# The Impact of Supplier Inventory Service Level on Retailer Demand in the Supply Chain for Functional Apparel Items

Nathan Craig[1], Nicole DeHoratius[2], and Ananth Raman[3]

[1]The Ohio State University

[2]University of Chicago

[3]Harvard Business School

August 12, 2014

**Abstract**

To set inventory service levels, firms must understand how changes in inventory service level affect customer demand. While the effects of service level changes have been studied empirically at the level of the end consumer, relatively little is known about the interaction between a retailer and a supplier. Using data from a supplier of branded apparel, we show increases in inventory service level to be associated with statistically significant and substantial increases in retailer orders (i.e., demand, not just sales). Controlling for other factors that might affect demand, we find a 1 percent increase in historical inventory service level to be associated with a 13 percent increase in demand from retailers, where historical service level is the type 1 service level performance of the apparel manufacturer over the prior year. Further, retailers that order frequently exhibit a larger reaction to changes in service level, an outcome that is consistent with retailers learning about and reacting to changes in supplier inventory service level. Our study not only provides the first empirical evidence of the impact of changes in service level on demand from retailers but also illustrates a method for estimating this relationship in practice.

# 1 Introduction

To set inventory service levels, suppliers must understand how changes in inventory service level affect retailer demand. While the effects of service level[1] changes have been studied empirically in business-to-consumer settings, i.e., at the level of the end consumer (Fitzsimons, 2000; Swait and Erdem, 2002; Anderson et al., 2006; Gallino et al., 2013), there is little empirical knowledge about business-to-business relationships, such as the interaction between a supplier and a retailer. Analytical models that examine the impact of service level in this latter context predict that increases in a supplier's service level will result in either an increase or no change in the amount a retailer orders from the supplier when there are multiple suppliers selling identical products (Gerchak and Parlar, 1990; Anupindi and Akella, 1993; Dada et al., 2007; Federgruen and Yang, 2009).

In this paper, we estimate the relationship between supplier service level and retailer demand within the supply chain for functional apparel products.[2] This is a supply chain in which multiple unreliable suppliers sell imperfect substitutes to retailers, a context that, to the best of our knowledge, has yet to be explored. To conduct this study, we collected proprietary data from the bodywear division of Hugo Boss, a major supplier of branded apparel. We augmented our quantitative data collection efforts with field visits to the headquarters of Hugo Boss where we interviewed executives and team members responsible for all operational aspects of the bodywear division. The quantitative data include all orders placed by retailers in Germany for bodywear stock keeping units (SKUs) over the course of 148 weeks. Bodywear SKUs consist of socks, t-shirts, and undergarments of various sizes, colors, and fabrics. During our data observation window, Hugo Boss implemented a pilot program that increased the service level of a subset of its bodywear products. This change allows us to measure the effect of Hugo Boss's inventory service level on orders from retailers.

We find improvements in historical supplier service level, which we measure using a variety of common metrics, to be associated with substantial increases in retailer demand, where demand is a retailer's average weekly order quantity (i.e., the number of units requested) for a SKU. For

[1] We use service level to refer specifically to inventory service level.
[2] We use the term functional to refer to replenishment products with long life cycles, as in Fisher (1997).

example, we find a 1 percentage point increase in Hugo Boss's trailing 52-week aggregate (i.e., measured across all retailers) type 1 service level[3] to be associated with a statistically significant 13 % increase in demand among retailers, controlling for other plausible drivers of 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 cases, was approximately a 1 % increase in demand associated with a 1 percentage point increase in type 1 service level. These results persist for other measures of historical supplier service level, including proxies for type 2 service level[4] measured at the aggregate and retailer levels.

This observed empirical link between historical supplier service level and current retail demand demonstrates that retail buyers[5] track and react substantially to changes in a supplier's service level. Prior research on retailers purchasing from unreliable suppliers assumes buyers' decisions are driven by the fulfillment of exogenous end consumer demand. Our field research reveals that retailer demand for a supplier's product can be influenced independently of the end consumer through the actions of the retail buyer.

Supplier stockouts do not need to directly impact the end consumer in order to impact the retail buyer. 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. Moreover, retail buyers can influence end consumer demand in a number of ways. Buyers can determine the amount of marketing a product receives (Liu et al., 2009) and, in the store, can select which items to promote, display, offer additional floor space to, and advocate among store management teams. This differs from prior studies where the buyer cannot influence demand and instead simply responds to observed downstream demand. We argue that these factors—i.e., the retail buyer's incentives and ability to affect demand—are critical for understanding why we find the magnitude of the relationship between supplier inventory service level and retailer demand to be far greater than the magnitude

[3]Type 1 service level is the probability that all demand is filled (Nahmias, 2008).
[4]Type 2 service level, also called fill rate, is the expected portion of demand filled (Nahmias, 2008).
[5]Retail buyers are employees tasked with making purchasing decisions on behalf of the retailer.

identified in previous studies.

By studying the ordering patterns of retail buyers, our research describes the complex behavior of procurement professionals in practice. Researchers have studied the ordering decisions of experimental subjects representing procurement professionals as they manipulate a number of parameters, including service level. See, for example, Ho et al. (2010), Gurnani et al. (2013), and Nagarajan and Shechter (2013). The objective of these studies is to better understand and model actual human decision making in operational contexts. Gurnani et al. (2013), however, call for future work to explore more realistic procurement contexts in order to observe the behavior of procurement professionals within competitive situations. As discretionary decision makers, the behavior of procurement professionals such as retail buyers is difficult to predict. Hopp et al. (2009) note more research is needed to understand how discretionary decision making—such as retail ordering—works in practice. Retail buyers in our context are not making decisions for a single location and a single item (Ho et al., 2010) but rather are managing one or more product categories and multiple competing brands. Supplier service levels are unknown to the buyer and changing over time. In answering the call to explore ordering behavior in practice, our findings augment the work conducted by experimentalists within operations management.

Our study contributes to the literature in several ways. First, to the best of our knowledge, we are the first to establish a positive, significant, and managerially substantive link between aggregate type 1 service level and current retailer demand in practice. Our findings hold when we employ other definitions of historical supplier inventory service level, including proxies for type 2 service level measured at the retailer level and across all retailers. Further, we show that retailers that order more frequently exhibit a greater change in order quantity in response to a change in historical supplier service level, a finding consistent with retail buyers collecting information about and reacting to a supplier's service level (Tomlin, 2009; Chen et al., 2010). We identify mechanisms that drive the relationship between supplier inventory service level and retailer demand in practice, e.g., the buyer's ability to affect demand.

We also provide a straightforward method for suppliers to estimate the relationship between service level and retailer demand using data that is typically readily available. Hendricks and

Singhal (2005$a$,$b$) argue that operations researchers lack adequate metrics to track the impact of supplier reliability. Toward that end, we develop additional metrics of supplier reliability beyond inventory service level. These metrics reflect the multiple dimensions of supplier performance identified in existing practitioner and academic literature, namely, consistency and recovery (Peck, 2005; Che et al., 2012). We test the impact of these metrics on demand from retailers. By employing our methodology, suppliers can estimate how their retail customers have responded to changes in their reliability as a supplier. In so doing, suppliers can make informed decisions about the service level they ought to provide to their channel partners.

## 2 Retail Supply Chain for Functional Apparel Items

We study the relationship between supplier 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. Retailers emphasize the importance of supplier service level within these supply chains. Scorecards are a widely used mechanism for tracking supplier service level performance (Kulp et al., 2007), and retailers penalize suppliers that are even slightly late to deliver (Craig et al., 2013), even though such deliveries are unlikely to affect end consumers. Survey-based research has shown that 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). A key to understanding these behaviors is the retail buyer: as Bonoma (1982) notes (p. 4), "companies don't buy, people do."

In order to better understand the role of retail buyers, we conducted interviews with more than 20 buyers across a variety of retail chains as well as with numerous supplier salespeople and managers. Retail buyers play a critical role within many supply chains by selecting the type and quantity of merchandise to buy. They typically control the merchandising strategy for a product portfolio—often a line or category—and face incentives that depend on the performance of that portfolio. They engage in a variety of activities, including selecting the product assortment as well as the quantity of merchandise to buy, allocating orders to suppliers, negotiating wholesale prices, providing planogram recommendations, scheduling promotional events, and determining markdowns. Further, they consistently monitor their product portfolios and adjust their decisions accordingly. Given the importance of the retail buyer, apparel suppliers often describe needing to understand not only the preferences of the end consumer but also the key factors that drive purchasing decisions among retail buyers. Although there are a number of elements that drive these decisions, our research focuses on the impact of supplier service level.

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. Retail buyers can also maintain the minimum assortments and facings required to provide the desired shopping experience with less safety stock when supplier service level is high (van Donselaar et al., 2010). Moreover, product availability is critical to retail buyers planning in-store promotions. In our field research, we studied one retailer whose buyer 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.

Lapses in supplier product availability can impact a retail buyer in other ways as well. For example, 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. Field interviews with buyers indicate that they perceive these activities to be an unproductive use of their time. Slone (2004) argues "the ability to get [product] there fast is important, but not as important as your ability to get it there when you said you would" (p. 116). According to Duffy (2004), suppliers' sales teams find it difficult to convince retail buyers to work with them in the face of poor historical inventory service level. 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.

Moreover, end consumer demand is not exogenous to the retail buyer's decisions, since the retail buyer has a number of tools for driving consumer demand (van Nierop et al., 2008). For example, buyers can feature (e.g., display prominently) chosen products. Thus, although the branded luxury fashion products we study clearly do not have perfect substitutes, the retail buyer can use promotion to encourage customers to make substitution decisions and to select a particular supplier's products. A buyer may thus elect to favor a supplier's product if the supplier provides high inventory service level. In fact, Nike initiated an "Always Available" program to ensure the reliable supply of key items to retail buyers in part because poor historical supplier service level led retail buyers to offer additional shelf space to a competing brand with better service level. The ability of a buyer to influence the end consumer's purchasing and substitution decisions will affect the magnitude of the buyer's response to changes in supplier service level.

Nonetheless, a variety of beliefs exist among suppliers regarding the behavior of retail buyers. An informal survey of 141 managers within a single division of a branded consumer packaged goods manufacturer highlights this heterogeneity. When asked how the quantity of a specific product ordered by a retail buyer would change following a 1 percentage point increase in service level, the majority of managers, 79 (56 %), predicted that there would be no change in orders from retail buyers. Of the remaining managers, 44 (31 %) predicted that demand would increase, and 18 (13 %) predicted that demand would decrease. Field interviews with supplier salespeople reveal that they recognize the negative effects of poor service level; however, many salespeople contend that increased supplier service level often leads to a reduction in orders for several reasons. If the retail buyer does not consider a product to be an adequate substitute for other products, the buyer will reduce its order for the product 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 also reduce the shelf space dedicated to the product when the retailer's storage space is constrained.[6] In the common case of no backordering, retail buyers need fewer reorders as supplier service level increases. Further, lapses in product availability can lead to rationing games and order inflation when retailers do not observe a supplier's inventory (Lee et al., 1997). Our study of the link between supplier service level and retailer demand informs this debate.

In sum, our field interviews reveal two important aspects of the retail buyer's decision context.

[6] See van Donselaar et al. (2010) for a discussion of storage space restrictions in retailing.

First, whereas prior research on a retailer purchasing from unreliable suppliers assumes that the retailer's incentives are solely driven by the fulfillment of end consumer demand, we find that the decision maker's—i.e., the retail buyer's—incentives can be affected by poor supplier service level in a number of other ways. Second, prior research typically assumes that end consumer demand is exogenous. In contrast, we find that retail buyers have considerable influence over demand, whether by promoting a particular supplier's products or by encouraging customers to purchase substitutes for said supplier's products.

# 3 Research Objectives and Hypotheses

Our research objective is to empirically measure the relationship between current retailer demand and historical supplier inventory service level. In this section, we survey related research and present our hypotheses. We concentrate on two areas of the literature: research on how product availability affects end consumer demand as well as on the interaction between a retailer and unreliable suppliers.

## 3.1 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. Researches have also suggested that high product stocking quantity may drive demand by acting as a sort of 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 evaluate the relationship between product availability and demand. However, we investigate this relationship between a supplier and a retailer, where findings at the end-consumer level—e.g., the signaling role of the amount of inventory on display as well as the costs of unavailability—do not directly apply. As noted in §2, an end consumer and a retail buyer face very different costs when they are unable to procure a product.

### 3.2 Inventory Service Level and Retailer Demand

Supplier unreliability is often modeled using the concept of random yield or stochastically proportional supply. In this arrangement, a customer places an order $r$, and the supplier delivers $Yr$, where $Y$ is a random variable that represents the supplier's yield distribution. Prior research has examined a retailer's orders 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 within the EOQ and newsvendor models. In these cases, the retailer adjusts its order quantity to accommodate the possibility of receiving a smaller quantity than requested. Yano and Lee (1995) provide a survey of related models. Lee et al. (1997) and Cachon and Lariviere (1999) predict the same relationship between service level and orders for a different reason: product shortages can lead to rationing games that increase demand, e.g., through phantom orders. 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. 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, 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 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 the set of potential suppliers. Gurnani et al. (2013) conduct laboratory experiments in a multi-supplier setting in which subjects order from one unreliable and one perfectly reliable supplier, finding that subjects do increase orders for the unreliable supplier as its service level improves.

Our empirical context differs from the aforementioned research in several respects. First, the products sold by different suppliers in our setting are imperfect substitutes, and the supplier's service level is unknown to the buyer and changes over time. Second, the retail buyer can affect end consumer demand. As described in §2, retail buyers have considerable influence on end consumer demand and can promote products from suppliers that provide high inventory service level and encourage substitution for products from suppliers that provide lower inventory service level. Third, the retail buyer's incentives are not solely drive by end consumer demand. Specifically, the effects of a supplier stockout on the retail buyer include not only costs due to the impact on the end consumer but also effort costs and reputation-based risks. We therefore expect retail buyers to use their discretion to shift their orders substantially towards suppliers that offer higher inventory service level in order to reduce the cost of supply risk. Thus, we hypothesize:

**Hypothesis 1.** *Current retailer demand for a supplier's products is positively associated with historical supplier inventory service level. (H1)*

The aforementioned articles also assume that the buyer knows its suppliers' yield distributions but, in practice, retailers must 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 have a different response to changes in supplier service level than retailers that amass less information. Further, a similar 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 therefore hypothesize:

**Hypothesis 2.** *Retailers with higher order frequencies exhibit a different change in demand in response to changes in historical supplier inventory service level than retailers with lower order frequencies. (H2)*

To the best of our knowledge, we are the first to conduct a field-based test of how changes in supplier inventory service level affect a retailer's orders in terms of both direction and magnitude. A related study conducted by Terwiesch et al. (2005) uses data from a manufacturer of customized capital goods (i.e., semiconductor equipment) to study how the purchasing forecasts of the manufacturer's customers (who are themselves manufacturers) affect lead time as well as how lead time affects the customers' order cancellations. While we also study many buyers purchasing from a single supplier, our focus is on how historical service level, rather than lead time, affects orders from retailers.

## 4 Data and Measures

To study the relationship between supplier service level and retailer demand empirically in the context of the supply chain for functional apparel items, we collected a proprietary dataset from Hugo Boss, a European fashion house known for men's and women's fashion apparel, shoes, and accessories (Raman et al., 2009). We focus on products from Hugo Boss's bodywear and hosiery division, which supplies functional apparel products such as undergarments, and on Hugo Boss's retailer customers in Germany.

The bodywear and hosiery 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. These products were sold to retailers such as Peek & Cloppenburg and Anson's Herrenhaus. The SKUs comprised Hugo Boss-branded men's undershirts, boxers, underwear, and socks. Retailers could determine the availability of a product only by placing an order. If Hugo Boss lacked the product to satisfy all retailer demand, orders were filled 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 program aimed at testing the effects of increased service level on its direct customers' (i.e., retailers) orders, Hugo Boss increased the service level to retailers for 45 of the bodywear and hosiery division's 513 BOSS Black replenishment SKUs. The SKUs were not selected due to sales velocity, margins, or other known variables that would lead to a large increase in orders. Instead, Hugo Boss managers selected one line in a production facility to use as a test bed. Prior to the pilot program, the Hugo Boss warehouse ordered from Hugo Boss's factory monthly. To implement the pilot program, the Hugo Boss warehouse began ordering the selected production line's 45 SKUs weekly, allowing the warehouse to provide a higher service level to retailers. While the ordering process within Hugo Boss—i.e., the ordering process between the Hugo Boss warehouse and factory—changed for the 45 pilot SKUs, the ordering process between retailers and the Hugo Boss warehouse did not change, and the non-pilot SKUs were completely unaffected. Moreover, the Hugo Boss managers that conducted the pilot assert that prices for the pilot group and non-pilot group products were held constant with respect to prices for similar products from competitors.

We collected weekly data from Hugo Boss about the 45 replenishment bodywear SKUs affected by the pilot project. These data comprise weekly order quantities from all retailers in Germany as well as end-of-week inventory at Hugo Boss's warehouse. The data cover 148 weeks beginning on the 31$^{st}$ week of 2004, and the pilot program began during the 31$^{st}$ week of 2005.

The order quantity data for the pilot group of 45 SKUs records the quantity of each SKU ordered by each retailer during each week. For the pilot group of SKUs, there are 58,787 orders for 509,800 units from 693 retailers. To eliminate the possible effect of entry and exit, we keep 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.[7] Therefore, we have a total of 632,700 observations at the SKU per retailer per week level (i.e., 45 SKUs, 95 retailers, and 148 weeks). Let the variable $\textit{Order Quantity}_{rst}$ be the quantity of SKU $s$ ordered by retailer $r$ during week $t$.

[7] In 5.1, we consider a less restrictive sample.

The inventory data captures end-of-week, i.e., end-of-cycle, inventory at Hugo Boss's warehouse for each SKU in the pilot group. Therefore, end-of-week inventory allows us to identify cycles in which the warehouse ran out of a particular SKU. The inventory data we collected does not cover a total of five weeks in our sample. When fitting our empirical models, we exclude these weeks from our sample, leaving 40,152 orders for 382,926 units from 95 retailers and a total of 611,325 observations. Imputing the inventory levels in these missing weeks does not affect the results of our analyses.

## 4.1 Operationalization of Key Variables

To test whether Hugo Boss's historical inventory service level impacts current retailer demand (H1), we use two common metrics: type 1 and type 2 inventory service levels. We begin by calculating these metrics across all retailers two reasons. First, prior research on unreliable suppliers employs the distribution of a supplier's service level performance. Service level calculated across all retailers incorporates all realizations of the supplier's performance and thus best represents the distribution of Hugo Boss's service level performance. Second, we use aggregate measures since service level targets are often set and managed at this level (Slone, 2004). For robustness, we also test our hypotheses using proxies for type 2 service level calculated for each individual retailer.

We calculate the service level metrics at the product line level. 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). Similarly, 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—for example, a project is unlikely to improve the availability of the small size of a particular jacket while not improving the availability of the medium size of said jacket. Estimating beliefs at the pilot group level has the distinct advantage of allowing us to employ less data in initializing the historical service level measures. We discuss the results of calculating the service level metrics at the SKU level in §5.1.

Our service level metrics are calculated over the prior year on a running basis. We employ service level performance calculated over the prior year as opposed to from the beginning of the data observation window to abet interpretation of the results, since annual measures are commonly used in practice. We equally weight all historical delivery data within the prior year. As discussed in §5.1, the findings reported herein are not affected if the measures are calculated from the start of the data observation window.

To present the metrics of supplier service level, beginning with those calculated across all retailers, we introduce several definitions. Let $S$ be the set of 45 pilot-group SKUs, let $R$ be the set of 95 retailers, and let $T$ be the ordered set of weeks in our data. Let $\mathit{Ending\ Inventory}_{st}$ be the amount of SKU $s$ in warehouse inventory at the end of cycle $t$. Let $\mathit{Stockout}_{st} = \mathbb{1}\left[\mathit{Ending\ Inventory}_{st} = 0\right]$, which takes the value 1 if SKU $s$ stocks out during cycle $t$ and the value 0 otherwise. The aggregate type 1 service level through week $t$ across all retailers is

$$\mathit{Aggregate\ Type\ 1\ Inventory\ Service\ Level}_t = \sum_{s \in S} \sum_{i=t-51}^{i=t} \frac{1 - \mathit{Stockout}_{si}}{52\,|S|}.$$

This metric can be used to illustrate the effect of the pilot project on service level. At the outset of the pilot project, aggregate type 1 service level was nearly 98.0 %. After the pilot project, it reached 99.9 %.

To calculate type 2 service level, we must determine the inventory available for sale during the week, i.e., the beginning inventory. The warehouse we study did not directly record the beginning inventory for each cycle or replenishment quantities, so we use the following proxy for week $t$ and SKU $s$, where $\mathcal{R}$ is the full set of 693 retailers in our sample:

$$\mathit{Beginning\ Inventory}_{st} = \begin{cases} \mathit{Ending\ Inventory}_{st} + \sum_{r \in \mathcal{R}} \mathit{Order\ Quantity}_{rst} & \mathit{Stockout}_{st} = 0 \\ \mathit{Ending\ Inventory}_{s,t-1} & \mathit{Stockout}_{st} = 1. \end{cases}$$

This variable is based on the assumption that replenishments did not arrive during weeks in which a stockout transpired and will underestimate the actual fill rate if this assumption is violated. To test this assumption, we determined the replenishment quantity during all non-stockout weeks for each SKU and compared that 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 $\textit{Replenishment Quantity}_{st} = \textit{Beginning Inventory}_{st} - \textit{Ending Inventory}_{s,t-1}$. Let $\textit{Replenishment Count}_s$ be the count of replenishments for SKU $s$ (i.e., the count of weeks during which $\textit{Replenishment Quantity}_{st}$ is strictly positive). Let $\textit{Average Replenishment Quantity}_s = \frac{\sum_{t \in T} \textit{Replenishment Quantity}_{st}}{\textit{Replenishment Count}_s}$. In all cases, the amount of SKU $s$ ordered by all retailers during a week in which a stockout occurred was less than $\textit{Average Replenishment Quantity}_s$. Given this definition of beginning inventory, aggregate type 2 service level is calculated as

$$\textit{Aggregate Type 2 Inventory Service Level}_t = \frac{\sum_{s \in S} \sum_{i=t-51}^{i=t} \min\left(\sum_{r \in R} \textit{Order Quantity}_{rsi}, \textit{Beginning Inventory}_{si}\right)}{\sum_{s \in S} \sum_{j=t-51}^{j=t} \sum_{r \in R} \textit{Order Quantity}_{rsj}}.$$

The service level metrics introduced thus far are calculated across all retailers. 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 service level metrics for specific retailers. Instead, we calculate three proxies for type 2 service level using three different assumptions.[8] The first assumption is that Hugo Boss follows its stated policy of filling orders on a first-come, first-served basis. The second assumption is that Hugo Boss fills smaller orders first, e.g., to maximize the number of retailers that receive their full order. The third assumption is that Hugo Boss fills 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: $\textit{Quantity Delivered}_{rst}^{FCFS}$, $\textit{Quantity Delivered}_{rst}^{Small}$, $\textit{Quantity Delivered}_{rst}^{Large}$. These variables are generated by distributing the amount of initial inventory, $\textit{Beginning Inventory}_{st}$, to each retailer based on the respective assumptions. The retailer-specific type 2 service level is then

$$\textit{Retailer Type 2 Inventory Service Level}_{rt}^{\Theta} = \frac{\sum_{s \in S} \sum_{i=t-51}^{i=t} \textit{Quantity Delivered}_{rsi}^{\Theta}}{\sum_{s \in S} \sum_{j=t-51}^{j=t} \textit{Order Quantity}_{rsj}},$$

[8]The authors benefitted immensely from discussions with and insight provided by Vishal Gaur pertaining to the development of these proxies.

where $\Theta \in (FCFS, Small, Large)$. Thus, we employ a total of five across-retailer and retailer-specific measures of service level.

When fitting our empirical models, we use lagged values of historical service level to avoid conflating the immediate effects of a stockout—e.g., in the next period, retailers may need to amend their inventory position by placing an abnormally large order, or they may forego ordering altogether because they have found a substitute—with the long-term effects. In our estimations, we lag the historical service level measures by two weeks; nonetheless, as discussed in §5.1, the results presented herein are robust to other reasonable lag specifications. Due to the two-week lag and the fact that the historical service level measures require one year of data to compute, we drop the first 54 weeks of our sample when estimating our models, leaving a total of 380,475 observations (45 SKUs, 95 retailers, and 89 weeks).

To test whether the response of retailers to changes in service level differs depending on the frequency with which a retailer orders (H2), we calculate each retailer's order frequency using data generated during the first 54 weeks of our data observation window (i.e., data not used in estimation).[9] We define a retailer's order frequency as the ratio of the number of orders the retailer places in those 54 weeks to the total number of opportunities to order in that time frame:

$$Order\ Frequency_r = \sum_{s \in S} \sum_{t=1}^{54} \frac{\mathbb{1}\left[Order\ Quantity_{rst} > 0\right]}{54\,|S|}.$$

For instance, over the course of 54 weeks, a retailer has 2,430 opportunities to order any of the pilot group products. A retailer that places 1,458 orders over that time frame would have an order frequency of 0.6.

## 4.2 Controls

There are a number of factors that could affect demand for the pilot group of SKUs besides Hugo Boss's historical inventory service level. For instance, the popularity of Hugo Boss's products may impact the demand of any individual product. To control for the popularity of the BOSS Black

[9]While we do not employ data used in estimation to calculate a retailer's order frequency, we note that a paired $t$-test of order frequency before and during our estimation window does not find a statistically significant difference in order frequency.

brand as a whole, we collected data on weekly orders from all retailers for a set of 219 non-pilot BOSS Black SKUs produced by the bodywear and hosiery division of Hugo Boss. The non-pilot SKUs are chosen for similarity to the pilot SKUs. Specifically, they differ only slightly from the pilot SKUs, e.g., in fabric or cut. Let $\textit{Non-Pilot Orders}_t$ be the total orders of these comparison products in week $t$, scaled down by a factor of 10,000 for regularity. We also include retailer, SKU, week-of-year, and year fixed effects to control for time-invariant retailer and SKU characteristics as well as for seasonality.

The broader economic conditions faced by retailers in Germany may also impact the demand for the pilot SKUs. Our analysis incorporates an index obtained from the Bundesbank Statistical Office as a control (Deutsche Bundesbank, 2010). 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 $\textit{Retail Sales Index}_t$ be the value of this index during week $t$.

In the case of a stockout, we would expect firms that place an order that is not filled to attempt to mitigate this lack of supply, e.g., by placing another order the following week. We control for such proximal effects of a stockout for SKU $s$ using the stockout indicator variable lagged by one week, i.e., $\textit{Stockout}_{s,t-1}$.

Figure 1 plots the orders for the pilot group of products against the orders of the non-pilot, comparison products, where the vertical line denotes the start of the pilot project. Table 1 provides summary statistics for the variables introduced in this section. Table 2 provides a correlation matrix. In §5, we present our empirical models, tests of our hypotheses, and our interpretation of the estimation results.

## 5 Estimation Strategy and Results

In this section, we introduce tests of the hypotheses developed in §3. To test the hypotheses, we use a Poisson panel regression framework. While the data are overdispersed relative to the Poisson distribution (see Table 1), the Poisson method generates consistent estimates even if the dependent variable does not actually take a Poisson distribution (Wooldridge, 1999). For all of our empirical tests, we use the control variables introduced in §4.2.

Our base model with only controls is

$$\begin{aligned}\ln\left[\mathrm{E}\left(\mathit{Order\ Quantity}_{rst}\right)\right] = \alpha &+ \delta_1 \times \mathit{Stockout}_{s,t-1} + \delta_2 \times \mathit{Non\text{-}Pilot\ Orders}_t \\ &+ \delta_3 \times \mathit{Retail\ Sales\ Index}_t + \mathit{Retailer\ Fixed\ Effect}_r \\ &+ \mathit{SKU\ Fixed\ Effect}_s + \mathit{Week\text{-}of\text{-}Year\ Fixed\ Effect}_t \\ &+ \mathit{Year\ Fixed\ Effect}_t + \epsilon_{rst},\end{aligned} \tag{1}$$

where $\alpha$ is a fixed intercept. The errors, denoted $\epsilon_{rst}$, are cluster-robust errors clustered by retailer (Wooldridge, 2002). To fit the model, we use the conditional fixed effects Poisson command implemented by Stata 12. Models 1 through 3 in Table 3 present the results of fitting Equation (1).

To test H1, we incorporate the measures of historical inventory service level. For brevity, we refer to the control variables and their coefficients as $\boldsymbol{\delta}'\boldsymbol{x}_{rst}$. The model is

$$\ln\left[\mathrm{E}\left(\mathit{Order\ Quantity}_{rst}\right)\right] = \alpha + \beta \times \mathit{Historical\ Inventory\ Service\ Level}_{r,t-2} + \boldsymbol{\delta}'\boldsymbol{x}_{rst} + \epsilon_{rst}, \tag{2}$$

where *Historical Inventory Service Level* is any one of the five metrics developed in §4. Models 4 and 5 in Table 3 present the results of this estimation for the aggregate inventory service level metrics. Hypothesis 1 is supported for both metrics of aggregate historical service level. The coefficient for aggregate type 1 service level (Model 4, Table 3) is positive and significant ($z = 15.34$). Holding all variables at their means, increasing aggregate type 1 service level by one standard deviation, or by nearly 1 percentage point, is associated with approximately a 13 % increase in a retailer's order quantity. Increasing aggregate type 2 service level by one standard deviation, or by half of a percentage point, under the same conditions yields an 8 % increase in a retailer's order quantity.

Table 4 presents the estimation of Equation (2) using retailer-specific fill rate metrics. The coefficients of Models 1, 2, and 3 reveal support for H1 among all three retailer-level measures. For example, the coefficient for retailer-specific type 2 service level under the first-come, first-served assumption (Model 1, Table 4) is positive and significant ($z = 8.53$). We find a single standard deviation increase in this measure of service level to be associated with a 12 % increase in a retailer's order quantity. Results for the other measures of retailer-specific service level are comparable.

To test H2, we add the interaction of the service level metric with the retailer's order frequency to Equation (2):

$$\begin{aligned}
&\ln\left[\mathrm{E}\left(\textit{Order Quantity}_{rst}\right)\right] = \\
&\qquad \alpha + \beta \textit{Historical Inventory Service Level}_{r,t-2} \\
&\qquad + \gamma \textit{Historical Inventory Service Level}_{r,t-2} \times \textit{Order Frequency}_{r} \\
&\qquad + \boldsymbol{\delta}' \boldsymbol{x}_{rst} + \epsilon_{rst}.
\end{aligned} \tag{3}$$

Although the main effect of a retailer's order frequency is subsumed into the retailer fixed effect, we only need to estimate the interaction to test the hypothesis. Models 6 and 7 (Table 3) present the estimation of Equation (3) for the aggregate service level measures while Models 5 and 6 (Table 4) present the coefficients for the retailer-level measures.

Hypothesis 2 is supported for each of the five service level metrics. For example, the coefficient on the interaction between aggregate type 1 service level and order frequency is positive and significant ($z = 3.02$). A retailer with an order frequency in the 85$^{\text{th}}$ percentile would be expected to exhibit a 30 % larger increase in order quantity in response to a single standard deviation increase in aggregate type 1 service level than a retailer with an order frequency in the 15$^{\text{th}}$ percentile, controlling for all other variables at their means.

Figure 2 illustrates the larger response of frequent orderers graphically by plotting the total order quantity by week for the pilot group of products for two sets of retailers: frequent orderers, defined as having an order frequency above the mean order frequency across the 95 retailers, and non-frequent orderers, defined as having an order frequency below the mean. The vertical line denotes the start of the pilot project, and the horizontal lines denote the means of each time series before and after the start of the pilot project. As the figure demonstrates, retailers that order more frequently exhibit a different response to changes in supplier service level than retailers that order less frequently.

In sum, we find historical supplier service level to be a statistically significant predictor of current retailer demand. This finding supports H1 and is consistent with retail buyers tracking and responding to a supplier's service level. Further, in support of H2, retailers that order frequently respond differently to changes in service level, suggesting that these retailers collect more information about a supplier's service level. Finally, the magnitude of the retailers' responses to the change in service level is substantial even after controlling for alternative drivers of retailer demand.

### 5.1 Robustness Tests

To evaluate the robustness of our empirical analysis, we ran several additional tests. First, to assess the effect of the lag of the historical service level measures, we estimated the model with lags of length two weeks, three weeks, four weeks, and thirteen weeks (one quarter). Second, to determine if the results are solely driven by large customers, we dropped the ten largest retailers by total order quantity from the sample. Third, we estimated the model using the service level measures calculated over the entire data observation window rather than over the prior year. Fourth, we fit the model with the aggregate service level metrics calculated at the SKU level rather than the product line level. Finally, to test if our sample restriction affected the results, we fit the aggregate service level models while including all 693 retailers. None of these alternate specifications yielded results that differed substantially from those discussed above.

## 6 Alternate Metrics

The practitioner literature offers additional perspectives on supplier reliability beyond the traditional type 1 and type 2 service levels. Duffy (2004), for example, presents several metrics firms can use to track reliability including case fill rate, on-time delivery, the perfect order, and order fulfillment lead time. Solomon (2012) reveals that supply chain leaders think of reliability in terms of predictability and consistency. On the other hand, Turner (2011) highlights the role of supply chain resilience, defined primarily as the ability to recover after service disruptions, in driving supply chain performance.

Operations management researchers also cite consistency and the ability to recover after a disruption as two key elements of supplier reliability. Christopher (2005) argues that consistency, within industrial markets, may be more influential for winning orders than product or technical features and is perceived as a key driver of customer loyalty and retention. Resiliency, or the ability of a supply chain to recover after a delivery failure, is also a noted driver of supply chain performance (Bakshi and Kleindorfer, 2007). We describe each of these supplier reliability dimensions—which we term consistency and recovery—in turn and discuss how each may impact retailer demand.

Researchers have shown consistency in product availability to influence consumer choice and consumer product evaluation. Swait and Erdem (2002) argue "consistency in availability will increase utility because product unavailability on the shelf may force the consumer to reevaluate their commitment to the SKU" (p. 306). Su and Zhang (2009) argue that consistent product availability stimulates consumer demand, since consumers are less willing to visit a store in the event of a high likelihood of a stockout. Dana and Petruzzi (2001) find that higher inventory levels attract customers.

Numerous studies explore the impact of stockouts on key dimensions of firm performance. Stockouts have been observed to reduce customer satisfaction and brand loyalty (Sloot et al., 2005), lower repeat purchase behavior (Heim and Sinha, 2001), and negatively impact current and future orders (Anderson et al., 2006). DeHoratius and Raman (2008) estimate the lost revenue due to stockouts induced by inventory inaccuracy to be 1.1 percent of retail sales.

This phenomenon exists not only among end consumers but also among suppliers and their retail partners. Malmbak and Albaum (2007), in a survey of retailers, find inconsistent product availability to be one of the top ten reasons for discontinuing a supplier's brand. As the vice president of Global Supply Chain at Whirlpool Corporation, Slone (2004) emphasized the importance of not just high levels of availability but also of having "consistently good" product availability performance (p. 119).

On the other hand, firms can mitigate the adverse effects of a stockout by quickly recovering after a service disruption (Bakshi and Kleindorfer, 2007). Craighead et al. (2007) identify a firm's recovery capability as critical to performance and derive the following proposition from their case-based research. They propose that a firm's recovery capability is negatively associated with the severity of such service disruptions. In other words, firms that have the capability to return to high levels of service after a stockout will not necessarily suffer all of the negative effects commonly associated with stockouts. Moreover, firms that have the ability to recover will win additional business from competitors (Sheffi, 2007; Christopher, 2005).

To operationalize the notions of consistency and recovery, we model Hugo Boss's delivery performance as though each SKU were governed by a binary Markov chain (Song and Zipkin, 1996; Parlar et al., 1995). State 1 represents full availability: orders placed by a customer when a supplier is in state 1 are filled completely. State 0 represents a stockout: orders placed by a customer when a supplier is in state 0 are not filled. Under this simplification, Hugo Boss is in state 1 for SKU $s$ during week $t$ when $Stockout_{st}$ is 0 and is in state 0 when $Stockout_{st}$ is 1.

The binary Markov representation allows the model to capture changes in both the probability and the persistence of a stockout, or changes in a supplier's consistency and recovery. Consistency is the probability that the supplier transitions from state 1 to state 1. Recovery is the probability that the supplier transitions from state 0 to state 1. Let $n_{abst}$ be the number of transitions from state $a \in \{0,1\}$ to state $b \in \{0,1\}$ starting in period $t-51$ and ending in period $t$ for SKU $s$. We calculate historical consistency and recovery as the maximum likelihood estimates of the transition probabilities using the prior year of data:

$$Consistency_t = \frac{1}{|S|}\sum_{s\in S}\frac{n_{11st}}{n_{10st}+n_{11st}}, \qquad Recovery_t = \frac{1}{|S|}\sum_{s\in S}\frac{n_{01st}}{n_{00st}+n_{01st}}.$$

Over the estimation period, $Consistency_{t-2}$ has a mean of 0.9967 with a standard deviation of 0.0029 while $Recovery_{t-2}$ has a mean of 0.7281 with a standard deviation of 0.2349.

To assess how these measures affect retailer demand, we use the following model:

$$\ln\left[\mathrm{E}\left(Order\ Quantity_{rst}\right)\right] = \alpha + \beta_1 Consistency_{t-2} + \beta_2 Recovery_{t-2} + \boldsymbol{\delta}'\boldsymbol{x}_{rst} + \epsilon_{rst}. \tag{4}$$

Table 5 presents the results when fitting this model. Model 1 includes only consistency, and Model 2 adds recovery. The coefficient on consistency in Equation (4) is positive and significant ($z = 28.18$). We find that, holding all other variables at their means, increasing consistency by one standard deviation is associated with an 18 % increase in retailer demand. The coefficient on recovery is also positive and significant ($z = 3.67$), and we find that increasing recovery by one standard deviation while keeping all other variables at their means is associated with a 3 % increase in order quantity.

While practitioners have noted conceptually the importance of consistency and recovery capabilities, few have attempted to quantify their impact. The metrics in this section provide a means for tracking these capabilities, and we find a statistically significant link between these measures and downstream demand. If used in practice, these measures can aid suppliers in better understanding their retail customers' responses to changes in supplier performance.

## 7 Conclusion

Our results demonstrate a positive, statistically significant, and substantial relationship between supplier historical inventory service level and retailer demand in the supply chain for functional apparel items. Through our research, we answer the call for field-based evidence regarding the presence and strength of this important relationship (Gurnani et al., 2013). We argue that retail buyers—employees that make purchasing decisions on behalf of the retailer—are an important driver of these results. Buyers determine which items a retailer should carry and in what quantity. Unlike the procurement professionals studied in prior laboratory experiments who only 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 demonstrate that an increase in service level can substantially increase orders even when service level is already high. Therefore, managers that 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. Nevertheless, determining the value of improved inventory service level in practice is nontrivial. In the case of Hugo Boss, we can establish that the actions the firm took to improve service level amounted to an additional €230,495 in margin,[10] or approximately 11 % of sales. We derived this number in the following way. Our empirical model

[10] The average margin for pilot group products sold was €4.89 during the time of our study and was calculated by dividing the total margin for the pilot group during the period of our study (i.e., total revenue minus total cost of goods sold) by the number of units sold.

allows us to predict what retailer demand would have been had aggregate type 1 inventory service level remained fixed at pre-pilot levels, namely, at 98.1 %, over the post-pilot period. If Hugo Boss had not improved its inventory service level and inventory service level had been fixed at 98.1 %, predicted retailer demand would be 219,030 units, or nearly 18 % lower than the actual observed retailer demand of 266,166 units.

Our empirical research augments the body of work on retailers ordering from unreliable suppliers. It examines the behavior of discretionary decision makers (Hopp et al., 2009), retail buyers, in practice. However, it 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. Therefore, further research is necessary in other contexts. Future analytical research could study buyers that can influence demand ordering from multiple unreliable suppliers. Future models could also 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. Further research could also track the adoption of the method we provide for estimating the impact of changes in service level. This method can be readily implemented in practice and can help managers accurately judge the benefits of a change in service level as well as the ramifications of their current service level targets. Finally, researchers could explore in more detail the metrics of supplier reliability—consistency and recovery—introduced herein.

**Acknowledgments** The authors thank research participants at Hugo Boss, including Constantine Moros, Katja Ruth, and Andreas Stockert. In addition, we thank Saravanan Kesavan, Laura Kornish, Bill Simpson, Dennis Yao, and seminar participants at Booth School of Business, Cambridge Judge Business School, Darden Business School, Georgia Tech Scheller School of Business, Harvard Business School, INSEAD, London Business School, UNC Kenan-Flagler Business School, University of Lausanne, University of Oregon, and Willamette University for their excellent insights and suggestions. We gratefully acknowledge the University of Portland's Arthur Butine Faculty Development Fund, the Pamplin School of Business, and the Sloan Industry Studies Program for their support of Nicole DeHoratius's research.

## References

Anderson, E. T., Fitzsimons, G. J. and Simester, D. (2006), 'Measuring and mitigating the costs of stockouts', *Management Science* **52**(11), 1751–1763.

Anupindi, R. and Akella, R. (1993), 'Diversification under supply uncertainty', *Management Science* **39**(8), 944–963.

Bakshi, N. and Kleindorfer, P. (2007), 'Co-opetition and Investment for Resilience in Global Supply Chains'. Working Paper, The Wharton School.

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 Revieww* **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.

Christopher, M. (2005), *Logistics and Supply Chain Management*, third edn, Prentice Hall, Harlow, England.

Craig, N., DeHoratius, N., Jiang, Y. and Klabjan, D. (2013), 'Inventory management with purchase order errors and rework', *Working Paper* .

Craighead, C. W., Blackhurst, J., Rungtusanathamn, M. J. and Handfield, R. B. (2007), 'The severity of supply chain disruptions: Design characteristics and mitigation capabilities', *Decision Sciences* **38**, 131–156.

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.

Dana, James D., J. and Petruzzi, N. C. (2001), 'Note: The newsvendor model with endogenous demand', *Management Science* **47**(11), 1488–1497.

DeHoratius, N. (2014), 'Global Sourcing at Walmart'. Chicago Booth Case.

DeHoratius, N. and Raman, A. (2008), 'Inventory record inaccuracy: An empirical analysis.', *Management Science* **54**(4), 627 – 641.

Deutsche Bundesbank (2010), 'Business statistics', Accessed September 2010.

Duffy, M. (2004), 'How gillette cleaned up its supply chain', *Supply Chain Management Review* **8**(3), 20–27.

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.

Hendricks, K. B. and Singhal, V. R. (2005*a*), 'Association between supply chain glitches and operating performance', *Management Science* .

Hendricks, K. B. and Singhal, V. R. (2005*b*), 'An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm', *Production and Operations Management* **14**(1), 35–52.

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.

Kulp, S., DeHoratius, N. and Kanji, Z. (2007), *Vendor Compliance at Geoffrey Ryans (A), Harvard Business School Case No. 9-108-022*, Harvard Business School Publishing, Boston, MA.

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’, *TextilWirthschaft* .

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.

Parlar, M., Wang, Y. and Gerchak, Y. (1995), ‘A periodic review inventory model with markovian supply availability’, *International Journal of Production Economics* **42**(2), 131 – 136.

Peck, H. (2005), ‘Drivers of supply chain vulnerability: an integrated framework’, *International Journal of Physical Distribution & Logistics Management* **35**(4), 210–232.

Raman, A., DeHoratius, N. and Kanji, Z. (2009), ‘Supply Chain Optimization at Hugo Boss (A)’. Harvard Business School Case Number 9-609-029.

Sheffi, Y. (2007), *The Resilient Enterprise*, MIT Press, Cambridge, MA.

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.

Sloot, L. M., Verhoef, P. C. and Franses, P. H. (2005), ‘The impact of brand equity and the hedonic level of products on consumer stock-out reactions’, *Journal of Retailing* **81**(1), 15 – 34.

Solomon, M. B. (2012), ‘Slow steam ahead’, *DC Velocity* pp. 37–39.

Song, J. S. and Zipkin, P. H. (1996), ‘Inventory control with information about supply conditions’, *Management Science* **42**(10), 1409–1419.

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.

Su, X. and Zhang, F. (2009), ‘On the value of commitment and availability guarantees when selling to strategic consumers’, *Management Science* **55**(5), 713–726.

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.

Terwiesch, C., Ren, Z. J., Ho, T. H. and Cohen, M. A. (2005), 'An empirical analysis of forecast sharing in the semiconductor equipment supply chain', *Management Science* **51**(2), 208–220.

Tomlin, B. (2009), 'Impact of supply learning when suppliers are unreliable', *Manufacturing & Service Operations Management* **11**(2), 192–209.

Turner, M. D. (2011), 'Be prepared to bounce back: Building a resilient supply chain', *Booz and Company* .

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.

van Nierop, E., Fok, D. and Franses, P. H. (2008), 'Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements', *Marketing Science* **27**(6), 1065–1082.

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

Figure 1: Orders of Pilot Group and Non-Pilot Group Products over Time

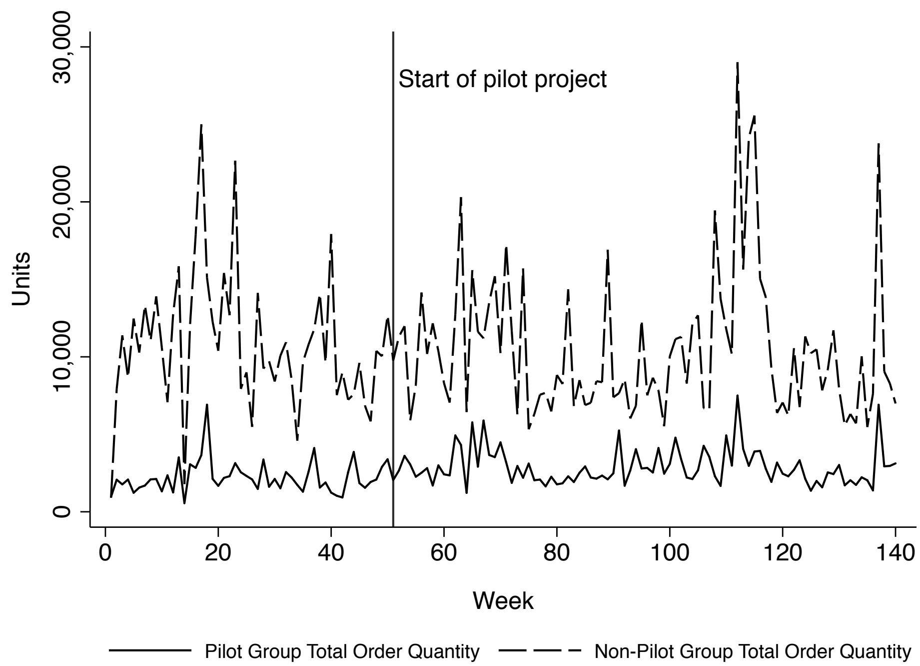



Table 1: Summary Statistics for Dependent and Independent Variables

| Variable | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Order Quantity | 0.6979 | 6.3391 | 0.0000 | 572.0000 |
| Aggregate Type 1 Service Level | 0.9936 | 0.0069 | 0.9808 | 0.9996 |
| Aggregate Type 2 Service Level | 0.9960 | 0.0044 | 0.9867 | 0.9996 |
| Retailer Type 2 Service Level FCFS | 0.9903 | 0.0244 | 0.7500 | 1.0000 |
| Retailer Type 2 Service Level Small | 0.9923 | 0.0178 | 0.8500 | 1.0000 |
| Retailer Type 2 Service Level Large | 0.9888 | 0.0247 | 0.7500 | 1.0000 |
| Order Frequency | 0.0674 | 0.1310 | 0.0018 | 0.7843 |
| Stockout Indicator | 0.0040 | 0.0631 | 0.0000 | 1.0000 |
| Non-Pilot Order Quantity | 1.0496 | 0.4682 | 0.5323 | 2.9009 |
| Retail Sales Index | 1.0100 | 0.0830 | 0.8680 | 1.2370 |

These statistics are calculated over the 89-week estimation period. Order quantity is measured in units per week. See §4 for definitions of the remaining variables. The estimates of historical inventory service level are lagged by two weeks.

Table 2: Correlation Matrix for Dependent and Independent Variables

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) Order Quantity | 1.00 | | | | | | | | | |
| (2) Aggregate Type 1 Service Level | -0.00 | 1.00 | | | | | | | | |
| (3) Aggregate Type 2 Service Level | -0.00 | 1.00*** | 1.00 | | | | | | | |
| (4) Retailer Type 2 Service Level FCFS | 0.01*** | 0.08*** | 0.08*** | 1.00 | | | | | | |
| (5) Retailer Type 2 Service Level Small | 0.00** | 0.09*** | 0.09*** | 0.93*** | 1.00 | | | | | |
| (6) Retailer Type 2 Service Level Large | 0.02*** | 0.08*** | 0.09*** | 0.98*** | 0.91*** | 1.00 | | | | |
| (7) Order Frequency | 0.36*** | 0.00 | 0.00 | 0.04*** | 0.02*** | 0.04*** | 1.00 | | | |
| (8) Stockout Indicator | 0.00 | 0.04*** | 0.04*** | 0.01** | 0.01*** | 0.01*** | 0.00 | 1.00 | | |
| (9) Non-Pilot Order Quantity | 0.03*** | -0.05*** | -0.06*** | -0.00 | -0.00 | 0.00 | 0.00 | -0.01*** | 1.00 | |
| (10) Retail Sales Index | 0.02*** | -0.13*** | -0.13*** | -0.00** | -0.00** | -0.00* | 0.00 | -0.04*** | 0.44*** | 1.00 |

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

This correlation matrix is calculated over the 89-week estimation period. The estimates of historical inventory service level are lagged by two weeks.

Table 3: Aggregate Inventory Service Level and Order Frequency

| Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|---|---|---|---|---|---|---|---|
| Aggregate Type 1 Service Level | | | | 19.65*** | | 18.27*** | |
| | | | | (1.28) | | (1.36) | |
| Aggregate Type 1 Service Level × Order Frequency | | | | | | 2.81** | |
| | | | | | | (.93) | |
| Aggregate Type 2 Service Level | | | | | 22.89*** | | 21.30*** |
| | | | | | (1.78) | | (1.92) |
| Aggregate Type 2 Service Level × Order Frequency | | | | | | | 3.23* |
| | | | | | | | (1.46) |
| Stockout Indicator | | | .16*** | .14*** | .14*** | .14*** | .14*** |
| | | | (.03) | (.03) | (.03) | (.03) | (.03) |
| Non-Pilot Sales Quantity | | .42*** | .42*** | .45*** | .44*** | .45*** | .44*** |
| | | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) |
| Retail Sales Index | | –8.34*** | –8.37*** | –8.20*** | –7.99*** | –8.20*** | –7.99*** |
| | | (.30) | (.30) | (.30) | (.30) | (.30) | (.30) |
| Retailer Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| SKU Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Week-of-Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 380,475 | 380,475 | 380,475 | 380,475 | 380,475 | 380,475 | 380,475 |
| Log Likelihood | –334,288.25 | –331,891.87 | –331,878.39 | –331,760.43 | –331,795.32 | –331,755.88 | –331,792.88 |

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

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

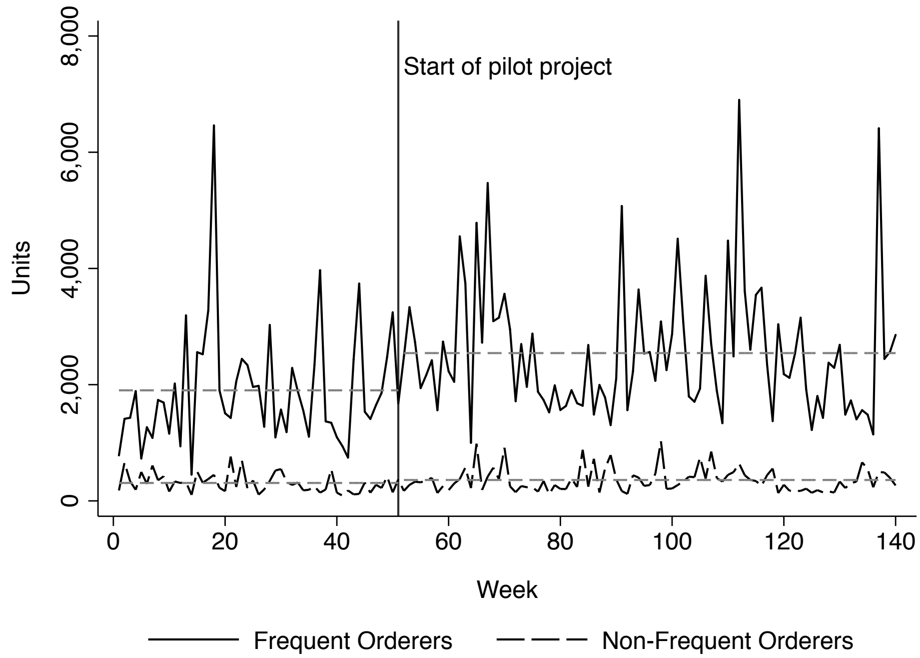
Table 4: Retailer-Specific Inventory Service Level and Order Frequency

| Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| Retailer Type 2 Service Level FCFS | 5.07*** | | | 3.06*** | | |
| | (.59) | | | (.66) | | |
| Retailer Type 2 Service Level FCFS × Order Frequency | | | | 18.10*** | | |
| | | | | (2.60) | | |
| Retailer Type 2 Service Level Small | | 5.05*** | | | 2.79*** | |
| | | (.61) | | | (.67) | |
| Retailer Type 2 Service Level Small × Order Frequency | | | | | 19.33*** | |
| | | | | | (2.44) | |
| Retailer Type 2 Service Level Large | | | 5.18*** | | | 3.05*** |
| | | | (.58) | | | (.69) |
| Retailer Type 2 Service Level Large × Order Frequency | | | | | | 19.42*** |
| | | | | | | (3.36) |
| Stockout Indicator | .16*** | .16*** | .16*** | .16*** | .16*** | .16*** |
| | (.03) | (.03) | (.03) | (.03) | (.03) | (.03) |
| Non-Pilot Sales Quantity | .42*** | .42*** | .42*** | .43*** | .43*** | .42*** |
| | (.01) | (.01) | (.01) | (.01) | (.01) | (.01) |
| Retail Sales Index | –8.30*** | –8.31*** | –8.31*** | –8.21*** | –8.21*** | –8.25*** |
| | (.30) | (.30) | (.30) | (.30) | (.30) | (.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 order quantity per SKU per retailer per week. Standard errors are in parentheses.

Figure 2: Pilot Group Total Order Quantity for Frequent and Non-Frequent Orderers




The horizontal lines denote the means of each time series before and after the start of the pilot project.

Table 5: Consistency and Recovery

| Variable | Model 1 | Model 2 |
|---|---|---|
| Consistency | 73.84*** | 66.22*** |
| | (2.62) | (3.29) |
| Recovery | | .11*** |
| | | (.03) |
| Stockout Indicator | .18*** | .16*** |
| | (.03) | (.03) |
| Non-Pilot Order Quantity | .48*** | .48*** |
| | (.01) | (.01) |
| Retail Sales Index | –6.08*** | –6.66*** |
| | (.31) | (.34) |
| 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 | –331,480.89 | –331,473.49 |

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

The dependent variable is order quantity per SKU per retailer per week. Standard errors are in parentheses. Retailer, SKU, week-of-year, and year fixed effects are omitted from this table.