Understanding and Managing Customer-Induced Negative Externalities in Congested Self-Service Environments

Hyun Seok (Huck) Lee
College of Business, Oregon State University, Corvallis, OR 97331
{HyunSeok.Lee@oregonstate.edu}
Saravanan Kesavan, Vinayak Deshpande
University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
{skesavan@unc.edu, Vinayak_Deshpande@kenan-flagler.unc.edu}

Managing congestion in self-service environments such as fitting rooms in apparel retailers is vital as retailers increasingly rely on their customers to perform many tasks independently. In this paper, we demonstrate the presence of *thwarting* behavior, defined as a systematic change in customers' behavior during congestion that imposes negative externalities on other customers. Using point-of-sale (POS), traffic, and labor data obtained from a retail technology platform firm, RetailNext, and one of its retail clients, we demonstrate an inverted-U relationship between fitting room traffic and sales that is consistent with the presence of thwarting behavior. To provide direct evidence of thwarting behavior, we conduct a field study observing customer behavior at another retailer. Our field study demonstrates that customers indeed change their behavior during congestion resulting in increased waiting time and phantom stockouts. Finally, we propose a solution to mitigate phantom stockouts and test it using a field experiment. We show that providing a timely backend recovery operation to mitigate phantom stockouts through a dedicated associate in the fitting room area increases hourly sales by 22.7%.

Keywords: Retail operations; Empirical; Field experiment; Fitting room; Phantom stockout

1. Introduction

Congestion management is an important endeavor in service settings. Since congestion occurs when there is variability in inter-arrival times or service times, researchers have identified several ways to reduce or accommodate this variability to relieve congestion (Hassin and Haviv 2003; Lu et al. 2013). In these settings, customers are typically assumed to be either passive – simply waiting for service and receiving it – or strategic, but only to the extent of deciding to balk or renege from queues (Kulkarni 2009; Aksin et al. 2013; Dong et al. 2015; Batt and Terwiesch 2015; 2016). However, customers could show other types of strategic behavior during congestion different from balking and reneging. More importantly, some of those strategic behaviors may exacerbate congestion by imposing negative externalities on other customers, unlike balking and reneging that have a positive externality on other customers. We call such strategic behavior as *thwarting* behavior and define it as a systematic change in customers' behavior during congestion that imposes negative externalities on others.

Anecdotal evidence of such thwarting behavior in other settings can be found in popular press and literature. Transportation researchers find that drivers tend to exhibit aggressive behaviors such as honking at other cars and tailgating during congestion, which leads to accidents and exacerbates congestion (Hennessy and Wiesenthal 1997). In call centers, irritated customers due to long wait have been known to demand longer service (Dong et al. 2015) that could increase waiting time for others. However, it is unclear whether retail customers exhibit any thwarting behavior when stores become congested.

In this paper, we test for the presence of thwarting behavior in fitting rooms of apparel retailers. We choose fitting rooms as they are usually self-service environments in most retail stores with little or no monitoring where we may be able to observe thwarting behavior. During congestion, customers may exhibit thwarting behavior by taking more clothes to try on in a fitting room because they do not want to make multiple visits or are afraid of losing items. This requires them to occupy fitting rooms longer (service slowdown) causing an increase in waiting time for other customers who want to use fitting rooms. In addition, they could leave behind more clothes in fitting rooms resulting in phantom stockouts. Lost sales could increase due to thwarting behavior, as it would lead to more balking/reneging due to longer waiting time for other customers and lower product availability due to phantom stockouts (DeHoratius and Ton 2015).

We first analyze archival data to test for the presence of an inverted-U relationship between fitting room traffic and sales; its presence would be consistent with thwarting behavior. To rule out alternate explanations for the inverted-U relationship such as selection bias, crowded checkout counters, phantom stockouts in the rest of the store, and overall poor service level due to understaffing, we conduct a field study to provide direct evidence for the presence of thwarting behavior. In the field study, we directly observe the behaviors of customers entering the fitting room area at a major retailer. We measure the extent to which customers change their behavior during congestion by taking more clothes into the fitting room. We also measure the increase in waiting time to other customers due to thwarting behavior and the incidence of phantom stockouts in fitting rooms. Finally, we perform a field experiment to demonstrate a solution that can potentially mitigate the negative impact of thwarting behavior on sales.

Our data were obtained through collaborations with three companies. RetailNext is a leading in-store analytics provider to retailers. It collects traffic information from video cameras in retail stores to codify customer-arrival patterns as well as customer pathways in retail stores. In addition, they collate customer traffic information and point-of-sale (POS) data. These data were obtained from one of its clients, a large U.S.-based retailer (retailer A). This retailer's stores are about 50,000 sq. ft. in size and primarily carry men's, women's, and children's apparel, along with some home furnishing goods. The data obtained from RetailNext and retailer A helped us provide initial evidence for the presence of thwarting behavior and its impact on store sales. We further observe customers directly through a field study and conduct a field experiment at another department store retail chain (retailer B).

Our primary findings are as follows. First, we observe an inverted U-shaped relationship between fitting room traffic and sales consistent with the presence of thwarting behavior among customers during congestion. Sales initially increase with fitting room traffic as more customers intend to purchase. Beyond a certain point, however, we observe a decline in sales when fitting rooms become congested. Contrary to the conventional wisdom that more traffic drives more store sales, we identify that too much traffic in fitting rooms can hurt store sales. This observation shows that managing congestion in fitting rooms is critical for brick-and-mortar apparel retailers.

Second, by observing 209 customers who used fitting rooms for 24 non-continuous hours spanning 3 weekends, we find that customers, on average, take additional 1.38 (23.96%) clothes into the fitting room when they experience congestion. This requires them to occupy fitting rooms for an additional 1.87 (18.61%) minutes. This behavior imposes a negative externality on other customers as waiting time increases. An additional negative externality arises due to increase in misplaced inventory in the fitting rooms that could result in phantom stockouts. We observe 38.6% of phantom stockout rate among items left behind in the fitting room area. Further analysis shows that those misplaced items were high-price merchandise that contributes to significant opportunity costs.

Finally, our field experiment shows that providing a timely backend recovery operation through a dedicated associate in the fitting room area can significantly increase sales by reducing phantom stockouts. By comparing the treatment to control groups, we find that an addition of a backend recovery operation for items left behind in the fitting rooms using a dedicated associate increases checkout-counter-level hourly sales by 22.7%.

Our research makes several contributions to the operations management (OM) literature. Our first contribution is the identification and demonstration of customers' thwarting behavior. Researchers have identified that customers exhibit strategic behaviors, e.g., balking or reneging, in the queue based on their service experience. While balking and reneging lead to positive externalities for other customers, none of the papers have empirically identified this specific strategic behavior, thwarting, which imposes a negative externality on others. We provide the first direct evidence for the presence of thwarting behavior in retail stores. Identification of thwarting behavior enables us to show a new dimension of customer impact in service operations as it reveals that *customer*-induced service slowdown, in contrast to server-driven slowdowns (Kc and Terwiesch 2009; Anand et al. 2011; Tan and Netessine 2014; Batt and Terwiesch 2016), can be an important reason for service slowdowns in retail settings.

Second, we measure instore-based phantom stockouts in retail stores, which is different from backroom-based phantom stockouts identified in the prior literature (DeHoratius and Raman 2008; Ton and Raman 2010). While phantom stockouts are expected to result in lost sales, the magnitude of the lost sales has not been estimated before. By using a field experiment we show that mitigating phantom stockouts in the fitting room area alone can lead to significant increase in store sales.

Third, our paper adds to the limited literature in OM that has conducted field experiments (e.g., Fisher and Rajaram 2000) and to the expanding literature on retail labor (Fisher et al. 2007; Ton 2009; Netessine et al. 2010; Kesavan et al. 2014; Mani et al. 2015) by conducting the first field experiment in this line of research. While prior literature has argued for increasing store labor to drive sales, we use a field experiment to show that specifically using labor to perform a timely backend recovery operation in fitting rooms can increase sales significantly. In addition, this paper uses a field study to observe customer's thwarting behavior rather than lab experiments. Field studies are preferred over lab studies when the treatment effects are expected to interact with participant characteristics (Al-Ubaydli and List 2015). Since thwarting behavior causes negative externalities on other customers, it is likely that this behavior is harder to study in a lab setting. Participants in the lab may either change behavior when they are aware of being observed or may result in systematically opting out, causing biases in our inferences. Due to the covert nature of the field, we can observe customers in their natural environment and understand the degree to which customers engage in thwarting behavior.

2. Prior Literature

Researchers have identified several ways to relieve congestion. In these settings, customers are typically assumed to be either passive or strategic, but only to the extent of deciding to balk or to renege from queues, which have a positive externality for other customers. Examples of analytical works with passive arrivals are Deo and Gurvich (2011) and Lee et al. (2012) while those of empirical works include Chan et al. (2014) and Batt and Terwiesch (2016). Models with balking and reneging can be found in many books, e.g., Kulkarni (2009). Recent empirical works on this topic include Aksin et al. (2013) and Batt and Terwiesch (2015). However, customers show other types of strategic behavior in retail settings different from balking and reneging, which impose negative externalities on other customers. For example, Allon and Hanany (2012) analytically studied customer's behavior of cutting the line that can increase waiting time for other customers. To the best of our knowledge, our paper is the first paper that empirically identifies and demonstrates customers' behavior change, when they experience congestion, in a way that imposes a negative externality on other customers; we call it thwarting. Thwarting is distinctive from a negative externality caused by joining the queue that has been studied in the prior literature. Thwarting deals with additional negative externalities on other customers, beyond joining the queue, due to the change in customers' behavior when they experience congestion.

Sasser (1976) identified customers as a "mixed blessing" in operating environments. On the one hand, customers could be a source of productive capacity by providing labor for self-service; on the other hand, they could introduce variability and uncertainty into operations. Frei (2006) further classified such customer-introduced variability into five categories (i.e., arrival, request, capability, effort, and subjective preference). This literature does not identify congestion-induced changes in customer behavior as a source of additional variability. It also does not recognize phantom stockouts as an outcome of this variability. We show that this type of variability is driven by willful behavioral changes on part of strategic customers who are responding to the congestion they face and quantify the impact of this variability on phantom stockouts.

Staffing problems have been studied in other service systems such as call center and healthcare settings (Gans et al. 2003; Green 2004). Recently Batt and Terwiesch (2016) empirically found that service rate is dependent on workload in a hospital's emergency department. Earlier papers have argued that high congestion levels may require servicepersons (hereafter, "servers") to multitask in parallel, which involves a cognitive switching cost (Kc 2013; Batt and Terwiesch 2016) and fatigue (Kc and Terwiesch 2009) that lead to server slowdown. Because customers themselves perform most activities in a fitting room, a self-service environment in general, servers do not cause an increase in service time in our research setting. Our paper therefore emphasizes *customer*-induced service slowdowns as opposed to *server*-driven service slowdowns.

The importance of labor as a key execution issue in stores has been highlighted in Raman et al. (2001). Other examples of store execution issues are inventory record inaccuracy (DeHoratius and Raman 2008) and phantom stockouts (Ton and Raman 2010), which were found to significantly impact retail store performance negatively. Using survey data collected from a small-appliance and furnishings retail store, Fisher et al. (2007) showed that labor issues, as an example of store execution issues, considerably affect both customer satisfaction and sales. In an apparel setting, overcrowded fitting rooms often result in temporary phantom stockouts within the store. Such customer-induced misplaced inventory *within the store*, different from phantom inventory *in the backroom* (Ton and Raman 2010), is expected to be a driver of the inverted-U relationship that we observe in our setting.

Ton and Huckman (2008) provided further evidence for the importance of store labor by demonstrating a link between an increase in employee turnover and a decrease in profit margin and customer service. In addition, Ton (2009) showed that an increase in store labor is associated with higher profits. Our paper is consistent with this work in demonstrating that a dedicated associate in the fitting room area can help improve store performance. More importantly, we conduct a field experiment involving labor intervention that has not been conducted in this stream of research thus far. While the prior literature has argued for increasing store labor to drive sales, our field experiment shows that specifically using labor to perform a backend recovery operation of misplaced items in fitting rooms can increase sales significantly.

Recently, several authors have used retail traffic data in their empirical analyses. Perdikaki et al. (2012) studied relationships between sales, traffic, and labor for apparel retail stores. By decomposing sales into conversion rate and basket value, they found that, at an aggregate level, store sales have an increasing concave relationship with traffic; conversion rate decreases nonlinearly with increasing traffic; and labor moderates the impact of traffic on sales. Our research results suggest that, beyond plateauing, sales *decline* when a self-service environment, a fitting room area in particular, is highly congested.

Tan and Netessine (2014) studied the impact of workload on servers' performance in a restaurant chain. They found that servers exert more effort on sales by sacrificing service speed when the overall workload is small, whereas servers start to reduce sales effort and increase service speed as workload increases. Our paper differs from previous papers in two ways. First, we focus more on how customer behavior affects service rates; previous work has emphasized servers. Second, while previous studies mostly used overall store traffic data, we demonstrate the value of in-store traffic data by showing the negative impact of congestion in fitting rooms on sales. Furthermore, we provide evidence for thwarting behavior through a field study directly observing customer behavior, which can explain the inverted U-shaped relationship observed in our data.

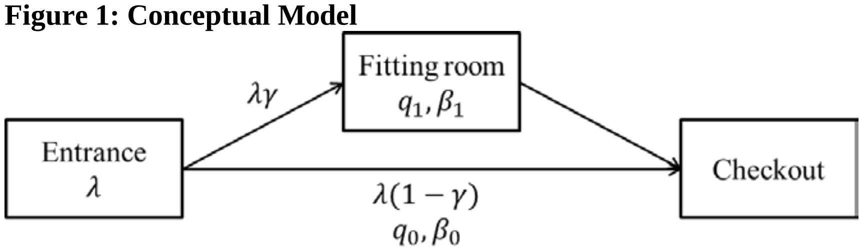
Finally, a stream of literature has focused on the self-service setting. Given traditional service (e.g., a teller in banking), most prior papers have studied the impact of introducing self-service technology (e.g., online banking) on customer satisfaction and retention in a number of settings (Buell et al. 2010). Buell and Norton (2011) revealed that engaging in operational transparency, which they termed labor illusion, is sufficient to increase perceived value. While those papers focus on the impact of operational changes on consumer behavior, our work emphasizes the impact of customers' behavior changes on operations. Also, as some prior literature (Moon and Frei 2000; Frei 2006; 2008; Hu et al. 2016) has consistently pointed out the importance of customers' impacts on operations, our paper adds to this stream by identifying another type of customers' impacts on operations during congested periods that we call thwarting behavior.

3. Theory Development and Hypotheses

We formalize our definition of thwarting behavior using a simple conceptual model in Figure 1 and then explain the hypotheses we plan to test. Consider a store where store traffic is λ per hour. Assume that a fraction (γ, where $0 \leq \gamma \leq 1$) of those customers enter the fitting room area. We call this $\lambda\gamma$ as fitting room traffic. The remaining customers, $\lambda(1-\gamma)$, leave the store without entering a fitting room. Assume that the conversion rate of customers who use fitting rooms to be q_1 and those who do not to be q_0. Let β_1

and β_0 be the basket values of the purchases made by customers who use fitting rooms and those who do not, respectively. We assume $q_1 > q_0$ as recent paper (Musalem et al. 2016) finds higher conversion rate of fitting room users than that of non-users. We also assume $\beta_1 > \beta_0$ as anecdotal evidence suggests that customers might need to use fitting rooms when they purchase high price clothes (e.g., formal suit) whereas they might not need to try on when they purchase low price clothes (e.g., basic white t-shirt). Finally, we assume that γ depends upon the waiting time in the fitting room area for impatient customers while conversion rate and basket value parameters are affected by product availability. The store sales in any given hour are the sum of sales from customers who used fitting rooms and those who did not, i.e., $Sales = \lambda\gamma q_1\beta_1 + \lambda(1-\gamma)q_0\beta_0$.

Figure 1: Conceptual Model



Now we construct hypotheses to investigate the relationship between fitting room traffic and store sales using this simple setup. We consider three possible scenarios.

Scenario 1: Traditional OM literature assumes that customers are passive (i.e., patient) or have limited strategic behavior where they can only balk or renege (i.e., impatient). We consider the implications of this assumption on the relationship between fitting room traffic and sales in this scenario.

Fitting rooms play a critical role for apparel retailers as customers experience the product and evaluate alternatives there to complete their purchase. These steps have been identified as part of the core pathway in the consumer purchase decision process. Hence, as fitting room traffic increases, sales are expected to increase initially, because more shoppers reveal their intention to purchase. In the conceptual model, we see that sales increases with γ. If customers are assumed to be passive, sales are expected to increase linearly with fitting room traffic till fitting room capacity is reached. Once it's reached, customers may passively queue outside the fitting rooms so sales will plateau. In the case of limited strategic behaviors such as balking and reneging, sales would have an increasing concave relationship with fitting room traffic. For example, our model can capture balking by setting γ to decrease with increases in the level of congestion. Those customers who balk or renege will end up not using fitting rooms. As the expected sales are higher for fitting room users than non-users, we anticipate sales to be increasing at a diminishing rate with higher fitting room traffic.

Since we expect customers to balk/renege and not remain passive, we propose the following relationship based on the OM literature:

Hypothesis 1a: *There is an increasing concave relationship between fitting room traffic and store sales.*

Scenario 2: In this scenario, we consider the case where customers exhibit thwarting in addition to reneging and balking. Specifically, we assume that customers change their behavior when they experience congestion in ways that impose negative externalities on other customers that causes sales to decline.

Customers could become strategic by bringing more clothes into the fitting room when they experience congestion. When the store is congested, customers might anticipate longer waiting time to secure a fitting room. So they may take more alternatives with them to reduce the chance of returning back to the selling floor. They may also take more alternatives for fear of stocking out when many customers are around. Taking more clothes into the fitting room would increase waiting time for the rest of the customers who want to use fitting rooms as the service time would increase. In this case, when service time is state-dependent, the waiting time for other customers could grow much faster than usual, which will increase abandonments significantly (much smaller γ). This could lead to decline in sales at higher levels of fitting room traffic.

Furthermore, bringing more clothes into the fitting room could affect product availability in the store leading to lower sales at higher levels of congestion. Customers typically leave behind unwanted clothes in the fitting room so bringing more clothes could result in more misplaced items in the fitting room. The prior literature in psychology has shown that customers in crowded stores feel disoriented (Dion 1999), less satisfied (Eroglu and Machleit 1990), more stressed, and tenser (Langer and Saegert 1977). So, even those customers who secure a fitting room but are unhappy with the fit, color, or design of their initial selections may be reluctant to go back into the store to find alternatives due to the crowd and may just leave those unwanted clothes in the fitting room. Though store associates may eventually return these clothes to the proper locations, temporary phantom stockouts (Ton and Raman 2010) are likely to occur and affect sales, especially during congested periods because associates are likely to be preoccupied with other customer-facing tasks and many customers in the selling floor are likely to take clothes into their shopping bags or carts. So, phantom stockouts could increase lost sales as customers cannot find the items they are looking for, resulting in lower conversion rates and basket sizes for both fitting room users as well as non-users.

These misplaced items in the fitting room could also make rooms dirty. It might cause customers to avoid some fitting rooms and increase waiting time for other rooms. This could lead to an increase in abandonments (smaller γ). Accordingly, we propose the following relationship based on thwarting:

Hypothesis 1b: *There is an inverted U-shaped relationship between fitting room traffic and store sales.*

Although Hypothesis 1b is based on thwarting behavior, there are other reasons that could drive inverted-U shaped relationship between fitting room traffic and store sales. For example, the rest of the store is likely to be congested when the fitting room is crowded so the store sales could decline due to decreases in q_0 and β_0 as a result of poor customer service in the rest of the store, phantom stockouts in the rest of the store, or long waiting times in the checkout counter. So, the presence of an inverted-U relationship is not by itself a direct evidence for thwarting behavior.

Scenario 3: Next we argue for an alternate relationship between fitting room traffic and sales in which sales have an increasing convex relationship with fitting room traffic based on a different underlying consumer behavior. This could happen if customers who visit the store during congested hours are unable to find an unoccupied fitting room to try on the clothes at the store so they decide to purchase multiple items with an intention of returning some later (q_0 and β_0 increase as traffic increases). Further, even customers who find a fitting room may decide to purchase more as they try on more items in the fitting room when they face congestion (q_1 and β_1 increase as traffic increases). Both explanations would increase sales non-linearly when the fitting rooms are busy. Accordingly, we propose the following relationship:

Hypothesis 1c: *There is an increasing convex relationship between fitting room traffic and store sales.*

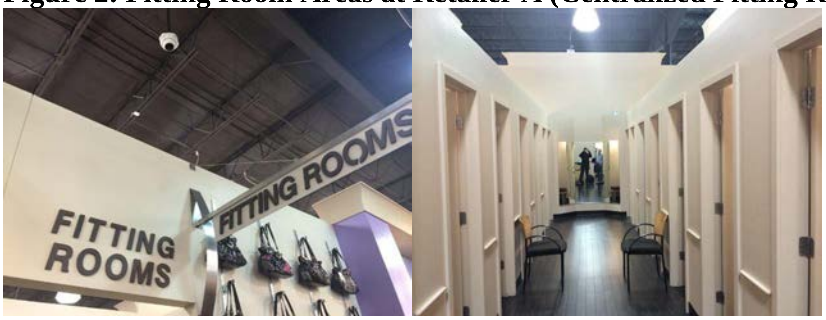
4. Data and Methodology for Archival Data Analysis

4.1. Data Sources

We test our hypotheses using archival data obtained from RetailNext and one of its retail clients. RetailNext is a leading in-store analytics provider to retailers such as Sears. It collects traffic information from video cameras in retail stores to codify customer arrival patterns and customer pathways in the stores. In addition to traffic data, we further obtain POS and labor data from one of their clients, a large U.S.-based retailer (retailer A). Retailer A is an apparel retail chain that sells primarily women's, men's, and children's apparel, along with some home décor goods. We worked closely with both companies by interacting with the retailer's senior management, corporate planners, and store management team. The study period is from July 2012 to October 2013.

We obtain the following data for retailer A during the study period. First, we obtain all POS information recorded through scanner. We aggregate it to hourly level to possess store sales volume per hour ($). Second, we obtain labor data which allow us to calculate the number of employees in the store. Finally, we possess traffic data. Retailer A installed video cameras at store entrances to count the number of visitors. All stores of retailer A had this entrance camera during our study period. Only one store, however, had additional cameras installed within the store to track customer movement. We therefore focus on this store to study the impact of fitting room traffic on store sales. These cameras capture only the entrance of the fitting room area and nothing within fitting rooms' interiors. Figure 2 shows the fitting room area at retailer A used in this study, located at the center of the store for ease of approach. The store has 16 fitting rooms in total in this area.

Figure 2: Fitting Room Areas at Retailer A (Centralized Fitting Room Layout)



Note. Fitting rooms are located at the center of the store. This store has 16 fitting rooms in total.

Traffic cameras used in this study were able to differentiate between incoming and outgoing traffic by tracking the direction of customers' movements. Figure 3 shows how this technology can distinguish incoming from outgoing traffic. Each camera has two sensors, and if a customer goes through both, she is counted. The camera captures the direction of movement by determining the order in which a customer's motion is detected by the two sensors. Consider the two outlines around the entrance door in Figure 3. If a customer goes through the outside line (i.e., farther from the entrance) and then the inside line (i.e., closer to the entrance), in that order, then she is categorized as an "out-count," and vice versa. We had this time-stamp records for every individual who passes the two sensors. We aggregate the time-stamp data to the hourly level, to match the other data. RetailNext audits the data regularly by manually counting the number of visitors and comparing that count to the numbers from the automated sensors, ensuring that the accuracy is at least 95%.

Figure 3: Distinguishing Incoming from Outgoing Traffic

4.2. Variables

We conduct our econometrics analysis at the hourly level.

4.2.1. Dependent variables. We measure the store performance on day *t* at hour *h* by sales in dollars ($Sales_{th}$). We find that store hours are not fixed. For example, before the Christmas holiday, the store is open until midnight. In order to avoid the spurious correlation that could arise between variables as a result of systematic differences in business hours, we use only data between 9 AM and 10 PM, which are the normal store hours.

4.2.2. Key variables of interest. We have in-count and out-count traffic measures for fitting room traffic ($Fit\_Traffic_{th}^{IN}$ and $Fit\_Traffic_{th}^{OUT}$). We use an average between the in- and out-counts ($A\_Fit\_Traffic_{th} = (Fit\_Traffic_{th}^{IN} + Fit\_Traffic_{th}^{OUT})/2$) as it would help mitigate concerns about potential measurement errors in the respective in- and out-count variables. As a robustness check, we repeat analysis by using in-count or out-count measures separately instead of their average and obtained consistent results.

4.2.3. Control variables. We next describe the rest of the controls used in our analysis. First, we need to control for store traffic to distinguish the impact on sales of fitting room traffic from of overall store traffic. We use the average store traffic ($A\_Traffic_{th}$), similar to fitting room traffic, to control for the number of customers visited the store. Second, we control for store labor ($Labor_{th}$), defined as the number of employees working in the store, as that is known to affect sales (Perdikaki et al. 2012). We eliminate backroom labor from our main model since that does not affect sales directly, though our results are similar when we include backroom employees in $Labor_{th}$. Since the retailer collected time-stamp data on when each associate started to work, $Labor_{th}$ can be fractional. For example, if one employee starts to work at 9:30 AM in the store, then we have a data point of 0.5 for $Labor_{th}$ at *h* = 9.

Third, store sales depend on promotions (Lam et al. 2001). We have information regarding the store's promotional activities. Based on the information, we create a dummy variable $Promotion_t$ that is set to one on days, *t*, when a promotion was ongoing, and set to zero, otherwise. Finally, we control for seasonality by introducing hourly dummies, day-of-the-week dummies, and monthly dummies.

We trim our data by excluding extreme values to ensure that our analyses are not influenced by extreme outliers and to obtain more robust statistics and estimators, though all results are consistent when we use full data. We remove all data with standardized residuals more extreme than 3. This resulted in a drop of 1.5% of observations. We perform all further analysis on this data set. We check the robustness of our analysis with cutoffs for standardized residuals at 2 and 2.5 and obtain consistent results.

Table 1 provides summary statistics for all variables used in our analysis. Subscript *t* denotes each date and *h*, ranging from 9 to 21, denotes each hour. The average hourly sales volume is $1,887. The average hourly store traffic is 111, while the average fitting room traffic is 66, indicating that majority of the customers use fitting rooms. Table 2 shows the Pearson correlation coefficients among all variables used in our analysis. Correlations between predictors are generally quite low, except for the correlation between $A\_Fit\_Traffic_{th}$ and $A\_Traffic_{th}$, which is quite high (0.79). This may pose multicollinearity problem. We explain how we deal with it in the next section. After taking care of it, we find the variance inflation factors (VIFs) to be below 10, indicating we are not likely to have multicollinearity problems.

Table 1: Summary Statistics of the Variables (Retailer A)

Variable name	Mean	Std. dev.	P5	P25	P50	P75	P95
$Sales_{th}$	1887.41	1317.21	271.37	986.5	1627.89	2480.47	4403.71
$A\_Traffic_{th}$	111.31	70.67	26	65.25	96.5	141.5	244
$A\_Fit\_Traffic_{th}$	65.97	41.13	16	37	57	86	143.5
$Labor_{th}$	8.64	3.41	5	6	8	10	14
$Promotion_{t}$	0.24	0.43	0	0	0	0	1

Note. Number of observations is 5312.

Table 2: Pearson Correlation Coefficients (Retailer A)

	(1)	(2)	(3)	(4)	(5)
(1) $Sales_{th}$	1.00				
(2) $A\_Traffic_{th}$	**0.87**	1.00			
(3) $A\_Fit\_Traffic_{th}$	**0.71**	**0.79**	1.00		
(4) $Labor_{th}$	**0.54**	**0.61**	**0.39**	1.00	
(5) $Promotion_{t}$	**0.09**	**0.14**	**0.10**	**0.20**	1.00

Note. Bold denote statistical significance at the 1% level.

4.3. Test for the Inverted-U Relationship

There are many ways proposed in the literature to test for the inverted-U relationship, though none of them is perfect. The most common method is a quadratic functional form regression using the coefficient estimates of the linear and quadratic terms. Two confirmatory tests for the inverted-U relationship can be performed by ensuring the sign flips within the range of the data. One is to verify that the slopes of either side of the peak point are significant and of opposite signs (Aiken and West 1991) and the other is to confirm that the confidence interval of the peak point lies within the sample (Lind and Mehlum 2010). While these tests have been performed by thousands of papers, they suffer from the disadvantage that they are all based on the assumption of quadratic form. We may avoid this quadratic assumption and test for the inverted-U relationship using spline regressions, a simple two-line test proposed by Simonsohn (2016), or including higher order powers. In this paper, we use all of these methods to test for the inverted-U relationship. Our primary results are based on the quadratic model and the robustness checks consider the various alternate approaches.

We propose the following model to relate store sales to fitting room traffic, with control variables:

$$Sales_{th} = \beta_0 + \beta_1 A\_Fit\_Traffic_{th} + \beta_2 A\_Fit\_Traffic_{th}^2 + \beta_3 A\_Traffic_{th} + \beta_4 A\_Traffic_{th}^2 + \beta_5 Labor_{th} + \boldsymbol{W_{th}}'\boldsymbol{\beta_6} + \varepsilon_{th} \quad (1)$$

where $\boldsymbol{W_{th}}$ is a column vector of control variables that includes a promotion indicator, hourly dummies, day-of-the-week dummies, and monthly dummies. The availability of promotion variable is especially valuable as it enables us to mitigate endogeneity concerns between store traffic and sales.

An important consideration in the choice of our model specification is multicollinearity. This concern is especially significant in our model because we not only use store traffic and fitting room traffic which are highly correlated (0.79) but also use quadratic terms of these variables. An additional source of multicollinearity concern is that labor schedules are typically highly correlated with the hour of the day. Using all of these variables in (1) results in some variables' VIFs exceeding 10. So, we mitigate multicollinearity concerns in the following way. First, we mean center $A\_Fit\_Traffic_{th}$ and $A\_Traffic_{th}$ with their quadratic terms because mean centering can potentially alleviate multicollinearity issues (Aiken and West 1991). This significantly reduces VIFs, but some of them are still above 10. So, we further remove hourly dummies and replace with hour-block dummies. We clustered 9 AM–12:59 PM, 1 PM–5:59 PM, and 6 PM–10 PM based on common traffic patterns across hours in those blocks. The average store traffic across these three blocks of hours are 97.23, 147.49, and 79.87. So, the second block captures the store's peak hours. By using mean centered variables and hour-block dummies, we ensure VIF of all variables to be below 10 in all models. In addition, we interact the hour-block dummies with day-of-the-week dummies to capture the heterogeneity in customers across different days of the week during the same periods. Finally, we also report results with hourly dummies separately, even though the VIF of some variables is above 10.

We test Hypothesis 1a-1c, the relationship between fitting room traffic and store sales, using estimates of coefficients β_1 and β_2. Since some customers may purchase without entering the fitting room, we control for store traffic in the model. We also add the quadratic term of store traffic, $A\_Traffic_{th}^2$, as Perdikaki et al. (2012) found non-linear relationship between store traffic and sales.

Finally, the coefficient of labor (β_5) is subject to endogeneity bias. So, we use an instrumental variable two-stage least squares (2SLS) technique to estimate this model where we consider two sets of instruments. Since labor is one of control variables instead of a key variable of interest, we assess the validity of the instrument and strength of it, and report the results using 2SLS in the Online Appendix A.

5. Results of Archival Data Analysis

We run (1) to test Hypothesis 1a-1c using ordinary least squares (OLS) methodology and results are provided in Table 3. Column (1) comprises only control variables. We then enter the fitting room traffic and its square in column (2) so this model can serve as full model.

The results from column (2) support our conjecture that fitting room traffic has an inverted-U relationship with sales, consistent with the presence of thwarting behavior. In this model, we find that the coefficients of $A\_Fit\_Traffic_{th}$ (1.69, p<0.01) and $A\_Fit\_Traffic^2_{th}$ (-0.04, p<0.01) are both statistically significant. The coefficients imply that the peak point of sales is about 1.69/(2*0.04) ≈ 24.02 of mean-centered fitting room traffic. This is about 58% of one standard deviation (41.13) above the sample mean (65.97) and lies well within the support of the data, indicating an inverted-U relationship. In other words, for low levels of fitting room traffic, sales increase with an increase in fitting room traffic; however, beyond the peak point, we find that an increase in fitting room traffic is associated with lower sales. This reveals the negative impact of congestion in the fitting room. Too much traffic in the fitting room can hurt store sales (Hypothesis 1b).

We find that the estimated coefficients of the control variables are in the expected direction. In column (1) we find that store traffic has an increasing concave relationship with sales. This is consistent with Perdikaki et al. (2012). While Perdikaki et al. (2012) claim that decline in service quality is a driver of the diminishing return to store traffic, they do not provide any evidence. In our setting, when we add fitting room traffic and its square (column (2)) we find that the quadratic term of the store traffic is no longer significant. It implies that the diminishing return to store traffic that was observed in column (1) is largely driven by the congestion-driven phenomenon in the fitting room area for this retailer. This further supports our claim that managing congestion in fitting rooms is crucial at this retailer.

Store labor is positively associated with sales (17.40, p<0.01). As we discussed earlier, the coefficient of labor is biased due to endogeneity issues, so we use an instrumental variable 2SLS regression to address this in the Online Appendix A. As expected, we find significant seasonality in store sales. For both models, we observe statistically significant monthly dummies and interactions between day-of-the-week dummies and hour-block dummies. We examine the individual dummy variables to ensure that it is consistent with prior expectations of seasonality. Consistent with the prior literature (Perdikaki et al. 2012), promotions are negatively associated with sales (-106.81, p<0.01), meaning that for a given level of store traffic a randomly chosen shopper in the store is likely to spend less during promotional period. However, because of much higher traffic during the promotion, overall sales are higher. The adjusted R^2 is 82.04%, indicating that our model fit is good.

Finally, we replace the hour-block dummies with hourly dummies and report results in column (3). We find that the coefficients are similar to those reported in column (2). However, the VIF of the linear and quadratic terms of store traffic are greater than 10 in this model, indicating multicollinearity issue.

Table 3: Inverted-U Relationship (Retailer A)

Dependent Variable:	**Sales ($Sales_{th}$)**		
	(1)	**(2)**	**(3) Hourly dummies**
$A\_Fit\_Traffic_{th}$		1.69***	1.51***
		(0.45)	(0.46)
$A\_Fit\_Traffic^2_{th}$		-0.04***	-0.03***

		(0.003)	(0.003)
$A\_Traffic_{th}$	14.33***	13.71***	12.09***
	(0.21)	(0.29)	(0.34)
$A\_Traffic_{th}^2$	-0.006***	0.001	0.006***
	(0.0008)	(0.001)	(0.001)
$Labor_{th}$	22.63***	17.40***	14.38***
	(3.49)	(3.48)	(3.48)
$Promotion_t$	-105.96***	-106.81***	-80.23***
	(24.29)	(24.03)	(23.97)
Controls	Yes	Yes	Yes
Observations	5312	5312	5312
Adjusted R^2	0.8156	0.8204	0.8240

Note. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The following control variables were included in all of the regressions: interactions between hour-block dummies and day-of-the-week dummies, and monthly dummies. Column (3) has hourly dummies instead of hour-block dummies.

5.1. Robustness Checks: Alternative Ways to Test for the Inverted U-Shaped Relationship

Up to this point, our conclusion about the presence of the inverted U-shaped relationship was based on two criteria: (1) negative and significant coefficient of $A\_Fit\_Traffic_{th}^2$ and (2) the peak point, defined as the value of fitting room traffic at which sales are highest, lies within the data range. However, a quadratic approximation of a concave unimodal relationship could be erroneous in the presence of extreme observations. So, as we mentioned in §4.3., we use various alternative ways to test for the inverted U-shaped relationship between fitting room traffic and sales. The first two methods are based on the quadratic assumption while the remaining three methods avoid such assumption.

First, we perform a robustness check based on Aiken and West's (1991) procedure for testing curvilinear relationship. Herein we compute the slope of the curve for different values of the variable and ensure that the slope differs significantly from zero and has different signs on either side of the peak point. In model (1), the slope and the standard error can be obtained by $\beta_1 + 2\beta_2 A\_Fit\_Traffic_{th}$ and $\sqrt{\sigma_{11} + 4A\_Fit\_Traffic_{th}^2\sigma_{22} + 4A\_Fit\_Traffic_{th}\sigma_{12}}$, respectively. Here σ_{11} and σ_{22} are the variance of β_1 and β_2, respectively, and σ_{12} is the covariance between β_1 and β_2. Table 4 shows that tests of the slopes for the (mean centered) fitting room traffic at the peak point, and ±1 SD, minimum, and maximum value in the sample. We find that the slopes are positive and significant in the region below the peak point while the slopes are negative and significant in the region above the peak point, confirming Hypothesis 1b.

Table 4: Robustness Checks for the Inverted-U Relationship (Retailer A)

	Value	**Slope**	***p*-value**
Minimum value	-64.65	6.24	0.000
Peak point – 1 SD	-17.11	2.89	0.000
Peak point	24.02	0	0.99
Peak point + 1 SD	65.15	-2.89	0.000

Maximum value	309.85	-20.11	0.000

Note. The *p*-values are based on one-tailed test on whether the slope is >0 or <0.

Second, we calculate the confidence interval of the peak point and ensure that it lies within the sample, following Lind and Mehlum (2010). We use the Fieller (1954) method, which was recommended by Staiger et al. (1997) as the delta method is severely biased for the finite sample. Since the 95% confidence interval lies well within the range of the sample, [12.44, 34.49], we again confirm the inverted U-shaped relationship.

Third, we run spline regressions as a further validation of the inverted U-shaped relationship. Spline regressions allow us to check if sales increase first and then drop as fitting room traffic increases. In spline regressions, we choose knots that split $A\_Fit\_Traffic_{th}$ into equal-sized groups. For example, one knot splits $A\_Fit\_Traffic_{th}$ into two equal-sized groups, then estimate a spline regression to fit piecewise linear functions of $A\_Fit\_Traffic_{th}$ *1* (the lower 50%) and $A\_Fit\_Traffic_{th}$ *2* (the higher 50%). The same idea applies to two knots case. In the one knot case (column (1) in Table 5), we find that the coefficient of the first spline is positive and significant (p<0.01), whereas the coefficient of the second spline is negative and significant (p<0.01). It implies that as fitting room traffic increases, sales first increase and then drop, supporting the inverted U-shaped relationship. The conclusions are similar when we consider two knots, as shown in column (2).

Fourth, we perform the two-line test (Simonsohn 2016). Herein, we estimate two separate regression lines, one for 'low' (i.e., below the peak point) and one for 'high' (i.e., above the peak point) values of $A\_Fit\_Traffic_{th}$. The inverted-U shape is present if the 'low' slope is positive and significant; and the 'high' slope is negative and significant. In column (3) of Table 5, we find that $A\_Fit\_Traffic_{th}$ *Low* is positive and significant (p<0.05) and $A\_Fit\_Traffic_{th}$ *High* is negative and significant (p<0.01). It again supports the inverted U-shaped relationship.

Finally, we include a cubic power for fitting room traffic in (1) to test for higher order effects in our model. In column (4) of Table 5, we still find that $A\_Fit\_Traffic_{th}^2$ is negative and significant (p<0.01) whereas $A\_Fit\_Traffic_{th}^3$ is insignificant, again supporting the inverted U-shaped relationship. Including a cubic term does not improve adjusted R^2 indicating that the quadratic functional form is parsimonious. We do not add a cubic power of store traffic in this model as the VIF increases above ten. Nonetheless, addition of this variable does not change our conclusions about the inverted-U relationship between fitting room traffic and sales.

In conclusion, we identify the negative impact of congestion by showing the inverted U-shaped relationship between fitting room traffic and store sales consistent with the presence of thwarting behavior.

Table 5: Robustness Checks without Quadratic Functional Form Assumption (Retailer A)

Dependent Variable:	Sales ($Sales_{th}$)			
	Spline Regressions		(3)	(4)
	(1) One knot	(2) Two knots	Two-Line Test	Higher Order Power
$A\_Fit\_Traffic_{th}$ *1*	4.30***	6.60***		
	(0.88)	(1.23)		
$A\_Fit\_Traffic_{th}$ *2*	-2.54***	0.02		
	(0.42)	(0.96)		
$A\_Fit\_Traffic_{th}$ *3*		-2.82***		
		(0.48)		
$A\_Fit\_Traffic_{th}$ *Low*			1.58**	
			(0.63)	
$A\_Fit\_Traffic_{th}$ *High*			-4.07***	
			(0.59)	
$High$			20.57	
			(31.06)	
$A\_Fit\_Traffic_{th}$				1.64***
				(0.46)
$A\_Fit\_Traffic_{th}^2$				-0.03***
				(0.006)
$A\_Fit\_Traffic_{th}^3$				-0.00002
				(0.00002)
Controls	Yes	Yes	Yes	Yes
Observations	5312	5312	5312	5312
Adjusted R^2	0.8176	0.8181	0.8175	0.8203

Note. The following control variables were included in all of the regressions: store traffic and its square, store labor, a promotion indicator, interactions between hour-block dummies and day-of-the-week dummies, and monthly dummies.
*, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

5.2. Alternate Explanations

While the archival data help us demonstrate the presence of the inverted-U relationship between fitting room traffic and sales, the data do not help us draw any causal inferences between thwarting behavior and sales. In fact, there are possible alternate explanations why we may observe the inverted-U relationship between fitting room traffic and sales even when there is no thwarting behavior. The rest of the store is likely to be crowded when the fitting room is congested. So the store sales could decline as a result of poor service in the rest of the store, phantom stockouts in the rest of the store, long waiting times in the checkout counter, or customers may decide to purchase less because of the crowded environment (Eroglu and Machleit 1990). It is also possible that customers who use a fitting room when it is congested are systematically different from others who use it when the rooms are not congested. It is impossible to rule out all alternative explanations without detailed data of all events concurrently occurring in the store during congested periods and of customer type. So, we conduct a field study to provide direct evidence for the thwarting behavior. In the field study we directly observe the behaviors of customers to determine if they exhibit thwarting behavior. Based on observations made from the field study, we subsequently design and run a field experiment to provide a solution that can potentially mitigate the negative effects of thwarting behavior on store sales.

6. Field Study: Evidence for the Presence of Thwarting Behavior

In this section, we describe a field study where we observe behaviors of customers who experience congestion and those who do not to collect direct evidence for the presence of thwarting behavior. We use an observational study rather than a field experiment to study the impact of congestion on customer behavior due to the challenges involved in randomizing congestion across subjects (i.e., customers). We explain the ideal field experiment to study the problem and then present our research design, its limitations, and how we overcome potential selection biases in the next sections.

6.1. Theory: Ideal Experiment

An ideal experiment would require randomizing the level of congestion perceived by each customer entering the fitting room area and their behavior noted. In other words, the treatment would be the level of congestion perceived by each customer and the outcome would include the number of clothes taken inside the fitting room and amount of time spent in the fitting room. We wish to measure the amount of time spent in the fitting room to rule out the alternate explanation that customers who take more clothes to fitting rooms may speed up their trials as they are aware of waiting customers. We follow Harrison and List (2004) to formalize the ideal experiment below. Let y_{i1} and y_{i0} be the outcomes of customer *i* under treatment ($T = 1$) and control ($T = 0$), respectively. Then the average treatment effect on the treated can be defined as $ATT = \mathbb{E}(y_{i1} - y_{i0}|T = 1)$. Since one of counterfactuals is missing, we do not observe $\mathbb{E}(y_{i0}|T = 1)$ and $\mathbb{E}(y_{i1}|T = 0)$ in our data. Therefore, the reported treatment effect on the treated is $Reported\ ATT = \mathbb{E}(y_{i1}|T = 1) - \mathbb{E}(y_{i0}|T = 0) = \mathbb{E}(y_{i1} - y_{i0}|T = 1) + \{\mathbb{E}(y_{i0}|T = 1) - \mathbb{E}(y_{i0}|T = 0)\}$. The *Reported ATT* differs from the true *ATT* due to the presence of selection bias, i.e., $\{\mathbb{E}(y_{i0}|T = 1) - \mathbb{E}(y_{i0}|T = 0)\}$. The normal approach used by experimentalists is to get rid of the selection bias using randomization.

In our case, we could not randomize congestion in the retail store. The decision on how many items to carry into the fitting room would be made based on whether an individual perceived congestion or not. The perceived level of congestion may depend upon many factors including checkout queues, the number of customers entering the store along with them, and the number of customers shopping in the same section. Manipulating all of these factors to change individual consumers' perception of congestion is hard for several reasons. First, the retailer we worked with insisted that we do not affect customer operations in anyway. Second, our analysis with archival data shows that congestion could lead to lower sales; hence, getting permission to run such an experiment would be difficult. So, we do not randomize congestion but let it be exogenously determined by store traffic. Since selection bias could be an issue in the absence of randomization of the treatment, we carefully design the field study to handle potential selection biases as explained next.

6.1.1. Handling selection biases. A selection bias could occur when customers in the two groups, i.e., those who experience congestion and those who do not, are systematically different. This is possible as congestion is correlated with factors such as day-of-the-week and promotion as traffic tends to be higher during weekends and during promotional periods. Also, prior research shows that weekday (Monday-Friday) customers are different from weekend (Saturday-Sunday) customers (Lam et al. 2001). So, we design our field study to observe customers under treatment and control conditions during the same day so that we remove any day-specific effects. In addition, it is possible that customers who arrive during peak hours, when there is congestion, could be different from those who arrive during non-peak hours, when it is less likely to have congestion. So, instead of defining congestion based on the time of arrival, we define it based on the state of congestion perceived by each customer when they decide to take clothes into the fitting room. The amount of congestion perceived by each customer could depend upon factors such as the number of visible customers in the store who were entering or leaving it, the length of the cashier line, and the number of customers who were shopping alongside them in the same area of the store. Since these factors are hard to measure, we adopt an alternate metric as a proxy for the level of congestion perceived by each customer in the store. We track the number of occupied fitting rooms when each customer enters the fitting room and use the utilization rate as a proxy for the level of congestion experienced by the customer. So, even within the same peak hour, as defined by the hour-of-the-day rather than traffic-based metrics, it is possible for some customers to experience congestion and for others to not. This measure of congestion also helps us mitigate another type of selection bias due to customer balking. If customers balk during highly congested periods, then the observed behavior is only of patient (or the most motivated) customers who may be different from the overall population. Since balking is more likely to be severe when there are long queues, we avoid an all-or-nothing approach to congestion but measure it at different levels based on the room occupancy. Our assumption is that at lower levels of congestion, customers could still exhibit thwarting behavior but they are less likely to balk.

It is also possible that customers shopping in typically congested stores might be different from customers shopping in typically uncongested stores due to heterogeneity across stores such as store management team, associates, and display of items. So, we observe customers in the two groups within the same store. Demographic differences such as gender and age might drive differences in customer behavior as well. So, we observe only a particular women's fitting room area where fitting room users are all female and in similar ages. Finally, the selection bias could occur when customers shopping in different part of the store are different populations. For example, one store manager at Retailer B informed us that customers who shop for formal dresses are generally different from those who shop for swimsuits or casual wear. So, we observe differences in behavior between customers who experience congestion and those who do not in the same fitting room area.

6.2. Field Study at Retailer B

We could not conduct a field study at retailer A due to lack of access, so we sought out another retailer, B, to conduct this part of the study. Retailer B permitted us to position a female research assistant (RA) in the women's fitting room area who was able to collect customer-level and item-level data. After confirming the presence of the inverted-U relationship at retailer B (see Online Appendix B), we use the observational data collected through our field study to provide direct evidence for thwarting behavior. Next, we provide a background about retailer B and explain the field study design in detail.

We conducted the field study observing customer behavior at one of Retailer B's stores located in North Carolina in April and May 2015. Retailer B is a U.S.-based department store like retailer A, but larger. This retailer operates around 300 stores in the U.S. with the average store size being 100,000 sq. ft. The store in which we conducted the field study had two floors. The first floor was mainly for men's, children's apparel, and home goods and the second floor was primarily for women's apparel, accessories, shoes, and cosmetics. The studied store of retailer B had multiple smaller fitting room areas (Figure 4) with 3–4 rooms inside each area. For example, Polo Ralph Lauren in the men's apparel section had its own fitting room area, with 4 rooms inside.

We chose two fitting room areas, one on the first floor and the other on the second floor, where the apparel categories and brands around these areas were similar. We chose weekends for our study to observe both customers who experienced congestion and those who did not because we found limited evidence of congestion during weekdays in our preliminary observation. This store had on average higher traffic on Saturdays compared to Sundays. Both fitting room areas had their own checkout counters (see Figure 4), where we collect POS information.

Figure 4: Fitting Room Areas at Retailer B (Decentralized Fitting Room Layout)

Note. Fitting rooms are located near each brand. This fitting room area has 3 rooms in total.

6.2.1. Data collection. A female RA stationed in the fitting room area collected customer-level data during the hours 12 PM–6 PM for four days spanning three weekends. She observed customers entering the fitting room area on two Saturdays and two Sundays. She recorded the times of entry and exit of customers from the fitting room and the number of clothes they were carrying. Since the RA could not interact with customers, the number of clothes carried by customers was only an estimate based on visual inspection. Because customers typically bring clothes still attached to their hangers, the RA could easily count the number of clothes.

We also measure the magnitude of phantom stockouts among items left behind in the fitting rooms to examine the impact of this thwarting behavior on lost sales. We do so by scanning each item left behind in the fitting rooms using a POS scanner to check the in-store availability of those SKUs.

6.3. Results

We observe a total of 209 customers over 24 non-continuous hours spanning three weekends. Using the definition of congestion based on the occupancy rate of the fitting rooms as a proxy for the congestion perceived by each customer (explained in §6.1.1.), Table 6 shows results of the two-sample *t*-test with an assumption of both equal and unequal variances between two groups. We find that customers brought an average of 4.38 items into the fitting room when they did not experience congestion, whereas they brought an average of 5.76 items into the fitting room when they experienced congestion. So, customers who face congestion bring 1.38 additional clothes into the fitting room (p<0.01). It is possible that customers who enter the fitting room area during congestion might speed-up their trials as they are aware of others waiting to use the room so more items carry into the fitting room may not convert into longer occupancy time. Our results show that although customers who experienced congestion slightly speed-up their trials (104.62 seconds per cloth during congestion vs. 111.97 seconds per cloth during non-congestion), they occupy fitting rooms 1.87 minutes longer (112 seconds, p<0.05) than those who enter the fitting rooms when there was no congestion. Thus, we find support that the waiting time for other customers increases during congested periods as customers take more clothes into the fitting rooms to try on.

Table 6: Impact of Congestion on the Number of Items and Time Spent in the Fitting Room (Retailer B)

		Mean (Std. dev.)	
	# obs (*N*=209)	**Number of items brought into the fitting room**	**Time spent in the fitting room (Seconds)**
Non-congestion	143	4.38 (2.87)	490.45 (373.58)
Congestion	66	5.76 (3.56)	602.61 (332.58)
Difference		-1.38	-112.16
t-test		***p*-value of H_1 (Diff < 0)**	
Equal (unequal) variances		0.0016 (0.0034)	0.0191 (0.0155)

Congestion: Fitting room occupancy of 3 & 4 for the first floor and 2 & 3 for the second floor.
Non-congestion: Fitting room occupancy of 0, 1, & 2 for the first floor and 0 & 1 for the second floor.

We perform several robustness checks. First, even though we compare customers in the same store on the same day at the same hour, we also run a regression with hour and date controls to see whether our finding is affected by these factors. We still find that customers brought 1.42 extra clothes into the fitting rooms (p<0.01) and occupied fitting rooms for additional 1.87 minutes (112 seconds, p=0.057) when they experienced congestion. This result indicates that our model-free evidence does not suffer from selection biases due to hour and date because we dealt with them as explained in §6.1.1.

Second, to mitigate concerns about the selection bias due to balking, we test for the presence of thwarting behavior only among customers who used fitting rooms when balking was less likely. We compare the average number of items brought by customers into the fitting room and time spent by them in the fitting room at lower rates of fitting room occupancy. Compared to low fitting room occupancy (i.e., 0 or 1 occupied rooms when a customer entered a fitting room), we still find that customers who entered the fitting room at the medium level of fitting room occupancy (i.e., 2 occupied rooms and either one or two empty fitting rooms) carried 0.89 more items (p=0.025) and occupied fitting rooms 115 seconds longer (p=0.018).

Third, we rule out fitting room specific effects by conducting a two-sample t-test on each floor separately and obtain consistent results. Finally, we simply compare the average number of clothes brought by customers into the fitting room and time spent by them in the fitting room on Saturday with those on Sunday because we observe on average higher level of congestion on Saturdays than Sundays. By obtaining similar results, we can mitigate concerns that the way we operationalize congestion may drive our inferences.

6.3.1. Measuring phantom stockouts. When customers bring more clothes into the fitting room, they increase the risk of phantom stockouts if they leave them behind in the fitting room, where it is not accessible for other customers. We test our assumption that customers are likely to leave behind more clothes in the fitting rooms during congestion in the Online Appendix C.

An additional assumption required to argue that phantom stockouts can lead to lost sales is that these misplaced items in the fitting room happen to be the last available items in the store. To check the magnitude of phantom stockouts among misplaced items in the fitting room, we scanned 559 items moved from fitting rooms to the recovery rack over another 12 hour period across two days (Table 7). Among 559 misplaced items, we find that 38.6% (216 items) were unique items in the store. In other words, 38.6% of the items left behind in the fitting rooms experienced phantom stockouts. We also find that 34.7% of these misplaced items had only one additional unit in the store according to the POS system. Since inventory records are updated once a day at this retailer, it is likely that many of these items could have been the last inventory in the store, making our estimate of 38.6% phantom stockouts conservative. When we reported this result to the senior management at Retailer B, one executive commented "*Our most popular items are either in our customers' shopping bags or lying in the fitting room*."

Table 7: Phantom Stockouts (Retailer B)

# of items (%)	**Total items**	**# of available items in the store**				
		1	**2**	**3**	**4**	**Over 5**
Day 1	223 (100%)	78 (34.98%)	88 (39.46%)	25 (11.21%)	16 (7.17%)	16 (7.17%)
Day 2	336 (100%)	138 (41.07%)	106 (31.55%)	55 (16.37%)	22 (6.55%)	15 (4.46%)
Total	559 (100%)	216 (38.64%)	194 (34.7%)	80 (14.31%)	38 (6.8%)	31 (5.55%)

Note. Detailed information (e.g., price) is obtained in day 2.

The high volume of phantom stockouts that we observe can be especially costly to retailers if the misplaced items have higher values, so we next compare the prices of items misplaced in the fitting rooms with the prices of other items in the immediate sales area. We collect price information for items left behind in the fitting rooms over a six hour period in day 2. In addition, we also collect information on size, clearance item or not, and the number of swimsuits because the store manager suspected that phantom stockouts could be driven by these factors. Of the 336 misplaced items in day 2 (Table 7), we found 138 items experienced phantom stockouts. Among these 138 items, 7% were clearance items (10 items) and 15% were swimwear (21 items). In addition, phantom stockouts occurred in all sizes, so they were not restricted to extreme sizes. The average price of items left behind in the fitting rooms was $53.77 and the average price of phantom stockout items was $61.53. Using price data for all SKUs in the immediate selling floor near the fitting room area, we found the average price of 3,200 SKUs to be $46. This indicates that the items left behind in fitting rooms were indeed high-price items and the phantom stockout items are even more expensive items with significant opportunity costs. This is consistent with our assumption on higher basket value for fitting room users compared to that for non-users in the conceptual model (Figure 1). Further details about phantom stockouts are available in the Online Appendix (see Tables A4 and A5).

To summarize, we find evidence that customers who face congestion carry extra items into the fitting rooms compared to those who did not face congestion. Such thwarting behavior of customers induces negative externality as it increases waiting time and phantom stockouts for others.

7. Field Experiment: Mitigating the Negative Impact of Phantom Stockouts

In this section, we describe a field experiment to provide a solution that can potentially mitigate the negative consequence of thwarting behavior on store performance. The prior literature (Frei 2006) stated that customer-introduced variability can be managed through accommodation or reduction strategies. In the same vein, we argue that thwarting behavior can either be accommodated or reduced. Reduction strategy would entail limiting the number of clothes that customers can take into the fitting rooms. On the contrary, accommodation of thwarting behavior would require managing the increase in waiting time and phantom stockouts through other means. We may do so by speeding up customer operations to reduce waiting time for others and by providing a timely backend recovery operation in fitting rooms to move misplaced inventory to the recovery rack where it can be re-shelved. Retailer B did not wish to pursue any options that involve direct interactions with the customers so we could neither impose a limit on the number of items brought into the fitting rooms nor speed-up customer operations. Instead we focus on an intervention aimed at reducing phantom stockouts by recovering misplaced items in the fitting room back in the shelves using a dedicated associate.

7.1. Experimental Design

The goal of our experiment is to demonstrate that mitigating phantom stockouts in the fitting rooms, through a timely backend recovery operation using labor, can significantly improve store sales. This can help us provide one possible solution which can potentially mitigate the negative effect of thwarting behavior. So, in this experiment, the treatment condition involves use of a dedicated associate in the fitting room area to restock the left-behind merchandise after each customer exits. The control condition does not have a dedicated associate in the fitting room area but has the fitting room operating under the normal condition where an associate in the selling floor cleans up fitting rooms regularly (every 30 minutes) if she is available. As she prioritizes other customer-facing tasks, however, she couldn't follow 30 minutes policy during congested periods. We observed that fitting rooms were occasionally not cleaned up for more than 2 hours when store is congested.

We use the same two fitting room areas that we conducted a field study which identifies direct evidence for the thwarting behavior and its negative externalities. They are located in the first and the second floors and we use one as the treatment and the other as the control during the same period. So, we track the sales during the intervention hours at both the treatment and control areas. This ensures that any time-specific effects such as weather, day, peak-hour, or promotion are similar across customers in the treatment and control conditions. By choosing treatment and control in the same store, our design also allows us to control for store-specific factors such as the management team, assortment, and competition. We choose two fitting room areas among 20 alternative areas where the apparel categories and brands around them were similar, to mitigate area-specific factors. Nonetheless, it is still possible for the customers using the treatment fitting room area to be different from those using the control fitting room area. So, we rotate the treatment and control between the fitting room areas in the first and the second floors to mitigate any remaining area-specific effect. An added advantage of this approach is that we can minimize any spillover effects that can happen with customers drifting from treatment to control or vice versa as we choose the treatment and control areas in different floors.

7.2. Implementation & Methodology

The implementation of our experiment presents a number of challenges. Enduring adherence to experimental design is harder to obtain in the case of experiments with labor as store associates may not follow the instructions we provide. For instance, we require the associate in our field experiment to perform a recovery operation of merchandise left behind in the fitting rooms but not interact with customers as it could confound our results. Further, we find that store managers placed greater emphasis on customer-facing tasks so they tended to move the associate from the fitting room area to tasks that involve directly helping customers. So, we had to be present in the store during the entire duration of the experiment to ensure adherence. This also limits the amount of time we can run the experiment.

We run the experiment for three weekends in April and May 2015 from 12 PM to 6 PM. On one of the days, we could find the dedicated associate for only 5 hours, so our total number of hourly (treatment) observations was 35. Table 8 shows the design of the experiment, which lasted for a total of four weekends, including post-experimental period. We collect our last week of data in the absence of the associate, to further confirm that we were not capturing an overall time-trend effect. We obtained hourly sales information for two fitting room areas from their own checkout counters during the intervention period.

Table 8: Experiment Design (Retailer B)

	Fitting room on the 1st floor	**Fitting room on the 2nd floor**
Weekend 1	Treatment	Control
Weekend 2	Control	Treatment
Weekend 3	Treatment	Control
Weekend 4	Control	Control

Control: Current staffing practice.
Treatment: Having a dedicated associate to restock items left behind in the fitting room area.

Apart from comparing the sales across the treatment and control rooms, we also used the difference-in-differences (DiD) estimation to control for other factors. The critical assumption underlying the DiD estimator is the existence of a parallel trend. That is, the two groups would follow the same trend in the absence of treatment. In our context, because the control and treatment fitting room areas were very similar in terms of products displayed around them and they were located in the same store, it was highly likely that the parallel trend assumption held. We confirm this assumption by performing statistical tests of difference in trends across these fitting room areas.

If the null hypothesis is that, without treatment, sales generated by the fitting rooms on the 1st floor (or 2nd floor) over sales generated by the fitting rooms on the 2nd floor (or 1st floor) is constant, then we could use the following regression to examine the effect of treatment.

$$Sales_{fdh} = \alpha_0 + \alpha_1 Treatment_{fdh} + \alpha_2 SecondFloor\ Dummy + \alpha_3 Hour\ Dummies + \alpha_4 Date\ Dummies + \varepsilon_{fdh} \quad (2)$$

where subscript f denotes floor, d denotes date, and h denotes hour. $Treatment_{dh} = 1$ in the first floor for weekends in weeks 1 and 3 and for the second floor for a weekend in week 2 from 12 PM to 6 PM; otherwise, it equals zero. The coefficient of interest is α_1, which can be interpreted as the sales change due to treatment. We control for other factors such as hour-of-the-day, date, day-of-the-week, and month.

7.3. Results

Using the raw data during the first three-weekend period, we first compare sales from the treated fitting rooms to sales from the control fitting rooms. We find that sales for the checkout counter allocated to the treated fitting room area is $127.3 higher compared to sales from the control fitting room area (667.64 vs. 540.34, $p<0.1$). This result indicates that providing a timely backend recovery operation to mitigate phantom stockouts through a dedicated associate in the fitting room area can increase store sales considerably. Even though we compare the treatment with the control in the same store on the same day at the same hour, we also run a regression, i.e., equation (2), with hour and date controls to see whether our finding is affected by these factors. In this analysis, we include the last week of data in the absence of the associate, to further confirm that we were not capturing an overall time-trend effect.

Table 9 shows that an increase in sales for the checkout counter allocated to the treated fitting room area due to a timely backend recovery operation provided by a dedicated associate in the fitting room area varied from $187.6 ($p<0.05$, Model 1) to $165 ($p<0.01$, Model 2), depending on control variables included in the regression. We found that the $187.6 increase in sales constituted a 22.7% increase in average hourly sales at this checkout counter during peak hours. Given that the wage rate was less than $15 per hour and the gross margin was about 40%, we show that having a dedicated associate in the fitting room area would be profitable.

To summarize, we find that providing a timely backend recovery operation through a dedicated associate to mitigate phantom stockouts in the fitting rooms can improve store sales significantly. This field experiment demonstrates one potential solution to mitigate the negative consequences of thwarting behavior due to phantom stockouts.

Table 9: Impact on Sales of Mitigating Phantom Stockouts by Fitting Room Associate (Retailer B)

Dependent Variable: Hourly Sales	**Model (1)**	**Model (2)**
Treatment	187.6**	165***
	(57.63)	(43.53)
Second Floor	125.14	122.74
	(68.43)	(65.84)
Hour dummies	Yes	Yes
Date dummies	Yes	No
Week, Day-of-the-week, & Month dummies	No	Yes

Number of observations	94	94
Number of treatments	35	35
Adjusted R^2	0.4238	0.3235

Note. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. We use hourly-checkout-counter-level sales information for two studied fitting room areas for 6 hours a day (from 12 PM to 6 PM; except one of days observing 5 hours only) during four-weekend (Saturday and Sunday) period.

8. Managerial Implications and Conclusion

To summarize, our archival data analysis shows an inverted-U relationship between fitting room traffic and store sales, consistent with thwarting behavior. Our field study shows that consumers indeed change behavior during congestion that could lead to increased waiting time for others and to phantom stockouts. Finally, our field experiment provides a solution to mitigate the negative impact of thwarting behavior due to phantom stockouts and discusses its implementation.

Our study has direct managerial impact at retailers. For example, we demonstrate the need for excellence in store execution by showing a large magnitude of phantom stockouts at retailer B when the fitting rooms were not well attended. Although retailer B's management was aware of the challenges posed by misplaced inventory, they did not realize the large impact on lost sales. Retailer B had therefore followed a policy of re-shelving items on the recovery racks just before the store closed when store is not crowded. Our study shows that such a policy would lead to significant phantom stockouts during congested periods, leading to lower sales. In response to our findings, retailer B changed its policy to continuously monitor fitting rooms and to pay special attention to them during congested hours.

Other aspects of our findings can be further generalized to other retailers and other organizations in different industries. First, we identify thwarting behavior, defined as a change in customer's behavior when they experience congestion in a way that induces negative externalities on other customers. Thwarting behavior is especially problematic if organizations have self-service operating systems with little or no monitoring, where customers determine their service speed. According to the recent report of Allied Market Research (Shende 2015), global self-service technology market will expand to \$31.75B by 2020 with an annual growth rate of 13.98% during the forecast period 2015–2020. Some of the self-service environments include ticketing kiosks at movie theaters, bus and train stations, and parking lots; ATMs; food-ordering kiosks at restaurants such as McDonald's and Chili's; self-checkout kiosks at retail outlets; DVD rental kiosks at Redbox; gasoline stations; and bike-sharing programs. Future research can test for the presence of thwarting behavior in these settings. For example, thwarting behavior in bike-sharing programs can manifest if customers had rather pay extra to retain bikes during congested periods rather than return and pick-up later for fear of losing bikes when they need them. In this particular case, while the sales may not be impacted in the short-term such strategic behavior may lead to long-term adoption problems. So, unlike recent studies (e.g., Gavett 2015 and references therein) show that self-service technologies can benefit sales, our study argues for caution and a need for understanding customer behavior especially during congestion.

Second, we demonstrate that providing a timely backend recovery operation through a dedicated associate in the fitting room area increases store sales significantly, as it is effective in accommodating thwarting behavior by mitigating phantom stockouts. Clearly, this is not the only approach to do so. Automation is another way to accommodate thwarting behavior. For example, Hointer, a U.S.-based retail start-up, is using robots to increase its service rate by sending customers' selections directly to the fitting room on a steel cable. Unused items are removed when customers drop them down a chute. Another approach to counter thwarting behavior would be to train customers to be expedient and to put unwanted items on the recovery rack instead of leaving them in the fitting room. Frei (2008) states that there are many instrumental (i.e., "carrot and stick" approach) and normative (i.e., using shame, blame, and pride) techniques available to train customers. However, as pointed out by Moon and Frei (2000), this approach must be used with caution, as it can overwhelm or annoy customers, leading them to avoid self-service options. One common approach followed by many retailers is a reduction strategy to limit the number of clothes that can be taken into the fitting room. Further research can examine and compare the effectiveness of each strategy for handling thwarting behavior.

Third, our paper offers insights on how labor can be used to boost store sales. Retailers tend to be budget-driven and view labor as a short-term expense rather than as a driver of sales (Ton 2009; Fisher and Raman 2010), resulting in understaffing chronically during peak hours (Mani et al. 2015). Under the environment of limited labor resource, using labor in a way to increase store sales is important. Our study demonstrates that the cost of having an associate in the fitting room area performing a backend recovery operation of misplaced merchandise in the fitting rooms could pay for itself and even more from the increase in sales. Using gross margin of 40% for retailer B and less than $15 hourly wage, we can conclude that having a dedicated associate in the fitting room area yields a significant positive ROI.

Like other empirical studies, our research has limitations related to data availability. Our store traffic and fitting room traffic data capture the total number of visits, not the number of visits by unique customers. In other words, if fitting room users make multiple trips to the store, they would be counted as multiple visits. Our results, however, are conservative, as we currently assume one visit per customer. If customers must make multiple visits before making their purchase, the impact of congestion on sales would be larger. Also, no extant technologies allow retailers to distinguish group shoppers, who may arrive with friends or family members to simply observe and offer opinions rather than to make purchases themselves, from single shoppers. This could cause measurement error in our variables for retailer A, but is not relevant for retailer B, where we observed every customer and identified group shoppers. Thus, our data analysis at retailer B relaxes some of these concerns.

Future research could have a number of venues. It would be interesting to understand the causes of thwarting behavior and ways to manage it. For example, how is the thwarting behavior linked to the patience threshold of customers? Researchers have pointed out that it is important for the OR and OM community to incorporate findings in behavioral literature in the design of queueing systems for service firms (Bitran et al. 2008). We hope that thwarting behavior is considered in the design of queueing systems.

References

Aiken LS, West SG (1991) *Multiple Regression: Testing and Interpreting Interactions*. Sage, Newbury Park, CA.

Aksin Z, Ata B, Emadi S, Su CL (2013) Structural estimation of callers' delay sensitivity in call centers. *Management Sci.* **59**(12): 2727-2746.

Allon G, Hanany E (2012) Cutting the line: Social norms in queues. *Management Sci.* **58**(3): 493-506.

Al-Ubaydli O, List JA (2015) Do natural field experiments afford researchers more or less control than laboratory experiments? A simple model. *Working paper*.

Anand KS, Pac MF, Veeraraghavan S (2011) Quality-speed conundrum: Trade-offs in customer-intensive services. *Management Sci.* **59**(1): 157-171.

Batt RJ, Terwiesch C (2015) Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Sci.* **61**(1): 39-59.

Batt RJ, Terwiesch C (2016) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Sci. forthcoming*.

Bitran GR, Ferrer J, Oliveira PR (2008) Managing customer experiences: Perspectives on the temporal aspects of service encounters. *Manufacturing Service Oper. Management* **10**(1): 61-83.

Buell WR, Campbell D, Frei FX (2010) Are self-service customers satisfied or stuck? *Production Oper. Management* **19**(6): 679-697.

Buell WR, Norton MI (2011) The labor illusion: How operational transparency increases perceived value. *Management Sci.* **57**(9): 1564-1579.

Chan CW, Yom-Tov G, Escobar G (2014) When to use speedup: An examination of service systems with returns. *Oper. Res.* **62**(2): 462-482.

Deo S, Gurvich I (2011) Centralized vs. decentralized ambulance diversion: A network perspective. *Management Sci.* **57**(7): 1300-1319.

DeHoratius N, Raman A (2008) Inventory record inaccuracy: An empirical analysis. *Management Sci.* **54**(4): 627-641.

DeHoratius N, Ton Z (2015) The role of execution in managing product availability. Agrawal N, Smith S, eds. *Retail Supply Chain Management*. Springer, 53-77.

Dion D (1999) A theoretical and empirical study of retail crowding. *Eur. Advances Consumer Res.* **4**: 51-57.

Dong J, Feldman P, Yom-Tov GB (2015) Service systems with slowdowns: Potential failures and proposed solutions. *Oper. Res.* **63**(2): 305-324.

Eroglu SA, Machleit KA (1990) An empirical study of retail crowding: Antecedents and consequences. *J. Retailing* **66**(2): 201-221.

Fieller EC (1954) Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B.* 175-185.

Fisher ML, Krishnan J, Netessine S (2007) Retail store execution: An empirical study. *Working paper*.

Fisher ML, Raman A (2010) *The New Science of Retailing*. Harvard Business Press, Boston.

Fisher ML, Rajaram K (2000) Accurate retail testing of fashion merchandise: Methodology and application. *Marketing Science* **19**(3): 266-278.

Frei FX (2006) Breaking the trade-off between efficiency and service. *Harvard Bus. Rev.* **84**(11).

Frei FX (2008) The four things a service business must get right. *Harvard Bus. Rev.* **86**(4).

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

Gavett G (2015) How self-service kiosks are changing customer behavior. *Harvard Bus. Rev.* March 11.

Green LV (2004) Capacity planning and Management in Hospitals. Brandeau ML, Sainfort F, Pierskalla WP, eds. *Operations Research and Health Care: A Handbook of Methods and Applications.* KIluwer Academic Publishers, Norwell, MA, 15-42.

Harrison GW, List JA (2004) Field Experiments. *J. Econ. Lit.* **42**(4): 1009-1055

Hassin R, Haviv M (2003) *To Queue or Not To Queue: Equilibrium Behavior in Queueing Systems*. Kluwer Academic Publishers, Boston.

Hennessy DA, Wiesenthal DL (1997) The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. *Ergonomics* **40**(3): 348-361.

Hu K, Allon G, Bassamboo A (2016) Understanding customers retrial in call centers: Preferences for service quality and service speed. *Working paper*.

Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Sci.* **55**(9): 1486-1498.

Kc DS (2013) Does multitasking improve performance? Evidence from the emergency department. *Working Paper*.

Kesavan S, Staats BR, Gilland W (2014) Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Sci.* **60**(8): 1884-1906.

Kulkarni VG (2009) *Modeling and Analysis of Stochastic Systems*, 2nd ed. Chapman & Hall, London, UK.

Lam SY, Vandenbosch M, Hulland J, Pearce M (2001) Evaluating promotions in shopping environments: Decomposing sales response into attraction, conversion, and spending effects. *Marketing Sci.* **20**(2): 194-215.

Langer EJ, Saegert S (1977) Crowding and cognitive control. *J. Personality Soc. Psych.* **35**(3) 175-182.

Lind JT, Mehlum H (2010) With or without U? The appropriate test for a U-shaped relationship. *Oxford Bull. Econom. Statist.* **72**(1): 109-118.

Lee H-H, Pinker EJ, Shumsky RA (2012) Outsourcing a two-level service process. *Management Sci.* **58**(8): 1569-1584.

Lu Y, Olivares M, Musalem A, Schilkurt A (2013) Measuring the effect of queues on customer purchases. *Management Sci.* **59**(8): 1743-1763.

Mani V, Kesavan S, Swaminathan JM (2015) Estimating the impact of understaffing on sales and profitability in retail stores. *Production Oper. Management* **24**(2): 201-218.

Moon Y, Frei F (2000) Exploding the self-service myth. *Harvard Bus. Rev.* **78**(3): 26-27.

Musalem A, Olivares M, Schilkrut A (2016) Retail in high definition: Monitoring customer assistance through video analytics. *Working paper*.

Netessine S, Fisher ML, Krishnan J (2010) Labor planning, execution, and store performance: An exploratory investigation. *Working paper*.

Perdikaki O, Kesavan S, Swaminathan JM (2012) Effect of traffic on sales and conversion rates of retail stores. *Manufacturing Service Oper. Management* **14**(1):145-162.

Raman A, DeHoratius N, Ton Z (2001) Execution: The missing link in retail operations. *California Management Rev.* **43**(3): 136-152.

Sasser W (1976) Match supply and demand in service industries. *Harvard Bus. Rev.* **54**(6): 133-140.

Shende S (2015) Global self service technologies market – opportunities and forecasts, 2015-2020. *Allied Market Research Report*, May 2015.

Simonsohn U (2016) Two-lines: The first valid test of U-shaped relationships. *Working paper*.

Staiger D, Stock JH, Giland WG (1997) The NAIRU, unemployement and monetary policy. *J. Econom. Perspect.* **11**(1): 33-49.

Tan TF, Netessine S (2014) When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Sci.* **60**(6): 1574-1593.

Ton Z, Huckman RS (2008) Managing the impact of employee turnover on performance: The role of process conformance. *Organ. Sci.* **19**(1): 56-68.

Ton Z (2009) The effect of labor on profitability: The role of quality. *Working paper*.

Ton Z, Raman A (2010) The effect of product variety and inventory levels on retail store sales: A longitudianl study. *Production Oper. Management* **19**(5): 546-560.

Online Appendix

Online Appendix A: 2SLS regression to address endogeneity between labor and sales

As we discussed, the coefficient of labor (β_5) in (1) is subject to endogeneity bias. So, we use an instrumental variable two-stage least squares (2SLS) technique to estimate this model where we consider two sets of instruments. The first instrument is a lagged labor variable from 7 days before. We argue lagged labor variables are valid instruments for the following reasons. First, the labor schedules do not change drastically from week to week so the current schedule is a good predictor of one-week ahead schedule for workers. Second, lagged labor variables satisfy the exclusion condition since they do not impact current period's sales. Lastly, lagged values of labor have been commonly used as instruments in many settings (Bloom and Van Reenen 2007; Siebert and Zubanov 2010; Tan and Netessine 2012). This instrument is not ideal in the present of common demand shocks that are correlated over time. So, we add a second set of instruments.

The second instrument is the labor variable of other store located in the same county under the same retail chain. As the labor cost within a county would be highly correlated, this market-based instrument serves as an exogenous cost-based labor-supply shifter. We chose county-level, as opposed to state-level because labor costs would be different across stores in different counties resulting in weak instruments. As stores in the same county face the same labor market condition, the labor schedule at one store could well predict the other store's schedule. This instrument also satisfies the exclusion restriction since it does not impact the focal store's sales. In addition, this market-based instrument has been commonly used in other settings (Nevo and Wolfram 2002).

The results are provided in Table A1. Our main finding about the inverted-U shaped relationship between fitting room traffic and store sales is present, supporting Hypothesis 1b. All coefficients are quite similar with OLS estimation. After correcting for the endogeneity bias, the coefficient of labor variable is 112.85 (p<0.01).

To assess the validity of the instruments, we perform several statistical tests to examine whether they meet the relevance criteria. First, the R^2 value from the first stage regression (Table A2) of the endogenous labor variable is 0.69, indicating that the instruments have significant explanatory power. In this first stage regression, the coefficients of the instrumental variables are statistically significant with expected sign. We also check the simple correlation. The correlation between labor and other store's labor is 0.62 and the correlation between labor and 7-day-lagged labor is 0.74. Second, the F-statistics of the excluded instruments in the first stage regression is well over 10 in our regressions, indicating that the instruments are not "weak" in the sense of Staiger and Stock (1997). Using lagged labor may not be an ideal instrument in the event of common demand shocks that are correlated over time. Although we use other instrument together with lagged labor, we adjust for these common demand shocks, which are basically trends (Villas-Boas and Winer 1999), in our models by adding monthly dummy variables, thus mitigating this concern. We further rule out serial correlation in our model by adding trend as a robustness check and obtain qualitatively the same results. Although not conclusive, these test statistics build our confidence that our instruments are valid.

Table A1: Inverted-U Relationship Using 2SLS (Second Stage Regression)

Dependent Variable:	**Sales ($Sales_{th}$)**
$A\_Fit\_Traffic_{th}$	1.53***
	(0.58)
$A\_Fit\_Traffic_{th}^2$	-0.03***
	(0.004)
$A\_Traffic_{th}$	11.81***
	(0.38)
$A\_Traffic_{th}^2$	-0.001
	(0.001)
$Labor_{th}$	112.85***
	(12.50)
$Promotion_t$	-66.67**
	(28.71)
Controls	Yes
Observations	3696
Adjusted R^2	0.8294

Note. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The following control variables were included in all of the regressions: interactions between hour-block dummies and day-of-the-week dummies, and monthly dummies. 2SLS has fewer observations due to paucity of data on instruments.

Table A2: Inverted-U Relationship Using 2SLS (First Stage Regression)

Dependent Variable:	**Labor ($Labor_{th}$)**
$Labor\ (last\ week)_{th}$	0.21***
	(0.01)
$Labor\ (other\ store\ in\ the\ same\ county)_{th}$	0.26***
	(0.02)
Controls	Yes
Observations	3696
Adjusted R^2	0.6914
F-value	224.72 ($p<0.01$)

Note. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The following control variables were included in the regressions: (mean centered) fitting room traffic and its squared term, (mean centered) store traffic and its squared term, promotion indicator, interactions between hour-block dummies and day-of-the-week dummies, and monthly dummies.

Online Appendix B: Validating the inverted U-shaped relationship at retailer B

To see whether our result is robust to other retailers, we further test the inverted U-shaped relationship between fitting room traffic and store sales at another retailer (retailer B), where the fitting room layout is different from that of retailer A. We have only 24 hourly data points (6 hours per day for four days) that we collected by observing each customer in the period of a field study, as retailer B had not installed technology to obtain traffic data.

Table A3-a shows the inverted U-shaped relationship between fitting room traffic and sales at the nearest checkout counter from the studied fitting room area. Model (1) does not control for any time effect as we have limited sample size. Model (2) includes hour dummies, Model (3) includes a Saturday indicator, and Model (4) includes both. Throughout different model specifications, even though the sample size is small, we find that the coefficient of $A\_Fit\_Traffic_{th}$ is positive and statistically significant (p<0.1) and $A\_Fit\_Traffic_{th}^2$ is negative and statistically significant (p<0.1) and the peak point is within the data range (0, 35); again, this supports Hypothesis 1b. We also conduct several robustness tests for the inverted U-shape. For example, Fieller's interval of [10.26, 34.27] for the peak point is within the data range. The spline regressions in Table A3-b shows that the coefficient of the first spline is positive and significant (p<0.05), whereas the second is negative and weakly insignificant (p=0.103), partially supporting the inverted U-shaped relationship between fitting room traffic and sales. The conclusions are similar when we consider the case with two knots, as shown in column (2). Since we have limited sample size, coefficients are not significant, but they are all on the right direction.

In conclusion, we find the inverted U-shaped relationship between fitting room traffic and store sales for both retailers A and B. Since these two retailers are different in terms of location, carried products, gross margin, and notably fitting room layouts, this finding could be robust to other retailers that use fitting rooms.

Table A3-a: Inverted-U Relationship (Retailer B)

Dependent Variable: $Sales_{th}$	**Model (1)**	**Model (2)**	**Model (3)**	**Model (4)**
$A\_Fit\_Traffic_{th}$	128.06***	121.82***	96.44**	84.69*
	(31.43)	(33.63)	(40.87)	(43.05)
$A\_Fit\_Traffic_{th}^2$	-4.54***	-4.27***	-3.49**	-3.03*
	(1.31)	(1.44)	(1.57)	(1.68)
Saturday dummy	No	No	Yes	Yes
Hour dummies	No	Yes	No	Yes
Observations	24	24	24	24
Adjusted R^2	0.4417	0.4193	0.4527	0.4462

Note. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A3-b: Robustness Checks Using Spline Regressions (Retailer B)

Dependent Variable: $Sales_{th}$	(1) One knot	(2) Two knots
$A\_Fit\_Traffic_{th}$ *1*	72.64**	54.42
	(25.82)	(38.33)
$A\_Fit\_Traffic_{th}$ *2*	-26.90	32.47
	(15.75)	(29.56)
$A\_Fit\_Traffic_{th}$ *3*		-48.30
		(42.18)
Saturday dummy	Yes	Yes
Observations	24	24
Adjusted R^2	0.4965	0.3829

Note. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Online Appendix C: Customers leave behind more clothes in the fitting rooms during congestion

In the field study, we were able to observe the number of items that customers bring into the fitting room because customers typically bring clothes still attached to their hangers. However, it was difficult for us to track the number of clothes carried back by customers when they exited fitting rooms because they had typically removed clothes from their hangers. Furthermore, we could not count the clothes left-behind by each customer in the fitting room as this would require us to visit the fitting room each time a customer exited and cause disruption to other customers. Therefore, we use an alternative approach to obtain a proxy for the number of items left-behind by each customer. We track the number of items brought by associates from fitting rooms to a recovery rack. Because we knew the number of customers who used the fitting rooms between two consecutive recovery operations by an associate, we could calculate the average number of items left-behind by each customer.

Specifically, we track an associate's recovery operation of items left behind in the fitting room to the recovery rack from 12 PM to 6 PM on one weekend. In this store, an associate in the selling floor was assigned to clean the fitting rooms periodically (every 30 minutes), if she is available, and to bring the misplaced items to a recovery rack located outside the fitting room area. So we have information about (1) the time at which an associate brought out misplaced items from the fitting room to the recovery rack and (2) the number of those items. We divide the total number of items that an associate brought out from the fitting rooms by the total number of customers who used fitting rooms during associate's two consecutive cleanups to obtain the average number of items left behind in the fitting rooms by each customer for a given period. We classify those periods into two categories, congestion and non-congestion, based on whether majority of customers who visited the fitting rooms during those periods experienced congestion or not, as explained in §6.1.1. This approach yields seventeen observations, admittedly a small sample size, as we could not track the number of items left behind by each customer as explained earlier. Then we perform the two-sample *t*-test. Consistent with our argument, we find that, on average, customers left behind 2.84 items more in the fitting room when more than half of fitting room users during the same period experience congestion compared to when less than half of fitting room users during the same period do (5.67 vs. 2.83, p=0.072). We obtain similar results when we group periods on Saturday and compared against the periods on Sunday, mitigating concerns about the way we operationalize congestion.

Table A4: Phantom Stockouts (Details on day 1)

Time stamp (each visit)	Total items brought by associate (each visit)	# items available in the store					Phantom Stockouts (each visit)
		1	2	3	4	Over 5	
11:45	4	0	2	0	0	2	0.00
12:15	18	5	9	2	0	2	0.28
12:40	1	0	1	0	0	0	0.00
12:48	1	0	1	0	0	0	0.00
13:02	2	1	0	1	0	0	0.50
13:10	4	2	2	0	0	0	0.50
13:13	1	1	0	0	0	0	1.00
13:25	5	1	4	0	0	0	0.20
13:35	7	1	4	0	2	0	0.14
13:50	5	0	2	1	2	0	0.00
13:56	12	7	4	1	0	0	0.58
14:24	3	2	1	0	0	0	0.67
14:40	11	6	1	3	0	1	0.55
14:52	3	2	1	0	0	0	0.67
14:55	6	1	4	1	0	0	0.17
15:24	11	4	6	1	0	0	0.36
15:35	18	7	6	1	3	1	0.39
15:50	24	8	12	1	1	2	0.33
16:00	8	3	4	0	1	0	0.38
16:18	24	7	6	6	3	2	0.29
16:30	13	7	4	2	0	0	0.54
16:48	4	2	2	0	0	0	0.50
17:05	1	1	0	0	0	0	1.00
17:20	5	1	1	0	0	3	0.20
17:25	14	5	6	1	2	0	0.36
17:40	13	3	3	3	2	2	0.23
17:58	5	1	2	1	0	1	0.20
Total	223	78	88	25	16	16	0.37

Note. Phantom stockouts (each visit) is defined as "#items with availability of 1 (each visit) / Total items brought by associate (each visit)," where the availability of 1 means that the misplaced item in the fitting room is the only available item in the store, which is subject to the phantom stockout.

Table A5: Phantom Stockouts with Price, Size, and Category (Details on day 2)

			# items available in the store				
		Total	**1**	**2**	**3**	**4**	**Over 5**
	# items	336	138	106	55	22	15
	(%)	1.00	0.41	0.32	0.16	0.07	0.04
	Petite	45	23	14	8	0	0
	(%)	0.13	0.17	0.13	0.15	0.00	0.00
	Clearance	18	10	6	1	0	0
	(%)	0.05	0.07	0.06	0.02	0.00	0.00
	Swimming	53	21	20	8	2	2
	(%)	0.16	0.15	0.19	0.15	0.09	0.13
Price	mean	53.77	61.53	53.29	45.62	37.27	39.73
	s.d.	28.52	34.37	26.05	13.33	16.96	13.87
	median	49	50	48.5	49	40	42
	min	9.97	11.97	9.97	9.99	11.97	11.97
	max	160	160	151.95	69	69	58
Size	2	2	1	1	0	0	0
	4	1	0	1	0	0	0
	6	9	6	1	2	0	0
	8	12	5	6	0	1	0
	10	7	3	4	0	0	0
	12	17	10	3	3	1	0
	14	23	4	10	7	1	1
	16	14	9	3	0	0	2
	18	1	1	0	0	0	0
	XS	6	6	0	0	0	0
	S	36	17	11	6	1	1
	M	46	19	14	4	7	2
	L	72	15	28	18	5	6
	XL	45	19	10	7	6	3
	4P	1	1	0	0	0	0
	6P	1	1	0	0	0	0
	8P	3	2	0	1	0	0
	10P	2	1	1	0	0	0
	12P	1	0	0	1	0	0
	PP	1	1	0	0	0	0
	PS	13	4	6	3	0	0
	PM	12	4	5	3	0	0
	PL	8	7	1	0	0	0
	PXL	3	2	1	0	0	0

Note. Average price of 3200 SKUs in this section, the immediate sales floor from the studied fitting room area, is $46.

References

Bloom N, Van Reenen J (2007) Measuring and explaining management practices across firms and countries. *Quart. J. Econom.* **122**(4): 1351-1408.

Nevo A, Wolfram C (2002) Why do manufacturers issue coupons? An empirical analysis of breakfast cereals. *RAND J. Econom*. **33**(2): 319-339.

Siebert WS, Zubanov N (2010) Management economics in a large retail company. *Management Sci*. **56**(8): 1398-1414.

Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica* **65**(3): 557-586.

Villas-Boas JM, Winer RS (1999) Endogeneity in brand choice models. *Management Sci.* **45**(10): 1324-1338.