refresher training restaurants exhibit larger and more persistent gains in sales from the monitoring technology. Thus, not all the managers optimized refresher training investments despite the very similar environments in which they managed and the identical tools available to them. In this sense, these results emphasize the critical role managers play in realizing the benefits of the monitoring technology both initially and perhaps most dramatically over the longer run.

Table 6: Effects of technology implementations on interaction performance and follow-up training

	Total Sales	Units Sold	Num. Tickets
Stores with > median refresher training pre-implementation	21,762***	8,014***	3,033***
	(801.6)	(399.5)	(106.2)
Drive-thru monitoring technology (first 6 weeks)	7,900***	4,368***	1,100***
	(2,114)	(1,120)	(290.8)
Drive-thru monitoring technology (first 6 weeks) * Stores with > median refresher training pre-implementation	4,451*	2,163*	541.0
	(2,437)	(1,286)	(384.8)
Drive-thru monitoring technology (after 6 weeks)	5,441	3,061*	768.3*
	(3,283)	(1,639)	(418.3)
Drive-thru monitoring technology (after 6 weeks) * Stores with > median refresher training pre-implementation	6,244**	2,963*	818.0*
	(2,930)	(1,534)	(420.8)
Observations	2,074	2,073	2,069

Notes: Table 6 shows the results for the estimation of equation (2), splitting the post-treatment dummy into two post-treatment dummies: the first one is equal to 1 for the first six weeks after the implementation, and the second is equal to 1 after six weeks of the implementation. Treatment is defined as stores implementing the performance monitoring technology. To split our sample, we computed the ratio between the number of workers trained in refresher training in a biweekly period and the average number of employees per store. Next, we split our sample between stores where managers invested above and below the median average refresher training pre-implementation standardized by the average number of employees per stores. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Standard errors (in parenthesis) are clustered at the store level. * significant 10%, ** significant 5%, *** significant 1%

5 Robustness

In this section we conduct a set of alternative estimation exercises to demonstrate the robustness of our baseline event-study specification and sample.

5.1 Balanced panel

We replicate the baseline event study analysis on a version of the ever-treated sample balanced in relative time. When there are no never-treated units but with a panel balanced in calendar time, we need to exclude at least two relative period indicators (Borusyak and Jaravel, 2017; Sun and Abraham, 2020). Thus, we normalize $\delta_{-1} = 0$ and $\delta_{+6} = 0$. We control for biweekly periods and store fixed effects. We again cluster standard errors at the store level for inference.

Results are also similar when we estimate the regressions on a balanced sample in event time. Figures C.8, C.9, and C.10 show the event-study coefficients for sales, units sold, and number of tickets, respectively, for the whole store, drive-thru, and counter (Panels A, B, and C, respectively). We compare the patterns for this restricted sample and outcomes with those from the baseline event-study sample (Figures 5, C.1, C.2, and C.3). Reassuringly, we see an absence of systematic pre-trends. Moreover, all three sets of results display patterns of the treatment effects with the same takeaways as those from the baseline sample (i.e., we observe some effects the same week of the event, but the largest impacts occur three biweekly periods after the implementation). In terms of magnitudes, We find larger effects (by around one percentage point) for this balanced sample as compared to our main results. Overall, it is encouraging that these alternative balanced sample results are consistent with our baseline results.

5.2 Heterogeneous effects

Recent literature in econometrics has raised concerns about the possibility of negative weights in multiple-period difference-in-difference estimators when treatment timing is staggered and there exists heterogeneity in treatment effects within-unit over time or between groups of units treated at different times (Athey and Imbens, 2021; Callaway and Sant'Anna, 2020; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2020). The latter case may also 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 group-time average treatment effect, where a group is defined by the time period when stores are first treated suggested by Callaway and Sant'Anna (2020). The key assumption in our main sample without any "never-treated" stores is the conditional parallel trends between stores treated in period g and groups that are "not-yet-treated" by time t. Second, we compute a Oaxaca-Blinder-Kitagawa decomposition to measure how much of the coefficient change comes from changes in the weights or changes in the 2×2 DD terms (Goodman-Bacon, 2021). Third, we estimate the cohort-specific average treatment effect suggested by Sun and Abraham (2020), which translates the Callaway and Sant'Anna (2020) group-time average treatment effect from calendar time into relative periods, allowing us to compare cohorts while holding their level of exposure to the treatment constant.

- **Callaway and Sant'Anna (2020)**. We estimate the disaggregated causal parameter suggested by Callaway and Sant'Anna (2020) called, the group-time average treatment effect, that is, the average treatment effect for group g at time t, where a group in this context is defined by the time period when stores implemented the new technology. We consider a "static" and "dynamic" (event study) aggregation schemes for the group average treatment effects. We follow the "cluster-robust" multiplier bootstrap procedure, clustering at the store level with 1,000 repetitions.

Figures D.1 ans D.2 present the average treatment effects by the length of exposure (event study) to the monitoring technology on sales and units sold, respectively, for the whole store, drive-thru, and counter (Panels A, B, and C, respectively). Note that the results are very similar to the ones presented in Figures 5, C.1, and C.2. Finally, Figure D.3 presents the average treatment effects by the length of exposure (event study) to the monitoring technology on new skilled employees at kitchen and non-kitchen stations. Again, the results are very similar to the event study-type regression with homogeneous effects across stores (Figure 8).

- **Oaxaca-Blinder-Kitagawa decomposition (Goodman-Bacon (2021)**. Tables D.2 and D.3 present the Oaxaca-Blinder-Kitagawa decomposition suggested by Goodman-Bacon (2021) for total sales and unis sold, respectively. For total sales, we find a small difference between earlier- to later-treated stores (average DD estimates equal to 3287) and later- to earlier-treated stores (average DD estimates equal to 4246), with similar weights 0.48 and .45, respectively. The comparisons of earlier- to later-treated states and later- to earlier-treated states also reveal a small difference in the average DD estimate for units sold for these two timing groups: 2081 and 2359, with weights 0.488 and 0.455, respectively. Figures D.4 and D.5 plots the 2×2 DDs against their weight, and confirm that there are not systematic differences across timing groups.

- **Sun and Abraham (2020)**. Finally, we estimate the cohort average treatment effects on the treated as the cohort-specific average difference in outcomes relative to last cohort to be treated (control group), suggested by Sun and Abraham (2020). Figures D.6 and D.7 present the results for sales and units sold, respectively, for the whole store, drive-thru, and counter (Panels A, B, and C, respectively). Note that the results are almost identical to the event-study regressions with homogeneous effects presented in Figure 5, Figures C.1, and C.2.

6 Alternative Mechanisms and Additional Evidence

We presented evidence in Figure 5 that the largest impacts of the monitoring technology on performance measures occur during the 6 weeks after the introduction of the technology. After this period, the impacts diminish though they remain positive. Table 5 shows that managers increased overall training by 22% and in new skills by 37%, concentrated among the key kitchen stations (Figure 8). Finally, we document that restaurants in which managers do not utilize refresher training account for the decline over time observed on the impact of the technology (on average) after six weeks.

Taken together, we interpret this pattern as indicative of the importance of managerial on-the-job human capital investments in actualizing the value of monitoring technology. However, several alternative interpretations are possible and we now discuss each in light of the full set of results discussed above.

6.1 Demand

First, we note that the analysis above focuses purely on the supply side. That is, we assume implicitly that there is sufficient residual demand such that if stores indeed achieve a faster pace of serving customers, sales will indeed go up because additional customers exist to be served. In practice, this requires that there are customers who are waiting in the drive-thru queue or are considering joining the queue but leave due to long wait times. This was certainly the case for all of the stores we visited (and at various times throughout the day) but is also consistent with the information provided by the managers with whom we spoke.

To explore the degree to which variation in residual demand might impact the effects we estimate, we check if the effects on sales of the monitoring technology vary across peak and non-peak hours. Peak hours are defined as those hours for which the average sales per hour are higher than the median, and non-peak hours are those below the median. Following this threshold, peak hours are between 11 AM and 9 PM, and non-peak hours are between 10 PM and 10 AM. Figure B.5 shows the average hourly sales for the 55 stores in our sample.

Panels A and B of Figure C.4 show the results of the event-study coefficients for the peak and non-peak hours, respectively, for the whole store. Panels A, B, C, and D of Figure C.5 show the event-study coefficients for peak and non-peak hours for the two points of sales—drive-thru and counter. We do not find significant differences between peak and non-peak hours for the whole store nor for either of the two points of sale.[20]

[20] Table C.4 compares the pooled (before vs. after) change in sales for peak and non-peak hours and confirms no significant differences between the two.

6.2 Worker selection

It is possible that the monitoring technology implementation at a given store might either attract workers who are faster or more conscientious or repel workers who work slowly and do not want the time pressure. This response in the composition of workers was shown to be important in a recent study of a change in pay schemes in a Peruvian egg producer (Amodio and Martinez-Carrasco, 2019). Such a change in the composition of workers to include more productive workers might increase workers' average productivity and as a result stores' performance.

However, the fact that restaurants did not hire new employees is inconsistent with this interpretation. The coefficient on hiring (presented in Table C.3) is both statistically insignificant and an order of magnitude smaller (roughly 3% change from the mean) than the marked increase in training for new skills (37%).

6.3 Workers' effort

Finally, we address the potential importance of workers' effort response to the introduction of the monitoring technology in mediating gains in sales. While it is indeed plausible that worker effort explains some part of the observed impacts of the technology, we believe it is unlikely that the majority of impacts are due to this mechanism for four reasons.

1. Anecdotal evidence. Upper level managers from the partner firm did not believe that the monitoring technology increased the existing level of pressure to fill orders as quickly as possible. They reported, "*The pressure to get out food out as fast as you can has been constant before and after the introduction of the monitoring technology.*"

2. Size of production area. Most of the kitchens range from 500 to 1,000 square feet in size. There are 15 to 20 crew members working at a time and 6 to 12 of them working in the kitchen, where the manager also spends the majority of her time throughout the day. Accordingly, managerial oversight of shirking or distraction among the crew members is relatively easy given the small, densely staffed production area. This feature of the kitchen and station-staffing layout did not change with the introduction of the monitoring technology, so it is less likely that workers' propensity to shirk or level of distraction (e.g., using their phones) would have changed after the introduction of the monitoring technology.

3. Coincident timing of effects on training. The introduction of the technology caused a substantial increase in managerial effort, as measured by training investments, particularly for kitchen tasks. Moreover, the increase in performance followed a similar timing as this increase in training for new skills. That is, if the effects on performance were primarily mediated by the provision of greater worker effort, we would expect this to be most salient immediately after the implementation. However, we find that the effects at the time of implementation were subtle and grew over the ensuing 6 weeks, exactly coincident with the rise in training responses. It is difficult to think of a hypothesis about workers' effort consistent with these timing patterns.

4. No effect on hiring. Relatedly, if the effects were primarily driven by pressure to extract more effort from the workers (rather than a greater investment in improving their skill stock), we would expect that workers might be more likely to leave and need to be replaced. However, as we mentioned before, we find that restaurants did not hire new employees as a result of the performance monitoring technology, reflecting no impact on worker turnover.

7 Conclusion

In this study, we aim to peer inside the "black box" of the management–productivity relationship, and highlight the critical role of on-the-job human capital investment in realizing and sustaining productivity gains from better performance monitoring in the context of quick-service restaurants in Puerto Rico. We find that implementation of the monitoring technology generates substantial changes in managerial effort, which drive productivity gains on the order of 4-5 percent. These gains are achieved without hiring new employees, but rather by managers investing effort (as well as better targeting their effort) toward training employees in new food production skills to gain more manpower in kitchen production. Some of these gains in productivity decay on average, a fact that our results suggest may be due to a depreciation of skill over time. Indeed we show that managers who show a commitment to investing in refresher training prior to the implementation of the technology exhibit large impacts on this type of investment, and as a result those restaurants see sustained gains, while restaurants in which managers do not invest in refresher training exhibit strong attenuation in sales gains after about six weeks with the new technology.

Our results speak to the academic literature on managerial responses to performance monitoring technologies; they are also relevant to the monumental changes currently ongoing in the QSR industry as a result of the COVID-19 global pandemic. Many fast food restaurant chains have experienced considerable growth as a result of the pandemic, and drive-thru interactions have become the default preferred mode of sales for most customers. Optimizing performance in drive-thru is thus of paramount importance to the industry (as well as to consumers), and understanding how managerial inputs might be better used in service of that mission is thus essential.

References

Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J., and Zilibotti, F. (2007). Technology, information, and the decentralization of the firm. *The Quarterly Journal of Economics*, 122(4):1759–1799.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2021a). Management and shocks to worker productivity. Technical report.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. A. (2021b). Managerial quality and productivity dynamics. Technical report.

Amodio, F. and Martinez-Carrasco, M. (2019). Inputs, incentives, and self-selection at the workplace.

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

Athey, S. and Stern, S. (2002). The impact of information technology on emergency health care outcomes. *RAND Journal of Economics*, pages 399–432.

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.

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *Journal of human resources*, pages 728–751.

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

Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). Ceo behavior and firm performance. *Journal of Political Economy*, 128(4):1325–1369.

Bandiera, O., Prat, A., Sadun, R., and Wulf, J. (2014). Span of control and span of attention. *Harvard Business School strategy unit working paper*, (12-053):14–22.

Bartel, A., Ichniowski, C., and Shaw, K. (2007). How does information technology affect productivity? plant-level comparisons of product innovation, process improvement, and worker skills. *The quarterly journal of Economics*, 122(4):1721–1758.

Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics*, 118(4):1169–1208.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., and Van Reenen, J. (2019). What drives differences in management practices? *American Economic Review*, 109(5):1648–83.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics*, 128(1):1–51.

Bloom, N., Garicano, L., Sadun, R., and Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12):2859–2885.

Bloom, N., Sadun, R., and Van Reenen, J. (2012). Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102(1):167–201.

Bloom, N., Sadun, R., and Van Reenen, J. (2016). Management as a technology? Technical report, National Bureau of Economic Research.

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

Bloom, N. and Van Reenen, J. (2010). Why do management practices differ across firms and countries? *Journal of economic perspectives*, 24(1):203–24.

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

Borusyak, K. and Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Bresnahan, T. F., Brynjolfsson, E., and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics*, 117(1):339–376.

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

Chetty, R., Saez, E., and Sándor, L. (2014). What policies increase prosocial behavior? an experiment with referees at the journal of public economics. *Journal of Economic Perspectives*, 28(3):169–88.

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

Duflo, E., Hanna, R., and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78.

Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy*, 21(3):673–705.

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., and Zubanov, N. (2022). Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain. *Management Science*, 68(1):211–229.

Fryer, R. G., Holden, R. T., and of Economic Research, N. B. (2012). Multitasking, learning, and incentives: a cautionary tale.

Glewwe, P., Ilias, N., and Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–27.

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

Gosnell, G. K., List, J. A., and Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128(4):1195–1233.

Halac, M. and Prat, A. (2016). Managerial attention and worker performance. *American Economic Review*, 106(10):3104–32.

Hansman, C., Hjort, J., León-Ciliotta, G., and Teachout, M. (2020). Vertical integration, supplier behavior, and quality upgrading among exporters. *Journal of Political Economy*, 128(9):3570–3625.

Hoffman, M. and Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy*, 129(1):243–285.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7:24.

Hong, F., Hossain, T., List, J. A., and Tanaka, M. (2018). Testing the theory of multitasking: evidence from a natural field experiment in chinese factories. *International Economic Review*, 59(2):511–536.

Hortaçsu, A., Luco, F., Puller, S. L., and Zhu, D. (2017). Does strategic ability affect efficiency? evidence from electricity markets. Technical report, National Bureau of Economic Research.

Hubbard, T. N. (2000). The demand for monitoring technologies: The case of trucking. *The Quarterly Journal of Economics*, 115(2):533–560.

Hubbard, T. N. (2003). Information, decisions, and productivity: On-board computers and capacity utilization in trucking. *American Economic Review*, 93(4):1328–1353.

Jackson, C. K. and Schneider, H. S. (2015). Checklists and worker behavior: A field experiment. *American Economic Journal: Applied Economics*, 7(4):136–68.

Kahn, L. B. and Lange, F. (2014). Employer learning, productivity, and the earnings distribution: Evidence from performance measures. *The Review of Economic Studies*, 81(4):1575–1613.

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics*, 33(4):823–861.

Mullen, K. J., Frank, R. G., and Rosenthal, M. B. (2010). Can you get what you pay for? pay-for-performance and the quality of healthcare providers. *The Rand journal of economics*, 41(1):64–91.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of economic literature*, 37(1):7–63.

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

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

Online Appendix

A Model

In this section, we present a theoretical framework in the spirit of Holmstrom and Milgrom (1991) to guide our empirical analysis. We use the model to study the mechanisms through which better monitoring affects managerial inputs and productivity in a multitasking environment.

A.1 Manager's Problem

The principal (in our context, central corporate management) provides monetary incentives to increase stores' sales. The store manager invests time training employees in different production tasks: counter, drive-thru, and kitchen, denoted by t_c, t_d, and t_k, respectively. Let $t \equiv (t_c, t_d, t_k)$, denote the vector of training investments. This vector of investments generates a cost, $C(t)$, to the store manager, where $C(t)$ is a strictly convex function.

The vector t determines the level of production at both the counter and drive-thru. Particularly, if the store manager invests in training (t_c, t_k), the production at the counter is $q_c \equiv \mu(t_c, t_k)$. Similarly, if the store manager invests in training input (t_d, t_k), the production at the drive-thru is $q_d \equiv \nu(t_d, t_k)$. We assume that both production functions, $\mu, \nu : \mathbb{R}^2 \longrightarrow \mathbb{R}$, are linear.[21]

After the store manager chooses the vector of training inputs, the principal observes noisy signals of the net production (sales) at the counter and the drive-thru, x_1 and x_2, respectively defined as

$$x_1 \equiv q_c + \varepsilon_c = \mu(t_c, t_k) + \varepsilon_c, \tag{A.1}$$

and

$$x_2 \equiv q_d + \varepsilon_d = \nu(t_d, t_k) + \varepsilon_d, \tag{A.2}$$

where ε_c and ε_d are two random variables independently and normally distributed (i.e., $\varepsilon_i \sim N(0, \sigma_i^2)$ for $i \in \{c, d\}$). These errors, ε_i for $i \in \{c, d\}$, capture unobserved factors of the production process (e.g., worker effort).

The principal offers a base wage, β, and a pair of linear piece-rate, $\{a_1, a_2\}$, contracts. Store managers have constant absolute risk aversion preferences over their total reward,

[21] That is, we assume that $\mu(t_c, t_k) = s_c t_c + s_k t_k$ and $\nu(t_d, t_k) = s_d t_d + s_k t_k$, where the constants (s_c, s_d, s_k) are positive and represent the marginal productivity of each training input. In Appendix A.6, we present a model that considers CES production functions. Our simulations suggest similar qualitative results.

$$w(x_1, x_2) = a_1 x_1 + a_2 x_2 + \beta, \tag{A.3}$$

and their net training input cost, $C(t)$. It is equal to

$$u(CE) = \mathbb{E}\{u[w(x_1, x_2) - C(t)]\}, \tag{A.4}$$

where $u(w) = -e^{-rw}$, r is the coefficient of absolute risk aversion, and CE is the store manager's certainty equivalent money payoff. Taking the compensation plan as given, the problem of the manager is to determine the optimal vector of training input t^* that maximizes her utility:

$$\max_{t' \in \mathbb{R}^3_+} CE(t'), \tag{A.5}$$

where $CE(t)$ is equal to

$$CE(t) = \beta + a_1 \mu(t_c, t_k) + a_2 \nu(t_d, t_k) - C(t) - \frac{1}{2} r \left(a_1^2 \sigma_c^2 + a_2^2 \sigma_d^2\right). \tag{A.6}$$

A.2 Principal's Problem

We assume that the principal's net sales are linear in x_1 and x_2, so that the gross benefits of the principal are:

$$B(t) \equiv \mathbb{E}[b_1 x_1 + b_2 x_2], \tag{A.7}$$

where b_1 and b_2 are positive constants representing the marginal benefit of the production at the counter and drive-thru, respectively.[22]

The principal aims to find the incentive efficient linear contract; that is, the contract that solves

$$\max_{a_1, a_2} TCE(t, a_1, a_2) \quad \text{s.t.} \quad t \in \arg\max_{t' \in \mathbb{R}^3_+} CE(t'). \tag{A.8}$$

where

$$TCE(t, a_1, a_2) = B(t) - C(t) - \frac{1}{2} r \left(a_1^2 \sigma_c^2 + a_2^2 \sigma_d^2\right). \tag{A.9}$$

A.3 Equilibrium

In this subsection, we show that there exists a unique vector of efforts, t^*, maximizing the CE given by (A.6), and study the effects of the monetary incentives (a_1 and a_2) on the optimal levels of training efforts and production at the counter and drive-thru, q_c^* and

[22]The expectation is taken with respect to the joint density of $\{\varepsilon_c, \varepsilon_d\}$.

q_d^*, respectively. Finally, we show that there is a unique compensation plan that solves principal problem. For tractability, we assume that the store manager's cost, $C(t)$, is represented by a quadratic function of the effort levels.

Assumption 1. *We assume that $C(t)$ is equal to*

$$C(t) = t^T C t = \begin{bmatrix} t_c & t_d & t_k \end{bmatrix} \begin{bmatrix} c_{cc} & c_{cd} & c_{ck} \\ c_{cd} & c_{dd} & c_{dk} \\ c_{ck} & c_{dk} & c_{kk} \end{bmatrix} \begin{bmatrix} t_c \\ t_d \\ t_k \end{bmatrix},$$

where C is a symmetric positive definite matrix with real-valued entries. Moreover, we assume $c_{cd} = 0$, $c_{ck} < 0$, $c_{dk} < 0$ and $c_{ii} > 0$, for $i \in \{c, d, k\}$.[23]

From A1, it follows that $C(t)$ is a strictly convex function. The condition $c_{cd} = 0$ implies that counter and drive-thru tasks are perfect substitutes. Similarly, if $c_{ck} < 0$ and $c_{dk} < 0$ then the pairs of tasks (counter-kitchen) and (drive-thru-kitchen) are complements.[24]

Proposition 1. *There exists a unique $t^* \in \mathbb{R}_+^3$ that maximizes the certainty equivalent money payoff, CE. Moreover, t^* is given by*

$$t^* = \frac{1}{2} C^{-1} S, \tag{A.10}$$

where $S = \begin{bmatrix} a_1 s_c, & a_2 s_d, & (a_1 + a_2) s_k, \end{bmatrix}^T$.

From Proposition 1, it follows that the optimal vector of efforts, t^*, is a linear combination of the monetary incentives (a_1, a_2), weighted by the entries of the inverse matrix of C. Note that the positive definiteness of C and the fact that $c_{cd} = 0$, $c_{ck} < 0$ and $c_{dk} < 0$ implies that $t_i^* > 0$ for each $i \in \{c, d, k\}$.[25] From (A.10), we get the following corollary.

Corollary 1. *In equilibrium:*

(i) $\frac{\partial t_i^*}{\partial a_j} > 0$ *for each $i \in \{c, d, k\}$ and $j \in \{1, 2\}$;*[26]

[23]Note that if $c_{ij} > 0$ (< 0), for $i, j \in \{c, d, k\}$, the tasks i and j are substitutes (complements), which implies that if the store manager increases the training input t_j, it will be more (less) costly for her to increase the input t_i. If $c_{ij} = 0$, the tasks i and j are perfect substitutes.

[24]Note that in this version of the model, complementarities between the various tasks are included in the cost function of the store manager and not in the production functions. In Appendix A.6, we present a model in which the production functions are represented by CES production functions and the store manager's cost is linear in training input.

[25]Second order conditions follow from the assumption that $C(t)$ is strictly convex and the production functions μ and ν are linear.

[26]Note that if $c_{ck} = 0$, then $\frac{\partial t_c^*}{\partial a_2} = 0$. Thus, when the counter and kitchen tasks are perfect substitutes, changes in the monetary incentive a_2 do not affect the optimal training effort at the counter.

(ii) $\frac{\partial q_i^*}{\partial a_j} > 0$ *for each* $i \in \{c, d\}$ *and* $j \in \{1, 2\}$.

Corollary 1(i) shows that the principal can incentivize the store manager to increase her training efforts by increasing a_1 or a_2. Similarly, Corollary 1(ii) shows that the principal can incentivize the optimal production at the counter and drive-thru by increasing a_1 or a_2.

Given a vector of efforts, $t^*(a_1, a_2)$, the principal needs to find a compensation plan $\{a_1^*, a_2^*\}$ that maximizes her profits (A.8). From (A.8), (A.9), and (A.10), the principal solves,

$$\begin{aligned} &\max_{a_1, a_2} B(t^*) - C(t^*) - \frac{1}{2} r \left(a_1^2 \sigma_c^2 + a_2^2 \sigma_d^2 \right) = \\ &\max_{a_1, a_2} \left\{ b_1 \mu \left(t_c^*, t_k^* \right) + b_2 \nu \left(t_d^*, t_k^* \right) - \left(t^* \right)^T C t^* - \frac{1}{2} r \left(a_1^2 \sigma_c^2 + a_2^2 \sigma_d^2 \right) \right\}. \end{aligned} \tag{A.11}$$

The following proposition characterizes the optimal compensation plan $\{a_1^*, a_2^*\}$ as a function of the variances of the unobserved production factors ε_c and ε_d.

Proposition 2. *There exists a unique compensation plan* $\{a_1^*, a_2^*\}$ *solving the principal's problem given by*

$$\begin{aligned} a_1^*(\sigma_d, \sigma_c) &= 2 \frac{M_1 \left(2r\sigma_d^2 + \delta_2 \right) - \beta_2 M_2}{\left(2r\sigma_d^2 + \delta_2 \right) \left(2r\sigma_c^2 + \beta_1 \right) - \beta_2^2}, \text{ and} \\ a_2^*(\sigma_d, \sigma_c) &= 2 \frac{M_2 \left(2r\sigma_c^2 + \beta_1 \right) - \beta_2 M_1}{\left(2r\sigma_d^2 + \delta_2 \right) \left(2r\sigma_c^2 + \beta_1 \right) - \beta_2^2}, \end{aligned} \tag{A.12}$$

for β_1, β_2, δ_2, M_1 *and* M_2 *positive constants. Moreover,* $a_i^*(\sigma_d, \sigma_c) > 0$ *for each* $i \in \{1, 2\}$ *and for any* $\sigma_c, \sigma_d \geq 0$. *Additionally, if* $\sigma_d = \sigma_c = 0$, *then* $a_1^* = b_1$ *and* $a_2^* = b_2$.

From Proposition 2, we have a unique compensation plan solving the principal's problem. Note that the optimal compensation plan is a function of the variances $\{\sigma_d^2, \sigma_c^2\}$. Similarly, since t^*, the optimal vector of efforts of the store manager, is a function of a_1^* and a_2^*, then $t^* = t^*(\sigma_d, \sigma_c)$. Moreover, these compensations are always positive and when there is no distortion in the production signals x_1 and x_2 (i.e., $\sigma_d = \sigma_c = 0$), the optimal compensation plan is given by the marginal benefits of production b_1 and b_2.

A.4 Monitoring Technology

In Proposition 1 we characterized the manager's optimal vector of training inputs, t^*, in terms of the monetary incentives a_1 and a_2. Also, in Proposition 2, we characterized the principal's optimal compensation plan $\{a_1, a_2\}$ as a function of the variances of the unobserved production factors ε_c and ε_d. In this subsection, we explore how the optimal managerial investments change after the introduction of the monitoring technology, which makes the signals coming from the sales at the drive-thru window more precise, as their noise decreases. In particular, the following corollaries characterize how the optimal investments in training vary as the variance of those signals change.

Corollary 2. $\frac{\partial t_d^*}{\partial \sigma_d^2} < 0$ *and* $\frac{\partial t_k^*}{\partial \sigma_d^2} < 0$. *Moreover,* $\frac{\partial t_c^*}{\partial \sigma_d^2} < 0$ *if and only if* $\sigma_c^2 > \frac{s_c s_k}{-2rc_{ck}}$.[27]

Corollary 2 shows that the training input at the kitchen and the the drive-thru increase as the monitoring technology in the drive-thru improves (i.e., as σ_d^2 decreases).[28] Moreover, the training input at the counter increases if and only if the variance of ε_c is higher than a $1:2$ ratio between the marginal productivities $s_c s_k$ and the absolute value of c_{ck} weighted by the manager's risk aversion coefficient r. That is, if the noise of the signal at the counter is large relative to marginal productivities, a decrease in σ_d^2 increases the training input at the counter and in total.

Corollary 3. *If* $s_c > s_k$, *then* $\frac{\partial t_c^*}{\partial \sigma_d^2} + \frac{\partial t_d^*}{\partial \sigma_d^2} + \frac{\partial t_k^*}{\partial \sigma_d^2} < 0$.

Corollary 3 provides an alternative sufficient condition for an increase the total training investment, which is to have a higher marginal productivity at the counter than in the kitchen. Furthermore, from Corollary 2 and Proposition 2, it follows that net sales for the principal increase as the noise at the drive-thru window decreases, which is directly linked to the increase in the optimal production at the counter and the drive-thru seen in Corollary 4.

Corollary 4. *Let* $q(t) \equiv q_c(t_c, t_k) + q_d(t_d, t_k)$ *and* $q^* \equiv q(t^*)$, *then* $\frac{\partial q^*}{\partial \sigma_d^2} < 0$.

Next, we impose stronger conditions over the matrix C that allow us to prove additional results.

Assumption 2. *Suppose that* $c_{ii} = c_1$ *for* $i \in \{c, d, k\}$, $c_{cd} = 0$, $c_{ck} = c_{dk} = c_2$, *where* c_1 *and* c_2 *are constants such that* $c_2 < 0$ *and* $c_1 + \sqrt{2}c_2 > 0$.

From Assumption 2, it follows that C is a positive definite matrix. Thus, Assumption 2 implies Assumption 1. The following corollary ranks the changes of the optimal training inputs t_c^*, t_k^*, and t_d^* with respect to increases in the monitoring technology in the drive-thru.

[27] From Corollary 2, as $c_{ck} \to 0^-$, $\frac{\partial t_c^*}{\partial \sigma_d^2}$ becomes strictly positive, since the right-hand side of the inequality $\sigma_c^2 > \frac{s_c s_k}{-2rc_{ck}}$ becomes infinity. Similarly, as $\sigma_c^2 \to \frac{s_c s_k}{-2rc_{ck}}$, $\frac{\partial t_c^*}{\partial \sigma_d^2} \to 0$.

[28] From Corollary 2, it follows that the optimal compensation for the counter production, a_1^*, decreases, while the optimal compensation for the drive-thru production, a_2^*, increases.

Corollary 5. *Under A2, then,* $\frac{\partial t_d^*}{\partial \sigma_d^2} < \frac{\partial t_c^*}{\partial \sigma_d^2}$ *and* $\frac{\partial t_k^*}{\partial \sigma_d^2} < \frac{\partial t_c^*}{\partial \sigma_d^2}$. *Moreover, if* $s_k \geq s_d$, *there exists* $M \in \mathbb{R}$ *such that* $\frac{\partial t_k^*}{\partial \sigma_d^2} < \frac{\partial t_d^*}{\partial \sigma_d^2}$ *for all* $\sigma_c^2 > M$.

From Corollary 5 it follows that as the monitoring technology decreases the noise of the signal at the drive-thru (σ_d decreases), managers will increase their inputs more at the kitchen and at the drive thru than at the counter.[29]

We test the predictions suggested by Corollaries 2 to 5 below.[30]

A.5 Proofs

Proof of Proposition 1: We show that there is a unique t^* solving the store manager's problem (A.5). First, we show that the expression for the store manager's certainty equivalent money payoff is given by (A.6). From (A.3) and (A.4),

$$\begin{aligned} e^{-rCE} &= \mathbb{E}\left\{e^{-r[w(x_1,x_2)-C(t)]}\right\} \\ &= e^{-r[\beta+a_1\mu(t_c,t_k)+a_2\nu(t_d,t_k)-C(t)]}\mathbb{E}\left\{e^{-r[a_1\varepsilon_c+a_2\varepsilon_d]}\right\}. \end{aligned} \tag{A.13}$$

Using the properties of the normal distribution

$$\mathbb{E}\left\{e^{-r[a_1\varepsilon_c+a_2\varepsilon_d]}\right\} = e^{\frac{1}{2}r^2\left(a_1^2\sigma_c^2+a_2^2\sigma_d^2\right)}. \tag{A.14}$$

From (A.13) and (A.14), it follows that

$$CE(t) = \beta + a_1\mu\left(t_c,t_k\right) + a_2\nu\left(t_d,t_k\right) - C\left(t\right) - \frac{1}{2}r\left(a_1^2\sigma_c^2 + a_2^2\sigma_d^2\right),$$

which proves (A.6). Now we show that there is a unique t^* that solves the first-order conditions associated with $CE(t)$. These first-order conditions are given by

$$\begin{aligned} \frac{\partial CE}{\partial t_c} &= a_1\frac{\partial \mu}{\partial t_c} - \frac{\partial C}{\partial t_c} = 0, \\ \frac{\partial CE}{\partial t_d} &= a_2\frac{\partial \nu}{\partial t_d} - \frac{\partial C}{\partial t_d} = 0, \quad \text{and} \\ \frac{\partial CE}{\partial t_k} &= a_1\frac{\partial \mu}{\partial t_k} + a_2\frac{\partial \nu}{\partial t_k} - \frac{\partial C}{\partial t_k} = 0. \end{aligned} \tag{A.15}$$

[29] In this subsection, we assumed stronger assumptions like A2. However, our simulations (Appendix Section A.5.1) show that this results follow under more general conditions like A1.

[30] From Corollary 2 and 5, if $s_k \geq s_d$ and $\sigma_c^2 > \max\{M, \frac{s_c s_k}{-2rc_2}\}$, it follows that

$$\frac{\partial t_k^*}{\partial \sigma_d^2} < \frac{\partial t_d^*}{\partial \sigma_d^2} < \frac{\partial t_c^*}{\partial \sigma_d^2} < 0.$$

Additionally, A1 implies that $\nabla C(t) = 2Ct$ for any $t \in \mathbb{R}^3$, which together with (A.15) results in $t^* = \frac{1}{2}C^{-1}S$, where $S = \left[a_1 s_c, a_2 s_d, (a_1 + a_2)\, s_k\right]^T$, and proves (A.10). To show that t^* is a unique maximum of $CE(t)$, we use the fact that C is strictly convex from A1, and note that

$$D^2 CE\,(t) = -D^2 C\,(t) = -2C.$$

That is, $CE(t)$ is a strictly concave function. Finally, to verify that $t^* \in \mathbb{R}^3_+$, we rely on:

(i) $c_{ii} > 0$ for each $i \in \{c, d, k\}$;

(ii) $c_{cc}c_{kk} - c_{ck}^2 > 0$, $c_{kk}c_{dd} - c_{dk}^2 > 0$ and $\det C \equiv |C| > 0$;

(iii) $c_{cd} = 0$, $c_{ck} < 0$ and $c_{dk} < 0$.

From (i)-(iii) and (A.10), it follows that

$$\begin{bmatrix} t_c^* \\ t_d^* \\ t_k^* \end{bmatrix} = \frac{1}{2\,|C|} \begin{bmatrix} c_{ck}\underbrace{\left(a_2 s_d c_{dk} - s_k(a_1+a_2)c_{dd}\right)}_{\leq 0} + a_1 s_c \underbrace{\left(c_{dd}c_{kk} - c_{dk}^2\right)}_{>0} \\ c_{dk}\underbrace{\left(a_1 s_c c_{ck} - s_k(a_1+a_2)c_{cc}\right)}_{\leq 0} + a_2 s_d \underbrace{\left(c_{kk}c_{cc} - c_{ck}^2\right)}_{>0} \\ c_{cc}\underbrace{\left(s_k(a_1+a_2)c_{dd} - a_2 s_d c_{dk}\right)}_{\geq 0} - a_1 s_c c_{dd} \underbrace{c_{ck}}_{\leq 0} \end{bmatrix}, \tag{A.16}$$

which proves that $t_i^* > 0$ for each $i \in \{c, d, k\}$.

Proof of Corollary 1: (i) From (A.10) and items (i)-(iii) in the proof of Proposition 1, it follows that

$$\begin{bmatrix} \frac{\partial t_c^*}{\partial a_1} \\ \frac{\partial t_d^*}{\partial a_1} \\ \frac{\partial t_k^*}{\partial a_1} \end{bmatrix} = \frac{1}{2\,|C|} \begin{bmatrix} s_c \underbrace{\left(c_{dd}c_{kk} - c_{dk}^2\right)}_{>0} - s_k c_{dd} \underbrace{c_{ck}}_{\leq 0} \\ \underbrace{c_{dk}}_{\leq 0} \underbrace{\left(s_c c_{ck} - s_k c_{cc}\right)}_{<0} \\ c_{dd} \underbrace{\left(s_k c_{cc} - s_c c_{ck}\right)}_{>0} \end{bmatrix},$$

and

$$\begin{bmatrix} \frac{\partial t_c^*}{\partial a_2} \\ \frac{\partial t_d^*}{\partial a_2} \\ \frac{\partial t_k^*}{\partial a_2} \end{bmatrix} = \frac{1}{2\,|C|} \begin{bmatrix} c_{ck} \underbrace{\left(s_d c_{dk} - s_k c_{dd}\right)}_{<0} \\ s_d \underbrace{\left(c_{cc}c_{kk} - c_{ck}^2\right)}_{>0} - s_k c_{cc} \underbrace{c_{dk}}_{\leq 0} \\ c_{cc} \underbrace{\left(s_k c_{dd} - s_d c_{dk}\right)}_{>0} \end{bmatrix}.$$

Checking these equalities component-wise verifies that $\frac{\partial t_i^*}{\partial a_j} > 0$ for each $i \in \{c, d, k\}$ and $j \in \{1, 2\}$.

(ii) Once we evaluate $\mu(t_c, t_k) = s_c t_c + s_k t_k$ and $\nu(t_d, t_k) = s_d t_d + s_k t_k$ in the optimal vector of training inputs, t^* given as in (A.16), and consider items (i)-(iii) in the proof of Proposition 1, we obtain that

$$\frac{\partial \mu}{\partial a_1} = \frac{1}{2\,|C|}\left[s_k c_{dd}\left(s_k c_{cc} - 2 s_c c_{ck}\right) + s_c^2\left(c_{kk} c_{dd} - c_{dk}^2\right)\right] > 0,$$

$$\frac{\partial \mu}{\partial a_2} = \frac{1}{2\,|C|}\left(s_k c_{cc} - s_c c_{ck}\right)\left(s_k c_{dd} - s_d c_{dk}\right) > 0,$$

$$\frac{\partial \nu}{\partial a_1} = \frac{1}{2\,|C|}\left(s_k c_{cc} - s_c c_{ck}\right)\left(s_k c_{dd} - s_d c_{dk}\right) > 0, \quad \text{and}$$

$$\frac{\partial \nu}{\partial a_2} = s_k c_{cc}\left(s_k c_{dd} - 2 s_d c_{dk}\right) + s_d^2\left(c_{cc} c_{kk} - c_{ck}^2\right) > 0.$$

Noticing that $q_c \equiv \mu(t_c, t_k)$ and $q_d \equiv \nu(t_d, t_k)$ completes the proof.

Proof of Proposition 2: We show that there is a unique compensation plan $\{a_1^*, a_2^*\}$ solving the principal's problem given by (A.8). From (A.8) and (A.10) we have that

$$TCE\left(a_1, a_2\right) = b_1 s_c t_c^* + \left(b_1 + b_2\right) s_k t_k^* + b_2 s_d t_d^* - \frac{1}{4} S^T C^{-1} S - \frac{1}{2} r\left(a_1^2 \sigma_c^2 + a_2^2 \sigma_d^2\right),$$

where $t^* = \frac{1}{2} C^{-1}\begin{bmatrix} a_1 s_c, & a_2 s_d, & \left(a_1 + a_2\right) s_k \end{bmatrix}^T$. Note that $TCE\left(a_1, a_2\right)$ is a quadratic function in a_1 and a_2. It follows that

$$\begin{aligned}
\frac{\partial TCE}{\partial a_1} &= \underbrace{b_1 s_c \frac{\partial t_c^*}{\partial a_1} + \left(b_1 + b_2\right) s_k \frac{\partial t_k^*}{\partial a_1} + b_2 s_d \frac{\partial t_d^*}{\partial a_1}}_{\equiv M_1 > 0} - \frac{1}{4} \frac{\partial}{\partial a_1}\left[S^T C^{-1} S\right] - r a_1 \sigma_c^2, \quad \text{and} \\
\frac{\partial TCE}{\partial a_2} &= \underbrace{b_1 s_c \frac{\partial t_c^*}{\partial a_2} + \left(b_1 + b_2\right) s_k \frac{\partial t_k^*}{\partial a_2} + b_2 s_d \frac{\partial t_d^*}{\partial a_2}}_{\equiv M_2 > 0} - \frac{1}{4} \frac{\partial}{\partial a_2}\left[S^T C^{-1} S\right] - r a_2 \sigma_d^2.
\end{aligned} \tag{A.17}$$

Let $(C^{-1})_{ij} \equiv d_{ij}$ (i.e., d_{ij} represents the (i, j)-entry of the inverse matrix of C), then

$$\frac{1}{2}\frac{\partial}{\partial a_1}\left[S^T C^{-1} S\right] = \underbrace{\left[d_{cc}s_c^2 + 2d_{ck}s_c s_k + d_{kk}s_k^2\right]}_{\equiv\beta_1} a_1 + \underbrace{\left[d_{cd}s_c s_d + d_{ck}s_c s_k + d_{dk}s_d s_k + d_{kk}s_k^2\right]}_{\equiv\beta_2} a_2,$$

$$\frac{1}{2}\frac{\partial}{\partial a_2}\left[S^T C^{-1} S\right] = \underbrace{\left[d_{cd}s_c s_d + d_{ck}s_c s_k + d_{dk}s_d s_k + d_{kk}s_k^2\right]}_{\equiv\delta_1} a_1 + \underbrace{\left[d_{dd}s_d^2 + 2d_{dk}s_d s_k + d_{kk}s_k^2\right]}_{\equiv\delta_2} a_2, \tag{A.18}$$

where $\beta_2 = \delta_1$. From (A.17) and (A.18), it follows that the first order conditions have a unique solution for a_1^* and a_2^* given by

$$\begin{aligned} a_1^*(\sigma_d, \sigma_c) &= 2\frac{M_1\left(2r\sigma_d^2 + \delta_2\right) - \beta_2 M_2}{\left(2r\sigma_d^2 + \delta_2\right)\left(2r\sigma_c^2 + \beta_1\right) - \beta_2^2}, \quad \text{and} \\ a_2^*(\sigma_d, \sigma_c) &= 2\frac{M_2\left(2r\sigma_c^2 + \beta_1\right) - \beta_2 M_1}{\left(2r\sigma_d^2 + \delta_2\right)\left(2r\sigma_c^2 + \beta_1\right) - \beta_2^2}. \end{aligned} \tag{A.19}$$

Let us show that a_1^* and a_2^*, defined by (A.19), are strictly positive. From items (i)-(iii) in the Proof of Proposition 1 we obtain

$$\begin{aligned} \delta_2\beta_1 - \beta_2^2 &= \frac{1}{|C|}\left[s_d^2 s_k^2 c_{cc} + s_c\left(s_c s_k^2 c_{dd} - 2s_d^2 s_k c_{ck} + s_c s_d\left(s_d c_{kk} - 2s_k c_{dk}\right)\right)\right] > 0, \\ M_1\delta_2 - \beta_2 M_2 &= \frac{b_1}{2|C|}\left[s_d^2 s_k^2 c_{cc} + s_c\left(s_c s_k^2 c_{dd} - 2s_d^2 s_k c_{ck} + s_c s_d\left(s_d c_{kk} - 2s_k c_{dk}\right)\right)\right] > 0, \quad \text{and} \\ M_2\beta_1 - \beta_2 M_1 &= \frac{b_2}{2|C|}\left[s_d^2 s_k^2 c_{cc} + s_c\left(-2s_d^2 s_k c_{ck} + s_c s_k^2 c_{dd} + s_c s_d\left(s_d c_{kk} - 2s_k c_{dk}\right)\right)\right] > 0. \end{aligned} \tag{A.20}$$

Notice that for a_1^* and a_2^* defined as in (A.19), the conditions in (A.20) are sufficient for $a_i^* > 0$ and $a_i^*(0,0) = b_i$, for each $i \in \{1,2\}$. Finally, for the global second-order conditions we have that

$$D^2TCE = -\frac{1}{2}\begin{bmatrix} \beta_1 + 2r\sigma_c^2 & \beta_2 \\ \beta_2 & \delta_2 + 2r\sigma_d^2 \end{bmatrix},$$

which implies that,

$$\begin{aligned} H_1 &\equiv -\frac{1}{2}\left(\beta_1 + 2r\sigma_c^2\right) < 0, \quad \text{and} \\ H_2 &\equiv \det D^2TCE = \frac{1}{4}\left[\left(\beta_1 + 2r\sigma_c^2\right)\left(\delta_2 + 2r\sigma_d^2\right) - \beta_2^2\right] > \frac{1}{4}\left[\beta_1\delta_2 - \beta_2^2\right] > 0. \end{aligned}$$

Thus, the second-order conditions are satisfied and the compensation plan $\{a_1^*, a_2^*\}$, defined by (A.19), uniquely maximizes $TCE(a_1, a_2)$.

Proof of Corollary 2: From (A.19) and the conditions seen in (A.20) we have that $\frac{\partial a_i^*}{\partial \sigma_d^2}$, for $i \in \{1, 2\}$, are:

$$\begin{aligned} \frac{\partial a_1^*}{\partial \sigma_d^2} &= \frac{4\beta_2 r \left[M_2\left(\beta_1 + 2r\sigma_c^2\right) - \beta_2 M_1\right]}{\left(\beta_2^2 - \left(\beta_1 + 2r\sigma_c^2\right)\left(\delta_2 + 2r\sigma_d^2\right)\right)^2} > 0, \quad \text{and} \\ \frac{\partial a_2^*}{\partial \sigma_d^2} &= -\frac{4r\left(\beta_1 + 2r\sigma_c^2\right)\left[M_2\left(\beta_1 + 2r\sigma_c^2\right) - \beta_2 M_1\right]}{\left(\beta_2^2 - \left(\beta_1 + 2r\sigma_c^2\right)\left(\delta_2 + 2r\sigma_d^2\right)\right)^2} < 0. \end{aligned} \tag{A.21}$$

Hence, we can characterize $\frac{\partial t_i^*}{\partial \sigma_d^2}$ for $i \in \{c, d, k\}$. First, note that from (A.21)

$$\frac{\partial a_1^*}{\partial \sigma_d^2} = -\frac{\beta_2}{\left(\beta_1 + 2r\sigma_c^2\right)} \frac{\partial a_2^*}{\partial \sigma_d^2}. \tag{A.22}$$

Then, replacing according to (A.19) and (A.22) in (A.10) we obtain that

$$\begin{bmatrix} \frac{\partial t_c^*}{\partial \sigma_d^2} \\ \frac{\partial t_d^*}{\partial \sigma_d^2} \\ \frac{\partial t_k^*}{\partial \sigma_d^2} \end{bmatrix} = \underbrace{\frac{1}{2(\beta_1 + 2r\sigma_c^2)}}_{>0} \underbrace{\frac{\partial a_2^*}{\partial \sigma_d^2}}_{<0} \begin{bmatrix} \left(d_{cd}s_d + d_{ck}s_k\right)\left(\beta_1 + 2r\sigma_c^2\right) - \left(d_{cc}s_c + d_{ck}s_k\right)\beta_2 \\ \left(d_{dd}s_d + d_{dk}s_k\right)\left(\beta_1 + 2r\sigma_c^2\right) - \left(d_{cd}s_c + d_{dk}s_k\right)\beta_2 \\ \left(d_{dk}s_d + d_{kk}s_k\right)\left(\beta_1 + 2r\sigma_c^2\right) - \left(d_{ck}s_c + d_{kk}s_k\right)\beta_2 \end{bmatrix}. \tag{A.23}$$

Notice that the conditions

$$\begin{aligned} &\left(d_{dd}s_d + d_{dk}s_k\right)\left(\beta_1 + 2r\sigma_c^2\right) - \left(d_{cd}s_c + d_{dk}s_k\right)\beta_2 > 0, \quad \text{and} \\ &\left(d_{dk}s_d + d_{kk}s_k\right)\left(\beta_1 + 2r\sigma_c^2\right) - \left(d_{ck}s_c + d_{kk}s_k\right)\beta_2 > 0 \end{aligned}$$

always hold and result in $\frac{\partial t_d^*}{\partial \sigma_d^2} < 0$ and $\frac{\partial t_k^*}{\partial \sigma_d^2} < 0$ respectively. However, for $\frac{\partial t_c^*}{\partial \sigma_d^2} < 0$ we need

$$\left(d_{cd}s_d + d_{ck}s_k\right)\left(\beta_1 + 2r\sigma_c^2\right) - \left(d_{cc}s_c + d_{ck}s_k\right)\beta_2 > 0,$$

which holds true if and only if

$$2r\sigma_c^2 > \frac{\left(d_{cc}s_c + d_{ck}s_k\right)\beta_2}{d_{cd}s_d + d_{ck}s_k} - \beta_1 = \frac{s_c s_k}{-c_{ck}},$$

concluding the proof.

Proof of Corollary 3: From (A.23) it follows that $\frac{\partial t_c^*}{\partial \sigma_d^2} + \frac{\partial t_d^*}{\partial \sigma_d^2} + \frac{\partial t_k^*}{\partial \sigma_d^2} = \frac{\kappa_1}{2(\beta_1 + 2r\sigma_c^2)} \frac{\partial a_2^*}{\partial \sigma_d^2}$, where

$$\kappa_1 = 2r\sigma_c^2\left(s_d(d_{cd}+d_{dd}+d_{dk})+s_k(d_{ck}+d_{dk}+d_{kk})\right)+s_c^2\left(s_d c_{kk}-s_k c_{dk}\right) \\ +s_d s_k(s_k c_{cc}-2s_c c_{ck})+s_c s_k c_{dd}(s_c-s_k)+s_c s_d c_{dk}(s_k-s_c). \tag{A.24}$$

Since $(\beta_1+2r\sigma_c^2)>0$ and $\frac{\partial a_2^*}{\partial\sigma_d^2}<0$, then $\frac{\partial t_c^*}{\partial\sigma_d^2}+\frac{\partial t_d^*}{\partial\sigma_d^2}+\frac{\partial t_k^*}{\partial\sigma_d^2}<0$ if and only if $\kappa_1>0$. Using items (i)-(iii) in the proof of Proposition 1 and the fact that $d_{ij}>0$ for all $i,j\in\{c,d,k\}$, we get directly that all the terms in the right-hand side of (A.24) are positive but $s_c s_k c_{dd}(s_c-s_k)+s_c s_d c_{dk}(s_k-s_c)$. However, the positivity of both such terms is achieved thanks to the hypothesis that $s_c>s_k$.

Proof of Corollary 4: As $q_c\equiv\mu(t_c,t_k)=s_c t_c+s_k t_k$ and $q_d\equiv\nu(t_d,t_k)=s_d t_d+s_k t_k$, it follows from (A.23) that $\frac{\partial q^*}{\partial\sigma_d^2}=s_c\frac{\partial t_c^*}{\partial\sigma_d^2}+s_d\frac{\partial t_d^*}{\partial\sigma_d^2}+2s_k\frac{\partial t_k^*}{\partial\sigma_d^2}=\frac{\kappa_2}{2(\beta_1+2r\sigma_c^2)}\frac{\partial a_2^*}{\partial\sigma_d^2}$, where

$$\kappa_2 = 2r\sigma_c^2 s_d^2(c_{cc}c_{kk}-c_{ck}^2)+s_c^2 s_d^2 c_{kk}+s_k^2 c_{dd}(s_c^2+4r\sigma_c^2 c_{cc})+s_d^2 s_k^2 c_{cc} \\ +2s_c c_{ck}(r\sigma_c^2 s_d c_{dk}-s_k(s_d^2+r\sigma_c^2 c_{dd}))-2s_d s_k c_{dk}(s_c^2+3r\sigma_c^2 c_{cc}). \tag{A.25}$$

Since $(\beta_1+2r\sigma_c^2)>0$ and $\frac{\partial a_2^*}{\partial\sigma_d^2}<0$, then $\frac{\partial q^*}{\partial\sigma_d^2}<0$ if and only if $\kappa_2>0$. Nevertheless, using items (i)-(iii) in the proof of Proposition 1 we conclude that the latter condition always holds, as all the terms in the right-hand side of (A.25) are positive.

A.5.1 Simulation

Assume that the principal's marginal compensations are given by $b_1=1$ and $b_2=1$, the marginal productivities are $s_c=2$, $s_d=2$, and $s_k=2$. Additionally, $r=1$, $\sigma_c=3$ and

$$C=\begin{bmatrix}1 & 0 & -0.5\\ 0 & 1 & -0.5\\ -0.5 & -0.5 & 1\end{bmatrix}. \tag{A.26}$$

Panel A of Figure A.1 plots a_1^* and a_2^* as functions of σ_d. Similarly, Panel B of Figure A.1 shows that as the performance monitoring technology decreases the noise at the drive-thru, the optimal level of training effort at the kitchen and at the drive-thru increase above the optimal level of training effort at the counter.

Note that as σ_d decreases the optimal compensation for the production at the counter, a_1^*, decreases, while the optimal compensation for production at the drive-thru, a_2^*, increases, as predicted by Corollary 2(i).

Figure A.1: Simulation

Panel A. Optimal compensation rates

Panel B. Optimal training efforts t^*

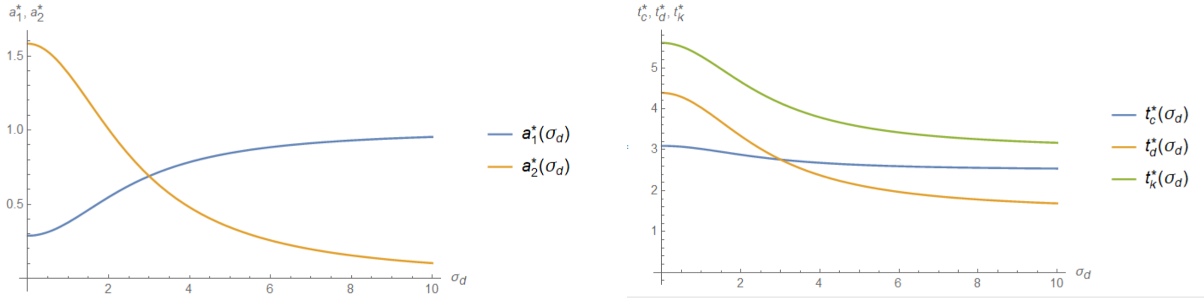
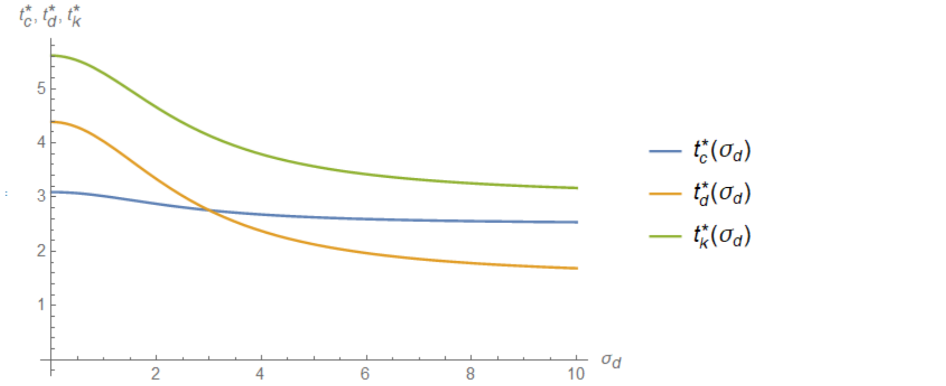


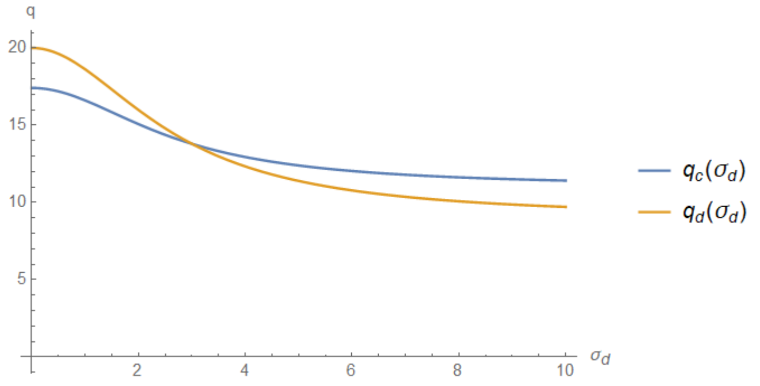
Figure A.2: Optimal training efforts t^*



Note: x-axis $= \sigma_d$, y-axis $=$ training efforts. The simulation used the following parameters: $b_1 = 1$, $b_2 = 1$, $s_c = 2$, $s_d = 2$, $s_k = 2$, $r = 1$, $\sigma_c = 3$ and C given by (A.26).

Figure A.2 shows that the optimal training efforts t_c^*, t_d^* and t_k^*, increase as σ_d decreases. Finally, Figure A.3 shows that the optimal production levels $q_c^* = \mu\left(t_c^*, t_k^*\right)$ and $q_d = \nu\left(t_d^*, t_k^*\right)$, increase as σ_d decreases.

Figure A.3: Optimal production levels q_c^* and q_d^*



Note: x-axis = σ_d, y-axis = Compensation rates. For this graph we have fixed parameters $b_1 = 1$, $b_2 = 1$, $s_c = 2$, $s_d = 2$, $s_k = 2$, $r = 1$, $\sigma_c = 3$ and C given by (A.26).

A.6 CES production functions and linear cost function

In this section, we modify some assumptions of the model presented in Section A. In particular, we assume that both production functions, $\mu, \nu : \mathbb{R}^2 \longrightarrow \mathbb{R}$, are CES functions (instead of linear functions). That is,

$$\begin{aligned} \mu\left(t_c, t_k\right) &= \mu_0 \left(\lambda t_c^{\rho} + (1-\lambda) t_k^{\rho}\right)^{\frac{1}{\rho}}, \quad \text{and} \\ \nu\left(t_d, t_k\right) &= \nu_0 \left(\gamma t_d^{s} + (1-\gamma) t_k^{s}\right)^{\frac{1}{s}}, \end{aligned} \tag{A.27}$$

where $\{\mu_0, \nu_0, \lambda, \gamma\}$ are positive constants. Similarly, for tractability, we assume that $C(t)$ is linear (instead of a quadratic function), i.e.,

$$C(t) = c_c t_c + c_d t_d + c_k t_k,$$

where $\{c_c, c_d, c_k\}$ are positive constants representing the marginal cost of each training effort. We assume that there is a maximum level of total training effort T, so that the constraint

$$t_c + t_d + t_k \leq T, \tag{A.28}$$

must be satisfied at the equilibrium. The following proposition characterizes the store manager's optimal vector of training efforts given the compensation rates $\{a_1, a_2\}$.

Proposition 3. *Suppose there is a unique solution w^* of the equation*

$$\frac{(1-\lambda)}{\lambda} \alpha\left(w, a_1\right)^{1-\rho} \left(c_c + w\right) + \frac{(1-\gamma)}{\gamma} \beta\left(w, a_2\right)^{1-s} \left(c_d + w\right) - c_k - w = 0, \tag{A.29}$$

where

$$\alpha\left(w, a_1\right) \equiv \left(\frac{1-\lambda}{\left[\frac{c_c + w}{\lambda a_1 \mu_0}\right]^{\frac{\rho}{1-\rho}} - \lambda} \right)^{\frac{1}{\rho}} \quad \text{and} \quad \beta\left(w, a_2\right) \equiv \left(\frac{1-\gamma}{\left[\frac{c_d + w}{\gamma a_2 \nu_0}\right]^{\frac{s}{1-s}} - \gamma} \right)^{\frac{1}{s}}. \tag{A.30}$$

Then, there is a unique vector t^* solving the FOC of the store manager's problem given by

$$\begin{aligned} t_c^* &= \alpha\left(w^*, a_1\right) t_k^*, \\ t_d^* &= \beta\left(w^*, a_2\right) t_k^*, \\ t_k^* &= \frac{T}{\alpha\left(w^*, a_1\right) + \beta\left(w^*, a_2\right) + 1}. \end{aligned} \tag{A.31}$$

Proof of Proposition 3. The first-order conditions of the store manager's problem are given by

$$\begin{aligned} a_1\mu_0\lambda t_c^{\rho-1}\left(\lambda t_c^{\rho}+(1-\lambda)t_k^{\rho}\right)^{\frac{1}{\rho}-1}-c_c-w=0,\\ a_2\nu_0\gamma t_d^{s-1}\left(\gamma t_d^{s}+(1-\gamma)t_k^{s}\right)^{\frac{1}{s}-1}-c_d-w=0, \end{aligned} \tag{A.32}$$

$$a_1\mu_0(1-\lambda)t_k^{\rho-1}\left(\lambda t_c^{\rho}+(1-\lambda)t_k^{\rho}\right)^{\frac{1}{\rho}-1}+a_2\nu_0(1-\gamma)t_k^{s-1}\left(\gamma t_d^{s}+(1-\gamma)t_k^{s}\right)^{\frac{1}{s}-1}-c_k-w=0, \tag{A.33}$$

and

$$T-\sum_i t_i=0. \tag{A.34}$$

From (A.32)-(A.34), it follows that

$$t_c=\underbrace{\left(\frac{1-\lambda}{\left[\frac{c_c+w}{\lambda a_1\mu_0}\right]^{\frac{\rho}{1-\rho}}-\lambda}\right)^{\frac{1}{\rho}}}_{\equiv\alpha}t_k \quad \text{and} \quad t_d=\underbrace{\left(\frac{1-\gamma}{\left[\frac{c_d+w}{\gamma a_2\nu_0}\right]^{\frac{s}{1-s}}-\gamma}\right)^{\frac{1}{s}}}_{\equiv\beta}t_k \tag{A.35}$$

Thus, $t_c=\alpha\left(w\right)t_k$, $t_d=\beta\left(w\right)t_k$ and $T-\alpha\left(w\right)t_k-\beta\left(w\right)t_k-t_k=0$, where w satisfies the equation

$$\frac{(1-\lambda)}{\lambda}\alpha\left(w\right)^{1-\rho}\left(c_c+w\right)+\frac{(1-\gamma)}{\gamma}\beta\left(w\right)^{1-s}\left(c_d+w\right)-c_k-w=0.$$

If (A.29) has a unique solution. From (A.9) and (A.27), the principal's problem becomes

$$\max_{a_1,a_2} TCE\left(a_1,a_2\right)=$$

$$\Big[b_1\mu_0\left(\lambda\alpha^{\rho}+(1-\lambda)\right)^{\frac{1}{\rho}}+b_2\nu_0\left(\gamma\beta^{s}+(1-\gamma\right)^{\frac{1}{s}}-\left(c_c\alpha+c_d\beta+c_k\right)t_k^*-\frac{1}{2}r\left(a_1^2\sigma_c^2+a_2^2\sigma_d^2\right). \tag{A.36}$$

Solving the principal's problem requires to solve a complex non-linear system of equations. In the next subsection, we present a simulation of the principal's problem solution.

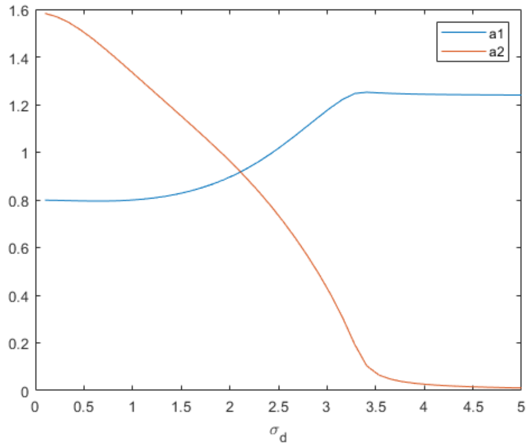
A.7 Simulation CES production functions

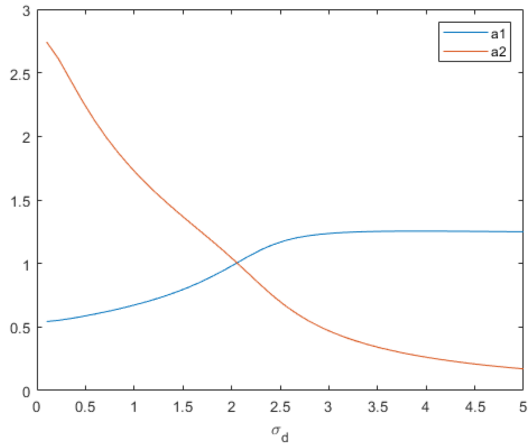
Suppose that $b_1=2$, $b_2=3$, $c_c=3$, $c_d=1$, $c_k=4$, $r=0.5$, $\sigma_c=3$, $T=4$, $\lambda=0.5$, $\gamma=0.5$, $\mu_0=4$, $\nu_0=3$. Figure A.4 shows the graphs of $a_1^*\left(\sigma_d\right)$ and $a_2^*\left(\sigma_d\right)$ as functions of σ_d for different values of ρ and s. Similarly, Figure A.5 and A.6 show the graphs of t^*,

q_c^* and q_d^* as functions of σ_d for different values of ρ and s.

Figure A.4: Optimal compensation rates a_1^* and a_2^*

$\rho = -1$, and $s = -1$ $\qquad$ $\rho = 0.2$, and $s = 0.1$





Note: x-axis $= \sigma_d$, y-axis = compensation rates. The simulation presented in Figure A.4 used the following parameters: $b_1 = 2$, $b_2 = 3$, $c_c = 3$, $c_d = 1$, $c_k = 4$, $r = 0.5$, $\sigma_c = 3$, $T = 4$, $\lambda = 0.5$, $\gamma = 0.5$, $\mu_0 = 4$, $\nu_0 = 3$.

Note that as σ_d decreases the optimal compensation for the production at the counter, a_1^*, decreases, while the optimal compensation for production at the drive-thru, a_2^*, increases. This result is true for both set of parameters, $(\rho, s) = (-1, -1)$, and $(\rho, s) = (0.2, 0.1)$.

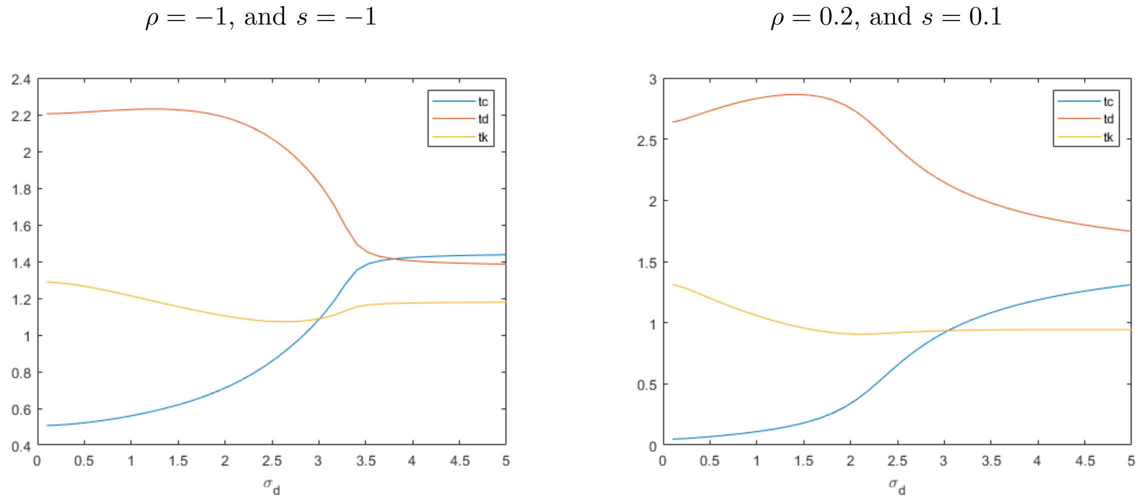


Figure A.5: Optimal training efforts t^*

Note: x-axis = σ_d, y-axis = Training efforts. For this graph we have fixed parameters $b_1 = 2$, $b_2 = 3$, $c_c = 3$, $c_d = 1$, $c_k = 4$, $r = 0.5$, $\sigma_c = 3$, $T = 4$, $\lambda = 0.5$, $\gamma = 0.5$, $\mu_0 = 4$, $\nu_0 = 3$.

Figure A.2 shows that the optimal training efforts t_c^* and t_d^* decrease as σ_d decreases and approaches to 0, while t_k^* increases as σ_d decreases and approaches to 0. Once again, this result is true for both set of parameters, $(\rho, s) = (-1, -1)$, and $(\rho, s) = (0.2, 0.1)$. Finally, Figure A.3 shows that the optimal production at the counter, $q_c^* = \mu\,(t_c^*, t_k^*)$, decreases as σ_d decreases, while the optimal production at the drive-thru increases as σ_d decreases.

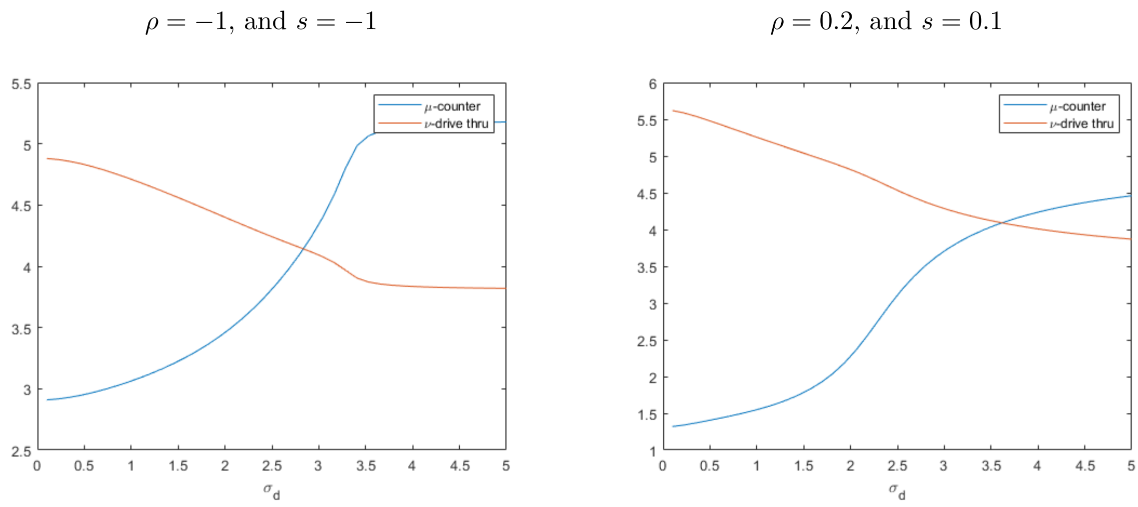


Figure A.6: Optimal production levels q_c^* and q_d^*

Note: x-axis $= \sigma_d$, y-axis $=$ Compensation rates. For this graph we have fixed parameters $b_1 = 2$, $b_2 = 3$, $c_c = 3$, $c_d = 1$, $c_k = 4$, $r = 0.5$, $\sigma_c = 3$, $T = 4$, $\lambda = 0.5$, $\gamma = 0.5$, $\mu_0 = 4$, $\nu_0 = 3$.

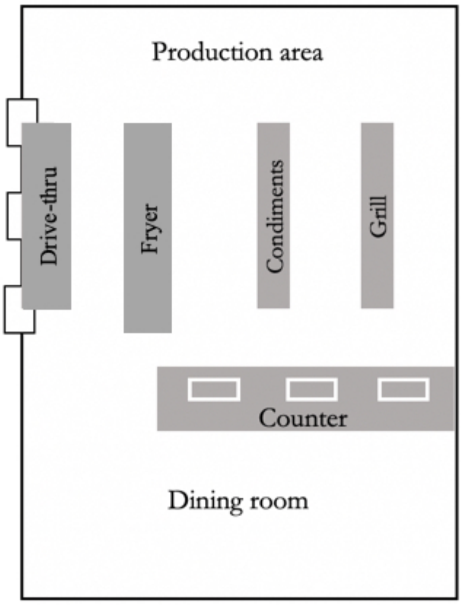
B Additional Statistics

Table B.1: Number of approved training programs per station

Station	Mean	Std. Dev
Opening and production transition	1.67	1.55
Opening and service transition	1.65	1.18
Drive-thru	2.67	2.10
Hot and cold drinks	2.21	2.05
Service closure	1.70	1.20
Production closure	1.69	1.33
Trainer	1.76	1.28
Hospitality	2.69	3.44
Cleaning and maintenance	1.96	2.73
Lobby, parking and playground	1.64	1.50
Breakfast menu preparation	1.38	0.84
Regular menu preparation	1.86	1.30
French fries and hash browns	2.39	2.73
Breakfast menu grill	1.40	0.85
Regular menu grill	1.85	1.35
Desserts and ice creams	2.17	2.09
Assembly	1.42	0.96
Personal preparation	1.69	1.32
Fried products	1.87	1.79
Security	2.91	4.79
Food security	3.00	4.74
Personal security	2.77	4.48
Cashier	2.12	1.58

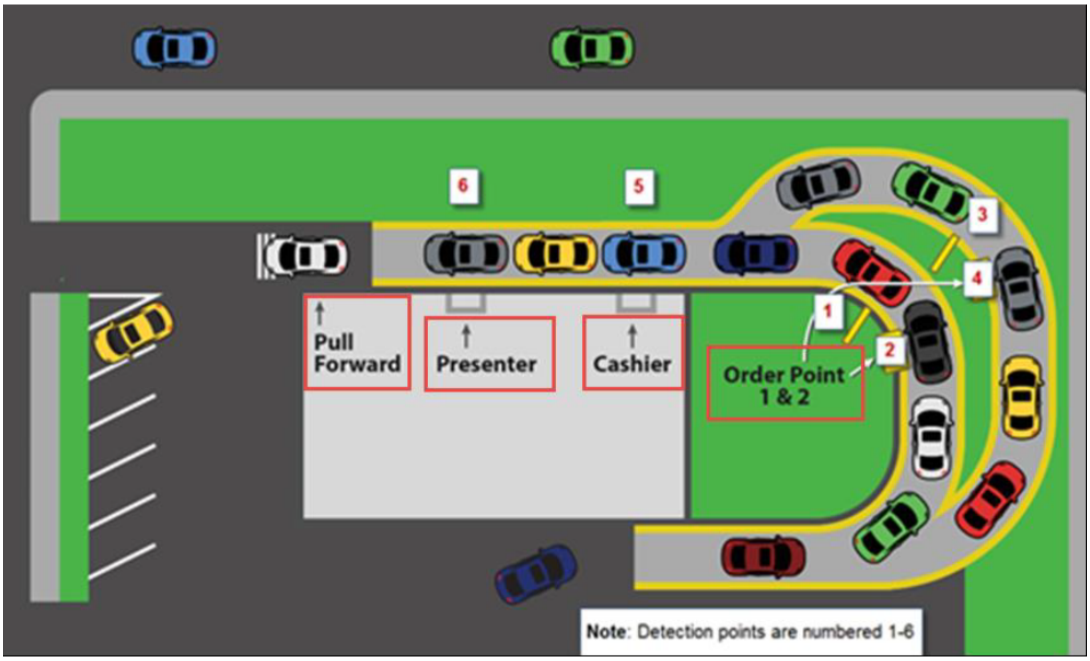
Note: Number of approved training programs biweekly per store by area

Figure B.1: Production process



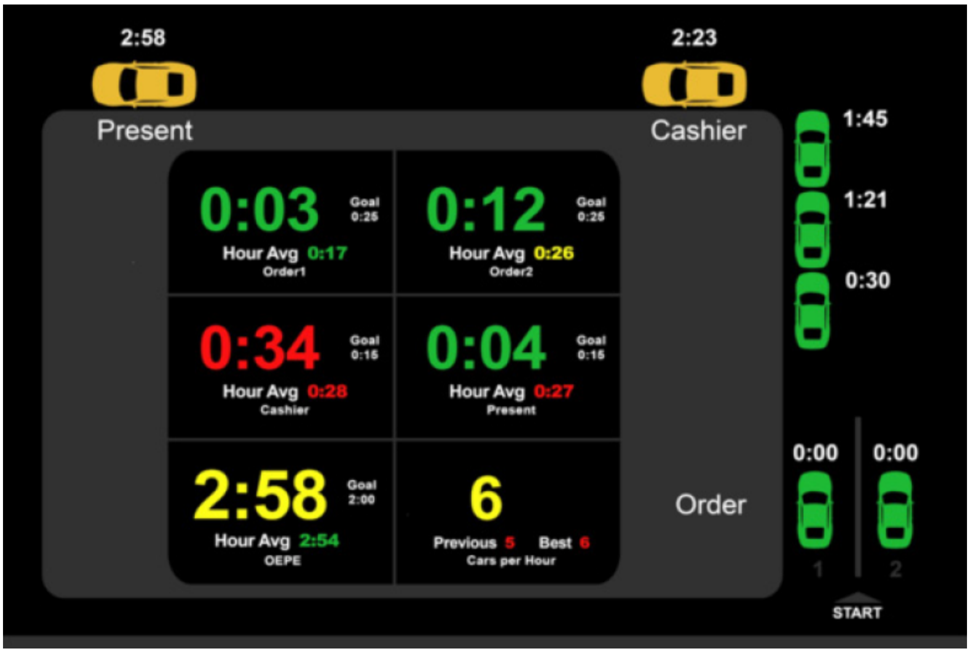
Notes: Figure B.1 shows the layout of typical store. In the kitchen, orders are completed in five stations: assembly (including condiments), grill, fryer, soda fountain, and desserts. When a burger is ordered, for example, the bun is first toasted, then placed in a box or wrapping paper, followed by the addition of condiments and vegetables. Afterwards, meat from the grill cooked following strict specifications is added and packaged for the customer's consumption. The final assembled product is then delivered at the counter or at the pick up window of the drive-thru.

Figure B.2: Loop points



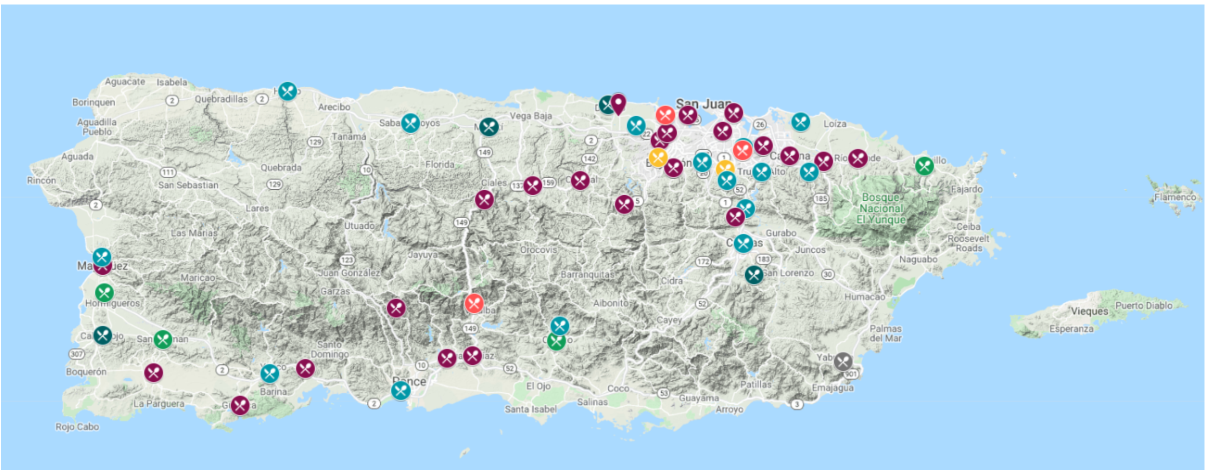
Notes: Figure B.2 shows the underground detectors installed at the order, cashier, presenter and put forward points, which record the exact time in which a car reaches each point.

Figure B.3: Drive Thru Timer Monitor



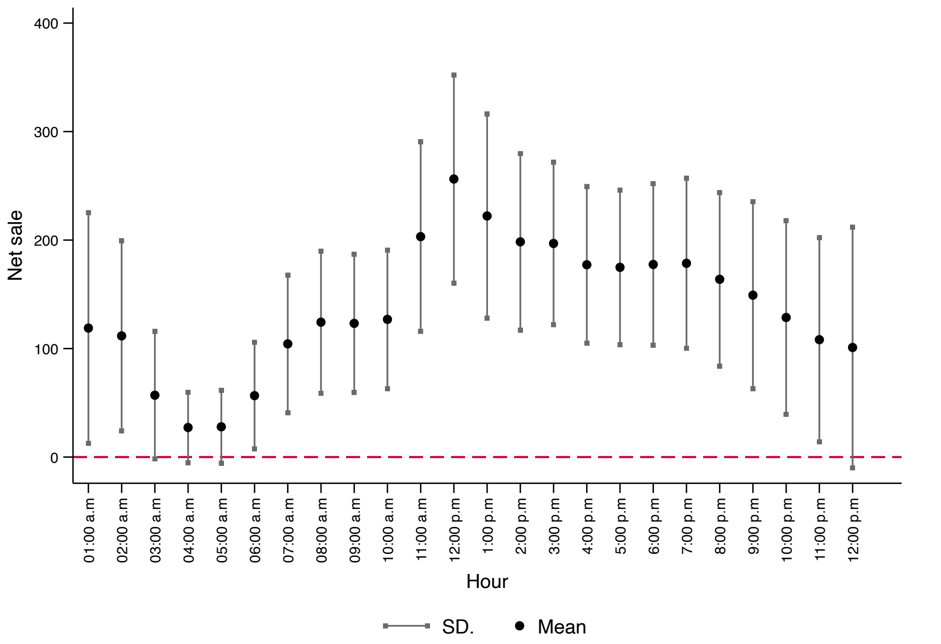
Notes: Figure B.3 presents an example of the dashboard; a monitor installed by the presenter window, and displays detection point times (time a car spends at the order, cashier and presenter windows), average times (updating after every car departure), and car counts (per hour, pace estimate based on the past 20 minutes speed of service).

Figure B.4: Implementations by geographical location



Notes: Figure B.4 shows the geographical pattern in the implementation of the technology across restaurants. It reassures that there is no geographical pattern in the implementation of the technology.

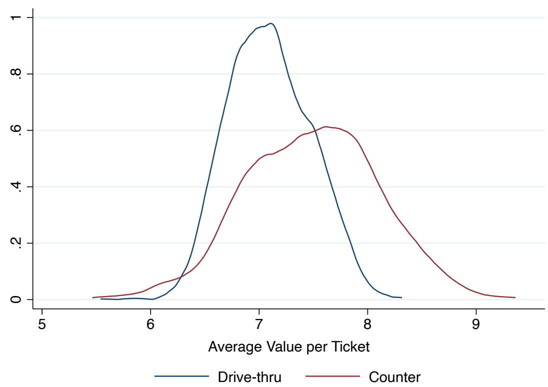
Figure B.5: Hourly sales



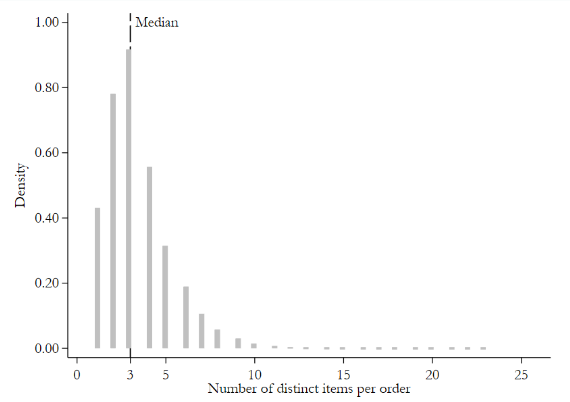
Notes: Figure B.5 plots the average of daily net sales for each hour of the day across biweekly periods.

Figure B.6: Order Composition

Panel A. Sales

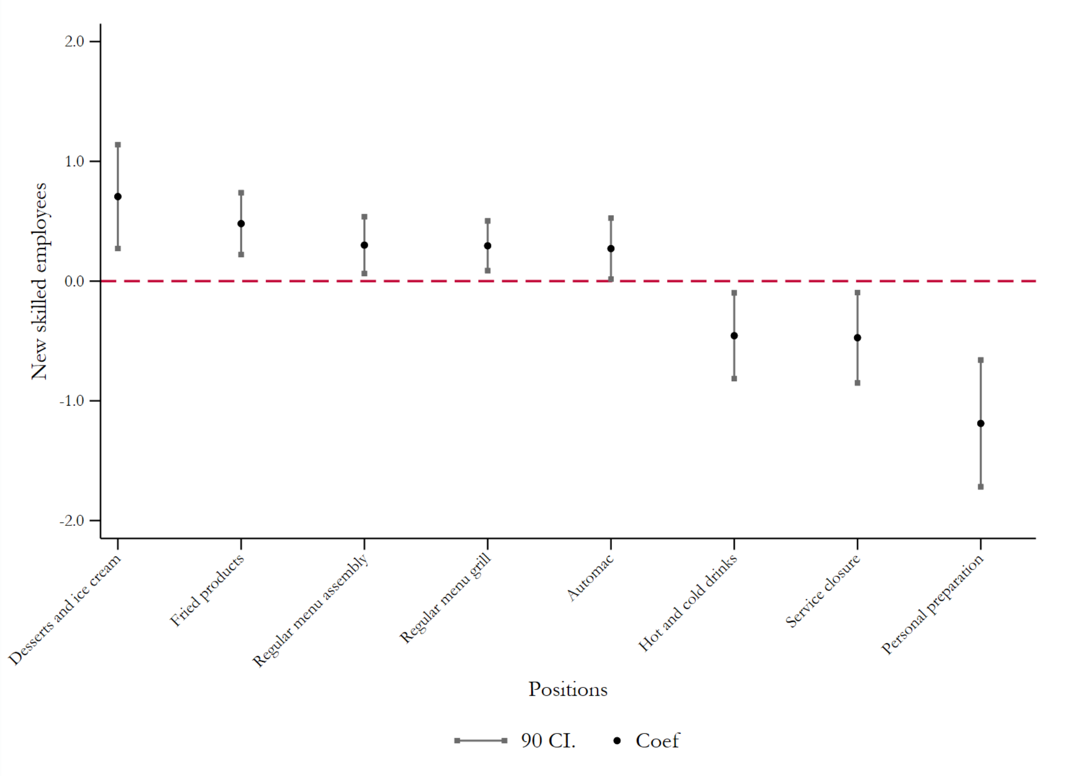


Panel B. Distinct items per ticket



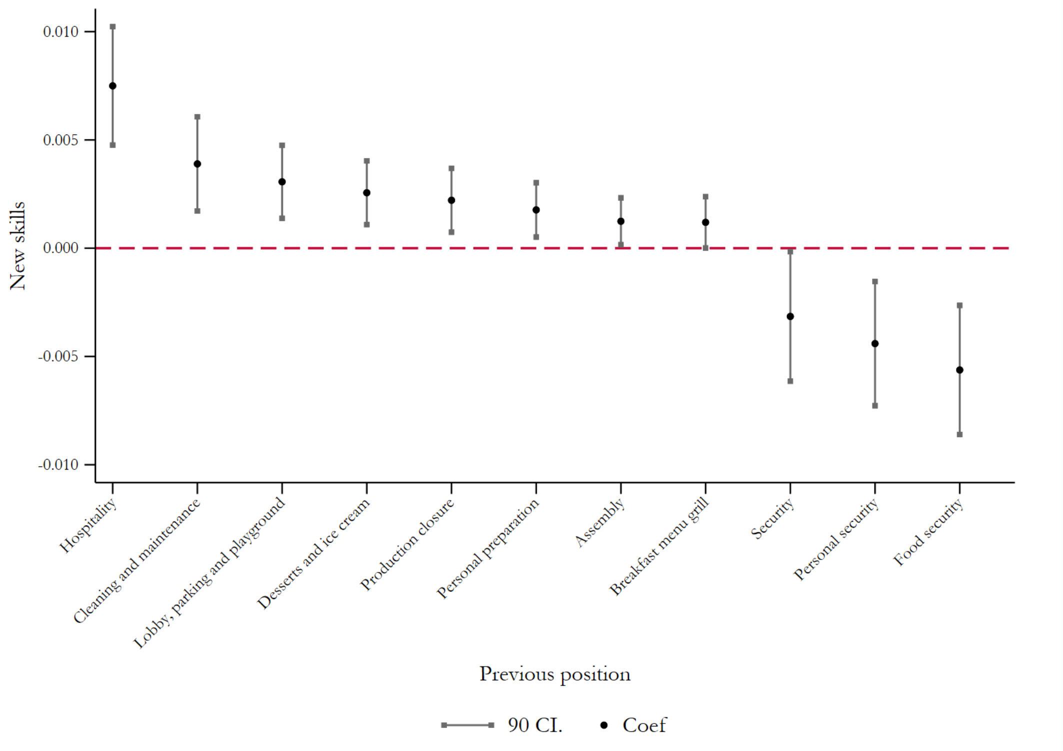
Notes: Figure B.6 plots the distribution of two measures of order composition. Panel A presents the average value per ticket. Panel B shows the number of distinct items per ticket, and the bar represents the median equal to 3 items per ticket.

Figure B.7: DD estimates of the implementation on training by program



Notes: Figure B.7 shows DD coefficients and 95 percent confidence intervals from estimation of equation (1), replacing all the time to treatment indicators with a post-treatment dummy (before and after), on each training station. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores implementing the new technology between March 2018 to October 2019.

Figure B.8: DD estimates of the implementation on training by previous skills



Notes: Figure B.8 shows DD coefficients and 95 percent confidence intervals from estimation of equation (1), replacing all the time to treatment indicators with a post-treatment dummy (before and after), on new skills conditional for each previous training station. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019.

C Additional Evidence

C.1 Difference in-differences estimator

Table C.1: Effects of technology implementations on key performance measures

	Whole store	**Drive-thru**	**Counter**
		Panel A. Sales	
Drive-thru monitoring technology	3,372**	2,877***	448.8
Stand. Err.	(1,647)	(987.5)	(813)
Mean of the dep. var.	76,080	50,734	25,432
Relative effect	4.40%	5.67%	1.76%
		Panel B. Units sold	
Drive-thru monitoring technology	1,938**	1,661***	242.9
Stand. Err.	(894.2)	(563.5)	(410.8)
Mean of the dep. var.	41,432	28,488	12,989
Relative effect	4.68%	5.83%	1.87%
		Panel C. Number of tickets	
Drive-thru monitoring technology	499.3**	440.5***	51.36
Stand. Err.	(221.4)	(137.5)	(107.8)
Mean of the dep. var.	10,531	7,138	3,405
Relative effect	4.74%	6.17%	1.51%
Observations	2,074	2,073	2,069

Notes: Table C.1 shows the results for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Column 1 presents the estimations for the whole store. Column 2 presents the estimations only for the orders placed at the Drive-thru section and Column 3 presents the estimations only for the orders placed at the Counter. Observation numbers change across sections due to unbalanced stores across the whole period. Numbers in parentheses are clustered standard errors at the store level. The table shows that the point estimates are consistently positive for sales, units sold and number of tickets and seem to be larger and more precisely estimated for the Drive-thru than the Counter. * significant 10%, ** significant 5%, *** significant 1%

Table C.2: Effects of technology implementations on (log) key performance measures

	Whole store	**Drive-thru**	**Counter**
		Panel A. Sales	
Drive-thru monitoring technology	0.0697	0.0656**	0.0550
Stand. Err.	(0.0419)	(0.0313)	(0.0607)
		Panel B. Units sold	
Drive-thru monitoring technology	0.0707*	0.0652**	0.0545
Stand. Err.	(0.0419)	(0.0315)	(0.0600)
		Panel C. Number of tickets	
Drive-thru monitoring technology	0.0775*	0.0729**	0.0586
Stand. Err.	(0.0417)	(0.0313)	(0.0606)

Notes: Table C.2 shows the results for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy using the log of key performance measures. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Column 1 presents the estimations for the whole store. Column 2 presents the estimations only for the orders placed at the Drive-thru section and Column 3 presents the estimations only for the orders placed at the Counter. Observation numbers change across sections due to unbalanced stores across the whole period. Numbers in parentheses are clustered standard errors at the store level. The table shows that the point estimates are consistently positive for sales, units sold and number of tickets and seem to be larger and more precisely estimated for the Drive-thru than the Counter. * significant 10%, ** significant 5%, *** significant 1%

Table C.3: Effects of technology implementations on new hires

	New hires
Drive-thru monitoring technology	0.0532
Stand. Err.	(0.0888)
Mean of the dep. var.	1.48
Relative effect	3.58%
Observations	1,287

Notes: Table C.3 shows the results for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy. Treatment is defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between January 2019 to Novemnber 2019. Table shows average treatment effects for new hires. Numbers in parentheses are clustered standard errors at the store level. * significant 10%, ** significant 5%, *** significant 1%

Table C.4: Effects of technology implementations on sales at peak and non-peak hours

	Whole store	Drive-thru	Counter
	Panel A. Peak hours		
Drive-thru monitoring technology	2,206*	1,868**	297.7
Stand. Err.	(1,245)	(714.7)	(642.8)
Mean of the dep. var.	58,150	37,518	20,700
Relative effect	3.79%	4.98%	1.44%
	Panel B. Non-peak hours		
Drive-thru monitoring technology	1,166**	1,009***	151.1
Stand. Err.	(466.3)	(326.6)	(195.5)
Mean of the dep. var.	17,930	13,216	4,732
Relative effect	6.50%	7.63%	3.19%
Observations	2,074	2,073	2,069

Notes: Table C.4 shows the results for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy. Treatment defined as stores implementing the performance monitoring technology. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Column 1 presents the estimations for the whole store. Column 2 presents the estimations only for the orders placed at the Drive-thru section and Column 3 presents the estimations only for the orders placed at the the Counter. Numbers in parentheses are clustered standard errors at the store level. * significant 10%, ** significant 5%, *** significant 1%

Table C.5: Effects of technology implementations on the composition of the order

	Whole store	Drive-thru	Counter
	Panel A. Average value per ticket		
Drive-thru monitoring technology	-0.0570**	-0.0531**	-0.0433
Stand. Err.	(0.0230)	(0.0247)	(0.0324)
Mean of the dep. var.	7.218	7.095	7.491
Relative effect	-0.79%	-0.75%	-0.58%
	Panel B. Share of complex orders		
Drive-thru monitoring technology	-0.00337**	-0.00417**	-0.00217
Stand. Err.	(0.00149)	(0.00168)	(0.00273)
Mean of the dep. var.	0.374	0.380	0.360
Relative effect	0.90%	1.10%	0.60%
Observations	2,074	2,073	2,069

Notes: Table C.5 shows the results for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy. Treatment defined as stores implementing the performance monitoring technology on complex orders, which are orders with more than 3 distinct items (see Figure B.6b). The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Column 1 presents the estimations for the whole store. Column 2 presents the estimations only for the orders placed at the Drive-thru section and Column 3 presents the estimations only for the orders placed at the Counter. Numbers in parentheses are clustered standard errors at the store level. * significant 10%, ** significant 5%, *** significant 1%

Table C.6: Effects of technology implementations on kitchen training

	New kitchen training skills
Drive-thru monitoring technology	1.047**
Stand. Err.	(0.424)
Mean of the dep. var.	2.9706
Relative effect	34.47%
Observations	1,149

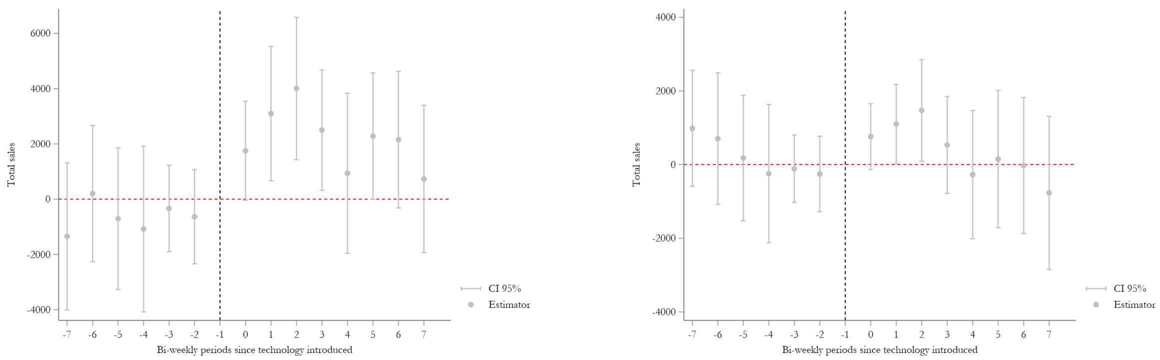
Notes: Table C.6 shows the results for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy. Treatment defined as stores implementing the performance monitoring technology on complex orders, which are orders with more than 3 distinct items (see Figure B.6b). Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. The Column presents the estimations for the kitchen skills of workers. Numbers in parentheses are clustered standard errors at the store level. * significant 10%, ** significant 5%, *** significant 1%

C.2 Additional Evidence using the Baseline Event-Study Sample

Figure C.1: Event study estimates on sales

Panel A. Event study estimates in Drive-thru

Panel B. Event study estimates in counter

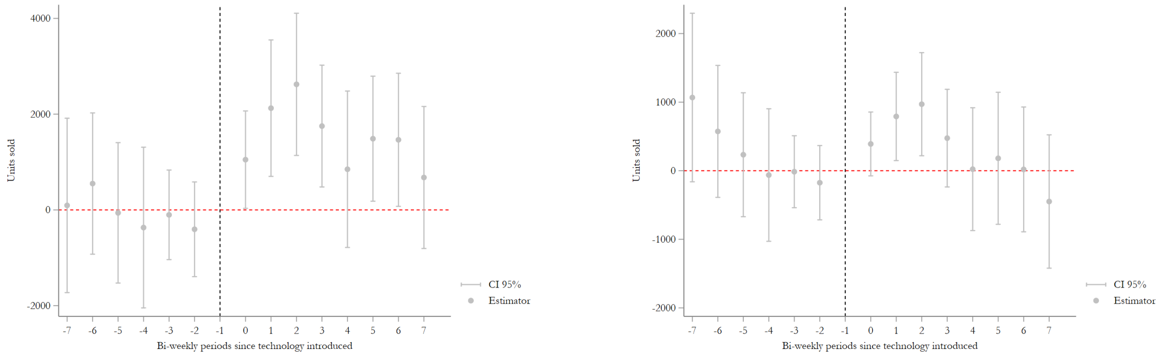


Note: Panels A and B of Figure C.1 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on sales. Treatment is defined as stores implementing the new technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows the estimations for the orders placed at the Drive-thru, and Panel B shows the estimations for the orders placed at the Counter. The vertical line represents the time of the treatment.

Figure C.2: Event study estimates on units sold

Panel A. Event study estimates in Drive-thru

Panel B. Event study estimates in counter

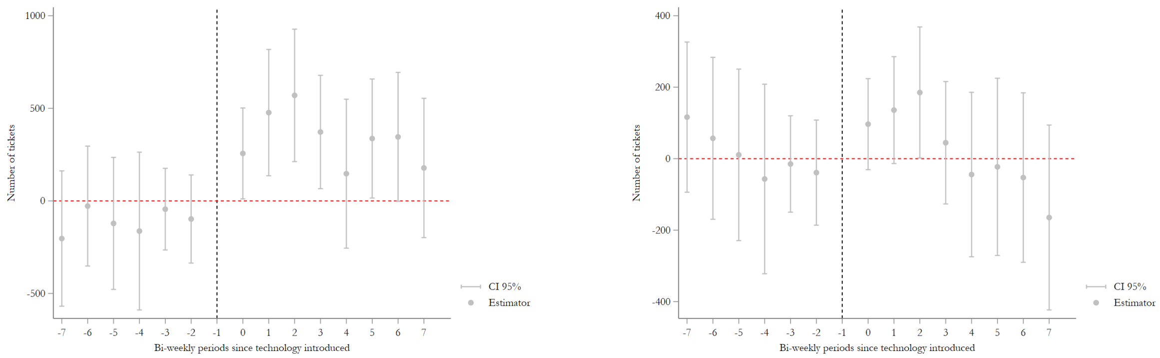


Note: Panels A and B of Figure C.2 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on units sold. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the orders placed at the the Drive-thru and Panel B shows estimations for the orders placed at the Counter. Vertical line represents time of treatment.

Figure C.3: Event study estimates on tickets

Panel A. Event study estimates in Drive-thru

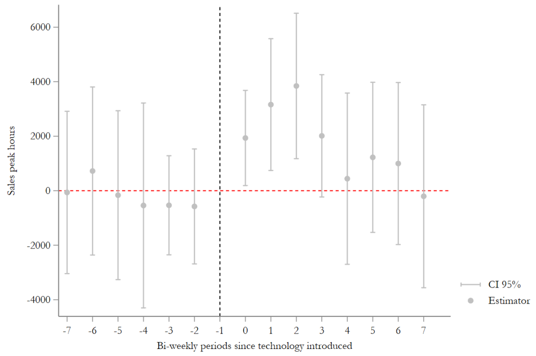
Panel B. Event study estimates in counter



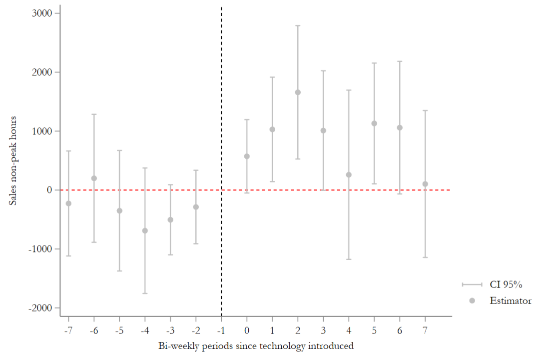
Note: Panels A and B of Figure C.3 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on tickets. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the orders placed at the the Drive-thru and Panel B shows estimations for the orders placed at the Counter. Vertical line represents time of treatment.

Figure C.4: Effects of technology implementations on sales at peak and non-peak hours

Panel A. Event study estimates on sales on peak hours



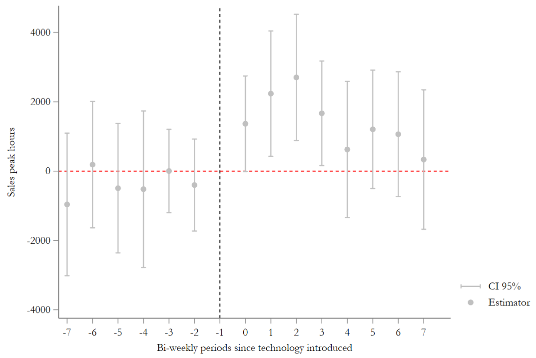
Panel B. Event study estimates on sales on non-peak hours



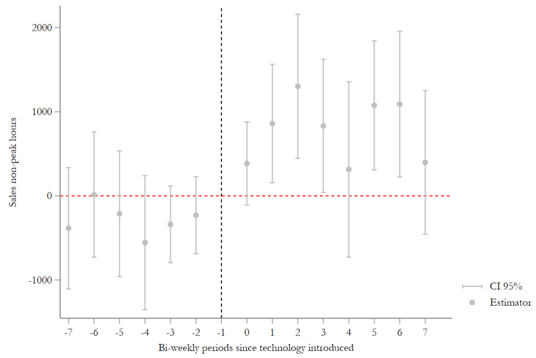
Note: Panels A and B of Figure C.4 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on sales at peak and non peak hours. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for orders placed on peak hours and panel B shows estimations for orders placed on non-peak hours. Vertical line represents time of treatment.

Figure C.5: Effects of technology implementations on sales at peak and non-peak hours

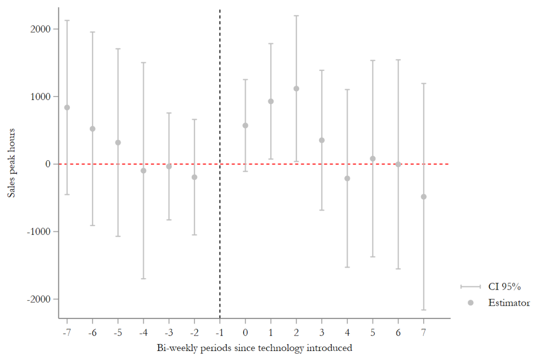
Panel A. Event study estimates drive-thru on peak hours



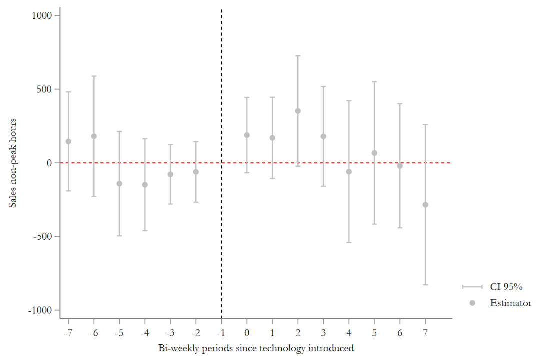
Panel B. Event study estimates in drive-thru on non-peak hours



Panel C. Event study estimates counter on peak hours



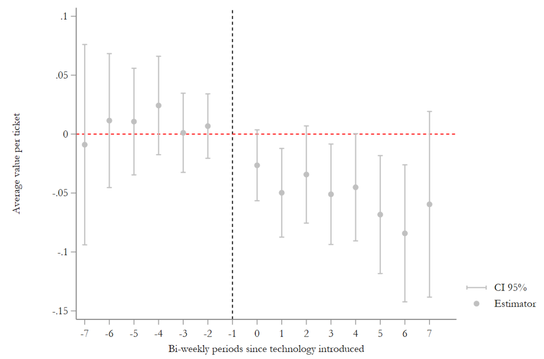
Panel D. Event study estimates in counter on non-peak hours



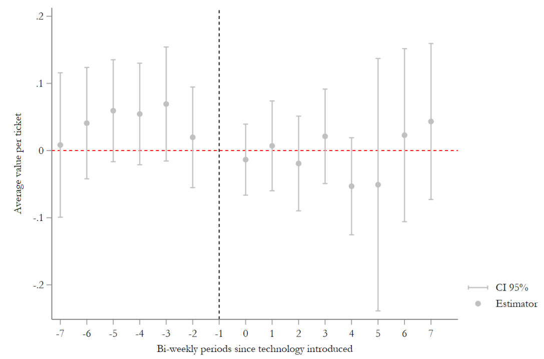
Note: Panels A, B, C and D of Figure C.5 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on sales at peak and non peak hours. Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A and B show estimations for sales for the orders placed at the Drive-thru and Panel C and D shows estimations for sales for the orders placed at the Counter. Vertical line represents time of treatment.

Figure C.6: Event study estimates on the composition of the order

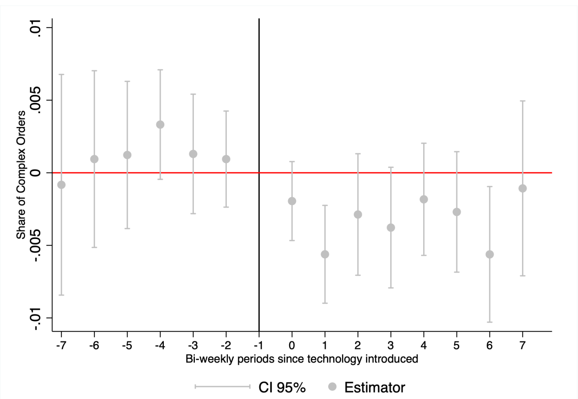
Panel A. Event study estimates on average value per ticket in Drive-thru



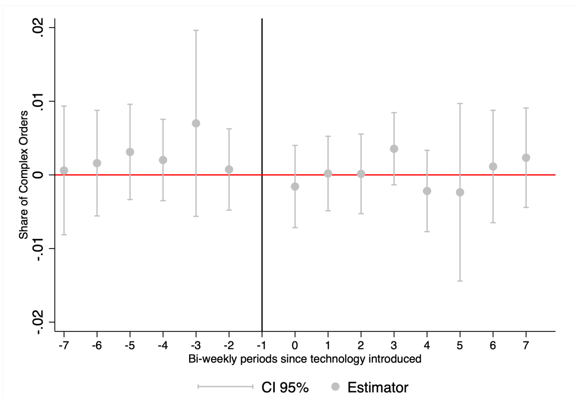
Panel B. Event study estimates on average value per ticket in counter



Panel C. Event study estimates on share of complex orders in Drive-thru

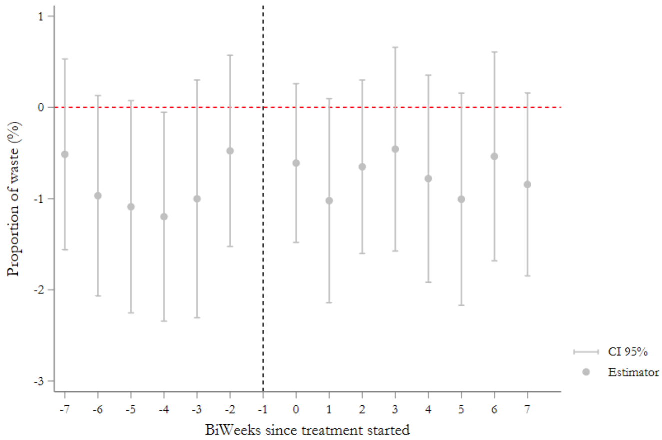


Panel D. Event study estimates on share of complex orders in counter



Note: Panels A, B, C, and D of Figure C.6 show the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on average value per ticket (A and B) and complex oders (C and D) . Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A and C show estimations for the orders placed at the Drive-thru and Panel B and D show estimations for the orders placed at the Counter. Vertical line represents time of treatment.

Figure C.7: Event study estimates of implementation on proportion of waste

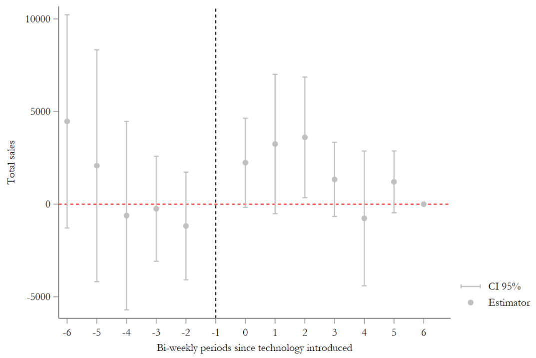


Notes: Notes: Figure C.7 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on the proportion of waste. Treatment defined as stores implementing performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Vertical line represents time of treatment

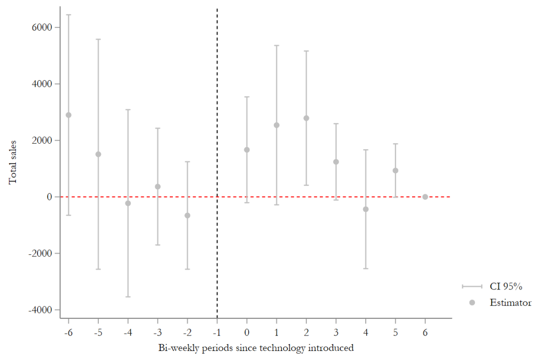
C.3 Balanced Panel

Figure C.8: Balanced sample: Total sales

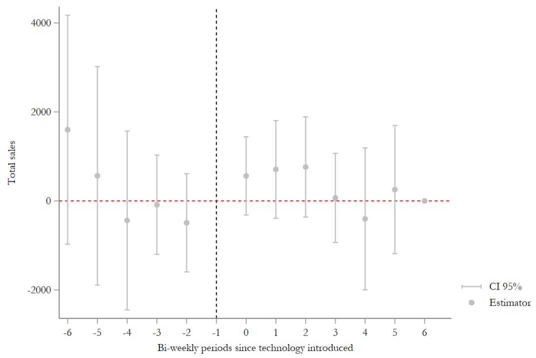
Panel A. Event study estimates in whole store



Panel B. Event study estimates in Drive-thru



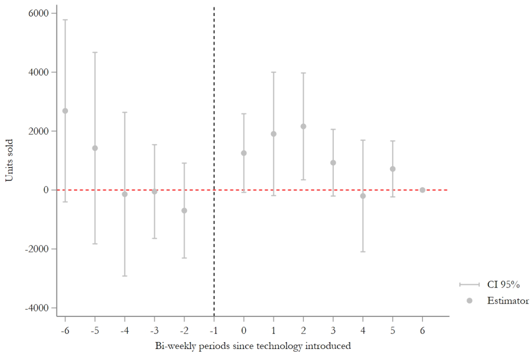
Panel C. Event study estimates in counter



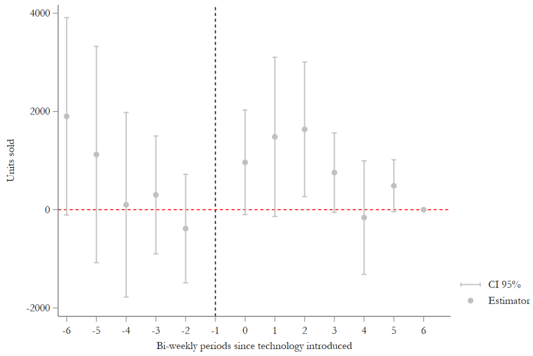
Note: Panel A, B and C of Figure C.8 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on total sales. Treatment defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the whole store, Panel B shows estimations for the orders placed at the Drive-thru and Panel C shows estimations for the orders placed at the Counter. Vertical line represents time of treatment

Figure C.9: Balanced sample: Units Sold

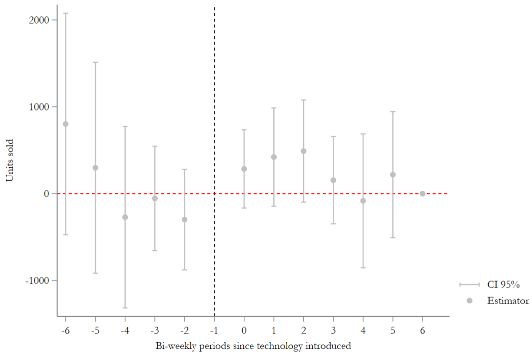
Panel A. Event study estimates in whole store



Panel B. Event study estimates in Drive-thru



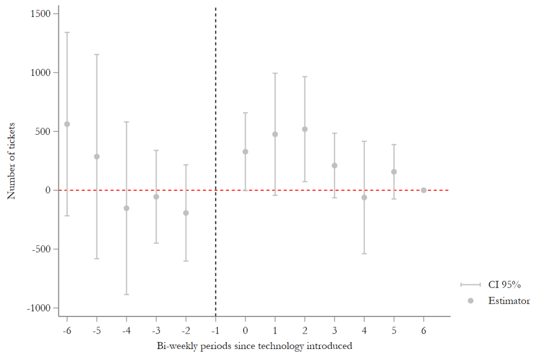
Panel C. Event study estimates in counter



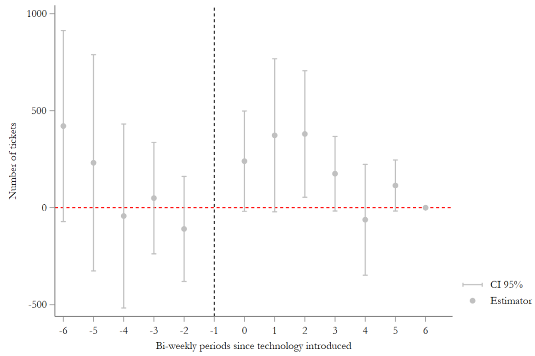
Note: Panel A, B and C of Figure C.9 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on units sold. Treatment defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the whole store, Panel B shows estimations for the orders placed at the Drive-thru and Panel C shows estimations for the orders placed at the Counter. Vertical line represents time of treatment

Figure C.10: Balanced sample: Number of tickets

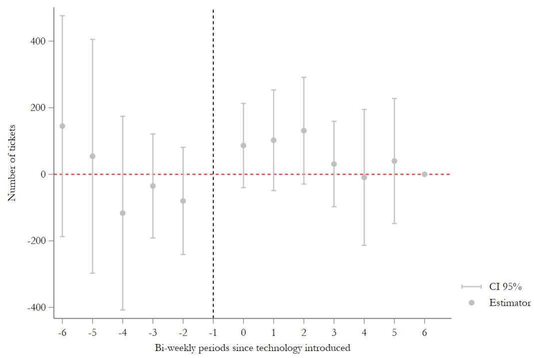
Panel A. Event study estimates in whole store



Panel B. Event study estimates in Drive-thru



Panel C. Event study estimates in counter



Note: Panel A, B and C of Figure C.10 shows the event-study coefficients and 95 percent confidence intervals from the estimation of equation (1) on the number of tickets. Treatment defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. Standard errors clustered by store. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the whole store, Panel B shows estimations for the orders placed at the Drive-thru and Panel C shows estimations for the orders placed at the Counter. Vertical line represents time of treatment

D Heterogeneous Results

D.1 Callaway and Sant'Anna (2020)

Table D.1: Callaway - Effects of technology implementations on key performance measures

	Whole store	**Drive-thru**	**Counter**
		Panel A. Sales	
Coefficient	4,028	3,046	893.7
Stand. Err.	(2,290)	(1,714)	(661.9)
Conf. Interval	[260.7, 7795.0]	[226.8, 5868.8]	[-195.1, 1982.5]
Relative effect	5.29%	6.00%	3.51%
		Panel B. Units sold	
Coefficient	2,572	1,931	589.0
Stand. Err.	(1,264)	(1,018)	(1,018)
Conf. Interval	[490.9, 4652.2]	[255.5, 3606.8]	[-20.06, 1157.9]
Relative effect	6.21%	6.78%	4.53%
Observations	2,074	2,073	2,069

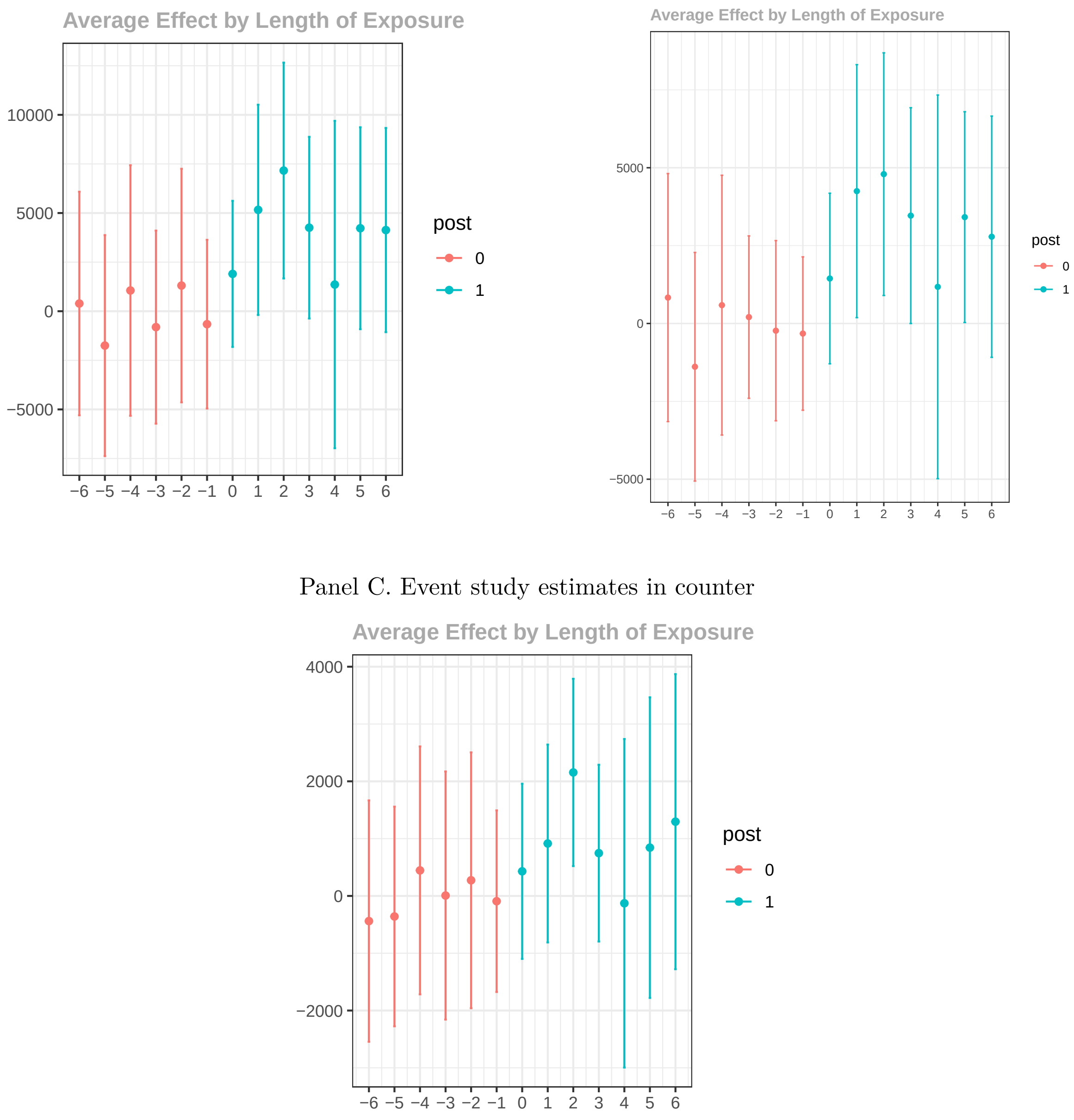
Notes: Table shows DD coefficients for estimation of equation (1), replacing the full set of time to treatment indicators with a simple post-treatment dummy, after correcting for staggered treatment timing per Callaway and Sant'Anna (2020) methodology. Treatment defined as stores implementing the technology. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Numbers in brackets are the confidence intervals.

Figure D.1: Effects of technology implementation on sales (Callaway and Sant'Anna, 2020)

Panel A. Event study estimates in whole store

Panel B. Event study estimates in drive-thru

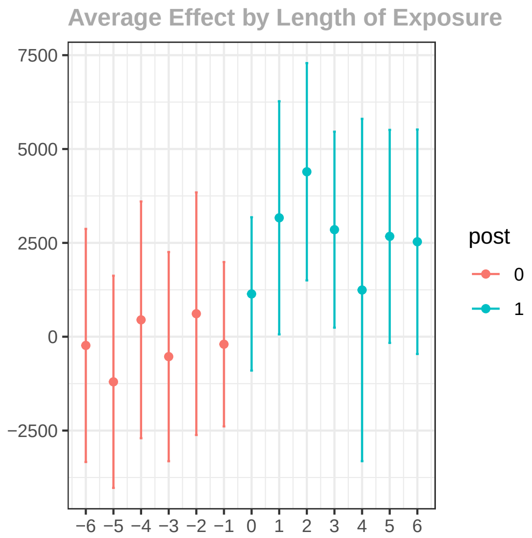
Panel C. Event study estimates in counter



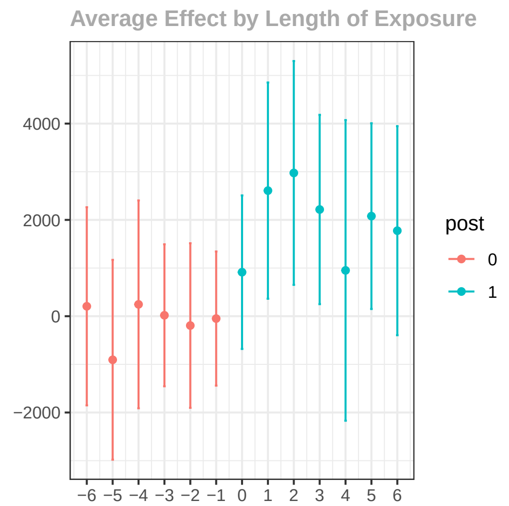
Notes: Figure D.1 shows the event-study coefficients and 95 percent confidence intervals on total sales, after correcting for staggered treatment timing per Callaway and Sant'Anna (2020). Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the whole store, Panel B shows estimations for the drive-thru and Panel C shows estimations for the counter.

Figure D.2: Effects of technology implementation on units sold (Callaway and Sant'Anna, 2020)

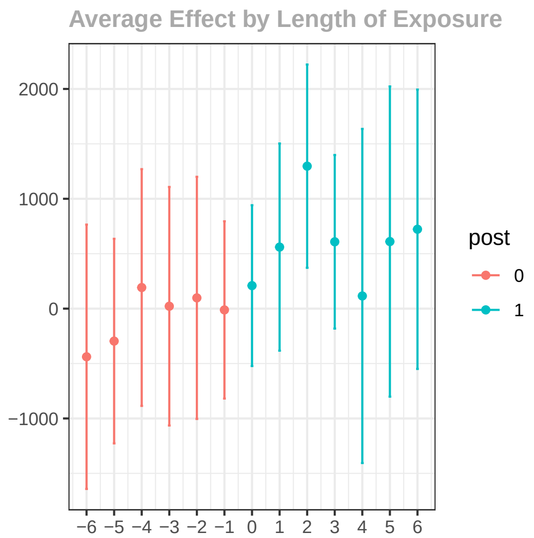
Panel A. Event study estimates in whole store



Panel B. Event study estimates in drive-thru



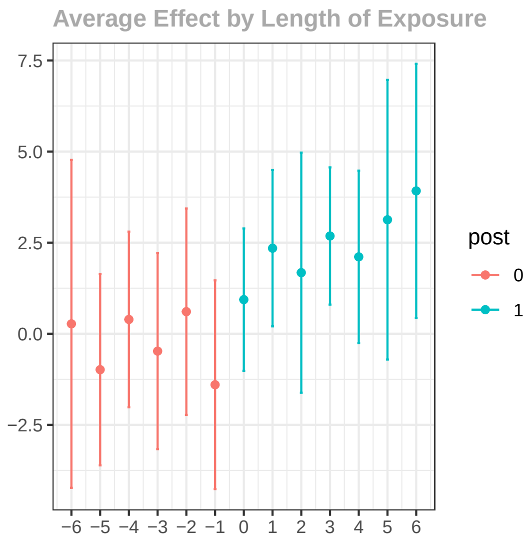
Panel C. Event study estimates in counter



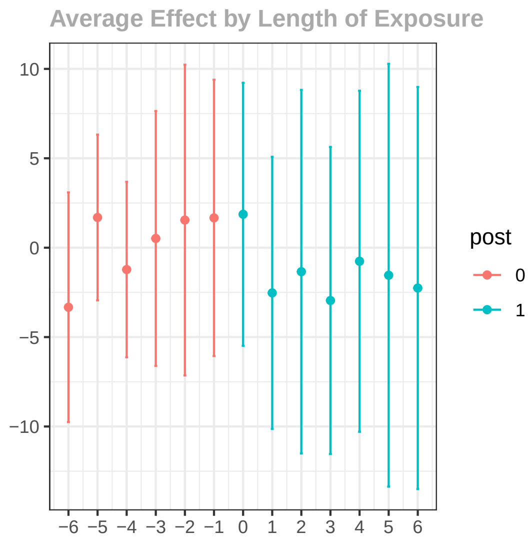
Notes: Figure D.2 shows the event-study coefficients and 95 percent confidence intervals on units sold, after correcting for staggered treatment timing per Callaway and Sant'Anna (2020). Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for the whole store, Panel B shows estimations for the drive-thru and Panel C shows estimations for the counter.

Figure D.3: Effects of IT on Training on New Skills (Callaway and Sant'Anna, 2020)

Panel A. Kitchen



Panel B. Non-Kitchen



Notes: Figure D.3 shows the event-study coefficients and 95 percent confidence intervals on training on new skills, after correcting for staggered treatment timing per Callaway and Sant'Anna (2020). Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for kitchen stations and panel B shows estimations for non-kitchen stations.

D.2 Oaxaca-Blinder-Kitagawa decomposition (Goodman-Bacon (2021))

Table D.2: Oaxaca-Blinder-Kitagawa decomposition, Total Sales

DD Comparison	Weight	Avg DD Est
Earlier T vs. Later C	0.488	3287.411
Later T vs. Earlier C	0.455	4246.946
T vs. Already treated	0.057	1911.047

Figure D.4: Oaxaca-Blinder-Kitagawa decomposition, Total Sales

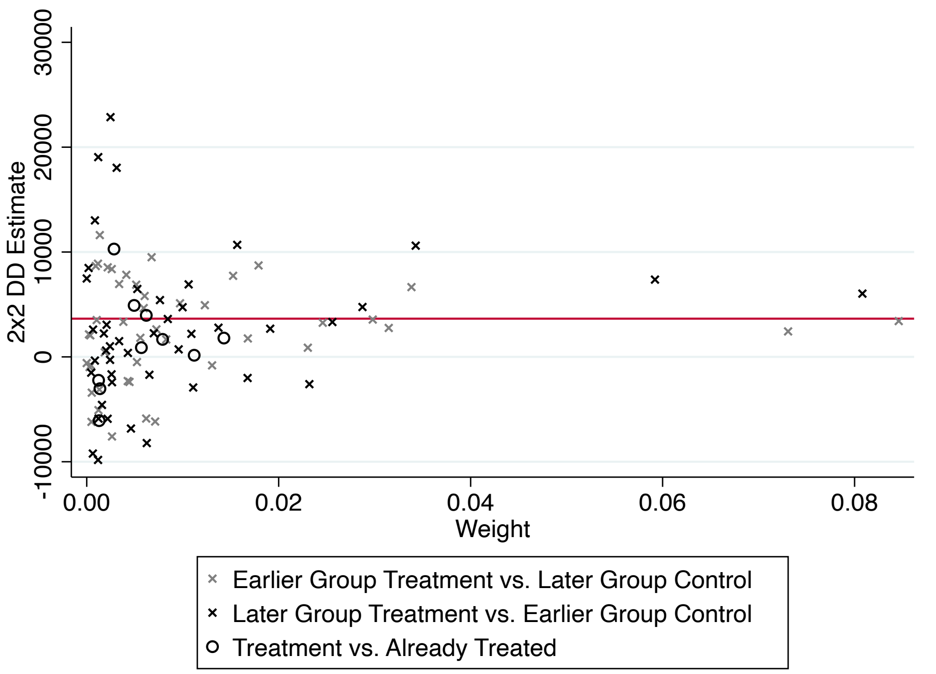
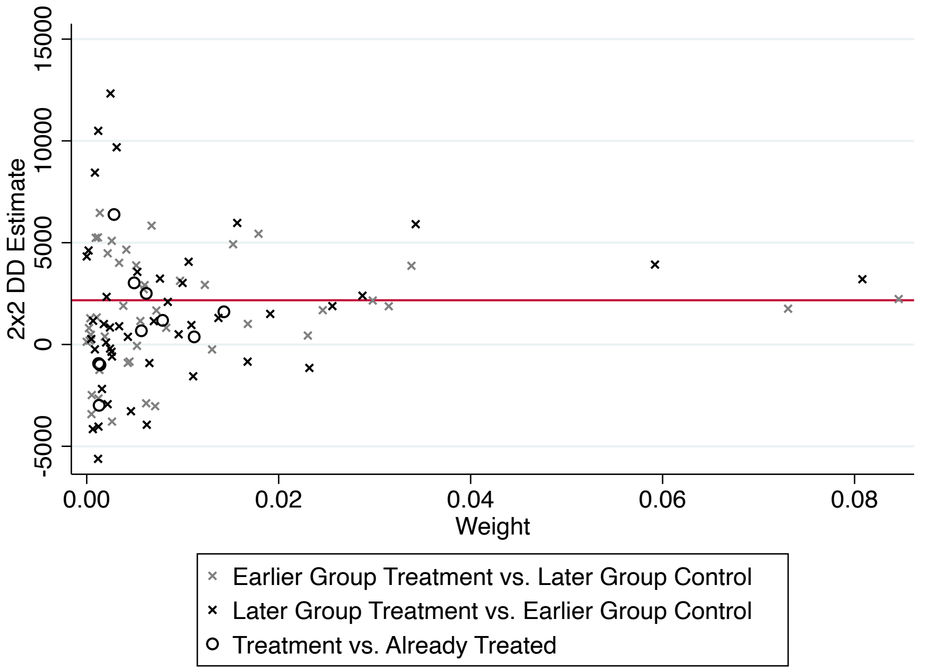


Table D.3: Oaxaca-Blinder-Kitagawa decomposition, Units Sold

DD Comparison	Weight	Avg DD Est
Earlier T vs. Later C	0.488	2081.985
Later T vs. Earlier C	0.455	2359.508
T vs. Already treated	0.057	1450.8

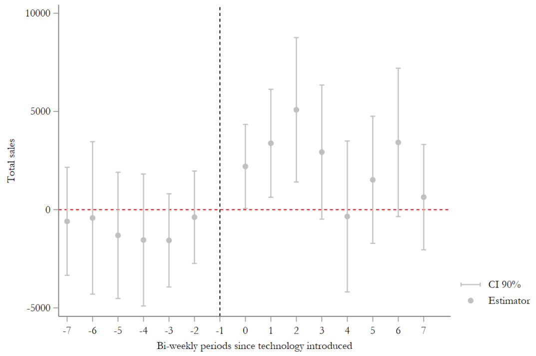
Figure D.5: Oaxaca-Blinder-Kitagawa decomposition, Units Sold



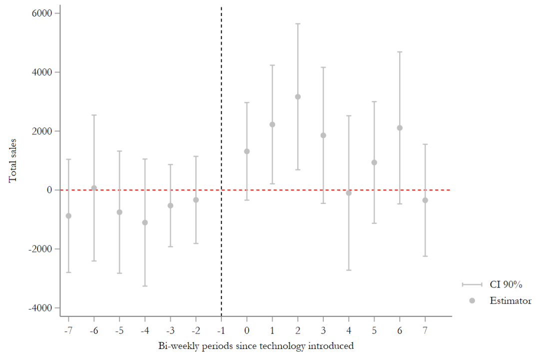
D.3 Sun and Abraham (2020)

Figure D.6: Effects of technology implementations on sales (Sun and Abraham, 2020)

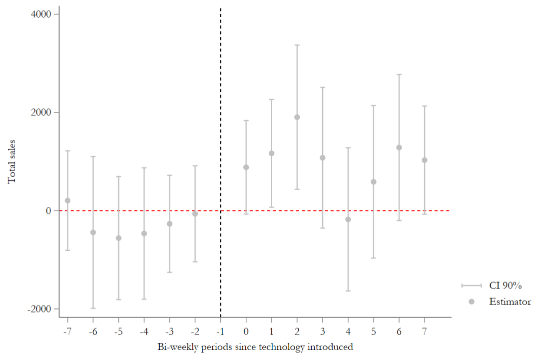
Panel A. Event study estimates in whole store



Panel B. Event study estimates in drive-thru



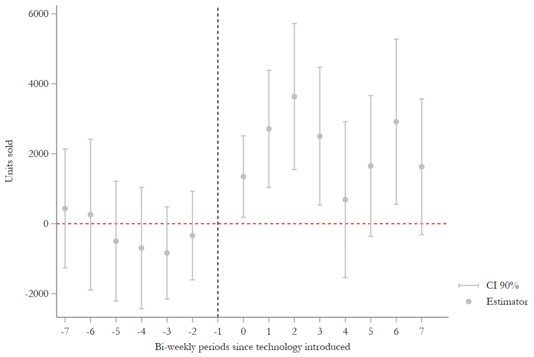
Panel C.Event study estimates in counter



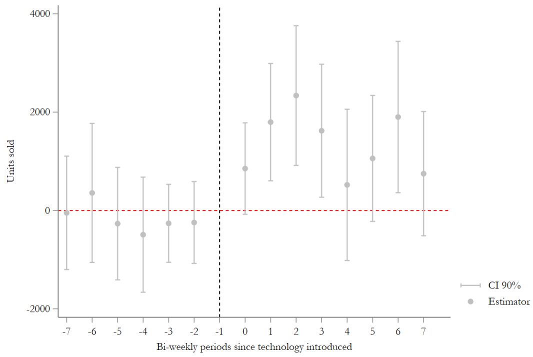
Note: Figure D.6 shows event-study coefficients and 95 percent confidence intervals from estimation of equation (1) on sales, after correcting for staggered treatment timing per Sun and Abraham (2020). Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations on sales in whole store, Panel B shows estimations on sales in drive-thru, and Panel C shows estimations on sales in counter. Vertical line represents time of treatment

Figure D.7: Effects of technology implementations on units sold (Sun and Abraham, 2020)

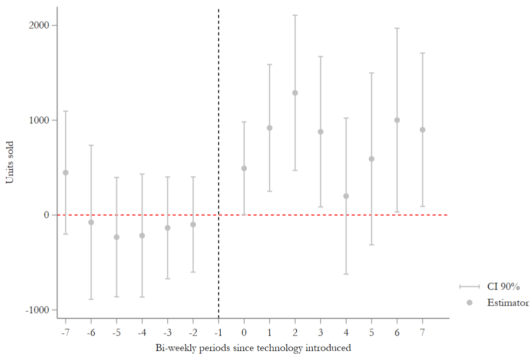
Panel A. Event study estimates in whole store



Panel B. Event study estimates in drive-thru



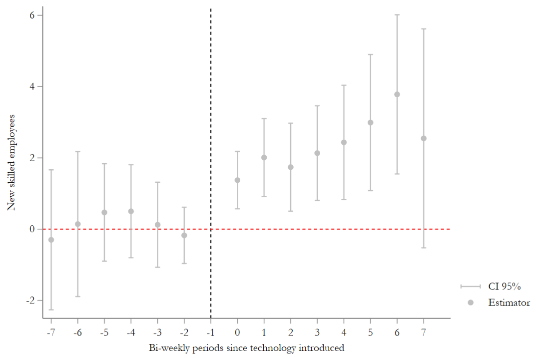
Panel C.Event study estimates in counter



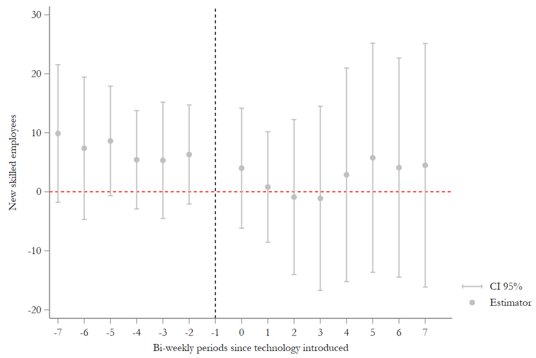
Note: Figure D.7 shows event study coefficients and 95 percent confidence intervals from estimation of equation (1) on units sold, after correcting for staggered treatment timing per Sun and Abraham (2020). Treatment defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations on sales in whole store, Panel B shows estimations on sales in drive-thru, and Panel C shows estimations on sales in counter. Vertical line represents time of treatment

Figure D.8: Effects of technology implementations on training on new skills (Sun and Abraham, 2020)

Panel A. Event study estimates on kitchen training



Panel B. Event study estimates on non-kitchen training



Note: Figure D.8 shows event-study coefficients and 95 percent confidence intervals from estimation of equation (1) on training on new skills, after correcting for staggered treatment timing per Sun and Abraham (2020). Treatment is defined as stores implementing the performance monitoring technology. Control stores are those not yet treated. The sample consists of store-biweekly panel data for the 51 stores between March 2018 to October 2019. Panel A shows estimations for kitchen stations and panel B shows estimations for non-kitchen stations. Vertical line represents time of treatment.