

The Impact of Supplier Inventory Service Level on Retailer Demand

January 4, 2016

Abstract

To set inventory service levels, suppliers must understand how changes in inventory service level affect demand. We build on prior research, which uses analytical models and laboratory experiments to study the impact of a supplier's service level on demand from retailers, by testing this relationship in the field. We analyze a field experiment at the supplier Hugo Boss to determine how the supplier's inventory service level affects demand from its retailer customers. We find increases in historical fill rate to be associated with statistically significant and managerially substantial increases in current retailer orders (i.e., demand, not just sales). Specifically, a one percentage point increase in fill rate, measured over the prior year, is associated with a statistically significant 11 % increase in current retailer demand, controlling for other factors that might affect retailer demand. We explore the drivers of this demand increase, including changes in retailer assortment and order frequency. We discuss features of a retail buyer's decision context identified through our field work that may explain the magnitude of the relationship we observe.

1 Introduction

To set inventory service levels, suppliers to retailers must understand how changes in service level affect the orders they receive. Increasing inventory service level often requires substantial investments in technology and process change. Whereas suppliers can measure these costs of increasing

service level, the benefits are not as straightforward. A supplier that increases its inventory service level may not only capture lost sales but may also change the demand it receives from retail buyers, the employees tasked with making purchasing decisions on behalf of a retailer.

In this paper, we measure the effect of changes in a supplier’s inventory service level on the demand the supplier receives from retailers within the supply chain for functional apparel products (Fisher, 1997). We examine potential drivers of the change in retailer demand, such as changes in retailer product assortment and order frequency. Although there are several empirical studies of the effects of inventory service level changes on demand in business-to-consumer settings (Fitzsimons, 2000; Swait and Erdem, 2002; Anderson et al., 2006; Gallino et al., 2013), there are few empirical studies of this relationship in business-to-business (B2B) settings.

To conduct this study, we collaborated with Hugo Boss, a major supplier of branded apparel. We tracked a pilot project designed to increase the inventory service level of a subset of Hugo Boss’s bodywear stock keeping units (SKUs) while leaving the service level of similar bodywear SKUs unchanged. Bodywear products are functional apparel items and consist of socks, t-shirts, and undergarments of various sizes, colors, and fabrics. The design of this pilot project allows us to measure the relationship between supplier service level and demand (i.e., orders) from retailers in our context and to test theories regarding this relationship.

We begin our study by estimating the overall effect of the pilot project on demand from retailers using a difference-in-differences (DID) approach (Angrist and Pischke, 2009). In this analysis, bodywear products affected by the pilot project are the treatment group. Managers at Hugo Boss chose the treatment group of SKUs in a pseudorandom approach based on the manufacturing location of the SKUs. We construct a control group via propensity score matching (Rosenbaum and Rubin, 1983) from similar bodywear products not affected by the pilot project. Fill rate, or the proportion of demand satisfied (Nahmias, 2008), for the treated SKUs increased whereas fill rate for the control SKUs remained relatively constant. The DID model shows that treated SKUs experienced a statistically significant increase in demand from retailers.

Having established the impact of the pilot project, we next present analyses that examine the drivers of the demand increase. First, we measure the relationship between fill rate and retailer

demand (i.e., retailer orders) to connect our findings to theoretical models and practitioner metrics. We find a one percentage point increase in Hugo Boss’s historical fill rate, measured over the prior year, to be associated with a statistically significant 11 % increase in retailer demand. The magnitude of this relationship is much larger than that observed by Gurnani et al. (2013) in laboratory experiments and by Anderson et al. (2006) among end consumers, which, in both studies, was approximately a one percent increase in demand associated with a one percentage point increase in service level.

We examine the long- and short-term effects of supplier inventory service level. We find that supplier stockouts increase demand from retailers in the short term and decrease demand in the long term. We observe that the relationship between historical inventory service level and retailer demand is stronger for those retailers that order more frequently and, therefore, gather more information about the supplier’s service level. This result is consistent with retailers learning about changes in supplier inventory service level (Tomlin, 2009; Chen et al., 2010).

We assess the degree to which the increase in demand following Hugo Boss’s increase in service level is due to retailers ordering more frequently, changing their assortments, or placing larger orders for products already in assortment. We find that retailers did not order more frequently in response to the increased supplier service level. Some retailers changed their assortment in response to the change in service level. However, the change in assortment explains only a small portion of the overall increase in retailer demand. Most of the increase in demand we observe is due to retailers placing larger orders for products that they ordered prior to the increase in service level. This suggests retail buyers placed larger (rather than more frequent) orders for products already in their assortment.

Our study contributes to the literature in several ways. To the best of our knowledge, we are the first to test the numerous theoretical models that link supplier inventory service level metrics and demand in the B2B context. We identify a positive, significant, and managerially substantive link between supplier service level and retailer demand in practice. We show that the reaction to changes in supplier service level is more pronounced among retailers that order more frequently. We characterize the relationship between supplier service level and retailer demand in terms of

long-term and short-term effects as well as assortment and order frequency effects. Overall, our study measures and tests an important relationship within supply chain management theory and informs suppliers making decisions about the service level they ought to provide to their channel partners.

This paper proceeds as follows. In §2, we survey the relevant literature and develop hypotheses. We discuss the Hugo Boss pilot project and present our research design in §3. In §4, we describe the data we collected from Hugo Boss. In §5, we measure the effect of the pilot project. Having established the impact of the pilot project, we next present analyses that examine the drivers of the demand increase in §6. We discuss the results as well as insights from our field research in §7.

2 Related Literature and Hypothesis Development

Our research objective is to study the relationship between supplier inventory service level and retailer demand within the functional apparel supply chain. Apparel supply chains, like many consumer product supply chains, consist of various branded suppliers selling numerous products to a set of retail chains. In this section, we survey related research and develop our hypotheses. First, we describe research on the relationship between product availability and end consumer demand. We then highlight what makes the relationship between supplier product availability and retailer demand, a B2B context, different from these previously explored business-to-consumer (B2C) contexts. We conclude by generating our hypotheses from studies that explore the interaction between a retailer and unreliable suppliers.

2.1 Inventory Service Level and End Consumer Demand

Prior empirical studies at the consumer level of the relationship between product availability and demand show that increased inventory may amplify or dampen demand. Heim and Sinha (2001) demonstrate that stockouts endanger customer loyalty. Customers facing stockouts have been observed abandoning their purchase, switching retailers, and substituting similar items in lieu of their first choice (Fitzsimons, 2000; McKinnon et al., 2007). Emmelhainz et al. (1991) and Anderson et al. (2006) investigate the long-term effect of stockouts and find that stockouts reduce long-run

demand in the case of consumers purchasing from a mail-order catalog. Researchers have also found that high product stocking quantity may drive demand by acting as a billboard (Balakrishnan et al., 2004, 2008).

Other research suggests that low product availability may drive demand. For example, end consumers that observe an item with low stock in a store’s assortment may infer that said item is more desirable. Gallino et al. (2013) explore this phenomenon empirically and demonstrate that increased inventory can lead to decreased sales. Similarly, Stock and Balachander (2005) argue that abundant inventory can signal an unpopular product.

Our work is related to these studies in that we also examine the relationship between product availability and demand. However, we investigate the interaction between a supplier and a retailer. In this context, many findings at the end-consumer level—e.g., the signaling role of the amount of inventory on display—do not apply. As we discuss below, an end consumer and a retail buyer face very different costs when they are unable to procure a product.

2.2 Retail Buyer’s Decision Context

To better understand the B2B relationship between a supplier and a retailer, we conducted interviews with more than 20 retail buyers across a variety of retail chains as well as with numerous supplier sales people and managers. We designed these interviews to identify the drivers of buyer behavior in the B2B context. This approach is similar to that taken by B2C researchers seeking to characterize the drivers of consumer behavior in the face of a stockout (see, for example, Emmelhainz et al. (1991)). Our interviews identify several distinguishing features of buyer behavior within the B2B context that differs from what is commonly presumed in existing literature on supplier-retailer relationships.

Retail buyers play a critical role within many supply chains. Retail buyers control the merchandising strategy for a product portfolio—often a line or category—and face incentives that depend on the performance of that portfolio. Managing these portfolios requires retail buyers to engage in a variety of activities, including selecting the type and quantity of merchandise to buy, allocating orders to suppliers manually or through replenishment systems, negotiating wholesale prices, pro-

viding planogram recommendations, scheduling promotional events, and determining markdowns. Given the importance of the retail buyer, suppliers describe needing to understand not only the preferences of the end consumer but also the key factors that drive purchasing decisions among retail buyers.

Supplier service level directly affects the performance of a retail buyer's portfolio of products. Higher supplier service level lets a buyer achieve a given service level to end consumers while holding less inventory. Moreover, when supplier service level is high, retail buyers can maintain the minimum assortments and facings required to provide the desired shopping experience with less safety stock. A supplier's product availability is also critical to retail buyers planning in-store promotions. We interviewed one buyer who elected to drop a planned advertising circular for a product category that was experiencing poor supplier service level at an estimated cost of \$14 million. This corroborates survey findings by Malmbak and Albaum (2007) that identify inconsistent product availability as one of the top ten reasons for discontinuing a supplier's brand.

We also observed that lapses in supplier product availability can impact a retail buyer even without affecting an end consumer. This is contrary to the assumptions commonly made in existing inventory management models (see §2.3 for a review of such models). How are retail buyers impacted by supplier unreliability? Consider, for example, a retail buyer who prepares an in-store promotion by designing circulars, selecting prices, developing planograms, communicating product placement and signage to store personnel, etc. The buyer's effort is wasted if the supplier is unable to deliver product, even if the promotion is never advertised to consumers. Similarly, when an order arrives off schedule, the buyer is typically involved in determining what went wrong with the order and how to fix any underlying problems. One retail buyer effectively summarized the negative impact of supply problems noting that buyers, as CEOs of their product portfolios, ultimately take all responsibility if anything goes wrong, since the buyers must explain, on a weekly basis, why targets are not being met (DeHoratius, 2014). Eliminating supply risk helps buyers avoid such reputational risks.

Finally, we observed that demand for a supplier's product can be influenced through the actions of the retail buyer. Prior research generally assumes that end consumer demand is exogenous to

the retail buyer. In contrast, we find that retail buyers exert considerable influence over demand, whether by promoting a particular supplier’s products or by encouraging customers to purchase substitutes for said supplier’s products. To drive consumer demand, buyers can feature (e.g., display prominently) chosen products. A buyer may thus elect to favor a supplier’s product if the supplier provides high inventory service level. Nike initiated an “Always Available” program to ensure the reliable supply of key items to retail buyers after poor service level led retail buyers to shift shelf space toward a competing brand with higher service level.

In sum, our field interviews reveal important aspects of the retail buyer’s decision context. To the best of our knowledge, this decision context has only been explored by experimentalists. Ho et al. (2010), Gurnani et al. (2013), and Nagarajan and Shechter (2013) study the ordering decisions of experimental subjects representing procurement professionals as they manipulate a number of parameters, including service level. The objective of these studies is to better understand and model human decision making in operational contexts. Although Gurnani et al. (2013) find that subjects increase orders for an unreliable supplier as its service level improves, they call for future work to explore more realistic procurement contexts in order to observe the behavior of procurement professionals within competitive situations. Hopp et al. (2009) note more research is needed to understand how discretionary decision making—such as retail ordering—works in practice.

2.3 Inventory Service Level and Retailer Demand

Existing research on the influence of product availability in the B2B context focuses on unreliable suppliers. Supplier unreliability is often modeled using random yields or stochastically proportional supply. Prior research has examined a retailer’s orders in the face of supplier unreliability under both single- and multi-supplier arrangements. For the single-supplier case, Silver (1976) and Shih (1980) find that increased supplier service level decreases a retailer’s orders, assuming exogenous demand, within the economic order quantity (EOQ) and newsvendor models. The retailer decreases its order quantity if it is unlikely to receive a smaller quantity than requested. Yano and Lee (1995) provide a survey of related models. Liu et al. (2009) study joint marketing and order quantity decisions for a single-period problem with a single, unreliable supplier, where the retailer may

increase demand for the product by incurring marketing costs, but they do not incorporate other buyer-related costs (e.g., reputational costs or effort costs). They find that increased supplier service level may increase or decrease demand depending on the model parameters.

When there are multiple suppliers selling an identical product, the retailer may mitigate its supply risk by spreading orders across these suppliers. For an EOQ model with two suppliers and exogenous demand, Gerchak and Parlar (1990) demonstrate that, when the suppliers have identical costs, the retailer's optimal order quantity for a supplier is increasing in the supplier's expected service level. Anupindi and Akella (1993) show the same result in single- and multi-period models with random, exogenous demand. Dada et al. (2007), Burke et al. (2009), and Federgruen and Yang (2009) find that increased supplier service level can increase orders for a supplier when the supplier has already been selected as a source from a set of potential suppliers.

Neither the single-supplier nor the multi-supplier model assumptions fit precisely the branded functional apparel context. As in many supply chains, the products sold by different suppliers in our setting are not perfect substitutes, and the supplier's service level is unknown to the buyer and changes over time. If a buyer does not consider a product to be an adequate substitute for others, the buyer may reduce orders for the product over time as the supplier's service level improves, an outcome predicted by single-supplier models (Silver, 1976; Shih, 1980). Moreover, improved service level may reduce the amount of safety stock a retailer carries, which can reduce the shelf space dedicated to the product within a retailer's limited storage space (see van Donselaar et al. (2010) for a discussion of storage space restrictions in retailing). Further, the retail buyer can affect end consumer demand through product promotion and placement. Which type of theoretical model better describes a retail buyer's long-term response to a change in service level is therefore an empirical question. We hypothesize:

Hypothesis 1A *Current retailer demand for a supplier's products is positively associated with historical supplier inventory service level in the long term. (H1A)*

Hypothesis 1B *Current retailer demand for a supplier's products is negatively associated with historical supplier inventory service level in the long term. (H1B)*

Whereas the models discussed thus far treat long-term, steady-state behavior, the literature also identifies distinct short-term effects of inventory service level. Specifically, researchers have suggested a short-term positive effect of stockouts on demand from retailers. Lee et al. (1997) and Cachon and Lariviere (1999) predict that product shortages lead to increased demand in the short term due to inventory rebalancing, rationing games, and phantom orders. We therefore hypothesize:

Hypothesis 2 *Current retailer demand for a supplier's products is positively associated with recent stockouts. (H2)*

In practice, retailers do not know their suppliers' yield distributions and must instead develop forecasts or beliefs about the service level of a supplier. Retailers may track changes in supplier service level informally, as in the case of a buyer's attitude toward a particular supplier, or formally, e.g., through the use of automated software and supplier scorecards. Tomlin (2009) and Chen et al. (2010) study how a retailer's orders change as it receives information and updates its beliefs about a supplier's service level. Yang et al. (2012) study a related model in which a supplier's service level is private information. If retailers build beliefs about the service level of their suppliers using historical supplier service level performance, then retailers that gather more information through more frequent ordering will exhibit a larger response to changes in supplier service level than retailers that amass less information. A related phenomenon has been studied at the consumer level. Che et al. (2012) find that consumers with short interpurchase times for a given product react more negatively to unavailability of that product than consumers with longer interpurchase times. We hypothesize:

Hypothesis 3 *Retailers that order more frequently exhibit a larger change in demand in response to changes in historical supplier inventory service level than retailers that order less frequently. (H3)*

3 Pilot Project and Research Design

To measure the relationship between supplier service level and retailer demand empirically in the context of the supply chain for functional apparel items, we collaborated with Hugo Boss, a Euro-

pean fashion house known for men’s and women’s fashion apparel, shoes, and accessories (Raman et al., 2009). We study products from Hugo Boss’s bodywear division, which supplies functional apparel products such as undergarments to retailers. In this section, we describe the pilot project implemented by Hugo Boss and discuss our research design.

3.1 Hugo Boss’s Pilot Project

The bodywear division of Hugo Boss produces bodywear SKUs comprising men’s undershirts, boxers, underwear, and socks. These products are sold by Hugo Boss to retailers such as Peek & Cloppenburg and Anson’s Herrenhaus. At the time of the field experiment, the bodywear division supplied a total of 711 SKUs for the BOSS Black brand. Of these SKUs, 513 were replenishment SKUs that did not change from year to year, which we classify as functional apparel products.

The fulfillment process at Hugo Boss operated on a weekly cycle. Retailers could determine the current availability of bodywear products only by placing an order and observing whether the order was scheduled for delivery during the week in which the order was placed. If Hugo Boss lacked the product to satisfy all retailer demand, it filled orders on a first-come, first-served basis. There was no backordering—i.e., orders not filled during the week in which they were placed were dropped.

In a pilot project designed to test the effects of increased service level on the orders of its direct customers (i.e., retailers) in Germany, Hugo Boss increased the service level to these retailers for 45 of the bodywear division’s 513 BOSS Black replenishment SKUs. The SKUs were purposefully not selected due to sales velocity, margins, or other known variables that would be expected to lead to large increases in retailer demand. Instead, Hugo Boss managers elected to use a single production line within a factory as a test bed, an approach compatible with their chosen method for increasing service level, which we describe below. As part of the pilot project, managers identified a comparison group of 219 replenishment BOSS Black SKUs that were most similar to the 45 treatment SKUs based on factors including type, cut, fabric, and color.

Although we describe the manner in which Hugo Boss increased its service level to retailers, we note that our research design does not depend on Hugo Boss’s approach to increasing service level. Hugo Boss increased its service level for the treatment group of SKUs to retailers with the following

operational change. Prior to the pilot, Hugo Boss production planners ordered the 45 treatment SKUs from Hugo Boss’s factory monthly. To implement the pilot, Hugo Boss production planners began ordering these SKUs from the factory weekly. This change allowed Hugo Boss to provide a higher inventory service level to retailers, since production planners had more opportunities to order SKUs in cases when inventory in its warehouse ran low. While the ordering process within Hugo Boss—i.e., the ordering process between Hugo Boss and its factory—changed for the 45 treatment SKUs, the ordering process between retailers and Hugo Boss did not change. Further, other SKUs were not affected since Hugo Boss altered its operations only for the pilot SKUs.

3.2 Research Design

The pilot project allows us to observe inventory service level and retailer demand both before and after an operational change that increased inventory service level for a subset of SKUs. Our first objective is to determine whether the pilot project had a measurable impact on retailer demand. To do so, we test for differences in retailer demand across treatment and control SKUs upon the implementation of the pilot.

A field experiment to measure the relationship between inventory service level and retailer demand would ideally divide all similar SKUs sold by a supplier randomly into a treatment group and a control group. The experimenter would then change the service level to retailers of the treatment group while holding the service level of the control group constant. All other supplier activities regarding both the treatment and control SKUs—such as pricing and promotion—would also be held constant. Under such a field experiment, a DID model would allow the experimenter to compare changes in demand for the treatment group to changes in demand for the control group. All SKUs would be subject to the same conditions beyond the service level the supplier provides to retailers.

The pilot project we study approximates this ideal field experiment. The treatment group of SKUs receives the treatment of increased service level while the similar SKUs that were not part of the pilot project are a comparison group. An important departure in the case of the pilot project is that the treatment SKUs were not chosen randomly. Instead, these SKUs were produced by the

same factory. Had the pilot SKUs been chosen at random, Hugo Boss would have required buy-in for their operational change from each of its partner factories, thereby impeding the execution of the pilot project. Hugo Boss therefore faced a trade-off common in field-based experimental design, namely that between designing ideal treatment and control groups and managerial expediency (Anderson and Simester, 2011).

One could argue that the treatment group was selected such that retailer demand for these items was growing irrespective of any changes in service level. Thus, any observed change in retailer demand would not be due to the change in service level but rather due to selection bias. To mitigate this concern, we use propensity score matching to select a control group from the broader comparison group. The control group is similar to the treatment group on several dimensions, including retailer demand and trend in retailer demand prior to the pilot project. In so doing, the pilot SKUs and the constructed control group of SKUs should not differ on observables that would be expected to lead to higher demand among the treated SKUs. Hereafter, the phrase *control group* refers to the set of control SKUs chosen through propensity score matching whereas *comparison group* refers to the 219 similar bodywear items identified by managers.

We rely on the qualitative data collected throughout our study to ensure the parallel trends assumption of the DID model is met for factors such as pricing and advertising, since these measures are unavailable to us. Specifically, our field interviews with numerous Hugo Boss managers revealed that Hugo Boss held advertising constant across all BOSS Black SKUs produced by the bodywear division during our observation period. No bodywear SKU received extra advertising or promotion to retailers. Moreover, managers held prices for all bodywear SKUs constant with respect to prices for similar products from competitors. Given that Hugo Boss’s pricing and advertising were not altered dramatically during our period of study, it is unlikely that these factors account for any observed changes in retailer demand for the SKUs we study.

4 Data and Measures

We collected from Hugo Boss weekly observations about the functional bodywear SKUs that compose the treatment and comparison groups. These data cover a total of 148 weeks starting one

year prior to the implementation of the pilot project. The data collected from Hugo Boss include information about retailer demand (orders) and Hugo Boss’s inventory level.

For the treatment group of 45 SKUs and the comparison group of 219 SKUs, we collected the overall quantity ordered by all retailers in Germany at the SKU level on a weekly basis. Let the variable $SKU\ Order\ Quantity_{st}$ be the total quantity of SKU $s \in S$, where S is the set comprising all treatment and comparison SKUs, ordered by all retailers during week $t \in T$. To match the treatment and comparison SKUs, we collected information about the fashion of each SKU. These characteristics include garment type, cut, color, fabric, and size. We describe the matching procedure in §5.

We use the data introduced thus far to match treatment and comparison SKUs and to fit the DID model, which estimates the overall effect of the pilot project. We also construct several further analyses to study the mechanism linking supplier inventory service level and retailer demand. These analyses have different structures than the DID model, since they incorporate information, such as Hugo Boss’s fill rate, calculated over the pre-treatment period. The data we collected from Hugo Boss for use in these analyses include inventory data from Hugo Boss’s warehouse for the treatment and comparison SKUs. The inventory data captures end-of-week inventory at Hugo Boss’s warehouse for each SKU. This information allows us to identify weekly cycles in which the warehouse ran out of a particular SKU. We define the variable $Stockout_{st}$ to take the value 1 if SKU s had an inventory level of zero at the end of week t . Because we are missing inventory observations for a total of eight weeks in our sample, we exclude these weeks when fitting our empirical models. Imputing inventory values for these missing weeks does not affect the results of our analysis.

In §6, we measure the relationship between Hugo Boss’s historical fill rate and demand from retailers. To do so, we define the following notation. Let $Ending\ Inventory_{st}$ be the amount of SKU s in warehouse inventory at the end of cycle t . To calculate fill rate, we must determine the inventory available for sale during the week. Our data does not directly record the beginning inventory for each cycle or replenishment quantities, so we use the following proxy for week t and

SKU s :

$$Beginning\ Inventory_{st} = \begin{cases} Ending\ Inventory_{st} + SKU\ Order\ Quantity_{st} & \text{If } Stockout_{st} = 0 \\ Ending\ Inventory_{s,t-1} & \text{If } Stockout_{st} = 1. \end{cases}$$

We assume that replenishments did not arrive during weeks in which a stockout occurred and will therefore underestimate the actual fill rate if this assumption is violated. To test this assumption, we calculated the replenishment quantity during all non-stockout weeks for each SKU and compared the result to the amount of each SKU ordered by retailers during weeks in which a stockout occurred. This allows us to assess whether it is plausible that no replenishments arrived during weeks in which a SKU was stocked out. The replenishment quantity for SKU s during week t is $Replenishment\ Quantity_{st} = Beginning\ Inventory_{st} - Ending\ Inventory_{s,t-1}$. In all cases, the amount of SKU s ordered by all retailers during a week in which a stockout occurred was less than the minimum non-zero value of $Replenishment\ Quantity_{st}$.

Given these variables, we can calculate fill rate for a given SKU, s , in a week, t . Fill rate across all retailers is as follows:

$$Weekly\ Fill\ Rate_{st} = \frac{\min(SKU\ Order\ Quantity_{st}, Beginning\ Inventory_{st})}{SKU\ Order\ Quantity_{st}}.$$

Fill rate is also defined over aggregate time periods and across multiple SKUs by summing the quantities ordered and filled. Fill rate for the treatment group of SKUs was 96.4% for the year prior to the pilot and 99.6% over the approximately two years after the pilot. Fill rate for the comparison group was 96.6% over the year before the pilot and 96.3% after the pilot. In §6, we calculate fill rate across all retailers as well as at the individual retailer level. We describe the different approaches as we introduce them.

For the treatment group of SKUs, we also collected orders placed by individual retailers to further analyze the retailers' responses to the change in service level. We collected information at the SKU level rather than the retailer level for the comparison group due to the difficulty of collecting the quantities ordered at the retailer level from Hugo Boss's information systems. The

retailer-level data comprise 58,787 orders for 509,800 units from 693 retailers. Many of these retailers are small customers that do not appear regularly throughout our data window. To focus the analyses in §6 on retailers that interact with Hugo Boss regularly, we retain only retailers that place their first order within two months of the start of the data observation window and their last order within two months of the end of the data observation window. This leaves 43,516 orders for 411,415 units from 95 retailers. We thus have a total of 598,500 order observations at the SKU per retailer per week level for the treatment group of SKUs (45 SKUs, 95 retailers, and 140 weeks). We define the variable $Order\ Quantity_{rst}$ to be the quantity of SKU s ordered by retailer $r \in R$, the set of 95 retailers, during week t .

To test whether the response of retailers to changes in service level differs depending on the frequency with which a retailer orders (H3), we calculate each retailer’s order frequency using data generated during the pre-pilot portion of our data observation window. We define a retailer’s order frequency as the ratio of the number of orders the retailer places in those 52 weeks to the total number of opportunities to order in that time frame. Let S^P be the set of 45 treatment SKUs. Then a retailer’s order frequency is

$$Order\ Frequency_r = \sum_{s \in S^P} \sum_{t=1}^{52} \frac{\mathbb{1}[Order\ Quantity_{rst} > 0]}{52|S|}.$$

For instance, over the course of 52 weeks, a retailer has 2,340 (52×45) opportunities to order any of the treatment group SKUs. A retailer that places 1,404 orders over that time frame would have an order frequency of 0.6.

When fitting models that incorporate historical fill rate in §6, we use the pre-pilot data to initialize the fill rate variable. Therefore, we do not rely on the before-and-after approach of the DID model to control for plausible drivers of changes in retailer demand beyond service level. Instead, we explicitly control for the broader economic conditions faced by retailers by incorporating an index obtained from the Bundesbank Statistical Office in our analysis (Deutsche Bundesbank, 2015). This index tracks retail sales excluding cars on a monthly basis. The retail sales index is normalized so that the level of retail sales in January 2003 equals 1. Let $Retail\ Sales\ Index_t$ be the

value of this index during week t . We control for the popularity of Hugo Boss bodywear products using orders for the comparison group of SKUs. Let *Comparison Group Order Quantity* $_t$ be the total orders for all comparison group SKUs during week t . We rescale comparison group orders by dividing the variable by its mean. Table 1 contains summary statistics for the variables introduced in this section.

5 Overall Effect of the Pilot Project

We begin our empirical analysis by estimating the overall effect of the pilot project to establish its impact on retailer demand. To do so, we use a DID approach with a control group constructed through propensity score matching. The propensity score matching approach creates a control group that is similar to the treatment group based on observable characteristics of the SKUs.

We collected observables regarding both demand and product fashion qualities to perform matching. The fashion observables include type, cut, fabric, color, and size. Our objective is to ensure that the control group SKUs do not differ from the treatment group SKUs on factors that could influence retailer demand. We match on pre-treatment demand measured in two ways: average weekly orders during the pre-pilot period and linear time trend in weekly orders during the pre-pilot period. We estimate the linear time trend in demand using ordinary least squares regression on deseasonalized demand data generated as follows. We divide each SKU's weekly pre-pilot order quantity, *SKU Order Quantity* $_{st}$, by its average value over the pre-pilot period. We then divide the result by the average demand across all SKUs in each pre-pilot week.

We perform nearest-neighbor matching without replacement using the PSMATCH2 module in Stata. We created control groups matched when including all observables, fashion and demand, in different combinations. We utilize the control group matched on only the two demand variables, average weekly orders and linear time trend in orders, as this control group yields the most conservative estimate within the DID analysis (i.e., the smallest estimate of the pilot's effect on demand from retailers).

Propensity score matching resulted in a total of 86 SKUs: 43 from the treatment group of 45 SKUs and 43 from the comparison group of SKUs. This yields an observation count of 12,040 (86

SKUs and 140 weeks). The mean of weekly orders during the pre-pilot period is 70.22 with standard deviation 81.72 for the matched treatment SKUs and 61.93 with standard deviation 71.33 for the control SKUs. The mean weekly linear time trend in demand is 0.02 units per week with standard deviation 0.31 for the matched treatment SKUs and 0.05 units per week with standard deviation 0.42 for the control SKUs. Hence, both average demand and time trend in demand are balanced across the treatment and control groups during the pre-pilot period. Table 2 presents demand information for the treatment, comparison, and control groups before the pilot. To further validate the control group, we used the regression framework introduced below to conduct a placebo test during the pre-pilot period and to fit a separate time-trend model for the control and treatment groups over the pre-pilot period. Both validation tests supported the parallel trends assumption of the DID model.

Fill rate for the control group was 96.5% over the year before the pilot and 96.4% for the approximately two years after the pilot. In comparison, fill rate for the treatment group was 96.4% over the year prior to the pilot and 99.6% after the pilot. Figure 1 plots weekly fill rate for the treatment and control SKUs before and after the pilot.

Given the count nature of the data, we use a Poisson panel regression framework for all empirical models (Wooldridge, 1999). Further, although Wooldridge (1999) finds that the Poisson method generates consistent estimates when the dependent variable is overdispersed relative to the Poisson distribution, we also fit negative binomial and linear models as a robustness test. We observe no substantive difference in the magnitude or the statistical significance of the coefficients across these models.

The dependent variable for the DID model is the total quantity ordered by all retailers for a given SKU in week t , $SKU\ Order\ Quantity_{st}$. The treatment indicator is $Treatment\ Indicator_{st}$, which takes the value 1 if SKU s is in the treatment group and week t is after the pilot went into effect. The model for estimating the effect of the pilot project is:

$$\begin{aligned} \ln [E(SKU\ Order\ Quantity_{st})] = & \alpha + \beta \times Treatment\ Indicator_{st} + SKU\ Fixed\ Effect_s \\ & + Week-of-Year\ Fixed\ Effect_t + Year\ Fixed\ Effect_t. \end{aligned} \quad (1)$$

To fit all models presented herein, we use the conditional fixed effects Poisson command implemented by Stata 13. We employ cluster-robust errors clustered by SKU (Wooldridge, 2002).

Model 1 in Table 3 contains the results when fitting the model using only the fixed effects. Model 2 in Table 3 presents the results of fitting the full model in Equation (1). We observe that the coefficient on the pilot indicator variable, β , is positive and significant ($z = 40.10$). Models 3 and 4 in Table 3 present results when using the full comparison group instead of the matched control group.

We find that a treated SKU experiences approximately a 17% increase in orders from all retailers as a result of the pilot project. In support of H1A, we conclude that the increase in service level due to the pilot project had a statistically significant, positive, and substantial effect on retailer demand. Given this result, we explore the mechanism linking supplier inventory service level and retailer demand.

6 Drivers of the Increase in Demand

While our analysis thus far establishes the effect of the pilot project on treated SKUs, our data allow us to further explore drivers of the observed effect. First, to relate our study to prior research and practitioner metrics, we estimate the relationship between historical fill rate and retailer demand. Through this analysis, we identify distinct long- and short-term effects of service level. Second, to test H3, we assess whether retailers that collect more information about Hugo Boss’s reliability by placing more frequent orders exhibit a different response to a change in fill rate than retailers that place less frequent orders. Third, we assess the degree to which the increase in demand from retailers is due to changes in assortment, changes in order frequency, and changes in order quantity on a constant order cycle for products already in assortment.

6.1 Fill Rate and Retailer Demand

We measure the relationship between Hugo Boss’s historical fill rate and current retailer demand by calculating fill rate both across all retailers and at the retailer level. Prior research on unreliable suppliers employs the distribution of a supplier’s service level. Inventory service level calculated

across all retailers incorporates all realizations of Hugo Boss’s performance and thus best represents the distribution of the supplier’s service level. Further, we study fill rate calculated across all retailers because service level targets are often set and managed across all customers (Slone, 2004). For robustness, we also develop and test proxies for fill rate measured at the individual retailer level.

We calculate annual historical service level across all retailers using the following metric.

$$\text{Annual Aggregate Fill Rate}_{st} = \frac{\sum_{s \in SP} \sum_{i=t-51}^t \min(\text{SKU Order Quantity}_{si}, \text{Beginning Inventory}_{si})}{\sum_{s \in SP} \sum_{j=t-51}^t \text{SKU Order Quantity}_{sj}}$$

From week 52 of the data onward, $\text{Annual Aggregate Fill Rate}_{st}$ has mean 0.984 and standard deviation 0.015.

We calculate fill rate at the product line level for three reasons. First, our interviews with buyers indicate that they form beliefs at the product line level. Buyers stated that they were unlikely to discern differences in service level between two items within a product line (e.g., two colors of the same shirt) but would be able to track differences between two product lines (e.g., bodywear for the business segment versus bodywear for the active segment). Second, our discussions with suppliers indicate that any project designed to improve the availability of a single SKU would most likely improve the availability of other SKUs within a line. Finally, calculating fill rate at the product line level has the advantage of allowing us to employ less data when calculating historical fill rate. For robustness, we fit models with fill rate calculated at the SKU level, which yielded results consistent with those presented herein. The results of the SKU-level fill rate models are found in Appendix A.

Calculating annual aggregate fill rate requires the use of all data generated prior to the pilot implementation. We must therefore develop a model that controls explicitly for plausible drivers of retailer demand, since we no longer can rely on the before-and-after structure of the DID model. To measure the effect of changes in inventory service level metrics on retailer demand, we use the following base model, which includes only controls. The dependent variable for this analysis is the

order quantity of retailer r for SKU s in week t .

$$\begin{aligned}
\ln [E(\text{Order Quantity}_{rst})] &= \alpha + \delta_1 \times \text{Comparison Group Order Quantity}_t \\
&+ \delta_2 \times \text{Retail Sales Index}_t + \text{Retailer Fixed Effect}_r \\
&+ \text{SKU Fixed Effect}_s + \text{Week-of-Year Fixed Effect}_t \\
&+ \text{Year Fixed Effect}_t,
\end{aligned} \tag{2}$$

where α is a fixed intercept. In this model, the orders from the comparison group control for the popularity of Hugo Boss bodywear products while the retail sales index controls for the broader retail climate. The results presented herein do not change significantly when using control group (see §5) orders to account for Hugo Boss’s popularity instead of comparison group orders. For brevity, we refer to the above control variables and their coefficients as $\boldsymbol{\delta}'\mathbf{x}_{rst}$.

We incorporate service level using the following model:

$$\begin{aligned}
\ln [E(\text{Order Quantity}_{rst})] &= \alpha + \beta \times \text{Annual Aggregate Fill Rate}_{s,t-2} \\
&+ \gamma \times \text{Stockout}_{s,t-1} + \boldsymbol{\delta}'\mathbf{x}_{rst}.
\end{aligned} \tag{3}$$

We lag *Annual Aggregate Fill Rate* by two weeks to avoid conflating the immediate effects of a stockout with the long-term effects. We include the stockout indicator to measure retailers’ immediate reaction to a stockout, testing H2. Given the lag and the fact that the historical fill rate measure requires one year of data, we omit data from the first 54 weeks of our sample when estimating this model, leaving a total of 380,475 observations (45 SKUs, 95 retailers, and 89 weeks).

Models 1 through 3 in Table 4 present the results of fitting Equation (2). Models 4 and 5 in Table 4 present the results of estimating Equation (3). The results support H1A, namely, the coefficient for aggregate fill rate (Model 4, Table 4) is positive and significant ($z = 7.62$). Increasing annual aggregate fill rate by one percentage point yields an 11 % increase in a retailer’s demand. We find that the overall impact of the increase in annual aggregate fill rate from week 52 onward, calculated by assuming annual aggregate fill rate remains constant at its value at the start of the pilot project and predicting the resulting retailer orders, is a demand increase of 21 %, consistent

with the results of §5. The coefficient for the lagged stockout indicator is positive and significant ($z = 4.67$), supporting H2, which predicts that recent stockouts will increase retailer demand in the short term.

One question of interest is whether the effect of annual aggregate fill rate persists over time. We examined this issue by fitting Equation (3) with annual aggregate fill rate lagged for time intervals longer than two weeks. We tested lags of 4, 8, 13, and 26 weeks. The coefficient for annual aggregate fill rate continued to be positive and significant for each of these lags.

For robustness, we also calculate fill rate at the retailer level. The warehouse we study did not record fulfillment data regarding which retailers received product in the event of a stockout. Thus, we are unable to directly calculate retailer-level fill rate. Instead, we calculate three proxies for fill rate using three different assumptions. The first assumption is that Hugo Boss followed its stated policy of filling orders on a first-come, first-served basis. The second assumption is that Hugo Boss filled smaller orders first, e.g., to maximize the number of retailers that receive their full order. The third assumption is that Hugo Boss filled larger orders first, e.g., to satisfy larger customers. We introduce three variables to represent the amount of SKU s delivered to retailer r during week t under each assumption: $Quantity\ Delivered_{rst}^{FCFS}$, $Quantity\ Delivered_{rst}^{Small}$, $Quantity\ Delivered_{rst}^{Large}$. These variables are generated by distributing the initial inventory, $Beginning\ Inventory_{st}$, to each retailer based on the respective assumptions. The retailer-specific fill rate is

$$Annual\ Retailer\ Fill\ Rate_{rt}^{\Theta} = \frac{\sum_{s \in SP} \sum_{i=t-51}^{i=t} Quantity\ Delivered_{rsi}^{\Theta}}{\sum_{s \in SP} \sum_{j=t-51}^{j=t} Order\ Quantity_{rsj}},$$

where $\Theta \in (FCFS, Small, Large)$. From week 52 of our data onward, $Annual\ Retailer\ Fill\ Rate_{rst}^{FCFS}$ has mean 0.989 and standard deviation 0.026. $Annual\ Retailer\ Fill\ Rate_{rst}^{Small}$ has mean 0.992 and standard deviation 0.019, and $Annual\ Retailer\ Fill\ Rate_{rst}^{Large}$ has mean 0.988 and standard deviation 0.027. Table 5 presents the estimation of Equation (3) when replacing $Annual\ Aggregate\ Fill\ Rate$ with each of the retailer-specific fill rate metrics. The coefficients of Models 1, 2, and 3 reveal support for H1A among all three retailer-level measures as well as continued support for H2.

6.2 Differences in Demand Responses to Fill Rate Changes Across Retailers

Prior literature predicts that a retailer’s response to a change in supplier inventory service level will depend on the amount of information the retailer collects regarding the supplier’s service performance (H3). To test this prediction, we add the interaction of the service level metric with the retailer’s order frequency, calculated over the pre-pilot period, to Equation (3):

$$\begin{aligned} \ln [E(\textit{Order Quantity}_{rst})] = & \\ & \alpha + \beta_1 \textit{Annual Aggregate Fill Rate}_{s,t-2} + \beta_2 \textit{Order Frequency}_r \\ & + \beta_3 \textit{Annual Aggregate Fill Rate}_{s,t-2} \times \textit{Order Frequency}_r \\ & + \gamma \times \textit{Stockout}_{s,t-1} + \boldsymbol{\delta}' \mathbf{x}_{rst}. \end{aligned} \tag{4}$$

We note that the main effect of a retailer’s order frequency (β_2) is subsumed into the retailer fixed effect and not directly estimable. Nonetheless, we only need to estimate the interaction to test H3. Models 6 and 7 of Table 4 present the estimation of Equation (4).

In support of H3, we find the coefficient on the interaction between annual aggregate fill rate and order frequency to be positive and significant ($z = 2.22$). We observe differential responses across retailers to changes in supplier fill rate. Specifically, a retailer with an order frequency in the 85th percentile would be expected to exhibit a 23% larger increase in order quantity in response to a single percentage point increase in annual aggregate fill rate than a retailer with an order frequency in the 15th percentile, controlling for all other variables at their means. For robustness, we also tested the interaction of order frequency and the retailer-level fill rate metrics. Models 5 and 6 of Table 5 present the coefficients for the retailer-level fill rate measures, in continued support of H3.

6.3 Components of Demand Increase

The increase in retailer demand we observe for the treatment group may be due to different responses from individual retailers. Retailers may change their assortment, increase their order frequency while ordering the same quantities of each item, and/or change order quantities while

maintaining a constant order cycle for products already in assortment. In this section, we assess the degree to which each of these potential responses explains the increase in retailer demand we observe.

First, we explore whether retailers ordered more SKUs from the treatment group of 45 SKUs after the increase in service level. For the constant set of 95 retailers, the mean number of SKUs ordered from the treatment group during the year prior to the pilot is 19.35 with standard deviation 12.05. After the treatment, the mean number of SKUs ordered is 23.20 with standard deviation 11.33. The mean difference (number of SKUs ordered after less number of SKUs ordered before) is 3.75 with standard deviation 7.13. A paired t -test finds that retailers increased their assortment from the group of 45 treatment SKUs ($p < 0.001$). To assess the extent to which this change in assortment accounts for the observed change in demand, we use the following procedure. At the retailer-SKU level, we define a SKU as new to a retailer's assortment if the retailer did not order that SKU before the treatment and did order the SKU after the treatment. After the pilot implementation, new SKUs account for 1.2% of retailer demand. Taking as a baseline the 17% increase in demand estimated by the DID model, we find that new SKUs explain 7.1% of the overall increase in demand after the pilot.

We next test whether retailers changed their order frequency after Hugo Boss's increase in service level. To do so, we compare retailers' order frequencies before and after the pilot project. For the constant set of 95 retailers, the mean order frequency during the year prior to the pilot is 0.07 with a standard deviation of 0.13. During the approximately two years after the pilot, the mean order frequency is also 0.07 with a standard deviation of 0.14. Hypothesis testing using a paired t -test to determine if there is a difference before versus after the pilot yields a p -value of 0.65. We therefore conclude that retailer order frequency did not change after the pilot implementation. This suggests that retailers did not respond to the increase in service level by changing their order cycles.

Based on these results, we conclude that retailers increased their demand following Hugo Boss's pilot project by increasing order quantity for products already in assortment while holding their order cycles constant.

7 Discussion

Our results demonstrate the existence of a statistically and managerially significant relationship between supplier inventory service level and retailer demand in the supply chain for functional apparel items. We find a positive relationship between historical fill rate and current retailer demand, and we observe that this relationship is larger for retailers that order more frequently. Further, we find that the relationship between service level and retailer demand in the context we study is primarily driven by Hugo Boss’s retailer customers placing larger orders, rather than by changes in retailer assortment or order cycles.

Our empirical research augments the body of work on retailers ordering from unreliable suppliers and answers the call for field-based evidence regarding the presence and strength of the relationship between supplier service level and retailer demand (Gurnani et al., 2013). We examine the behavior of retail buyers—employees that make purchasing decisions on behalf of the retailer—in practice and argue that retail buyers are an important driver of these results. As Bonoma (1982) notes (p. 4), “companies don’t buy, people do.” And, unlike the procurement professionals studied in prior laboratory experiments who respond to observations of downstream demand, retail buyers in practice can influence downstream demand. Retail buyers do so through their choice of which items to promote, display, and advocate internally. Buyers make these decisions in the face of pre-established product category performance targets as well as other incentives, including reputational risks and the costs of effort associated with managing supplier unreliability.

Our findings complement survey-based research that has shown retailers that trust suppliers are 12% more committed to the relationship between the two firms (measured by intent to carry the supplier’s products in the future) and 22% less likely to develop alternate sources of supply (Kumar, 1996). Moreover, our empirical approach allows us to determine the value of such trust. In the case of Hugo Boss, the average margin for products affected by the pilot project was €4.89 during the time of our study, which we calculated by dividing the total margin for the treatment group during the period of our study (i.e., total revenue minus total cost of goods sold) by the number of units sold. Using the overall effect of the pilot project calculated in §5, we estimate that the pilot project yielded an additional €300,002, or 29.9%, in margin. This increase in margin

substantially exceeds the cost of implementing the pilot project (Raman et al., 2009). Further, our findings demonstrate that an increase in service level can substantially increase orders when service level is already high. Therefore, managers who ignore the possibility of driving orders with service level increases will forego potentially profitable supply chain improvements as well as opportunities to capture market share.

7.1 Limitations and Future Research

Our research is not without limitations. The focus of our study is the supply chain for functional apparel items, and thus we are unable to generalize our findings to other supply chains. Further research is necessary to understand the effects of service level in other B2B contexts.

Further research is needed to better understand the role of retail buyers as discretionary decision makers within a supply chain. We observe a magnitude for the relationship between supplier inventory service level and retailer demand that exceeds the magnitudes measured in prior research involving end consumers (Anderson et al., 2006) and laboratory experiment participants (Gurnani et al., 2013). The magnitude of the relationship we measure reflects, at least in part, the influence of the retail buyer. Studying retail buyer behavior in more depth utilizing a variety of research approaches will improve our understanding of buyers' influence on supply chain management.

Other retailer-specific characteristics, in addition to order frequency, may influence the magnitude of a retailer's response to service level changes. Future research could explore the drivers of such differences across retailers. It is plausible, for example, that the relationship between demand and service level could be influenced by factors such as the portion of the retailer's assortment dedicated to a given supplier or the competitiveness of the channel within which the retail chain operates. Finally, researchers could build on our empirical finding that frequent orderers exhibit behaviors that differ from non-frequent orderers. Incorporating this observed differential effect into standard inventory models could lead to alternative results, for example, shorter order cycles in EOQ models with unreliable suppliers.

References

- Anderson, E. T., Fitzsimons, G. J. and Simester, D. (2006), ‘Measuring and mitigating the costs of stockouts’, *Management Science* **52**(11), 1751–1763.
- Anderson, E. T. and Simester, D. (2011), ‘A step-by-step guide to smart business experiments’, *Harvard Business Review* **89**(3), 98–105.
- Angrist, J. D. and Pischke, J.-S. (2009), *Mostly Harmless Econometrics*, Princeton University Press, Princeton, New Jersey.
- Anupindi, R. and Akella, R. (1993), ‘Diversification under supply uncertainty’, *Management Science* **39**(8), 944–963.
- Balakrishnan, A., Pangburn, M. S. and Stavrulaki, E. (2004), “‘Stack them high, let ’em fly’: Lot-sizing policies when inventories stimulate demand”, *Management Science* **50**(5), 630–644.
- Balakrishnan, A., Pangburn, M. S. and Stavrulaki, E. (2008), ‘Integrating the promotional and service roles of retail inventories’, *Manufacturing & Service Operations Management* **10**(2), 218 – 235.
- Bonoma, T. V. (1982), ‘Major sales: Who really does the buying?’, *Harvard Business Review* **60**(3), 111–119.
- Burke, G. J., Carrillo, J. E. and Vakharia, A. J. (2009), ‘Sourcing decisions with stochastic supplier reliability and stochastic demand’, *Production and Operations Management* **18**(4), 475–484.
- Cachon, G. P. and Lariviere, M. A. (1999), ‘Capacity allocation using past sales: When to turn-and-earn’, *Management Science* **45**(5), 685–703.
- Che, H., Chen, X. and Chen, Y. (2012), ‘Investigating effects of out-of-stock on consumer stock-keeping unit choice’, *Journal of Marketing Research* **49**, 502–513.
- Chen, M., Xia, Y. and Wang, X. (2010), ‘Managing supply uncertainties through bayesian information update’, *IEEE Transactions on Automation Science and Engineering* **7**(1), 24–36.
- Dada, M., Petruzzi, N. C. and Schwarz, L. B. (2007), ‘A newsvendor’s procurement problem when suppliers are unreliable’, *Manufacturing & Service Operations Management* **9**(1), 9 – 32.
- DeHoratius, N. (2014), ‘Global Sourcing at Walmart’. Chicago Booth Case.
- Deutsche Bundesbank (2015), ‘Business statistics’, <http://www.bundesbank.de/>. Accessed April 2015.
- Emmelhainz, L. W., Emmelhainz, M. A. and Stock, J. R. (1991), ‘Logistics implications of retail stockouts’, *Journal of Business Logistics* **12**(2), 129–142.
- Federgruen, A. and Yang, N. (2009), ‘Optimal supply diversification under general supply risks’, *Operations Research* **57**(6), 1451–1468.
- Fisher, M. L. (1997), ‘What is the right supply chain for your product?’, *Harvard Business Review* **75**(2), 105–116.

- Fitzsimons, G. J. (2000), ‘Consumer response to stockouts’, *The Journal of Consumer Research* **27**(2), 249–266.
- Gallino, S., Cachon, G. P. and Olivares, M. (2013), ‘Does Adding Inventory Increase Sales? Evidence of a Scarcity Effect in U.S. Automobile Dealerships’. Working Paper.
- Gerchak, Y. and Parlar, M. (1990), ‘Yield randomness, cost trade-offs, and diversification in the EOQ model’, *Naval Research Logistics* **37**(3), 341–354.
- Gurnani, H., Ramachandran, K., Ray, S. and Xia, Y. (2013), ‘Ordering behavior under supply risk: An experimental investigation’, *Manufacturing & Service Operations Management* **16**(1), 61–75.
- Heim, G. R. and Sinha, K. K. (2001), ‘Operational drivers of customer loyalty in electronic retailing: An empirical analysis of electronic food retailers’, *Manufacturing & Service Operations Management* **3**(3), 264.
- Ho, T.-H., Lim, N. and Cui, T. H. (2010), ‘Reference dependence in multilocation newsvendor models: A structural analysis’, *Management Science* **56**(11), 1891–1910.
- Hopp, W. J., Iravani, S. M. R. and Liu, F. (2009), ‘Managing white-collar work: An operations-oriented survey’, *Production and Operations Management* **18**(1), 1–32.
- Kumar, N. (1996), ‘The power of trust in manufacturer-retailer relationships’, *Harvard Business Review* **74**(6), 92–106.
- Lee, H. L., Padmanabhan, V. and Whang, S. (1997), ‘Information distortion in a supply chain: The bullwhip effect’, *Management Science* **43**(4), 546–558.
- Liu, S., So, K. C. and Zhang, F. (2009), ‘Effect of supply reliability in a retail setting with joint marketing and inventory decisions’, *Manufacturing & Service Operations Management* **12**(1), 19–32.
- Malmbak, V. and Albaum, M. (2007), ‘Market and brand analysis’, *TextilWirtschaft* .
- McKinnon, A. C., Mendes, D. and Nababteh, M. (2007), ‘In-store logistics: An analysis of on-shelf availability and stockout responses for three product groups’, *International Journal of Logistics: Research & Applications* **10**(3), 251 – 268.
- Nagarajan, M. and Shechter, S. (2013), ‘Prospect theory and the newsvendor problem’, *Management Science* **60**(4), 1057–1062.
- Nahmias, S. (2008), *Production and Operations Analysis*, 6 edn, McGraw-Hill, New York, NY.
- Raman, A., DeHoratius, N. and Kanji, Z. (2009), ‘Supply Chain Optimization at Hugo Boss (A)’. Harvard Business School Case Number 9-609-029.
- Rosenbaum, P. and Rubin, D. (1983), ‘The central role of the propensity score in observational studies for causal effects’, *Biometrika* **70**(1), 41–55.
- Shih, W. (1980), ‘Optimal inventory policies when stockouts result from defective products’, *International Journal of Production Research* **18**(6), 677–686.

- Silver, E. A. (1976), ‘Establishing the order quantity when the amount received is uncertain’, *Infor* **14**(1), 32–39.
- Slone, R. E. (2004), ‘Leading a supply chain turnaround’, *Harvard Business Review* **82**(10), 114–121.
- Stock, A. and Balachander, S. (2005), ‘The making of a “hot product”: A signaling explanation of marketers’ scarcity strategy’, *Management Science* **51**(8), 1181–1192.
- Swait, J. and Erdem, T. (2002), ‘The effects of temporal consistency of sales promotions and availability on consumer choice behavior’, *Journal of Marketing Research* **39**(3), 304–320.
- Tomlin, B. (2009), ‘Impact of supply learning when suppliers are unreliable’, *Manufacturing & Service Operations Management* **11**(2), 192–209.
- van Donselaar, K. H., Gaur, V., van Woensel, T., Broekmeulen, R. A. C. M. and Fransoo, J. C. (2010), ‘Ordering behavior in retail stores and implications for automated replenishment’, *Management Science* **56**(5), 766–784.
- Wooldridge, J. M. (1999), ‘Distribution-free estimation of some nonlinear panel data models’, *Journal of Econometrics* **90**(1), 77–97.
- Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.
- Yang, Z., Aydın, G., Babich, V. and Beil, D. R. (2012), ‘Using a dual-sourcing option in the presence of asymmetric information about supplier reliability: Competition vs. diversification’, *Manufacturing & Service Operations Management* **14**(2), 202–217.
- Yano, C. A. and Lee, H. L. (1995), ‘Lot sizing with random yields: A review’, *Operations Research* **43**(2), 311–334.

Figures and Tables

Table 1: Summary Statistics

Variable	Mean	S.D.	Minimum	Maximum
SKU Order Quantity	55.305	86.562	0.000	985.000
Weekly Fill Rate	0.981	0.070	0.000	1.000
Retailer Order Quantity	0.698	6.339	0.000	572.000
Order Frequency	0.067	0.131	0.002	0.784
Stockout Indicator	0.004	0.063	0.000	1.000
Comparison Group Order Quantity	1.000	0.468	0.532	2.901
Retail Sales Index	1.010	0.083	0.868	1.237

SKU Order Quantity is the total quantity ordered by all retailers for a specific SKU in a given week. Retailer Order Quantity is the order quantity from a single retailer for a specific SKU in a give week. See §4 for definitions of the remaining variables.

Table 2: Demand Information for Matched Treatment, Comparison, and Control SKUs Prior to the Pilot Project

Group	Number of SKUs	SKU Order Quantity		Time Trend in Demand	
		Mean	S.D.	Mean	S.D.
Matched Treatment	43	70.22	81.72	0.02	0.31
Comparison	219	49.31	77.48	0.06	0.53
Control	43	61.93	71.33	0.05	0.42

All variables are measured over the pre-pilot period. SKU Order Quantity is measured in units per week and is the quantity ordered by all retailers for a particular SKU in a given week. Time trend in demand is described in §5.

Figure 1: Weekly Fill Rate Aggregated Across Treatment and Control Groups of SKUs

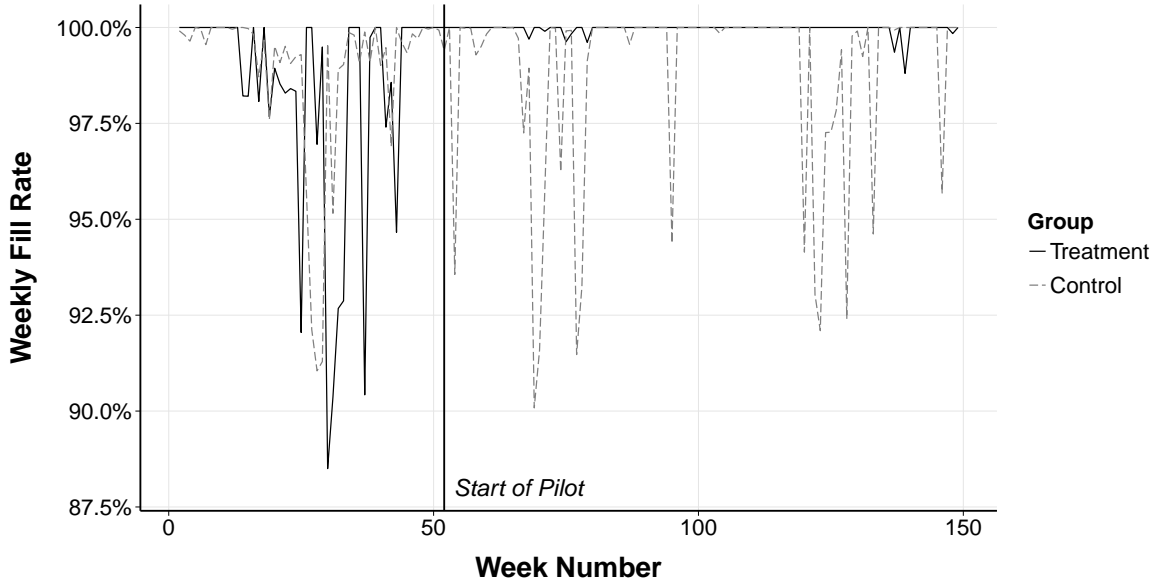


Table 3: Estimating the Overall Effect of the Pilot Project on Retailer Demand

Variable	<i>With Matching</i>		<i>Without Matching</i>	
	Model 1	Model 2	Model 3	Model 4
Treatment Indicator		.17*** (.00)		.30*** (.00)
SKU Fixed Effects	Yes	Yes	Yes	Yes
Week-of-Year Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
<i>N</i>	12,040	12,040	36,960	36,960
Log Likelihood	-207,582.12	-206,774.89	-568,593.38	-564,456.65

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable is quantity ordered per SKU from all retailers per week. Standard errors are in parentheses.

Table 4: Estimating the Effect of Annual Aggregate Fill Rate and Order Frequency on Retailer Demand

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Annual Aggregate Fill Rate				11.05*** (1.45)	9.76*** (1.56)
Annual Aggregate Fill Rate \times Order Frequency					2.63* (1.19)
Stockout Indicator			.16*** (.03)	.14*** (.03)	.14*** (.03)
Comparison Group Order Quantity		.42*** (.01)	.42*** (.01)	.45*** (.01)	.45*** (.01)
Retail Sales Index		-8.34*** (.30)	-8.37*** (.30)	-7.99*** (.30)	-7.99*** (.30)
Retailer Fixed Effects	Yes	Yes	Yes	Yes	Yes
SKU Fixed Effects	Yes	Yes	Yes	Yes	Yes
Week-of-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
N	380,475	380,475	380,475	380,475	380,475
Log Likelihood	-334,288.25	-331,891.87	-331,878.39	-331,740.61	-331,702.69

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable is quantity ordered per SKU per retailer per week. Standard errors are in parentheses.

Table 5: Estimating the Effect of Retailer-Level Fill Rate and Order Frequency on Retailer Demand

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Retailer Fill Rate FCFS	5.07*** (.59)			3.06*** (.66)		
Retailer Fill Rate FCFS \times Order Frequency				18.10*** (2.60)		
Retailer Fill Rate Small		5.05*** (.61)			2.79*** (.67)	
Retailer Fill Rate Small \times Order Frequency					19.33*** (2.44)	
Retailer Fill Rate Large			5.18*** (.58)			3.05*** (.69)
Retailer Fill Rate Large \times Order Frequency						19.42*** (3.36)
Stockout Indicator	.16*** (.03)	.16*** (.03)	.16*** (.03)	.16*** (.03)	.16*** (.03)	.16*** (.03)
Comparison Group Order Quantity	.42*** (.01)	.42*** (.01)	.42*** (.01)	.43*** (.01)	.43*** (.01)	.42*** (.01)
Retail Sales Index	-8.30*** (.30)	-8.31*** (.30)	-8.31*** (.30)	-8.21*** (.30)	-8.21*** (.30)	-8.25*** (.30)
Retailer Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
SKU Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Week-of-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	380,475	380,475	380,475	380,475	380,475	380,475
Log Likelihood	-331,842.00	-331,843.58	-331,839.03	-331,817.79	-331,812.13	-331,822.24

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable is quantity ordered per SKU per retailer per week. Standard errors are in parentheses.

A SKU-Level Fill Rate

In this appendix, we present the results when using SKU-level fill rate to estimate the models introduced in §6.1 and §6.2. We define the annual fill rate for SKU s in the pilot group, S^P , as follows:

$$\text{Annual SKU Fill Rate}_{st} = \frac{\sum_{i=t-51}^t \min(\text{SKU Order Quantity}_{si}, \text{Beginning Inventory}_{si})}{\sum_{j=t-51}^t \text{SKU Order Quantity}_{sj}}.$$

We employ the empirical models of §6.1 and §6.2 while replacing the annual aggregate fill rate measure with SKU-level fill rate. Model 1 in Table A.1 replicates the analysis of §6.1, which estimates the relationship between supplier fill rate and retailer demand. Model 2 in Table A.1 replicates the analysis of §6.2 by incorporating the retailer order frequency variable.

Table A.1: Estimating the Effect of Annual SKU-Level Fill Rate and Order Frequency on Retailer Demand

Variable	Model 1	Model 2
SKU Annual Fill Rate	13.30*** (.40)	10.38*** (.46)
SKU Annual Fill Rate \times Order Frequency		3.11*** (.48)
Stockout Indicator	.15*** (.03)	.15*** (.03)
Non-Pilot Sales Quantity	.42*** (.01)	.42*** (.01)
Retail Index	-8.29*** (.30)	-8.29*** (.30)
Retailer Fixed Effects	Yes	Yes
SKU Fixed Effects	Yes	Yes
Week-of-Year Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
N	380,475	380,475
Log Likelihood	-331,791.27	-331,710.64

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dependent variable is quantity ordered per SKU per retailer per week. Standard errors are in parentheses.