Title
Essays on Social and Behavioral Aspects of Apparel Supply Chains

Author
Saez de Tejada Cuenca, Anna

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Essays on Social and Behavioral Aspects of Apparel Supply Chains

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Anna Saez de Tejada Cuenca

2019

ABSTRACT OF THE DISSERTATION

Essays on Social and Behavioral Aspects of Apparel Supply Chains

by

Anna Saez de Tejada Cuenca

Doctor of Philosophy in Management

University of California, Los Angeles, 2019

Professor Felipe Caro, Chair

When we think about Operations Management and Business Analytics, we think about optimization, efficiency, algorithms, optimality, profits and costs, exact quantitative analyses, etc. However, the field was created by humans and for humans. So what is the role of humans in the process of operational decision-making? In this dissertation, we study two aspects of the interface between humans and operational decision making, in the context of the apparel industry. The first part (Chapter 1) is related to social responsibility, i.e., how operational decision making affects workers, communities, and society. The second part (Chapters 2 and 3) shows how human biases affect operational decision-making.

In the first chapter, we study unauthorized subcontracting, i.e., when suppliers outsource part of their production to a third party without the retailer's consent. This practice has been common practice in the apparel industry and it is often tied to non-compliant working conditions. Since retailers are unaware of the third party, the production process becomes obscure and cannot be tracked adequately. We present an empirical study of the factors that can lead suppliers to engage in unauthorized subcontracting. We use data provided by a global supply chain manager with over 30,000 orders, of which 36% were subcontracted without authorization. Our results show that there are different factory types, ranging from factories that used unauthorized third parties for all of their orders to factories that used none. Moreover, the degree of unauthorized subcontracting in the past is highly related to

the probability of engaging in unauthorized subcontracting in the future, which suggests that factories behave as if they choose a strategic level of unauthorized subcontracting. At the order level, we find that state dependence (i.e., the status of an order carrying over to the next one) followed by price pressure are the key drivers of unauthorized subcontracting. Buyer reputation and factory specialization can also play a role, whereas the size of an order shows no effect. We find that the main effect (state dependence) is tied to factory utilization. Finally, we show that unauthorized subcontracting can be predicted correctly for more than 80% of the orders in out-of-sample tests. This indicates that retailers can use business analytics to predict unauthorized subcontracting and help prevent it from happening.

In the second chapter, we study the adherence to the recommendations of a decision support system (DSS) for markdowns during clearance sales. The DSS was implemented at Zara, the Spanish fast fashion retailer. Managers' initial adherence was low, which motivated two interventions: 1. showing a revenue metric; and 2. showing a reference point for that metric. We use data collected by Zara during seven clearance sales campaigns to analyze the effect of the two interventions and the behavioral drivers of managers' adherence decisions. Intervention 1 did not significantly alter managers' adherence, but Intervention 2 increased it, and also decreased their likelihood to mark a product down when DSS recommended keeping its price unchanged. Managers were more likely to adhere to the DSS's recommendations when the suggested price was aligned with the heuristic they followed before the DSS was implemented. Managers' decisions were consistent with inventory minimization, as opposed to revenue maximization. Higher salvage values were related to higher adherence, but also to larger deviations when managers did not adhere. Managers were minimizing the number of different prices to set and basing their pricing decisions on metrics that were aggregated at the group level, instead of at the individual product level. These findings can be explained by preference for the status quo, salience of the inventory (compared to a revenue forecast), loss aversion, and inattention. Some of these biases were mitigated after the interventions. Our findings provide insights on how to increase voluntary adherence that can be used in any context in which a company wants an analytical tool to be adopted by its users.

In Chapter 3, we continue to study pricing decision making by country managers at Zara. We aim to disentangle managers' degree of loss aversion from other behavioral biases by building a structural model to replicate managers' price decision making process and fitting it using data collected by Zara prior to the DSS's implementation. In our model, managers choose prices to maximize their utility over the whole season, subject to a number of constraints given by the firm's pricing rules. The utility function consists of a revenue component and a loss aversion component that depends on a loss aversion parameter. Both components include demand uncertainty and contain all products of the same type (shirts, pants, etc.). This model is, therefore, a dynamic program over a finite horizon with a large state space and uncertainty set. We use a certainty equivalent for the demand function, and discretize the problem, given that the set of available prices is discrete. Our model thus becomes a mixed integer linear program. We find, for each value of the loss aversion parameter, its corresponding set of utility-maximizing prices, and then pick the value of the parameter that best fits the prices that managers implemented. We then compare their degree of loss aversion across product groups (e.g. fashion or basic products) and across country managers. Preliminary results using a very small dataset suggest that country managers at Zara are, indeed, loss averse, but some managers are more so than others. There seem to be no clear pattern on what product types trigger loss aversion in managers. Our model explains managers' observed prices 11% better than if we assume they are pure revenue maximizers, and several times better than if we model them as inventory minimizers.

The dissertation of Anna Saez de Tejada Cuenca is approved.

Melvin Keith Chen

Charles J. Corbett

Christopher Siu Tang

Felipe Caro, Committee Chair

University of California, Los Angeles

2019

Dedicated to all the brave individuals who are, right now, thinking about starting a great adventure, but are hesitant because they think they are not smart enough, not strong enough, not good enough. You are much smarter, stronger and better than you think.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my advisor, Felipe Caro, whose guidance has been the main key to my success in the PhD program. Felipe's mentorship goes beyond research, in which he also excels. He has taught me how academia works, introduced me to the community, made an effort to make our research visible, provided virtually unlimited conference support, and given me encouragement in the multiple times in which I was about to give up. His good heart, optimism, and laid back attitude are contagious and exemplary.

I am also thankful to the rest of members of my committee: Keith Chen, whose passion and enthusiasm led me to grow an interest in Behavioral Economics, and ultimately shaped the second half of my dissertation; Chris Tang, whose hard work, wisdom, and humility are and example for everyone in academia; and Charles Corbett, who always asks the right questions and gives the right feedback, and who helped me polish my job market materials.

To all the DOTM faculty for being such a nice group, for always being so available, and for the emotional and practical support during my PhD and my job market. To the BDM group at Anderson for having been so warm and welcoming, and for inviting me to present in their seminars and providing good feedback on my second paper. To Magali Delmas for the good feedback she gave me on my first paper when I presented it in her seminar courses, and for giving me the tools to reach an audience broader than the OM community.

To my academic siblings Taylor Cocoran, Araz Khodabashian, and Prashant Chintapalli, the best cohort ever. The endless hours studying together and our mutual support were key to my success. To Sandeep Rath, who mentored all of us like a big brother. To Christian Blanco, who made all of us happier and more social. To Bobby Nyotta for all the office banter and deep conversations. To Sean Brueggemann for being so friendly and thoughtful. To Zhenhuan Lei, my academic sister from Duke, whose strength and hard work are exemplary.

I want to give many thanks to Victor Martínez-de-Albéniz, who helped me discover and love Operations Management and first suggested that I pursue a PhD in the United States. This has been the greatest adventure of my life so far, and I owe it all to him.

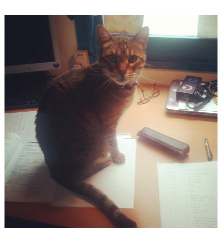
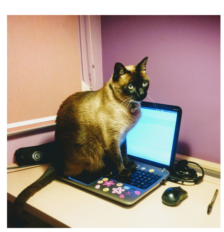
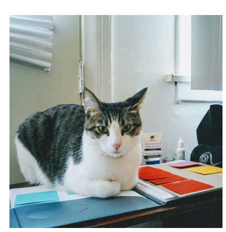
A warm thank you to Vicent Ribas, who believed in me and gave me my first job after college. Thanks to that opportunity, I grew a passion for applied research.

I am eternally grateful to my amazing partner, Jedidjah de Vries, who has been a constant source of support during the last stretch of my PhD and the academic job market. He has cooked nice meals for me, taken me to the airport at 4 AM multiple times, comforted me during my outbursts of anxiety, patiently listened to all my worries, done my laundry when I was exhausted, and always made sure I had hot tea ready while I was working on my papers or job market materials. I am looking forward to a whole lifetime together.

The biggest thank you is to my mother, Emília Cuenca, who has always believed I could do anything I wanted to, and to my father, Vicenç Sáez de Tejada, who raised me to be stronger and more independent than the gender roles of the time and place mandated. They have always put me first, enabled all my hobbies and interests, shown a huge amount of patience, supported me financially, and fought for me to have more opportunities than they did. They are my biggest fans, and I strive to make them prouder every day.

I am thankful to all the excellent professionals at the UCLA Counseling and Psychological Services for the important work they do for our academic community.

Finally, a special thank you to the very good boys who dutifully served as my research assistants at different stages of my career: Mixu, who assisted me while I was writing my master's thesis, studied English with me, and helped me prepare for the TOEFL. Rest in peace, little tiger. Panxo, who advised me to follow my passion and pursue an academic career, helped me prepare for the GRE, filled in many PhD applications, and agreed to stay home to comfort my parents after my departure. Spinoza, who gave so much feedback on my papers, helped me write my job market materials, filled in countless job applications, helped me practice my job talk, and is excited to move across the world to support my career.

VITA

Education

2003-2011	BSc in Mathematics, UPC-BarcelonaTech, Barcelona
2009-2012	MSc in Mathematical Engineering, UPC-BarcelonaTech, Barcelona

Employment

2010-2012	Junior R&D Engineer, Sabirmedical, Barcelona
2012-2014	Research Assistant, IESE Business School, Barcelona
2016	Research Scientist Intern, Amazon, Seattle
2015-2019	Teaching Assistant, UCLA Anderson School of Management
2018-2019	Graduate Student Researcher, UCLA Anderson School of Management

Honors and awards

2014-2017	UCLA Anderson Doctoral Fellowship
2015, 2017	UCLA Easton Center Summer Doctoral Fellows Program
2018	POMS College of Sustainable Operations Best Student Paper Award
2019	POMS College of Behavioral Operations Junior Scholar Paper Competition

PUBLICATIONS

Vicent J. Ribas, Jesús Caballero López, Anna Sáez de Tejada Cuenca, Juan Carlos Ruiz-Rodríguez, Adolfo Ruiz-Sanmartín, Jordi Rello, and Alfredo Vellido. *On the Use of Graphical Models to Study ICU Outcome Prediction in Septic Patients Treated with Statins.* In International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics, pp. 98-111. Springer, Berlin, Heidelberg (2011).

Vicent J. Ribas Ripoll, Anna Wojdel, Anna Sáez de Tejada Cuenca, Juan C. Ruiz-Rodríguez, Adolf Ruiz-Sanmartín, Miriam de Nadal, Enrique Romero, and Alfredo Vellido. *Continuous Blood Pressure Assessment from a Photoplethysmographic Signal with Deep Belief Networks.* The FASEB Journal, 28, no. 1 Supplement (2014): LB674.

Timothy G. Myers, Vicent Ribas Ripoll, Anna Sáez de Tejada Cuenca, Sarah L. Mitchell, and Mark J. McGuinness. *Modelling the Cardiovascular System for Assessing the Blood Pressure Curve.* Mathematics-in-Industry Case Studies, 8, no. 1 (2017): 2.

Felipe Caro, Leonard Lane, Anna Sáez de Tejada Cuenca. *Can Brands Claim Ignorance? Unauthorized Subcontracting in Apparel Supply Chains.* Under second-round major revision at Management Science.

Felipe Caro, Anna Sáez de Tejada Cuenca. *Believing in Analytics: Managers' Adherence to Price Recommendations from a DSS.* Under major revision at M&SOM.

Felipe Caro, Keith Chen, Anna Sáez de Tejada Cuenca. *Loss Aversion in Managers' Pricing Decision Making at a Fast Fashion Retailer.* Work in progress.

CHAPTER 1

Can Brands Claim Ignorance? Unauthorized Subcontracting in Apparel Supply Chains

1.1. Introduction

"Lack of transparency costs lives. It is impossible for companies to make sure human rights are respected and that environmental practices are sound without knowing where their products are made, who is making them and under what conditions. If you can't see it, you don't know it's going on and you can't fix it."(Moore et al. 2016)

The Rana Plaza collapse in Bangladesh in April 2013, which killed 1,129 people, showed to the international community the non-compliant working conditions of many apparel workers in developing countries. The major brands that had ongoing production at the Rana Plaza claimed that they had not placed their orders to any of those factories directly (Kerppola et al. 2014). This was not an isolated event. A few months earlier, a fire in the Tazreen Fashion factory in the outskirts of Dhaka, Bangladesh, killed more than 100 workers. Products for a large U.S. retailer were found at the Tarzeen facility, and though these garments were part of an order that had been placed at an authorized factory, they ended up at the location of the fire due to unauthorized subcontracting (Donaldson 2015). Over a decade, more than 1,800 apparel workers died in Bangladesh due to workplace disasters mostly at unauthorized factories (Clifford and Greenhouse 2013). Despite multiple efforts launched following these disasters, the intricate network of unauthorized subcontractors continues to exist (Drennan 2015), which is concerning for supply chains because of the risk that the subcontractors might be non-compliant.

The lack of visibility from unauthorized subcontracting allows for obscure practices, especially non-compliant working conditions. These practices have been endemic to the apparel industry in developing countries for decades. The issue got worldwide attention with the incidents in Bangladesh. However, many other incidents – possibly smaller in magnitude but no less grave – have been reported in almost every developing country where apparel manufacturing takes place. For instance, in China there was a recent case of a factory that diverted production of a major Australian brand to a facility in North Korea where forced labor was employed (McKenzie and Baker 2016). In Vietnam, a flurry of garments destined for suppliers of multinational companies were being produced under degrading conditions in detention centers (Theuws 2015). In Cambodia, a factory with unsafe working conditions was in the spotlight but the brands linked to the factory dodged the blame saying that the orders had been rerouted without their permission (Robertson and Heng 2014). These examples are limited to the cases that have been caught by the press. The actual number of incidents may be much larger.

Although most companies are already actively involved in improving suppliers' working conditions through codes of conduct, auditing, adherence to domestic and international regulations, etc., the complex nature of supply chains in the apparel industry makes it very difficult for retailers to track where their products are actually manufactured. It often occurs that a supplier subcontracts an order to a non-compliant third party without the buyer's authorization or even knowledge. As a consequence, despite the multiple efforts from retailers, many violations of building codes and labor standards remain unnoticed by the retailers. Of course, it is entirely possible that a supplier subcontracts an order to another factory in the retailer's network of trusted suppliers, and so the third party doing the work happens to be audited and compliant as well. However, once an order has been subcontracted without authorization, the buyer loses visibility and, therefore, it is impossible to monitor the actual conditions in which the order has been produced.

In this paper, we propose two research questions. First, what are the main drivers of unauthorized subcontracting? Second, can brands claim ignorance? In other words, when unauthorized subcontracting takes place, could retailers have anticipated it and, therefore, prevented it?

To find a response to the first question, we run an empirical study to analyze the factors that lead to unauthorized subcontracting using data from a middleman that matches retailers with suppliers in several developing countries. The data spans eight months in 2013-14 and consists of 32,477 orders from 226 factories delivered to 30 buyers. Unauthorized subcontracting is reported in 36.3% of the orders in our sample. However, we find that there are big differences among factories. We observe three consistent patterns. For some factories, there is no incidence of unauthorized subcontracting; for others, it occurs in 100% of the orders; and, for a third group of factories, it happens occasionally. Most importantly, the propensity of each firm to engage in unauthorized subcontracting emerges as an inherent characteristic or strategic choice that can be (partially) inferred from past behavior.

We argue that factories' unauthorized subcontracting behavior exhibits state dependence (Heckman 1981, Wooldridge 2010). Hence, a factory that just delivered an order that was subcontracted without authorization is more likely to do the same for the next order. We validate this hypothesis by testing whether unauthorized subcontracting is autoregressive. The result is large in magnitude and is statistically significant. We find that the chance of unauthorized subcontracting increases by 83%, i.e., almost doubles, when the previous order at the same factory was also subcontracted without authorization. Then we demonstrate that the state dependence effect is stronger when the factory utilization is higher. Based on the extant literature, we hypothesize three additional order-level characteristics that could drive unauthorized subcontracting: price pressure, buyer reputation, and factory specialization. We find ample empirical support for price pressure but only partial evidence for the other two. The effect of price pressure leads to more unauthorized subcontracting when the unit price of an order is below the historic baseline for a given category and factory. For instance, the likelihood of unauthorized subcontracting increases by 11% when the price is 25% lower than usual. Note that our data was collected after the Rana Plaza collapse, which increased the worldwide scrutiny on unauthorized subcontracting. Therefore, our estimates are most likely lower compared to a business-as-usual scenario.

For the second research question we use a simple linear prediction model to detect unauthorized subcontracting at the order level. The model leverages our findings on the drivers of unauthorized subcontracting. Hence, it includes a factory (fixed) effect and order characteristics such as price pressure. We partition the data in training and testing subsamples and we find that more than 82% of the test orders can be correctly predicted using the model calibrated with the training data set. Depending on the data availability, the prediction accuracy can be above 90%. Moreover, the proportion of false positives (type I errors) and false negatives (type II errors) in the out-of-sample tests are well balanced. These results are robust to the training and testing data split, which indicates that, to a great extent, unauthorized subcontracting is predictable, and therefore, can be managed.

Our work contributes to the literature on supplier compliance and supply chain visibility. The goal of the study is not to incite finger-pointing between brands and suppliers but rather demonstrate that the use of data and models can help in addressing the issue of unauthorized subcontracting. For instance, some variation of our models can be embedded in a decision support system and the likelihood of unauthorized subcontracting can be measured for each order-factory pair, which can help inform the decision on where to place orders. Our results also show the need for more collaboration between buyers and suppliers. Though price is always important, we find that working closely with the supplier to avoid situations (or states) that push the factory into unauthorized subcontracting can be even more effective. For instance, coordinated planning and scheduling can prevent periods of excessive workload. On the other hand, supply chain visibility can be increased by streamlining the process to authorize suppliers. In this paper we focus on building compliance but our findings could apply to other types of compliance that suffer from unauthorized subcontracting.

In the remainder of the paper, USC is used as short for unauthorized subcontracting. The sole term "subcontracting" will be used as a synonym of "unauthorized subcontracting" when the latter is implied by the context. In addition, "supplier" and "factory" will be used interchangeably as all of the factories in our data belong to different firms (suppliers).

1.2. Literature Review

Unauthorized subcontracting, before the Rana Plaza collapse, did not have much attention in the academic literature. However, two streams of literature are closely related to this topic: on the one hand, papers related to compliance with environmental and safety standards; on the other hand, papers related to supply chain design. For a general review on sustainable operations management, see Drake and Spinler (2013).

What determines factories' compliance level? Some factors listed in the literature are external to suppliers and policy-related, such as differences between countries, related to differences in regulations, inspections and freedom of press (Toffel et al. 2015). In addition to the social characteristics of the supplier's location and regulation levels, compliance levels are also related to the characteristics of supplier-buyer relationships and their degree of involvement, and to strategic decisions such as the adoption of lean manufacturing standards (Locke et al. 2009, Distelhorst et al. 2016). In China, severe price pressure from Western buyers, short-term opportunism, lack of observability, and corruption and passiveness in the government are factors that induce suppliers to adulterate their products (Tang and Babich 2014). Some operational drivers of compliance and violations of labor standards are price pressure, production complexity and supplier-buyer contract duration (Jiang 2009). We study unauthorized subcontracting as a special case of compliance, and focus mainly on its operational drivers, but on an order-to-order basis, to observe how subcontracting decisions change under different order characteristics and workload levels.

A large part of the literature on compliance is focused on inspections and audits. Inspections can decrease the incentive for the supplier to adulterate products (Babich and Tang 2012). However, they can also be counterproductive since they can induce suppliers to a lack of transparency (Locke et al. 2007). In some circumstances, scheduled inspections are more efficient than random ones as an incentive for the firm to disclose environmental violations (Kim 2015). Among the papers on the effect of inspections in compliance, one is particularly related to our work: Plambeck and Taylor (2016), who analyze the effects of audits on compliance with labor and environmental standards, and on unauthorized subcontracting. They show that the potential for a supplier to pass an audit through hiding reduces the optimal level of auditing for the buyer when either the margin or the buyer's damage for sourcing with violations are small. Their results hold for the case of unauthorized subcontrating: when the supplier will likely be able to hide that it is subcontracting, more auditing from the buyer can backfire and cause the supplier to exert more effort in hiding its subcontracting practices than in avoiding subcontracting altogether.

Incentives and penalties to increase environmental and social compliance have been vastly studied. Incentives have a higher effect than penalties in violation reduction (Porteous et al. 2015). Some mechanisms that can prevent suppliers from adulterating their products, in addition to inspections, are deferred payments (Babich and Tang 2012). Chen and Lee (2016) examine the interactions of three instruments to increase compliance (supplier certification, audits, and contingency payments) and find that they are complementary and, when used jointly, supplier screening becomes more effective, and sourcing costs decrease. The relationship type with the buyer, country, age, and degree of asset specialization are some factors that determine the incentives for suppliers to adhere to environmental standards when buyers request it (Delmas and Montiel 2009). If we focus on buyers, brand visibility and value, serving for European markets, and the density of NGOs in the country where the buyer has its headquarters increase its sustainable sourcing practices (Thorlakson et al. 2018). Suppliers are more likely to be compliant if they are in more regulated countries, with higher freedom of press, and producing for buyers which are based in wealthier countries with socially aware consumers (Toffel et al. 2015). In our paper, in addition to price, we find an effect of buyer's popularity, i.e., unauthorized subcontracting is less likely when the buyer is a major brand that can suffer reputation damage.

Regarding supply chain structure, Orsdemir et al. (2015) find the optimal vertical integration and compliance enforcement decision as a function of both the probability and the economic impact of the public disclosure of a violation of standards. They conclude that supply chain partnership can greatly improve an industry's compliance. Different sourcing policies and the influence of supplier competition on the supplier's willingness to switch to a sustainable process are studied in Agrawal and Lee (2016), who show that, when sourcing from a unique supplier, buyers should require them to switch but, in the presence of supplier competition, a the buyer should adopt a sustainability-preferred policy, as opposed to sustainability-required. Greater downstream competition, a more concentrated supplier base, and a less flexible supply chain make a retailer more likely to source responsibly (Guo et al. 2016). In a multi-tier supply chain, increased pressure from external stakeholders (consumers, NGOs, and governments) can decrease supply chain responsibility when the manufacturer shifts from direct monitoring all the supply chain to delegating such inspections (Huang et al. 2017). However, all these works refer to supply chains in which all the tiers are known to the buyer, as opposed to our work, in which buyers are not aware of the existence of second tier suppliers.

1.3. Research Setting

1.3.1 Data Collection.

The data for this paper was obtained from M., a global supply chain manager (or "middleman") located in Asia that links buyers to suppliers and supervises the whole production process. The middleman groups the buyer accounts in different lines of business. The dataset used in our analysis represents all the orders in one line of business B., which is mostly mass-market apparel products for (lesser-known) brands in North America. The buyers and product categories in this line of business are further described in Tables 1.3 and 1.4.

The middleman maintains a list of authorized factories. These are factories that have passed a building compliance inspection. The middleman also performs a conformity audit of each order to ensure that it meets the agreed specifications.[1] The conformity audits of the orders are independent from the compliance inspection of the facilities. The latter happen

[1] According to ISO 9000:2005 definition 3.6.1, conformity is fulfillment of a requirement. In this paper we reserve the term "conformity" for order-level requirements and we use the term "compliance" for factory-level adherence to (established) building regulations.

infrequently (e.g., once a year) and are performed by building inspectors. In contrast, the conformity audits are performed by employees of the middleman for each order. These audits are usually preannounced and scheduled during the order production or soon after its completion.

After the Rana Plaza collapse, the line of business B. decided to record when orders were subcontracted to unauthorized factories, i.e., factories that were not on the authorized list. This information was gathered during the routine conformity audit of each order. Suppliers were not aware that their subcontracting practices were being recorded and no penalty was applied.[2] Hence, it is safe to assume that the data is reliable as there was no incentive for factories to hide occurrences of unauthorized subcontracting. However, the data does not contain any information on who the subcontractor was and the date of the conformity audit was not recorded. There is no information either on any orders that the factories produced for other buyers outside our dataset, but Table 1.2 shows that the line of business B. used a large portion of the factory capacity available.

1.3.2 Data Description.

Our data consists of all the orders delivered by the line of business B. in the eight months between October 2013 and May 2014. There are 32,477 orders that span 226 factories, 30 buyers, and 34 categories. An order represents one observation in our dataset and it is defined as a purchase request of a SKU from a buyer to a supplier through the middleman. The data contains many characteristics of each order and of the factory that processed it, as well as the identification of the buyer. There is also a binary variable (USC) stating whether or not the order was subcontracted without authorization (YES=1). A sample of an observation (order) is given in Table 1.1.

Note that noncompliance can only happen at unauthorized facilities, as was the case in all the disasters described in §1.1, but not all unauthorized facilities are necessarily non-

[2] If the order was subcontracted, the supplier would tell the conformity auditor where it had been sent.

compliant. In other words, noncompliance implies USC but the converse implication is not necessarily true. Hence, USC is not a direct measure of noncompliance but rather should be seen as a proxy for increased compliance risk. Alternatively, an incident of USC can be seen as a lack of supply chain visibility.

An additional data set corresponding to orders delivered between November 2010 and October 2013 was also available. However, it did not contain any information on USC, but just order characteristics. This second dataset was used to compute the historical prices and specialization of each factory.

Factory ID	**Factory capacity**	**Factory country**	**Buyer ID**	**Product category**
XXXXXX	10^5 units/month	Vietnam	YYYYYY	Sleepwear
Quantity	**Price**	**PO date**	**Scheduled due date**	**Delivery date**
144 units	ZZZ USD/unit	2013-09-10	2013-12-12	2013-12-12
Unauthorized subcontracting (USC)				
YES				

Table 1.1: Example of one order. The factory, buyer and price information have been disguised.

Figure 1.1 shows a histogram of the percent of orders that each factory subcontracted without authorization. As shown in the figure, unauthorized subcontracting behavior is not evenly distributed among factories. Whilst some factories never subcontracted any order (Type A), others decided to subcontract on an order-to-order basis (Type B), and a third group of factories subcontracted all of their production (Type C). As we will explain in Section 1.4 and 1.5, the differences in unauthorized subcontracting behavior imply that we must account for managerial decisions at the factory level, and not only for the characteristics of the orders that the factories received. Table 1.2 reports summary statistics of the factories. From this table it can be seen that most of the factories are located in China and they produced orders for 1.4 buyers on average. Though we do not have information on other buyers that the factories might have work with, Table 1.2 shows that the workload due to orders from the middleman represents 42.6% of the factory's capacity on average, with a large standard deviation.

	Mean	St. dev.	Min.	Max.	% Factories	% Orders	% USC
Factory descriptive statistics:							
Capacity (monthly units)	388,929	864,181	10,000	10,000,000	-	-	-
Two-month utilization (%)	42.6	60.7	0.3	531.9	-	-	-
Number of categories produced	2.8	2.1	1	10	-	-	-
Number of buyers served	1.4	1.0	1	11	-	-	-
Orders delivered	143.7	270.6	1	2,772	-	-	-
Order delivery delay (days)	1.1	7.6	-84	141	-	-	-
Order size (units)	2,508	5,824	1	240,491	-	-	-
Orders from major brands (%)	1.9	10.4	0	100	-	-	-
Factory type:							
Type A - Never subcontracted	-	-	-	-	56.6	35.7	0.0
Type B - Sometimes subcontracted	-	-	-	-	31.9	57.9	51.6
Type C - Always subcontracted	-	-	-	-	11.5	6.4	100.0
Factory location:							
Bangladesh	-	-	-	-	10.6	12.8	0.1
Cambodia	-	-	-	-	5.8	5.4	52.0
China	-	-	-	-	59.7	54.8	47.9
Indonesia	-	-	-	-	6.6	13.1	16.0
Vietnam	-	-	-	-	9.3	9.1	54.4
Other	-	-	-	-	8.0	4.9	4.6

Table 1.2: Description of the factories. Two-month utilization is the workload divided by the factory's capacity for two-month subperiods, see 1.5.3. Order delivery delay is the difference between the actual and scheduled delivery dates. A major brand is one that has its own stores and website, see Section 1.4.4. The factory types are defined in Figure 1.1. %USC is the frequency of USC.

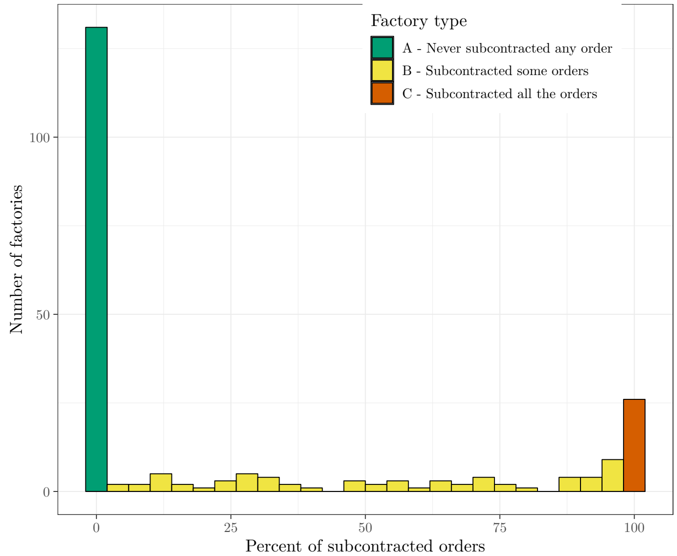


Figure 1.1: Histogram of percent of subcontracted orders by factory. The light bar, grey bars, and black bar correspond to 128, 72, and 26 factories, respectively.

Table 1.3 shows summary statistics of the buyers. It can be seen that, on average, the buyers ordered products in 3.4 categories and they worked with 10.7 different factories. The vast majority of the orders are coming from buyers located in North America, and most of the buyers represent brands that do not have their own stores or website. These smaller brands are more likely to use the middleman M. for most of their sourcing, though there is no exclusivity agreement in place. Finally, Table 1.4 contains a list of the top product categories in the data.

1.4. Hypotheses

The dependent variable in our analysis is the binary variable USC_{ij} that equals 1 if unauthorized subcontracting occurred, i.e., it equals 1 when order j of supplier i was subcontracted to a factory that was not on the authorized list of the middleman M. We propose four hypotheses to explain the main drivers of USC_{ij} that we test using the variables described below. Note that the academic literature on unauthorized subcontracting is meager, so we

	Mean	St. dev.	Min.	Max.	% Buyers	% Orders	% USC
Buyer descriptive statistics:							
Orders placed	1,082.6	2,387.5	1	9,652	-	-	-
Number of product categories ordered	3.4	3.3	1	11	-	-	-
Number of factories engaged with	10.7	17.6	1	61	-	-	-
Buyer recognition:							
Major brand (has own stores and website)	-	-	-	-	46.7	2.2	5.9
Not major brand	-	-	-	-	53.3	97.8	37.0
Buyer location:							
North America	-	-	-	-	63.4	99.3	36.6
Asia	-	-	-	-	13.3	0.2	0.0
Europe	-	-	-	-	10.0	0.2	0.0
Other	-	-	-	-	13.3	0.3	0.0

Table 1.3: Description of the buyers. The product categories are described in Table 1.4.

	% Orders	% Units	# Factories	# Buyers	% USC
Bottoms	6.8	9.7	72	9	44.8
Dresses	3.2	1.9	35	8	39.5
Jackets	1.6	1.2	32	8	52.4
Pants	8.3	7.1	79	8	34.7
Polos	6.2	6.1	25	7	60.5
Sets	11.8	11.1	42	2	65.0
Shirts	17.0	8.7	63	22	23.3
Sleepwear	6.8	11.5	12	4	21.0
Swimwear	3.0	2.8	13	2	21.8
Tops	19.8	23.7	126	11	32.0
T-shirts	6.5	3.9	33	8	19.4
Other	9.1	12.2	97	14	35.7

Table 1.4: Descriptive statistics of the top 11 product categories by number of orders. # Factories is the number of factories that produced products in a category. # Buyers is the number of buyers that placed orders in a category.

support some of our hypotheses drawing from the related literature on compliance of labor and safety standards.

1.4.1 State Dependence

Hypothesis 1. *Unauthorized subcontracting is more likely when the previous order in the same factory was subcontracted.*

A factory can be seen as a queueing system in which the state is given by the queue length or workload. When a factory is operating close to capacity there is pressure to engage in unauthorized subcontracting to satisfy scheduled delivery dates (Labowitz and Baumann-Pauly 2014, 2015). Hence, an order is more likely to be diverted to an unauthorized facility when the workload is high than when it is low. For this reason, we argue that unauthorized subcontracting exhibits state dependence.

The excessive workloads are caused by several factors. First, more orders means more business, which eventually translates into higher revenues. Second, rejecting an order can set a precedent and factories might fear that the buyer will not place orders there in the future. Third, the market is changing rapidly: tighter lead times, late sample approval and last-minute alterations to product specifications put increased pressure on factories, e.g., buyers will often delay their purchasing decisions until the last minute to be able to observe and react to what their competitors do (Ashida and Plinke 2004, Vaughan-Whitehead and Pinedo Caro 2017). Fourth, some factories exhibit poor production planning, such as lack of awareness of the critical path and inaccurate forecasting of the necessary time to complete an order (Hurst et al. 2005). Finally, buyers that have invested in improving the compliance of their suppliers inadvertently make those suppliers more attractive to other "free-riding" buyers that will also want to place their orders there (Greenhouse 2013, Plambeck and Taylor 2016).

In our dataset, we do not observe queueing levels at the arrival time of each order. However, we can expect unauthorized subcontracting to present an autoregressive behavior.

Indeed, consecutive orders face similar queue lengths, especially when the system utilization is high. For instance, in the classic $M/M/1$ queue, the chance that an incoming order finds a queue longer than the previous order is $\frac{\rho}{1+\rho}$ and this probability is increasing in the utilization level ρ. Therefore, if order $j-1$ at a given factory is subcontracted in an unauthorized manner, then we expect the likelihood of unauthorized subcontracting for order j to increase, all else remaining equal.

We test this hypothesis by introducing an autoregressive term $USC_{i,j-1}$ in our regression model. We also include factory fixed effects and several sets of dummies in order to distinguish between structural and spurious dependence (Heckman 1981). We acknowledge that, in addition to utilization, there can be other causes of state dependence. Suppliers could be reserving capacity for other buyers for which we lack visibility, or might be affected by the overall financial health of the company. Though we cannot completely rule out these alternative explanations, in §1.5.3 we strengthen our point by looking more closely at the role of factory utilization.

1.4.2 Price Pressure

Hypothesis 2. *Unauthorized subcontracting is more likely for orders that have a lower unit price.*

A lower price makes an order less profitable and could give the supplier incentives to divert it to an unauthorized third party which has lower production costs by being non-compliant. This is a common view in the public press (Goodman 2013) and is supported by industry-wide surveys (Vaughan-Whitehead and Pinedo Caro 2017). Several studies in the academic literature report results linking price pressure to non-compliant behavior. For instance, Jiang (2009) reports that price pressure is one of the main reasons that led to violations of supplier codes of conduct. Similarly, Tang and Babich (2014) mention that pressure from Western manufacturers to deliver products at persistently lower prices is a crucial factor that pushes Chinese suppliers to "cut corners" and adulterate products. In contrast, the work by Plambeck and Taylor (2016) shows that, under a certain backfiring condition, price pressure could lead to less unauthorized subcontracting, which could seem to run against Hypothesis 2. However, the backfiring condition is unlikely to hold in our setting since there was no additional compliance inspection effort when the data was collected and unauthorized subcontracting was not penalized.

We measure price pressure by comparing the unit price of an order with the moving-average in same factory and category. Specifically, let $Price_{ij}$ be the unit price of order j at factory i. Let $c(j)$ be the product category of order j and let $P_i^{c(j)}$ be the average price for category $c(j)$ at factory i. Then we define the continuous variable $RelPriceFC_{ij} = \frac{Price_{ij} - P_i^{c(j)}}{P_i^{c(j)}}$. To compute $P_i^c(j)$ we used all the orders of category $c(j)$ delivered by factory i during the calendar year prior to the delivery date of order j. Note that the definition of $RelPriceFC_{ij}$ is consistent with the notion of price pressure described in Jiang (2009) where factory managers felt pressure when the price they were paid for an order was lower than their usual price for that type of product. We test alternative measures of price pressure as robustness checks in §1.5.4.

1.4.3 Factory Specialization

Hypothesis 3. *Unauthorized subcontracting is more likely for orders of product categories other than the factory's main specialization.*

In the strategy literature it is well understood that specialization can be a source of competitive advantage (Dyer 1996). Moreover, production switchovers are costly in a multi-product manufacturing setting and are hard to manage (Rosa-Hatko and Gunn 1997, Cheng et al. 2000), which reinforces the specialization in a smaller set of products. Once a factory is specialized, activities that are outside its core are natural candidates to be outsourced to a third party. The tendency to outsource non-core activities is supported by the usual outsourcing frameworks in operations strategy (Van Mieghem 2008).

The factories in our dataset exhibit a high degree of specialization. Table 1.2 shows that the factories produce less than three product categories on average. Indeed, 73% of the factories produced at most three product categories, and only two factories produced 10 different product categories. Hence, we hypothesize that an order of a product category that a factory does not produce very often may have a higher probability of being subcontracted because it is outside the factory's core activities. A complementary view is that the factory might not have the technical capability or might simply want to avoid production switchovers, so this specific order is more complex for the factory. Note that the degree of complexity is specific to each factory-category pair. For instance, an order of pants may be complex for a factory that is specialized in sweaters, but not complex for a factory that regularly produces pants.

To measure factory specialization, we need to take into account what types of products each factory produces more often. Hence, we define $PropThisCategory_{ij}$ as the proportion of orders at factory i of the same category as order j during the calendar year prior to when order j was delivered. We use the continuous variable $PropThisCategory_{ij}$ to test Hypothesis 3.

1.4.4 Buyer Reputation

Hypothesis 4. *Unauthorized subcontracting is less likely when the buyer is a well-known brand.*

The evidence in practice has demonstrated that stronger brands face a higher cost and more reputation damage when disasters related to non-compliance occur. After the Rana Plaza collapse, some of the famous retailers whose products were being made there had to pay compensations to families of the victims (Butler 2014, Smithers 2015). There is also evidence that "name and shame" schemes are more effective with high-profile brands (Lee and Plambeck 2009). For instance, Apple agreed to monitor the pollution of its factories after being exposed by the Institute of Public and Environment Affairs (IPE), which is a nonprofit organization based in China (Nuttall 2012).

Given the evidence in practice, a standard assumption in supply chain compliance models is that the buyer faces a penalty or goodwill loss if the supplier commits a compliance violation (Plambeck and Taylor 2016, Guo et al. 2016, Caro et al. 2016, Huang et al. 2017). This penalty is higher for buyers that have more at stake and intuitively it should lead to actions that induce the supplier to be more socially responsible. In fact, Plambeck and Taylor (2016) in their model show that unauthorized subcontracting is less frequent when the buyer has more at stake as long as compliance auditing efforts do not backfire.[3] Similarly, Huang et al. (2017) find that greater pressure focused on the retailer decreases the chance of a violation in the supply chain.

We postulate that unauthorized subcontracting is less likely for an order placed by a well-known brand. This could be because the well-known brand will exert a higher effort in preventing irresponsible behavior, as in the single-period models cited above, or it could simply be because the damage cost of the supplier is correlated with the violation penalty faced by the buyer. Indeed, the supplier knows that, if a disaster happens, the attention of the press will be higher for orders of well-known brands and any penalty faced by these brands could spill over to her.

Some of the buyers placing orders to factories in our data correspond to major retailing brands, while others are distributors or brands that are not as well-known to the public. We define the binary variable $BuyerIsMajorBrand_{ij}$, which takes value 1 if the buyer that placed order j at factory i has its own website and its own network of stores, and 0 otherwise. The buyers in our data which we labeled as major brands correspond to retailers with hundreds of stores in multiple countries and an online store in their website, whereas the other buyers sold through third-party retail channels, e.g., department stores.[4] We use the variable $BuyerIsMajorBrand_{ij}$ to test Hypothesis 4.

[3]Proposition 12 in (Plambeck and Taylor 2016) requires that the supplier does not have incentives to hide his behavior, which was the case in our setting.

[4]It is common in the literature to identify major buyers as those that are publicly traded. However, such criteria does not discriminate in our setting because 99.8% of the orders were placed by publicly traded companies.

1.5. Methods and Results

1.5.1 Model Specification.

We propose the following linear regression model to explain unauthorized subcontracting:

$$USC_{ij} = \alpha USC_{i,j-1} + \beta x_{ij} + f_i + \varepsilon_{ij}, \tag{1.1}$$

where f_i are time-invariant factory fixed effects and ε_{ij} is the error term. The vector x_{ij} contains the independent variables used to test the hypotheses, i.e., $RelPriceFC_{ij}$, $PropThisCategory_{ij}$, and $BuyerIsMajorBrand_{ij}$, as well as the control $LogOrderSize_{ij}$, which is the logarithm of $OrderSize_{ij}$, the size of order j at factory i. Table 1.5 shows descriptive statistics of these variables. Collinearity is discussed in Appendix A.1.1.

	Mean	St. dev.	Min.	Max.
USC	0.363	0.481	0	1
RelPriceFC	-0.008	0.249	-0.879	4.103
PropThisCategory	0.610	0.351	0.001	1
BuyerIsMajorBrand	0.022	0.147	0	1
LogOrderSize	6.696	1.626	0	12.390

	Mean	St. dev.	Min.	Max.
USC	0.516	0.500	0	1
RelPriceFC	-0.011	0.263	-0.879	4.103
PropThisCategory	0.589	0.359	0.001	1
BuyerIsMajorBrand	0.019	0.135	0	1
LogOrderSize	6.665	1.605	0	12.390

Table 1.5: Descriptive statistics of the regression variables, using the full dataset (left) and the subsample of type B factories (right).

We use a linear model in order to introduce multiple fixed effects. An alternative probit model with random effects is described in Appendix A.1.4. The factory-level fixed effect f_i in Equation (1.1) captures the factory effect that is apparent in Figure 1.1 and it accounts for the idiosyncratic baseline of unauthorized subcontracting of each factory, i.e., the factory's managerial decision on whether or not to subcontract at all, and if so, how often on average. This (unobservable) fixed effect can also be viewed as the supplier's intrinsic ethical level, which is a common assumption in theoretical models (e.g., see Chen and Lee 2016). To mitigate possible sources of omitted variable bias, we introduce additional dummy variables for product category, delivery month, and buyer.[5] Finally, factories may be strategic about

[5] When the dummies per buyer are introduced we have to drop the binary variable $BuyerIsMajorBrand_{ij}$ because they are collinear.

which orders to subcontract depending on the order size. Therefore, we control for order size using the variable $LogOrderSize_{ij}$. We opt for a logarithmic form given the large variability of order quantities.

Since the right hand side in Equation (1.1) has a lagged variable and a fixed effect, we rely on the first difference equation $\Delta USC_{ij} = \alpha \Delta USC_{i,j-1} + \beta \Delta x_{ij} + \Delta \varepsilon_{ij}$ and the Arellano-Bond estimator (Wooldridge 2010, Greene 2003, Cameron and Trivedi 2005, 2010). Following the Arellano-Bond approach, the correlation between $\Delta USC_{i,j-1}$ and $\Delta \varepsilon_{ij}$ is addressed by using $USC_{i,j-2}$ as an instrumental variable (IV). This is a strong instrument, as $\Delta USC_{i,j-1}$ and $USC_{i,j-2}$ are highly correlated. It is also a valid one since it is uncorrelated with the error term, insofar the errors $\Delta \varepsilon_{ij}$ are not autocorrelated (we check the veracity of this assumption in Appendix A.1.2).

There is a possibility of simultaneity between price and unauthorized subcontracting. If either the buyer or the middleman have some visibility on a factory's past unauthorized subcontracting behavior, they will use this information to set the prices they offer, i.e., a factory's past subcontracting may affect the current prices of the orders it receives. For instance, the buyer can decide to offer a low price for an order to a supplier if it suspects that the supplier will likely subcontract it, and so it will be less costly to produce if not all safety standards are met. To account for the possible endogeneity of the price pressure variable, we also use an IV approach: $RelPriceFC_{i,j-2}$ will serve as an instrument for $\Delta RelPriceFC_{i,j}$. Note that, to be able to include second lags of the dependent variable and the price covariate as instruments, we need to remove the first two observations of each factory from the regression.

The coefficients of the first-differences IV model were computed using generalized method of moments in Stata 14.2. In Appendix A.1.4, we test the robustness of our results to a different estimation technique, the two-step least squares (also called the Hsiao estimator).

1.5.2 Main Results.

Table 1.6 shows the coefficient estimates of the Arellano-Bond regression. We show the results for two data sets: (i) including all the factories in our data, and (ii) including only type B factories (those that subcontracted some, but not all, of their orders). The reason for running our regression using this subsample is that these are factories whose subcontracting behavior was established on an order-to-order basis and not as a decision that determined all their orders, so individual order characteristics should have a more important effect on the decision to subcontract than for factories whose behavior is fixed and constant over time.

As shown in the left panel of Table 1.6, the coefficient for the lagged dependent variable is positive and highly significant, ranging from 0.295 to 0.299, robust to the inclusion of category, buyer and month dummies. The interpretation of this coefficient is that, when a factory subcontracts an order without authorization, the next order this factory delivers is about 30 percent points more likely to be subcontracted too than if the previous order had not been subcontracted (all else equal). The $USC_{i,j-1}$ coefficient is slightly larger if we only use the 72 type B factories, ranging from 0.296 to 0.302, and is still highly significant ($p < 0.001$). Hence, we find support for Hypothesis 1 as unauthorized subcontracting in our dataset exhibits a high and significant level of state dependence. The magnitude of this effect must be underscored. To see this, consider that the overall prevalence or baseline of unauthorized subcontracting is 36%. Then, the state dependence effect almost doubles the chance of unauthorized subcontracting with respect to the baseline. In §1.5.3 we show that this effect increases when the factory utilization is higher.

The second largest coefficient in magnitude is that of the price pressure. It ranges from -0.167 to -0.156 for the full sample and it is statistically significant at the 0.01 level. This results supports Hypothesis 2 and it means that, when a buyer offers a unit price for an order that is lower than the factory's usual price for that product category, the probability of unauthorized subcontracting increases. For instance, if the price pressure of an order is such that the unit price is 25% lower than the usual price, the chance of unauthorized subcontracting would increase by (at least) $(-15.6)\times(-25\%) = 3.9$ percentage points. When we consider only type B factories in the regression, the coefficients of the price pressure variable are larger in magnitude and range from -0.188 to -0.176.

The coefficients for the binary variable $BuyerIsMajorBrand_{ij}$ range from -0.0371 to -0.0353 ($p < 0.05$) for the full sample. Hence, an order for a major brand is about 3.7 percentage points less likely to be subcontracted without authorization, which in terms of magnitude is almost the same effect than paying a 25% price premium with respect to the average price. Since the overall subcontracting baseline is 36%, we find that orders from major brands in the full sample are 10% less likely to be subcontracted. The significance of this result is lost in the regression with only type B factories. Therefore, we only have partial support for Hypotheses 4, but this could be due to low percentage of orders from major brands in the dataset: only 2.2% in the full sample and 1.9% in the subsample, as shown in Table 1.5. We further explore the role of the buyer by considering a supplier-buyer panel structure in Section 1.5.5.

Finally, the coefficient estimates for the variable $PropThisCateg_{ij}$ were statistically insignificant in all cases. Hence, we find no empirical evidence to support Hypotheses 3 at the order level. However, in Appendix A.1.3 we show that the variable $PropThisCateg_{ij}$ is highly significant when it is considered jointly with the other order-level variables $RelPriceFC_{ij}$ and $BuyerIsMajorBrand_{ij}$. Moreover, in §1.5.6 we find evidence that factories that were less specialized did subcontract a larger proportion of their orders.

1.5.3 Relationship Between State Dependence and Factory Utilization.

The results in the previous section show that the main driver of unauthorized subcontracting in our dataset is state dependence. The state of a factory is multidimensional but clearly a key component is the utilization or workload (Vairaktarakis 2013). Our goal in this section is to relate our state dependence result to the factory utilization. One challenge is that the factory production schedule is not part of our dataset. Therefore, including the utilization directly in our regression model would create an endogeneity problem because any proxy variable for factory utilization is likely to have measurement error. To overcome this challenge, we repeat

	Unauthorized subcontracting			
	Arellano-Bond			
	(1)	(2)	(3)	(4)
$USC_{i,j-1}$	0.297***	0.295***	0.299***	0.295***
	(0.0278)	(0.0278)	(0.0243)	(0.0242)
RelPriceFC	-0.156**	-0.161**	-0.163**	-0.167**
	(0.0605)	(0.0594)	(0.0596)	(0.0616)
PropThisCategory	0.0293	0.00307	0.00514	0.00102
	(0.0343)	(0.0341)	(0.0342)	(0.0343)
BuyerIsMajorBrand	-0.0355**	-0.0353*	-0.0371*	
	(0.0120)	(0.0147)	(0.0152)	
LogOrderSize	-0.00356	-0.00365	-0.00374	-0.00391
	(0.00256)	(0.00260)	(0.00250)	(0.00269)
Category dummies	No	Yes	Yes	Yes
Month dummies	No	No	Yes	Yes
Buyer dummies	No	No	No	Yes
N	32028	32028	32028	32028

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Unauthorized subcontracting			
	Arellano-Bond			
	(1)	(2)	(3)	(4)
$USC_{i,j-1}$	0.299***	0.296***	0.302***	0.298***
	(0.0282)	(0.0283)	(0.0232)	(0.0228)
RelPriceFC	-0.176**	-0.183**	-0.185**	-0.188**
	(0.0617)	(0.0593)	(0.0591)	(0.0626)
PropThisCategory	0.0676	0.0189	0.0178	-0.0116
	(0.0566)	(0.0580)	(0.0575)	(0.0570)
BuyerIsMajorBrand	-0.0612+	-0.0594	-0.0661	
	(0.0317)	(0.0396)	(0.0402)	
LogOrderSize	-0.00506	-0.00534	-0.00480	-0.00531
	(0.00387)	(0.00394)	(0.00359)	(0.00388)
Category dummies	No	Yes	Yes	Yes
Month dummies	No	No	Yes	Yes
Buyer dummies	No	No	No	Yes
N	18646	18646	18646	18646

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.6: Coefficients of the Arellano-Bond regression, using all the data (left) and the subsample of type B factories (right).

the regression in the previous section for different subsets of the data when the utilization varied and then observe the changes in the coefficient for the lagged subcontracting variable $USC_{i,j-1}$.

We split the eight months of data in four two-month subperiods and we compute the utilization of each factory in the following way. For each two-month subperiod, we take the ratio of units delivered to factory capacity during that subperiod. The factory capacity in our data is reported in units per month, so in a two-month subperiod factory i can deliver $2Capacity_i$. Let J_m^i be the set of orders delivered by factory i during month m. Hence, the utilization of factory i during the two-month subperiod $[m, m+1]$ is given by $Utilization_i^m = \frac{\sum_{j \in J_m^i \cup J_{m+1}^i} OrderSize_{ij}}{2Capacity_i}$. Note that this workload metric is a lower bound of the factory's true workload. Indeed, factories may be producing for buyers that do not place their orders through the middleman M., in which case their true workload would be higher than the one computed here.

Figure 1.2 shows a box plot of the two-month utilization for the four subperiods. The figure suggests that the workload was higher in subperiods 2 and 4 (Dec 13-Jan 14 and Apr 14-May 14) than in subperiods 1 and 3 (Oct 13-Nov 13 and Feb 14-Mar 14). Indeed, the average workload in the former was 52% and 49% whereas in the latter it was 35% and 31%, respectively. The overall average was 43%. Therefore, subperiods 2 and 4 had above-average utilization and subperiods 1 and 3 had below average. Note from Figure 1.2 that several factories had a utilization above 100%, especially in subperiods 2 and 4. Quite often factories exceed their nominal capacity, which usually leads to overtime (Vaughan-Whitehead and Pinedo Caro 2017), but the fact that this is more prevalent in subperiods 2 and 4 confirms that these are the busier subperiods in our data.

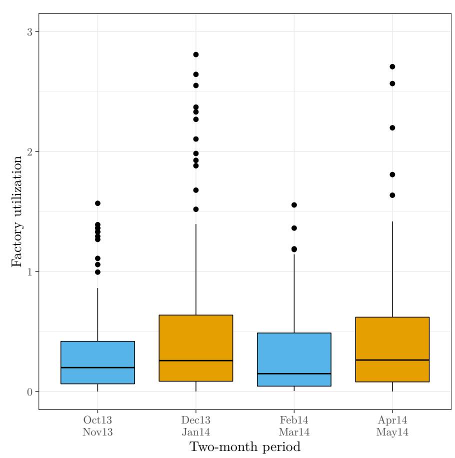


Figure 1.2: Box plot of the two-period utilization ($Utilization_i^m$) per subperiod.

Table 1.7 shows the results of the Arellano-Bond regression for each subperiod. We focus on the results for the full sample (left panel). The results for the type B subsample (right panel) are similar. From Table 1.7 we observe that the magnitude of the coefficient of the lagged variable $USC_{i,j-1}$ is larger and more significant when the (average) factory utilization is higher. The coefficient for subperiod 1 is 0.164, significant at the 0.01 level, and for subperiod 3 the coefficient is statistically insignificant. However, for the two subperiods with above-average utilization, this coefficient is 0.257 and 0.326, respectively, and significant at the 0.001 level. In other words, a subcontracted order is more likely to be followed by another subcontracted order when the workload is higher with respect to factory capacity. One plausible interpretation of this result is that when the factory is busier, it might enter a "subcontracting streak" in which several (consecutive) orders are diverted to unauthorized facilities. In contrast, in periods of low utilization the queue of orders is manageable so unauthorized subcontracting, if it happens, is more of an isolated event and is less likely to be followed by another case of subcontracting. In line with this result, in §1.5.6 we see that factories with a higher mean utilization showed a higher degree of unauthorized subcontracting. To recap, we find stronger support for Hypothesis 1 in subperiods when the factory utilization is higher.

1.5.4 Alternative Price Pressure Measures.

In this section use four alternative measures of price pressure to check the robustness of our result regarding Hypothesis 2. The first one is $RelPriceFCB_{ij}$, which is computed similarly to $RelPriceFC_{ij}$ described in §1.4.2 but taking into account the different buyers for which a factory produces. Let $cb(j)$ be the category-buyer pair of order j and let $P_i^{cb(j)}$ be the one-year price moving average at factory i of orders from the same product category and buyer than order j. Then, we define $RelPriceFCB_{ij} = \dfrac{Price_{ij} - P_i^{cb(j)}}{P_i^{cb(j)}}$.

The second alternative measure of price pressure is $RelPriceOrd_{ij}$. This variable compares an order's unit price with the expected price of orders of similar characteristics based on size, lead time, and product category. To compute this variable, for every delivery date t, we take all the orders delivered during the previous year (by any factory) and regress the unit price on the order size (in logarithmic scale), scheduled lead time, and product category dummies. With this regression, we compute the predicted price for all orders delivered on day t. Notice that this prediction is agnostic to factory differences because it does not include any factory dummy. Then, order j delivered on $t(j)$ will have predicted unit price $P_j^{Ord,t(j)}$, where the superscript Ord refers to "order characteristics". Note that $P_j^{Ord,t(j)}$ is essentially the prediction of a hedonic pricing model that captures volume discounts, price

	Unauthorized subcontracting			
	Arellano-Bond			
	(1)	(2)	(3)	(4)
Subperiod	Oct 13– Nov 13	Dec 13– Jan 14	Feb 14– Mar 14	Apr 14– May 14
Mean utilization	35%	52%	31%	49%
$\text{USC}_{i,j-1}$	0.164**	0.257***	-0.00307	0.326***
	(0.0600)	(0.0237)	(0.157)	(0.0659)
RelPriceFC	-0.0528	-0.119*	-0.00961	-0.0750+
	(0.0527)	(0.0483)	(0.0503)	(0.0452)
PropThisCategory	0.127+	0.0334	-0.00451	-0.0768
	(0.0718)	(0.0373)	(0.0630)	(0.112)
LogOrderSize	-0.00173	-0.00427	-0.000333	-0.00223
	(0.00170)	(0.00282)	(0.00363)	(0.00307)
Category dummies	Yes	Yes	Yes	Yes
Buyer dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
N	9752	13712	2718	5846

Robust standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Unauthorized subcontracting			
	Arellano-Bond			
	(1)	(2)	(3)	(4)
Subperiod	Oct 13– Nov 13	Dec 13– Jan 14	Feb 14– Mar 14	Apr 14– May 14
Mean utilization	44%	66%	31%	64%
$\text{USC}_{i,j-1}$	0.252***	0.272***	0.0638	0.315***
	(0.0366)	(0.0384)	(0.0991)	(0.0638)
RelPriceFC	-0.0929	-0.133**	-0.0147	-0.0832+
	(0.0858)	(0.0485)	(0.0703)	(0.0466)
PropThisCategory	0.367*	0.0617	-0.0682	-0.136
	(0.150)	(0.0704)	(0.137)	(0.159)
LogOrderSize	-0.00120	-0.00593+	-0.00119	-0.00307
	(0.00300)	(0.00303)	(0.00667)	(0.00381)
Category dummies	Yes	Yes	Yes	Yes
Buyer dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes
N	4792	7704	1642	4508

Robust standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.7: Arellano-Bond (all-dummies model) using different subsets of data: (1) months 1 and 2, in which the average factory utilization is low; (2) months 3 and 4, in which the average factory utilization is high; (3) months 5 and 6, in which the average factory utilization is low; (4) months 7 and 8, in which the average factory utilization is high. Left panel has all factories. Right panel is for type B factories.

premiums for rush orders, and the buyers' willingness to pay across categories. We use this predicted price as a reference point for the price pressure variable. Namely, we have $RelPriceOrd_{ij} = \dfrac{Price_{ij} - P_j^{Ord,t(j)}}{P_j^{Ord,t(j)}}$.

Next, the variable $RelPriceOrdF_{ij}$ is defined in the same way as $RelPriceOrd_{ij}$, but including factory dummies in the regression model. Therefore, $P_{ij}^{OrdF,t(j)}$ does take into account differences between factories, apart from differences between orders, and $RelPriceOrdF_{ij} = \dfrac{Price_{ij} - P_{ij}^{OrdF,t(j)}}{P_{ij}^{OrdF,t(j)}}$.

Finally, it is natural to associate price with quality. Our conversations with the middleman M. indicate that the quality within each factory in our dataset tends to be quite uniform, but we do not have information about the quality characteristics of the orders placed at each factory, so we are not able to validate this claim empirically. We try to address this issue by considering the variable $RelPriceFSub_{ij}$ that is computed in the same way as $RelPriceFC_{ij}$, but dividing each category into two subcategories depending on price. Specifically, for each factory i and order j, if $Price_{ij} \geq P_i^{c(j)}$, we classify order j into the subcategory "orders of $c(j)$ with high price"; otherwise, its subcategory is "orders of $c(j)$ with low price". The underlying assumption in this classification is that price correlates with quality, so products with a price above (below) average are more likely to be high (low) quality. Let $s(j)$ be the subcategory of order j and let $P_i^{s(j)}$ be the one-year moving average of price in that subcategory. Then, we measure price pressure within order j's subcategory, i.e., $RelPriceFSub_{ij} = \dfrac{Price_{ij} - P_i^{s(j)}}{P_i^{s(j)}}$.

The coefficient estimates of the full-dummies Arellano-Bond regression using different measures of price pressure is shown in Table 1.8. The first column uses $RelPriceFC_{ij}$, and is therefore the same as column (4) in Table 1.6. Columns (2)-(5) use the alternative measures described in this section. The variable $RelPriceFCB_{ij}$ is significant and has a coefficient in same order of magnitude than $RelPriceFC_{ij}$. The two price pressure measures that are based on order characteristics, $RelPriceOrd_{ij}$ and $RelPriceOrdF_{ij}$, are only significant in the regression using the type B subsample, which is consistent with the fact that type B factories are those that subcontract without authorization on an order-by-order basis. Finally, when we consider subcategories, the variable $RelPriceFSub_{ij}$ is highly significant in both samples, which provides further support for Hypothesis 2. It should be noted that, for the subsample with type B factories, all the coefficients of the different price pressure measures are remarkably in the same range.

1.5.5 Supplier-Buyer Panel Structure.

The panel structure in Table 1.6 is at the factory level and it includes an (unobservable) fixed effect for each factory. The latter is a valid model if a factory's baseline or intrinsic level of unauthorized subcontracting is the same regardless of the buyer placing the order. However, it is possible that the factories behave strategically depending on the buyer they are dealing with. For instance, if the order is placed by a preferred buyer, the factory might be less willing to engage in unauthorized subcontracting. We cannot observe the details of the buyer-supplier relationship, but we can redefine the panel structure of our data so every cross-section unit corresponds to a unique supplier-buyer pair. This panel structure allows us to include a fixed effect for each supplier-buyer pair.

Given that most suppliers worked with a small number of buyers (1.4 on average, see Table 1.2), there are 321 supplier-buyer pairs. Then, for every supplier-buyer pair, we compute variables equivalent to those described in §1.4 and repeat the Arellano-Bond regression now shown in Table 1.9. Note that in this case the model cannot include $BuyerIsMajorBrand_{ij}$ or buyer dummies, as they would be collinear with the fixed effects. The results with the supplier-buyer panel structure are consistent with those shown in Section 1.5.2. In particular, the coefficients and the significance of the state dependence ($USC_{i,j-1}$) and price pressure ($RelPriceFC_{ij}$) variables are very much the same, which provides further support for Hypotheses 1 and 2.

1.5.6 Unauthorized Subcontracting and Factory Characteristics.

Figure 1.1 shows that factories exhibit different propensity to engage in unauthorized subcontracting. Do any factory characteristics correlate with such differentiated behavior? To shed some light, we regress each factory's proportion of subcontracted orders on factory characteristics. Specifically, the dependent variable is $\overline{USC}_i = \sum_{j=1}^{n_i} USC_{ij}/n_i$, where n_i is the number of orders processed at factory i in our dataset. The names of the regressors are self-explanatory and are given in Table 1.10. We use two different models: a linear regression, estimated with OLS, and a Tobit regression, since the dependent variable is bounded between 0 and 1, with a large number of observations being zero. The Tobit regression represents factories' latent willingness to subcontract, as opposed to their observed level of subcontracting.

Table 1.10 shows that the most significant variable is the mean utilization, computed as the average of the utilization during the period in which each factory was active. Its coefficient estimates are 0.105 (in the linear model) and 0.191 (Tobit). This means that a factory would have subcontracted 5.3% more of its production if it had been 50% busier, all else equal. This observation strengthens our analysis on the relationship between factory utilization and unauthorized subcontracting (Section 1.5.3), measured through state dependence. The number of categories that each factory produced over the data collection period is also significant at the 0.05 level. For each additional category that a factory produced, its aggregate degree of unauthorized subcontracting increased by 3.04 percent points, i.e., less specialized factories subcontracted more. Factory capacity is weakly significant ($p < 0.1$) in the linear model and insignificant in the Tobit regression. A factory's proportion of orders that came from a major brand is weakly significant in the Tobit model: producing for a major brand decreases factories' willingness to subcontract, but this result does not carry over to the linear model, again possibly due to the low number of orders from major brands. The number of different buyers a factory produced for does not show a significant effect on its subcontracting practices. Factories located in China subcontracted more than in any other country and factories in Bangladesh exhibited a lower willingness to subcontract ($p < 0.05$). This last result should be taken with some distance since the data was collected after the Rana Plaza collapse when a large number of NGOs, media outlets, governments, etc., were monitoring Bangladesh closely.

Finally, notice that the R^2 for the linear model is 0.1965, and the pseudo-R^2 for the Tobit model is 0.1375. In other words, observable characteristics of factories explain a small amount of the variability in their subcontracting behavior. We infer that suppliers choose a baseline subcontracting level as a strategic decision, which is unobservable to us.

1.6. Prediction

In Section 1.5 we analyzed the factors driving unauthorized subcontracting. We now aim to predict which orders have a high probability of being subcontracted. In other words, would retailers be able to know that a certain order is likely to be subcontracted so to make better sourcing decisions? Predicting binary events is a vastly researched topic in finance, for instance to predict bankruptcy or financial distress (Sarkar and Sriram 2001, Alan and Lapré 2018), but now it is also becoming more common in other areas, including supply chain management. In fact, predicting unauthorized subcontracting was the original motivation of the middleman M. that sparked this research project. Our goal here is to show how well the middleman or a retailer could do with simple linear models based on the findings from Section 1.5. We acknowledge that more sophisticated models are likely to perform even better, but that is beyond the scope of this paper. We rather want to provide a competitive benchmark that is a good starting point.

To study how much future unauthorized subcontracting can be predicted, we divide our data in two subsets, training and test. The training data is used to estimate the coefficients of a linear regression model, whilst the test data is used to make out-of-sample predictions and quantify how accurate they are. To divide the observations between training and test we sort them by order delivery date and then split it; the reason for not randomizing them before partitioning is that the goal of this analysis is using knowledge of the past to predict the future, so we need to keep the chronological order. To check the robustness of our prediction, we repeat this procedure several times using different proportions of data in each subset (from 20% training and 80% test, to 80% and 20%).

Once the data is divided, we fit three different linear models with the training data set. First, a model containing the following variables: $Utilization_{ij}$ (factory i's utilization during the one-month period before delivering order j), $RelPriceFC_{ij}$, $PropThisCategory_{ij}$, $BuyerIsMajorBrand_{ij}$, $LogOrderSize_{ij}$, and a whole set of product category dummies. We call these *order-level variables*, since they vary order to order. Note that this model does not take into account any information about the factory's willingness to subcontract or not, but does contain, implicitly, information on the factory's reference prices, specialization, and amount of workload in the recent past. The second linear model we fit contains the previous variables, plus a set of factory dummies, which measure each factory's idiosyncratic baseline degree of unauthorized subcontracting during the training data set months, representing its managerial decision on whether or not to subcontract, and how much. Finally, we fit a linear model containing the order-level variables and the lagged variable $USC_{i,j-1}$ to capture the unauthorized subcontracting status of the previous order delivered by the factory.

After fitting the linear models, we compute the predicted probabilities of unauthorized subcontracting for the orders in the test dataset and apply the following classification rule: if the predicted probability of unauthorized subcontracting of a given order is greater than a threshold δ, we predict it will, indeed, be subcontracted; if it is δ or smaller, we predict it will not. Figure 1.3 shows the receiver operating characteristic (ROC) curve for each model, and Table 1.11 shows the area under each ROC curve, obtained by varying the threshold δ between 0 and 1. If we set the classification threshold to $\delta = 0.5$, which is arguably the most parsimonious choice, we obtain the accuracies shown in Table 1.12, and the Type I and II errors shown in Table 1.13.

Tables 1.11-1.13 show the importance of considering an idiosyncratic baseline. Indeed, including a factory fixed effect improves all metrics significantly; in particular, the accuracy of the prediction is greater than 0.82. Hence, the middleman can correctly predict whether or not unauthorized subcontracting will occur for more than 82% of the orders, and can do so using information at his disposal. Note that this accuracy level is a substantial improvement upon naive predictions. For instance, predicting that all orders will not be subcontracted has an accuracy of 64%. Such prediction is very extreme because the Type II error is 100%. A slightly more clever approach is to predict that subcontracting will occur for orders in factories type C (see Figure 1.1), will not occur for orders in factories type A, and then randomize with a 50% chance for orders in factories type B (the prevalence of unauthorized subcontracting in factories type B is roughly 50%, see Table 1.5). This approach has an accuracy of 72%, which is still quite below the accuracy of the linear model with order-level variables and fixed effects.

If it were possible to observe the factory's decision for every order it produced, then by adding a lagged dependent variable it would be possible to predict unauthorized subcon-

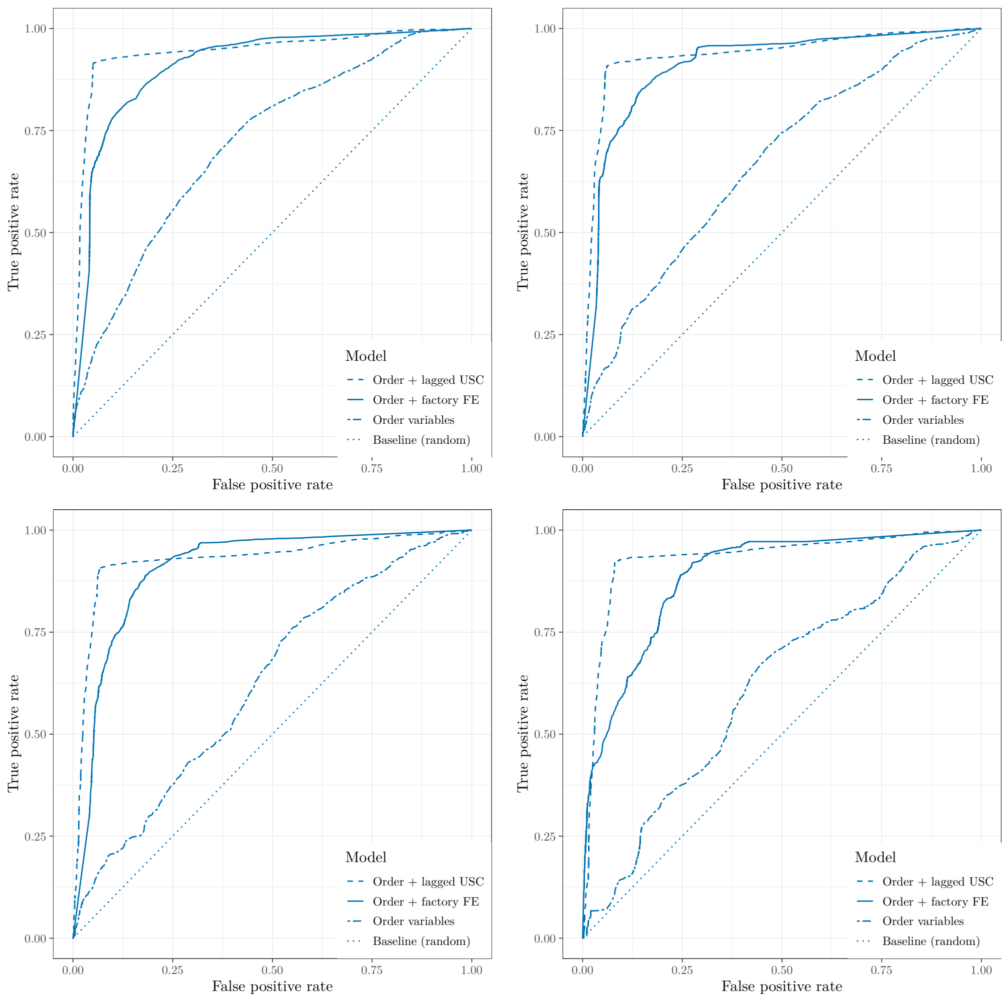


Figure 1.3: ROC curve of the linear prediction for our three different models (order variables, order variables + factory FE, order variables + lagged USC), using 20% of the data for training and 80% for test (upper left), 40%-60% (upper right), 60%-40% (bottom left), and 80%-20% (bottom right).

tracting with more than 92% accuracy. However, knowing every factory decision might not be realistic. Using only order-level variables, without fixed effects, leads to a lower accuracy due to the number of false negatives: of all the orders that were subcontracted, the model is able to predict only 20-30% of them (see Table 1.13). In other words, the model that does not include any factory information is quite conservative. Nevertheless, it could be useful when there is no past information available about a factory's subcontracting behavior.

One caveat of the linear prediction models presented here is that they are highly affected by seasonality patterns in unauthorized subcontracting that are driven by the peaks of workload in different times of the year. This makes the accuracy of the prediction actually decrease when the size of the training data set increases. Unfortunately, we only have eight months of data, which means that we cannot include any seasonality variable in our prediction. However, this issue would be solved if more data became available. An additional cause for the decrease in accuracy when the training dataset is larger is overfitting: the accuracy ratio between training and test is increasing in the size of the training data set. This is also true for the AUC ratio and for the sum of squared errors. A full implementation would have to take care of these details but the starting point provided here is already quite promising.

1.7. Discussion and Managerial Insights

Unauthorized subcontracting in the apparel industry is a problem that needs to be addressed: most of last years' industrial catastrophes and disclosed severe violations of labor standards occurred in second tier suppliers that retailers had not authorized. In this paper, we have also shown that this problem can indeed be addressed, as it is highly predictable, and related to some operational variables that buyers can act on.

We proposed two research questions. As the main drivers of unauthorized subcontracting, we first observe a great amount of autocorrelation between orders of the same factory. An order is 83% more likely to be subcontracted without authorization if the previous order by the same factory was also subcontracted than if it was not. This state-dependent subcontracting behavior is highly related to workload: its effect is larger when factories enter periods of high utilization. After state dependence, price pressure is the next main driver of unauthorized subcontracting: an order with a unit price 10% lower than the factory's average for that product category, which happens for one in every three orders in our data, is 4% more likely to be subcontracted, and an order with a price pressure of 25%, which happens 12% of the times, is 11% more likely to be subcontracted without the buyer's authorization.

Orders produced for a buyer which is a major brand are 10% less likely to be subcontracted, but this effect vanishes is some of the subsamples possibility because there are too few orders from major brands. We find no individual effect of factory specialization, but we do find a joint effect with other order characteristics, and we also find an effect at the factory level: less specialized factories subcontracted a larger proportion of their orders. In particular, for every extra product category that they produced, their average unauthorized subcontracting level increased by 14%.

At the supplier level, apart from its level of specialization, we find an effect of the supplier's location, with China being the country where unauthorized subcontracting occurs more often. We do not find any effect of the number of buyers or the size of the factory. Nevertheless, the factory-level regression only explains 20% of factories' differentiated subcontracting behavior, suggesting that suppliers choose to specialize in an amount of unauthorized subcontracting and this is a managerial decision that is not observable to us. However, as we have seen, we can use the observed supplier's past of unauthorized subcontracting as a proxy to measure this managerial decision. From an academic standpoint, our findings at the factory level provide empirical support to common assumptions used in theoretical models, e.g., Orsdemir et al. (2015) and Chen and Lee (2016).

The results in this paper show that buyers should study suppliers' past subcontracting behavior before starting to work with them but, in addition, they should minimize the probability of each individual order being subcontracted by not exerting price pressure and by monitoring suppliers' workload closely to avoid overutilization. Knowing that factories suffer from poor production planning (Hurst et al. 2005), buyers should collaborate with their suppliers and offer training and consulting to help them organize their production efficiently, as well as develop the technical capabilities that each product type requires (Kraft et al. 2017). On a larger scale, similar to Caro et al. (2016), there is an opportunity for buyer consortiums that jointly monitor their suppliers to avoid overutilization. In the case of intermediaries, they can also work closely with brands to help them plan their orders taking into account the workload upstream in the supply chain. This approach seems more plausible for basic apparel. For fashion products, buyers can try to emulate the strategy followed by the Spanish retailer Zara, which consists in operating reasonably below capacity on average to ensure rapid response times (Caro 2012).

For our second research question, we have proposed a simple prediction method to detect which orders will be subcontracted. The data is split between a training set, that is used to estimate the model's parameters, and a test set. By knowing the fraction of orders that the factory subcontracted during the training period, plus each order's specific characteristics, we can predict unauthorized subcontracting with more than 82% accuracy. Therefore, for at least 4 out of 5 orders, unauthorized subcontracting could have been predicted correctly and better sourcing decisions could have been made to prevent it. If subcontracting information for the most recent order is available, then one can predict unauthorized subcontracting correctly in more than 90% of the cases.

The data used in this work consists of variables that buyers (retailers or intermediaries) have already at their disposal. Hence, our findings could be implemented in a decision support system (DSS) to help buyers prevent unauthorized subcontracting among their suppliers. Such system could be used to monitor each factory's workload and queue of pending orders, in addition to tracking the product categories that the factory can produce together with the average price per category. Then, the system could detect and flag orders with a high chance of being subcontracted before they are placed to a given factory. The DSS could also suggest alternative factories that would be more suitable for that order or could require the user to justify their choice. There is ample literature on how to design systems to nudge users in a desirable direction (Meeker et al. 2016). We must note that our methods rely on time-varying variables. Hence, the DSS would have to update the data and estimations on a rolling horizon basis, which could pose some challenges if the recommendations of the DSS induces suppliers to hide information, similar to the backfiring situation described in Plambeck and Taylor (2016).

In addition to detecting risky orders and helping their current suppliers improve their efficiency, buyers should streamline the compliance certification process, as well as the authorization process to subcontract an order or a part of it. Subcontracting is a regular practice in the industry and it only poses a problem when the buyer loses visibility and an order ends up in a non-compliant factory. It is quite possible that some of the subcontractors have reasonable working conditions but lack a formal compliance qualification because the certification process is too cumbersome. Similarly, if getting the buyers's approval to subcontract an order is a lengthy process, then factories might skip it, even when the subcontractor is compliant. Regardless, the buyer might want to learn more about the network of informal factories to find ways to add as many as possible to the list of approved suppliers.

Our work provides opportunities for further research. On the empirical side, it would be important to see other studies that complement ours. For instance, our study is limited to the orders handled by the middleman M. Hence, we do not observe any other orders produced by the factories. Similarly, we are not aware of the sourcing portfolio of the different buyers. With more factory and buyer information and a proper costing model one could look at margins and its interactions, and study how they might relate to unauthorized subcontracting. Moreover, there are many open questions on subcontracting that our study is not able to answer, e.g., do ISO standards or other codes of conduct make a difference? Or, does the presence of competing factories in the same area make subcontracting more/less likely? There is also room for improvement in the prediction of subcontracting given the wide range of deep learning techniques that are becoming available. Note as well that our model predicts subcontracting when ideally one would want to predict noncompliance, or even better, the chance of a disaster. On the theoretical side, a crucial issue is finding mechanisms to ensure truthful disclosure of subcontracting without compromising the option of weeding out the factories that specialize in subcontracting a high percentage of their orders. Another interesting problem is how to assign orders to suppliers with the objective of minimizing cost or time subject to a chance constraint on subcontracting. Finally, more research is needed on how to engage brands and consumers in solving the unauthorized subcontracting problem.

Acknowledgments. First and foremost we thank the middleman M. for providing the data and for useful discussions that informed the problem. We also thank the participants at the Supply Chain Thought Leaders Roundtable held at the University of Cincinnati and seminar participants at Wharton, Penn State, Arizona State University, University of Utah, UNC Chapel Hill, UW-Madison, and the University of Calgary, for feedback on a preliminary version of this paper.

	Unauthorized subcontracting				
	Arellano-Bond				
	(1)	(2)	(3)	(4)	(5)
$\text{USC}_{i,j-1}$	0.295***	0.294***	0.309***	0.312***	0.303***
	(0.0242)	(0.0243)	(0.0240)	(0.0258)	(0.0250)
PropThisCategory	0.00102	-0.00398	0.00970	0.00814	0.00233
	(0.0343)	(0.0330)	(0.0354)	(0.0339)	(0.0336)
LogOrderSize	-0.00391	-0.00401	-0.00160	-0.00180+	-0.00274+
	(0.00269)	(0.00278)	(0.000981)	(0.000933)	(0.00146)
RelPriceFC	-0.167**				
	(0.0616)				
RelPriceFCB		-0.163**			
		(0.0597)			
RelPriceOrd			-0.0809		
			(0.0563)		
RelPriceOrdF				-0.0320	
				(0.0332)	
RelPriceFSub					-0.133***
					(0.0306)
Category dummies	Yes	Yes	Yes	Yes	Yes
Buyer dummies	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes
N	32028	32028	32028	32028	32028

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Unauthorized subcontracting				
	Arellano-Bond				
	(1)	(2)	(3)	(4)	(5)
$\text{USC}_{i,j-1}$	0.298***	0.296***	0.309***	0.306***	0.311***
	(0.0228)	(0.0224)	(0.0220)	(0.0226)	(0.0267)
PropThisCategory	-0.0116	-0.0257	0.0134	-0.0136	-0.00688
	(0.0570)	(0.0561)	(0.0613)	(0.0568)	(0.0572)
LogOrderSize	-0.00531	-0.00538	-0.00198	-0.00411	-0.00361+
	(0.00388)	(0.00395)	(0.00268)	(0.00319)	(0.00208)
RelPriceFC	-0.188**				
	(0.0626)				
RelPriceFCB		-0.181**			
		(0.0596)			
RelPriceOrd			-0.190***		
			(0.0573)		
RelPriceOrdF				-0.171*	
				(0.0695)	
RelPriceFSub					-0.135***
					(0.0337)
Category dummies	Yes	Yes	Yes	Yes	Yes
Buyer dummies	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes
N	18646	18646	18646	18646	18646

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.8: Arellano-Bond regression using the full data set (left) and the subsample of type B factories (right), with five different measures of price pressure: (1) with respect to the moving average per factory and category; (2) with respect to the moving average per factory, category, and buyer; (3) with respect to the predicted price using order size, lead time, and category; (4) with respect to the predicted price using order size, lead time, category, and factory; and (5) with respect to the moving average per factory and subcategory.

	Unauthorized subcontracting		
	Arellano-Bond		
	(1)	(2)	(3)
$\text{USC}_{i,j-1}$	0.297***	0.296***	0.300***
	(0.0278)	(0.0278)	(0.0243)
RelPriceFC	-0.158**	-0.162**	-0.165**
	(0.0607)	(0.0596)	(0.0598)
PropThisCategory	0.0311	0.00524	0.00737
	(0.0346)	(0.0343)	(0.0343)
LogOrderSize	-0.00371	-0.00380	-0.00390
	(0.00265)	(0.00268)	(0.00258)
Category dummies	No	Yes	Yes
Month dummies	No	No	Yes
N	32028	32028	32028

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Unauthorized subcontracting		
	Arellano-Bond		
	(1)	(2)	(3)
$\text{USC}_{i,j-1}$	0.300***	0.297***	0.303***
	(0.0282)	(0.0283)	(0.0231)
RelPriceFC	-0.178**	-0.185**	-0.187**
	(0.0620)	(0.0591)	(0.0591)
PropThisCategory	0.0905+	0.0359	0.0365
	(0.0534)	(0.0549)	(0.0544)
LogOrderSize	-0.00550	-0.00575	-0.00527
	(0.00403)	(0.00401)	(0.00369)
Category dummies	No	Yes	Yes
Month dummies	No	No	Yes
N	18646	18646	18646

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.9: Arellano-Bond regression using the full dataset (left) and the subsample of type-B factories (right) for a panel structure at the supplier-buyer level. The regression includes fixed effects for each supplier-buyer pair.

	Proportion of subcontracted orders			
	Linear		Tobit	
	(1)		(2)	
NumCategories	0.0304*	(0.0139)	0.0615*	(0.0297)
NumBuyers	-0.0138	(0.0211)	0.0678	(0.0664)
PropMajorBrand	-0.135	(0.113)	-2.116*	(0.837)
MeanUtilization	0.105*	(0.0426)	0.191*	(0.0934)
LogCapacity	0.0451+	(0.0261)	0.0999	(0.0737)
Bangladesh	-0.118	(0.0845)	-1.058*	(0.485)
Cambodia	0.0346	(0.132)	-0.0486	(0.459)
China	0.225*	(0.0874)	0.610*	(0.299)
Indonesia	-0.132	(0.0963)	-0.727+	(0.436)
Vietnam	0.0102	(0.111)	0.126	(0.367)
R^2	0.1965			
Pseudo R^2			0.1375	
N	226		226	

Robust standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.10: Linear and Tobit regression of factories' proportion of subcontracted orders on factory characteristics. The overall factory-level baseline is $\sum_i \overline{USC}_i/226 = 0.28$.

Training-test	20%-80%	40%-60%	60%-40%	80%-20%
Order variables	0.7195	0.6680	0.6238	0.6155
Order + factory FE	0.9099	0.9092	0.9059	0.8865
Order + lagged USC	0.9444	0.9349	0.9279	0.9302

Table 1.11: Area under the ROC curve of the linear prediction for our three different models.

Training-test	20%-80%	40%-60%	60%-40%	80%-20%
Order variables	0.6699	0.6476	0.5876	0.5554
Order + factory FE	0.8551	0.8480	0.8468	0.8247
Order + lagged USC	0.9360	0.9267	0.9210	0.9206

Table 1.12: Accuracy of the linear prediction for our three different models.

Training-test		20%-80%	40%-60%	60%-40%	80%-20%
Order variables	Type I	0.0561	0.1458	0.1815	0.1809
	Type II	0.7847	0.6708	0.7181	0.6906
Order + factory FE	Type I	0.1012	0.1222	0.1512	0.2743
	Type II	0.2175	0.1981	0.1558	0.0830
Order + lagged USC	Type I	0.0516	0.0609	0.0680	0.0817
	Type II	0.0846	0.0924	0.0936	0.0774

Table 1.13: Type I and II errors of the linear prediction for our three different models.

CHAPTER 2

Believing in Analytics: Managers' Adherence to Price Recommendations from a DSS

2.1. Introduction

"When it comes to implementation, the success of operations management tools and techniques, and the accuracy of its theories, relies heavily on our understanding of human behavior" (Bendoly et al. 2006)

A large portion of the Operations Management literature provides prescriptive solutions to practitioners' complex problems. In some cases, these solutions get successfully implemented, leading to higher revenue, lower costs, increased efficiency of the system, or other improvements.

Sometimes firms implement these solutions in the form of automated algorithms. However, leaving algorithms to make decisions unsupervised can be risky. For instance, algorithms that set prices automatically by tracking competitors have led to notorious failures, such as a $23 million biology textbook (Eisen Eisen), or thousands of products priced at £0.01, a mistake that brought losses of up to £100,000 to many small vendors (Neate 2014).

Other times, operational solutions are implemented in the form of a decision support systems (DSS), which is a tool that makes algorithm-based recommendations to human decision makers, who can implement the DSS' recommendations or deviate from them. The usage of DSSs can greatly improve the decisions being made (Sharda et al. 1988, Hoch and Schkade 1996), since humans are prone to a large number of cognitive biases (Kahneman 2003a,b, Hastie and Dawes 2010). Even highly trained managers can be biased and make suboptimal decisions (Bazerman and Moore 2008, Schweitzer and Cachon 2000).

When managers make operational decisions assisted by a DSS, their choices impact the performance of the firm. If the DSS's algorithm is recommending optimal decisions but managers deviate from them, the potential improvement that the DSS was supposed to cause may never be realized. Therefore, it is important to study how humans interact with these tools and what drives their adherence to its recommendations. Understanding the behavioral aspects of managers' usage of DSSs can provide insights on how to design their interfaces to ensure a successful implementation.

In this paper, we study the aftermath of the implementation of a DSS for clearance sales markdown optimization at Zara. A pilot test of a prototype in 2008 showed that the DSS increased revenue by almost 6%; see Caro and Gallien (2012). However, after the DSS was rolled out in 2010, managers' average adherence to its recommendations was only 46% for franchise stores and 61% for stores owned by Zara. In the three years that followed, the managers gained experience, but more importantly, Zara performed two key interventions to the DSS, which we study in detail in Section 2.4. Figure 2.1 illustrates the starting and ending points of this study. The horizontal axis is the adherence of each country manager and the vertical axis is the financial performance summarized by the metric Y, which is the ratio between clearance sales revenue and the initial inventory valued at regular season prices. As shown in the figure, after the interventions took place, the average adherence was higher and less disperse, and the behavior of franchises and own stores was very much aligned. The higher adherence translated into higher revenues as measured by the metric Y, therefore confirming what had been observed in the 2008 pilot.

The two interventions performed by Zara consisted in changes to the DSS's interface used by the country managers. The first intervention was showing the revenue metric Y that the algorithm was maximizing. The rationale was to provide timely feedback to the managers' pricing decisions. The second intervention consisted in showing a reference point for the Y revenue metric, so that it would be easier to interpret. We run a difference-in-differences analysis to study the effect that these interventions had on managers' adherence. We find

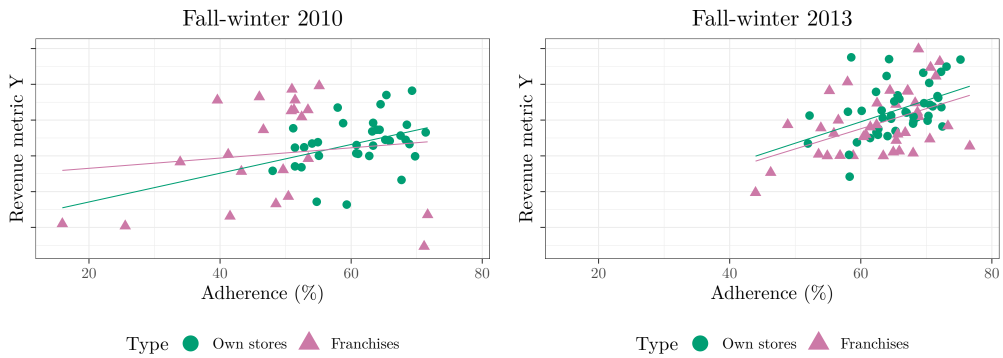


Figure 2.1: Relationship between adherence to the DSS's recommendations and revenue in fall-winter 2010, before the interventions (left) and fall-winter 2013, after the interventions (right). The green dots correspond to managers from countries in which Zara owns the stores. The pink triangles correspond to managers from franchise countries. The exact values of the revenue metric Y have been disguised.

that Intervention 1 had no significant effect on managers' adherence, but once combined with Intervention 2, the adherence increased by 9 percentage points. Overall, franchise managers' adherence increased by 25% when the second intervention took place, and their likelihood of deviating, i.e., marking down a product when the DSS recommended staying put, decreased by 48%.

In addition to the difference-in-difference analysis on the effect of the interventions, we perform a Heckman regression to further understand the drivers of managers' decision to deviate and the magnitude of their deviations. We find that managers were 45% more likely to adhere to a price recommendation if it was consistent with their pre-DSS rule of thumb (which was based on inventory minimization, not revenue maximization), but after the interventions this effect decreased to 17%. There was an even stronger relationship between adherence and their previous heuristic when it was computed using group-aggregate inventory metrics, as opposed to individual product's. For countries in which unsold inventory could be salvaged, a 10% increase in a product's salvage value was related to a 6% higher probability of implementing the price the DSS recommends, and this effect increased to 9%

after the interventions. Managers were more likely to deviate from the DSS's recommendations when the number of prices to set was high, as a company-specific pricing rule allowed them to reduce this quantity.

The speed at which a product was being sold affected the magnitude of their price deviations: an extra week in the predicted time until a product would be sold out was related to a 1.9% larger price deviation, but this number was 1.4% after the interventions. In countries in which unsold inventory could be salvaged, a 10% higher salvage value was related to 1.45% larger price deviations, but this effect was insignificant after the interventions.

Our results are consistent with a number of cognitive biases. First, preference for the status quo: managers expected the DSS to follow the heuristic that they used before it was implemented. Second, salience of the inventory compared to a revenue forecast: managers overestimated the importance of the speed of sales, and tried to sell all inventory out instead of maximizing revenue. Third, loss aversion: managers' inventory-minimizing behavior was less prominent when products had a large salvage value. Fourth, inattention: managers attempted to reduce the amount of prices to set every week, and often paid more attention to group-aggregate metrics instead of those for individual products. Finally, the effect of the interventions on adherence and the way in which they mitigated some of these cognitive biases suggests that the interventions effectively made revenue more salient than before, and that having a reference point for the abstract revenue metric helped managers interpret it.

Our findings provide insights on how to increase voluntary adherence that can be used in any context in which a company wants an analytical tool to be adopted organically by its users.

Some of the data presented in this paper has been disguised to protect its confidentiality, and we emphasize that the views presented here do not necessarily represent those of the Inditex Group.

2.2. Literature Review

Some related literature suggests that deviating from a DSS's recommendation is not always bad: it could be a way of enriching it with additional data. For instance, Cui et al. (2015) show that deviating from fixed inventory policies allows managers to incorporate private information that the policy does not take into account. In a laboratory experiment, Flicker (2018) finds that adding a noisy signal available to managers into the DSS's algorithm improves the outcomes of a newsvendor-type decision respect to those generated by the DSS when no additional information is included. In Tan and Staats (2016), managers in a casual dining restaurant deviate from the recommended routing policy by assigning more tasks to highly-efficient workers, or to those with low workload, and these deviations improve the restaurant's performance when they are not too large. In our work, we study adherence to a DSS that had been proven to improve the firm's revenue, and so deviating from its recommendations is detrimental to the firm's profit.

It may happen that managers' incentives are not aligned with the DSS's, and this leads them to deviate from its recommendations. In Van Donselaar et al. (2010), store managers at a supermarket chain receive inventory replenishment recommendations from a DSS. Managers were not penalized for holding costs, which were not salient to them, but they did take into account labor costs and workload, which the DSS was ignoring. In our paper, managers were incentivized to generate a high revenue, which was also the firm's objective. Furthermore, the DSS was already incorporating all the firm's pricing rules regarding the labor costs related to marking products down and the shelf space that was necessary during clearance sales.

Bearden et al. (2008) tackle cognitive biases in revenue management. In their study, decision makers in a laboratory setting do not set prices but, instead, decide whether to accept an offer for a product of which they have a fixed number of units to sell. The subjects' objective is to sell all inventory out, but they consistently deviate from the policies that would help them achieve their goal. In contrast, managers in our data were instructed and incentivized to maximize revenue, and this was also the DSS's objective.

In some contexts, deviations from prescribed policies is linked to worse outcomes for both decision makers and the firms for which they work. For instance, in Ibanez et al. (2017) doctors who examine randomly assigned radiological images have discretion to order the tasks in their queue. They find that doctors often deviate from the prescribed FIFO policy in favor of batching similar tasks or of completing shorter tasks first, and that experienced doctors deviate more. Although the quality of their work is not affected by their deviations from the prescribed order, their productivity decreases when they decide to deviate, i.e., their task completion times become longer. In our setting, managers' deviations from the DSS also lead to worse outcomes (lower revenue). Although we do not focus on experience and use it as a control, we also find that it is associated to lower adherence for some manager types.

The paper most closely related to our work is Elmaghraby et al. (2015), which studies how salespeople in a B2B setting decide prices for different grocery products assisted by a DSS. They find that salespeople use the price recommendation as an external reference point, but the extent to which this reference price influences their decision depends on the sign and magnitude of the recommended price change. They find characteristics of salespeople, customers, and products that affect how prices are set, as well as an important role of products' changes in cost on final changes in price. Our work differs from theirs in several aspects. First, the managers in our data set prices for the final consumer, as opposed to a B2B setting. Second, the products for which prices are being set are fashion products, with a much higher demand uncertainty than groceries. Third, in our setting costs are not an important driver of price changes during clearance sales, given that the cost of inventory has already been paid and does not change over the campaign. Finally, their focus is on what drives price changes, and the DSS's recommendations play a moderating role, while our focus is on the behavioral drivers of adherence decisions.

Our work contributes to the Behavioral Operations Management literature in multiple ways. First, the behavioral aspects of pricing have been studied in depth when consumers exhibit cognitive biases. However, to our knowledge, managers' behavioral biases in pricing decision making have been only studied in B2B contexts (Zheng and Özer 2012). Second, we not only show what the operational and behavioral drivers of adherence are, but we also study two interventions that had a large effect on adherence and that can help us understand how the DSS's interface impacts managers' adherence to its recommendations and, ultimately, how to design better DSSs. Third, while many papers in the OM literature describe tools, algorithms and methods which were implemented in practice, and report on the tool's performance after being implemented, few track the aftermath of a tool's implementation in terms of usage like this paper. Finally, we use data from a large fast fashion retailer that contains the actual pricing decisions made by the firm's managers. Using observational data to study adherence to a DSS and cognitive biases of real managers supplements our field's extensive findings in laboratory settings and can inform future modeling approaches to similar problems.

2.3. Empirical Setting

2.3.1 Clearance Sales at Zara

Spanish retailer Zara, like many other fashion retailers, sets a clearance sales period at the end of every season (fall-winter, or W, and spring-summer, or S). This period usually lasts around 12 weeks, set by the specific regulations of every country. Once the sales campaign begins, the prices of products change regularly (usually weekly) and these markdowns occur for all products in a country at the same time. The pricing decisions are the responsibility of one person for each country, called the *country manager*, with the support of a small pricing committee.

These markdowns are constrained by some company rules. To list them, we first need to define some company-specific terminology. A *group* is a set of products of the same kind targeted towards a customer type. For instance, young women's knitwear, basic women's shirts, etc. There are a fixed number of groups, which are consistent over the years and across countries, whereas the product assortment within every group changes every new season. Within a group, a *cluster* is a set of SKUs which had the same price during the regular season. For instance, all the basic tops which were 19.95€ during the regular season form a cluster, even though they were not the exact same SKU, and all the basic tops which were 14.95€ are part of a different cluster.

One of the firm's pricing rules is that, during the clearance sales campaign, the price of each cluster cannot increase. Another rule is that clusters cannot be split, i.e., the 19.95€ basic tops cluster may experience several markdowns during the sales campaign, but all products within that cluster will always be marked down by the same amount at the same time. In addition, different clusters can converge to the same price, but never cross: the 19.95€ basic tops cluster, even after several markdowns, must at least as expensive as than the 14.95€ basic tops cluster, because this was their order during the regular season. Finally, when two clusters' price converges, these two clusters' prices will be coupled for the rest of the campaign. For instance, if the 19.95€ cluster was marked down to 14.95€, and the 14.95€ cluster was unchanged, then both clusters would need to be marked down by the same amount for all the remaining sales weeks. A set of products containing one or more clusters whose prices have converged at some point is called a *category*. These rules are driven by legal and practical reasons detailed in Caro and Gallien (2012).

A particular characteristic of Zara's business model is that, in the countries where it is present, it either owns all the stores in the country (or a large majority) or all the stores there belong to a franchise. Some of the franchisee firms manage more than one country, while some others do so for only one. This distinction between country types will be important in our empirical analysis. More specifically, for own-store countries, pricing decisions were all made in Zara's headquarters in Spain, so the managers from those countries were arguably subject to peer effects regarding usage of the DSS. In contrast, pricing decisions for franchise countries were made where the franchisee company was (with the support of a country representative located in Zara's Spain headquarters), so the individuals who managed those countries were separate from each other. Additionally, any garments that remain unsold at the end of clearance sales in own-store countries can be salvaged, but franchises purchase the inventory and at the end of the campaign its salvage value is zero.

2.3.2 Implementation of the DSS

Until 2007, country managers made their markdown decisions based on weekly reports that contained metrics on inventory levels and speed of sales. Two metrics that were emphasized in those reports were the *success*, defined as the percentage of all the inventory for that category that had been sold by that week, and the *rotation*, or the remaining time for all the inventory of that category to be sold out at the speed of sales observed in the previous three days. A screenshot of one of those inventory reports can be seen in the appendix.

In 2008, Zara adopted a sales pricing optimization system (Caro and Gallien 2012). This system built a demand forecast and used it in a revenue maximization dynamic program which found optimal prices for the remainder of the clearance sales campaign. The algorithm, which was ran in a rolling horizon basis, included as constraints all the company's pricing rules, which reflected practical issues such as shelf space, labor, etc.

The company made the strategic choice to not automate the pricing process and keep managers responsible for the final prices implemented in each country, since Zara's operations are people-centric. For this reason, the algorithm was embedded in a DSS to assist the country managers with their markdown decisions. The DSS worked in the following way: every week, in addition to the inventory reports that managers used to see, the DSS included a recommended price for every category in a group, and the forecasted revenue corresponding to that price. In addition, managers could introduce another value of the price and the DSS would show them its forecasted revenue. With this information, managers confirmed the final price to be implemented for every category.

In S2008 a pilot test of the DSS in two countries resulted in a 5.8% increase in the revenue metric Y. After that, the DSS was implemented gradually (until 2012) in all other countries. It was obligatory for managers from own-store countries to use this tool and observe its recommendations, while franchise managers could set the prices using any other system, although the DSS was also available to them and they were equally trained to use it.

Managers were incentivized to maximize revenue, and the DSS's recommendations were proven to increase it. However, managers' adherence to such recommendations was initially low. For instance, in W2010 it was 57% (61% for own-store countries and 46% for franchises). More specifically, when managers deviated from the DSS's recommendations they usually did so by setting prices that were lower than the recommended ones (see Figure 2.2). A common way in which deviations occurred was that the DSS recommended keeping the price of a category unchanged, but the manager marked it down anyway. Observing managers' low adherence to the DSS's recommendations, particularly in franchises, Zara performed two interventions aimed to increase it. We study their effect in detail in Section 2.4. As we can see in Figure 2.2 (right), adherence was higher after the interventions, especially that of franchises.

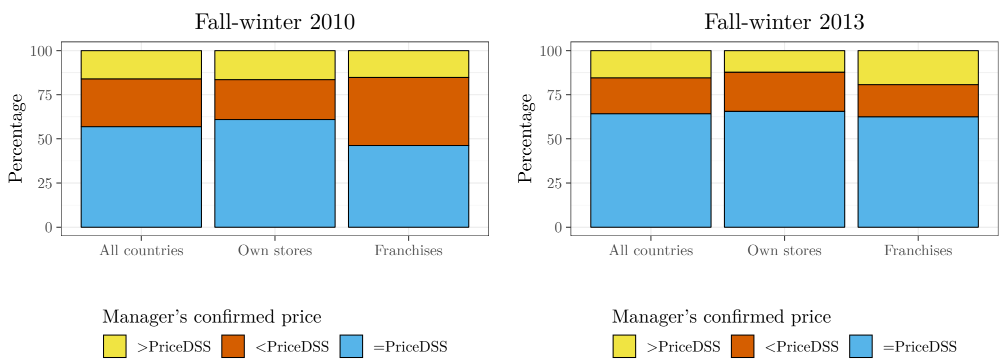


Figure 2.2: Pricing decisions in W2010 (left) and in W2013 (right), for all countries (first bar of each plot), countries in which Zara owns the stores (second bar), and countries in which stores are franchises (third bar).

2.3.3 Data Description

To study the effect of the two interventions and the drivers of managers' adherence to the DSS's recommendations, we use a dataset collected by Zara, which spans seven sales campaigns over 3.5 years, from W2010 to W2013. Our data contains 20 women's apparel groups. Due to Zara's constant expansion, the number of countries in our data increases from 53 in W2010 to 84 in W2013. Two countries were dismissed because Zara discontinued its presence there due to social and political conflicts. For each country, campaign, and group, the dataset contains all pricing decisions made by that country's manager for all clusters in that group and all weeks in that campaign.

We separately gathered information from Zara's annual reports regarding the number of stores in each country and year, as well as the type of country (own stores or franchises). Two countries in our data changed store ownership type in 2013. However, as we will see in the following sections, this does not affect our analyses.

Table 2.1 shows one observation in our original dataset. In this example, from W2010 in France (where Zara owns its stores, 113 in total), the cluster of women's outerwear that had been priced at X€ during the regular season (regular price column) was priced at Y€ during week 6 of clearance sales (previous price). In week 7, the DSS recommended keeping its price unchanged at Y€ (DSS price), but the manager decided to mark it down to Z€ (confirmed price). At the beginning of week 7, there were 183 units of this cluster in all the country, of which 93 were sold by the end of the week. All women's outerwear unsold by the end of W2010 would be salvaged at S€/unit.

Given that our data was collected at the cluster level, but managers made their pricing decisions at the category level (i.e., only one pricing decision for all clusters that belonged to the same category at a given time), we need to aggregate our data by categories. We identify a category, in a given week, as the set of clusters within a group whose price in the previous week was the same. In the example in Table 2.1, all clusters of women's outerwear with previous price equal to $Y would form a category in week 7. We then aggregate the category's inventory, sales, etc. and compute all the variables we need for our analysis from the category-aggregate dataset. After aggregating by category, our data contains 374,366 observations. Tables 2.2 and 2.3 contain summary statistics of the countries and categories in our data.

Country	**Campaign**	**Type**	**Num. stores**	**Group**	**Cluster**	**Week**
France	W2010	Own stores	113	Women's outerwear	C	7
Initial inventory	**Quantity sold**	**Regular price**	**Previous price**	**DSS price**	**Confirmed price**	**Salvage value**
183	93	X€	Y€	Y€	Z€	S€

Table 2.1: One observation in our original data set. The price information has been disguised.

	Mean	**St. dev.**	**Min.**	**Max.**	**% Countries**	**% Observations**	**% Adh. (W2010)**	**% Adh. (W2013)**
Country descriptive statistics:								
Number of stores (W2010)	15.94	41.74	0	333	-	-	-	-
Number of stores (W2013)	21.18	43.87	1	323	-	-	-	-
Number of campaigns using								-
the DSS (as of W2013)	7.12	0.85	3	11	-	-	-	-
Country type:								
Zara owns the stores	-	-	-	-	51.22	57.12	61.03	65.67
Stores are franchises	-	-	-	-	48.78	42.88	46.34	62.47

Table 2.2: Summary statistics of the countries in our data.

	Mean	**Median**	**St. dev.**	**Min.**	**Max.**	**% times**
Inventory at the beginning of the sales campaign	3,182.5	371	12,552.2	1	559,014	-
Weekly sales	539.7	44	2,526.3	0	217,809	-
The DSS recommends to markdown	-	-	-	-	-	39.60
% markdown recommended by the DSS						
(given that it recommends to markdown)	29.74	25.64	11.79	7.75	97.49	-
The country manager decides to markdown	-	-	-	-	-	47.38
% markdown chosen by the country manager						
(given that the manager decided to markdown)	27.67	25.03	10.97	-75.0	98.0	-
The country manager decides to deviate						
from the DSS's recommended price	-	-	-	-	-	39.10
% deviation from the DSS's recommendation						
(given that the manager chose to deviate)	30.73	25.01	37.40	0.226	3,052.6	-

Table 2.3: Summary statistics of the categories (observations) in our data.

2.4. Effect of Two Interventions

2.4.1 Context

As we have seen in Section 2.3.2, managers' initial adherence was low, and they often deviated by setting prices lower than the recommended ones, and by marking products down when the DSS recommended keeping their prices unchanged. Zara performed two interventions, in the form of changes to the DSS's interface, with the objective to increase managers' adherence. In this section, we analyze the effect of those interventions. The analysis in this paper was performed independently by the authors and it is not the responsibility of the Inditex group.

2.4.1.1 The Revenue Metric Intervention.

In S2011 the DSS started displaying the revenue metric Y that its price optimization algorithm was maximizing (detailed in Section 2.3.2). The purpose of this intervention was to give real-time feedback to managers on their performance and to shift salience from the inventory metrics they were used to seeing in the weekly report (and the physical inventory in the stores) to the revenue they were expected to maximize.

The positive role of feedback on decision making is well studied (Hogarth 1981, Kleinmuntz 1985, Bolton and Katok 2008), as well as the effect of making implicit quantities more salient. For instance, in Chetty et al. (2009), making taxes salient decreased consumer demand by 8%. In the context of utilities, the implementation of smart meters to show users real-time feedback (and so to increase the salience of their consumption) decreased demand and enticed resource conservation (Gilbert and Zivin 2014, Tiefenbeck et al. 2016).

We hypothesize that Intervention 1 had a positive effect on adherence because, by increasing the salience of revenue, managers' objective became more aligned with the DSS's.

2.4.1.2 The Reference Point Intervention.

In S2012 Zara made a second change in the DSS's interface: showing a reference point for the metric Y, to make the revenue metric easier to interpret. The reference point was the manager's own Y in the previous year for the same group and week, as well as the previous year's final Y for that group.

The ability to evaluate a quantity in relation to another is part of the number sense framework proposed by McIntosh et al. (1992). Numeracy research shows that decision makers may not use numbers until they are contrasted with available data to determine their meaning (Peters et al. 2006, Peters 2012). Barrio et al. (2016) show that reference points substantially improve people's ability to understand numerical measurements.

We hypothesize that Intervention 2 helped managers interpret the Y metric, which increased its weight in the decision making process. Therefore it aligned managers' objective with the DSS's further, increasing manager's adherence to the DSS's recommendations

2.4.2 Methods

2.4.2.1 Effect on Adherence.

The company rolled out the interventions for all country managers at the same time. However, we can exploit a specific characteristic in the structure of our data to study the effect of the interventions: managers' split between countries in which Zara owns the stores and countries where stores are franchises (see Section 2.3). In practice, this implied that own-store country managers actively provided feedback on the DSS's interface and helped improve it: both interventions were deployed after their feedback so, for those managers, interventions were mostly endogenous. Oppositely, franchise country managers did not participate in the development of the DSS, so the two interventions were exogenous to them, and mostly targeted to them, as they were the ones showing the lowest adherence. Therefore, we label own-store countries as the control group and franchise countries as the treatment group, and use the difference-in-differences (DiD) estimator to compute the effect of each intervention on franchise countries versus own-store countries. We use the binary variable $Franchise_i$, which takes value 1 for franchises, in our DiD regression. Two countries changed type in the period in which our data was collected: Finland and Kazakhstan used to be franchises and became own-store countries, but these changes took place in S2013, so they do not affect our analysis.

Given that the assortment of categories within groups changes every season, but groups do not, we need to aggregate our outcome of interest at the product group level to make the units of observation comparable in the pre- and post- intervention periods. We define the group-aggregate adherence of manager i and campaign t for group g as

$$Adherence_{itg} = \frac{\sum_{w \in \mathcal{W}(t)} \sum_{c \in \mathcal{C}(g)} \mathbb{I}(Price_{wc} = PriceDSS_{wc})}{|\mathcal{W}(t)| \times |\mathcal{C}(g)|},$$

where c indexes categories, $\mathcal{C}(g)$ is the set of categories in group g, w indexes weeks, $\mathcal{W}(t)$ is the set of weeks in campaign t, $PriceDSS_{wc}$ is the price suggested by the DSS, $Price_{wc}$ is the price implemented by the country manager, and $\mathbb{I}(\cdot)$ refers to the indicator function.

To measure the effect of each intervention separately, we use the data corresponding to the campaigns immediately before and immediately after it took place. For instance, to measure the effect of Intervention 1, which took place in S2011, we only use data from W2010 (pre-intervention, $Int1_t = 0$) and S2011 (post-intervention, $Int1_t = 1$). Additionally, given that the DSS was implemented gradually and some country managers started being able to use it after one or both interventions had taken place, we only use data from countries that were already using the DSS before each intervention.

We obtain our effect of interest by estimating the following linear regression using OLS

$$Adherence_{itg} = \alpha + \beta Int1_t + \gamma Franchise_i + \delta Int1_t \times Franchise_i + \varepsilon_{itg},$$

where the estimate of δ is the DiD coefficient. To study the effect of Intervention 2, we use $Int2_t$ instead if $Int1_t$. We cluster the standard errors at the country level. More details on using linear regression for DiD estimation can be found in Angist and Pischke (2009).

To strengthen our analysis, given the lack of randomization between treatment and control groups, we propose two alternative splits based on managers' adherence prior to each intervention. Indeed, the interventions' goal was to increase adherence. Hence, they were not aimed towards those who were already adhering the most, but towards those who were adhering the least. More specifically, we compute every manager's average adherence prior to each intervention. We then find the percentile 75 of this variable for each intervention, and label managers below these thresholds as low-adherence managers using the binary variable $LowAdherence75_i = 1$. We repeat the same procedure using percentile 90, and build the variable $LowAdherence90_i$. In both cases, managers who showed low adherence before each intervention are the treatment group.

2.4.2.2 Effect on Marking Down When the DSS Recommended Staying Put.

In addition to studying adherence, we want to measure the effect of the interventions on managers' likelihood to mark products down when the DSS was recommending keeping their price unchanged, as this was a very common way in which they deviated from the DSS's recommendations. We compute the probability of each manager marking a product down conditional on the DSS recommending keeping its price unchanged as

$$CMarkdown_{itg} = \frac{\sum_{w \in \mathcal{W}(t)} \sum_{c \in \mathcal{C}(g)} \mathbb{I}\left(Price_{wc} < Price_{w-1,c}\right) \times \mathbb{I}\left(PriceDSS_{wc} = Price_{w-1,c}\right)}{\sum_{w \in \mathcal{W}(t)} \sum_{c \in \mathcal{C}(g)} \mathbb{I}\left(PriceDSS_{wc} = Price_{w-1,c}\right)}$$

and repeat the previous DiD analysis with this dependent variable. Note that, for some countries and groups, the DSS always recommended marking down, so $CMarkdown_{itg}$ is not defined, and there are fewer observations of this variable than there are of $Adherence_{itg}$.

2.4.3 Results and Discussion

2.4.3.1 Effect on Adherence.

Figure 2.3 shows the distribution of managers' adherence per campaign, for own-store countries and franchises separately. It seems that the average adherence of these two country types did not change when Intervention 1 took place (S2011). However, after Intervention 2 was performed, in S2012, franchises' adherence increased dramatically.

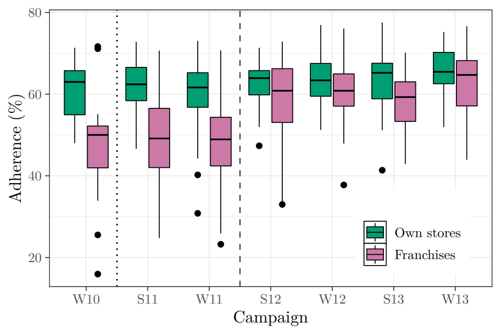


Figure 2.3: Distribution of managers' adherence to the DSS's recommendations by campaign, for own-store countries (green) and franchises (pink). The dotted line marks when Intervention 1 took place (S2011), and the dashed line marks Intervention 2 (S2012).

In Table 2.4 we can see the coefficients of the DiD estimation for both interventions, where the dependent variable is managers' adherence $Adherence_{itg}$. Intervention 1 did not have a significant effect on franchises respect to own-store countries. However, it did have a slightly significant effect ($p < 0.05$) on country managers with a past of low adherence. When we label $LowAdherence75_i$ as treated, their increase in adherence is 4.2 percentage points; the effect size is 4.99 percentage points if we we label $LowAdherence90_i$ as treated.

In contrast, Intervention 2 does show a positive effect, significant at the 0.001 level, for all three treated/control splits. When it took place, franchises' adherence increased by 9.1 percentage points. In countries where $LowAdherence75_i = 1$, adherence increased by 9.83 percentage points. If we use $LowAdherence90_i$ as treated, their adherence increased by 10.9 percentage points. Note that own-store countries' adherence also increased after Intervention 2 (by 2.78 percentage points), and this increase may be also explained by the intervention. Franchises, who were arguably the target of the interventions, experienced an increase in adherence which is more than 4 times that of own-store countries. In total, franchises' adherence increased by 25% after the second intervention.

	Intervention 1		
	Adherence		
	(1)	(2)	(3)
Int1	0.0109	-0.0166	-0.0332
	(0.00767)	(0.0118)	(0.0197)
Franchise	-0.145***		
	(0.0280)		
Int1×Franchise	0.00655		
	(0.0265)		
LowAdherence75		-0.171***	
		(0.0172)	
Int1×LowAdherence75		0.0420*	
		(0.0183)	
LowAdherence90			-0.163***
			(0.0194)
Int1×LowAdherence90			0.0499*
			(0.0230)
Constant	0.606***	0.674***	0.697***
	(0.0115)	(0.00708)	(0.0116)
N	2216	2216	2216

Robust standard errors in parentheses

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

	Intervention 2		
	Adherence		
	(1)	(2)	(3)
Int2	0.0278**	-0.00443	-0.0284
	(0.0103)	(0.0121)	(0.0183)
Franchise	-0.122***		
	(0.0222)		
Int2×Franchise	0.0910***		
	(0.0265)		
LowAdherence75		-0.149***	
		(0.0171)	
Int2×LowAdherence75		0.0983***	
		(0.0204)	
LowAdherence90			-0.170***
			(0.0160)
Int2×LowAdherence90			0.109***
			(0.0233)
Constant	0.593***	0.649***	0.690***
	(0.0130)	(0.0103)	(0.00960)
N	2999	2999	2999

Robust standard errors in parentheses

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

Table 2.4: Effect of Intervention 1 (left) and Intervention 2 (right) on adherence, using own-store countries as the control group and franchises are the treated (1), managers in the top quartile of pre-intervention adherence as the control group (2), and managers in the top decile of pre-intervention adherence as the control group (3).

2.4.3.2 Effect on Marking Down When the DSS Recommended Staying Put.

What was the effect of the interventions on $CMarkdown_{itg}$? As shown in Figure 2.4, after Intervention 2 the frequency at which franchise managers marked products down when the DSS did not recommend it decreased, but it is unclear if Intervention 1 had any effect.

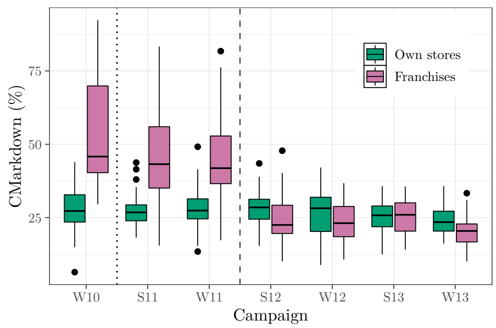


Figure 2.4: Distribution of managers' probability of marking a product down when the DSS recommended keeping its price unchanged, by campaign, for own-store countries (green) and franchises (pink). The dotted line marks when Intervention 1 took place (S2011), and the dashed line marks Intervention 2 (S2012).

Do the numerical results back up this claim? Indeed, for Intervention 1, the DiD coefficient is only statistically significant ($p < 0.01$) when the control is the set of managers such that $LowAdherence90_i = 1$ (column 3 of Table 2.5, left), but it is not for the other two splits. Nevertheless, Intervention 2 decreased franchise managers' likelihood to markdown when recommended to keep a price unchanged by 22.5 percentage points ($p < 0.001$), and this estimate is 15.8 and 14.5 for the other two control/treated splits, respectively.

To summarize, the Intervention 1 did not have a robust effect on either managers' adherence or managers' willingness to markdown when they should not, but the Intervention 2 did, and this effect is robust across three different control/treated splits. What these results suggest is that, consistent with the literature on numeracy and number sense, managers did not fully understand the meaning of the revenue metric Y in isolation, and thus they did not necessarily use this information when it was available to them. However, when they were

	Intervention 1		
	CMarkdown		
	(1)	(2)	(3)
Int1	0.00437	0.0215	0.0664***
	(0.0122)	(0.0138)	(0.0176)
Franchise	0.279***		
	(0.0432)		
Int1×Franchise	-0.0641		
	(0.0357)		
LowAdherence75		0.167***	
		(0.0329)	
Int1×LowAdherence75		-0.0458	
		(0.0241)	
LowAdherence90			0.142***
			(0.0315)
Int1×LowAdherence90			-0.0833**
			(0.0240)
Constant	0.279***	0.258***	0.249***
	(0.0148)	(0.00959)	(0.0163)
N	2172	2172	2172

Robust standard errors in parentheses

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

	Intervention 2		
	CMarkdown		
	(1)	(2)	(3)
Int2	-0.00230	0.0119	0.0234
	(0.0113)	(0.0151)	(0.0135)
Franchise	0.185***		
	(0.0257)		
Int2×Franchise	-0.225***		
	(0.0300)		
LowAdherence75		0.146***	
		(0.0256)	
Int2×LowAdherence75		-0.158***	
		(0.0274)	
LowAdherence90			0.154***
			(0.0282)
Int2×LowAdherence90			-0.145***
			(0.0247)
Constant	0.285***	0.262***	0.233***
	(0.0117)	(0.0175)	(0.0224)
N	2972	2972	2972

Robust standard errors in parentheses

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

Table 2.5: Effect of Intervention 1 (left) and Intervention 2 (right) on managers' probability of marking a product down when the DSS recommended keeping its price unchanged, using own-store countries as the control group and franchises are the treated (1), managers in the top quartile of pre-intervention adherence as the control group (2), and managers in the top decile of pre-intervention adherence as the control group (3).

shown a reference point for Y, which helped them make sense of this metric, their behavior significantly changed towards higher adherence to the DSS's recommendations.

2.5. Drivers of Adherence to the DSS's Recommendations

2.5.1 Context

In our data, variation in adherence behavior is not only driven by individual differences between country managers or product groups. In other words, we do not observe managers who systematically adhered or deviated, or high adherence for shirts but low for blazers, etc. Instead, any given country manager made different adherence decisions in different weeks. In this section, we analyze the behavioral drivers of their adherence decisions for each product and week.

2.5.2 Hypotheses

2.5.2.1 Preference for the Status Quo.

Before the implementation of the DSS, managers typically used the following rule of thumb: they checked the estimated time to sell the remaining stock of each category at its current price, which was computed using that category's speed of sales in the previous three days and, if there was a substantial risk of unsold inventory of that category at the end of the campaign, they would mark it down; otherwise, they would leave the current price unchanged. We hypothesize that, once the DSS was implemented, managers were (possibly implicitly) expecting it to follow the same heuristic and, hence, they were more likely to deviate when the DSS's recommendation was inconsistent with this rule. The interventions, by making it explicit that the objective was not to sell everything out but to maximize revenue, helped managers understand that their previous rule of thumb was not aligned with their objective, and so after they took place managers stopped expecting the DSS to follow that simple heuristic. This is an instance of status quo bias; people's preference for their current decisions or the current state has been widely documented (Samuelson and Zeckhauser 1988, Kahneman et al. 1991).

Hypothesis 5. *Managers were more likely to deviate from the price recommendation when it was not consistent with the heuristic they used before the DSS's implementation.*

2.5.2.2 Salience of the Inventory and the Speed of Sales.

Until the interventions took place, the inventory and sales metrics that they used in the past were also emphasized in the DSS's interface. Hence, these variables were more salient to managers than the revenue. Moreover, inventory is material and visible, so possibly more salient than a forecast of revenue, a more abstract concept. We hypothesize that managers set prices trying to sell all inventory out, instead of maximizing revenue. If salience of the inventory was driving the magnitude of their price deviations, then the interventions, which made the Y metric more salient, mitigated this effect. Salience is a cognitive bias well described in Tversky and Kahneman (1974), Bordalo et al. (2013).

Hypothesis 6. *Managers deviated from the DSS's recommendation by a smaller amount when a product had a high speed of sales and a low inventory level.*

2.5.2.3 Loss Aversion.

Loss aversion (Kahneman and Tversky 1979, Camerer 1998, Kőszegi and Rabin 2006) could also be leading managers to set prices with the objective of selling all inventory out. In this setting, loss aversion would manifest itself in the following way: managers would have an internal preferred price for each product, based on their own past training and experience; the DSS would recommend a different price, possibly higher, which managers would encode as a gain; however, when comparing it to the loss from having to salvage unsold inventory, they overweighted the expected loss over the expected gain, and they decided to deviate from the recommendation. If loss aversion was driving the probability and magnitude of managers' price deviations from the DSS, higher salvage values should decrease the perceived loss from unsold inventory, aligning managers' objective with revenue maximization, and hence making their pricing decisions more similar to the DSS's recommendations.

Hypothesis 7. *Managers were less likely to deviate from the DSS's recommendation or, if they decided to deviate, they deviated by a smaller amount, when a product's salvage value was high.*

2.5.2.4 Cognitive Limitations.

The large amount of product groups (20 corresponding to women's apparel, plus many others) and number of clusters per group (up to 25 in our data) implies that managers had hundreds of prices to set every week, based on an equally large amount of inventory and sales metrics (and revenue, after the interventions). However, research shows that humans have limited cognitive capacity in complex decision making settings (Kahneman 1973). Moreover, when information acquisition is costly it can be rational to deviate from optimal decisions to respond to the trade-off between the effort of analyzing new data and the cost of deviating from optimality (Sims 2003, Cheremukhin et al. 2011, Caplin and Dean 2015). In the DSS, managers were able to observe metrics of inventory and speed of sales for every category, but also the group-aggregate ones (which were emphasized). Given the huge amount of groups and clusters, ignoring the category-level metrics in favor of the group-level ones would greatly facilitate their work. In addition, it would make sense for them to reduce the number of pricing decisions to make (and of metrics to inspect) by grouping different clusters into categories. As described in Section 2.3.1, once two clusters' prices coalesce, they become one category, and the manager needs to make one common pricing decision, instead of two, in the subsequent weeks. The DSS's algorithm did not take the number of categories per group into account. Therefore, if they were trying to reduce this number (at the expense of revenue), they would need to deviate from the DSS's recommendations.

Hypothesis 8. *(a) Managers were trying to minimize the number of distinct prices to assign to products, i.e., the number of product categories.*

(b) Managers were paying more attention to group-aggregate metrics than to individual category ones when setting prices for each category.

2.5.3 Methods

To test our hypotheses, we need to use a regression model that accounts for the nature of manager's pricing decision process: first, the manager observes the DSS's price recommendation for that category, and decides whether to adhere or to deviate; second, in case they decide to deviate, they set their preferred price. For this reason, we use the Heckman model, also called Heckit. This is a two-step estimator that models a decision process of this type. In the first step, the manager decides whether to adhere or to deviate from the DSS's recommendation. This is called the *selection part* and it is modeled like a probit regression. In the second step, provided that the manager has decided to deviate, they decide by how much. This is called the *deviation part* and is estimated as a linear regression which contains a term that corrects for the selection decision of the first step. The standard errors are robust, following Heckman (1977). Additional details of this method can be found in Wooldridge (2010), Cameron and Trivedi (2005), and Greene (2007).

We test Hypotheses 5, 6, 7 and 8(a) using the covariates we describe below. Hypothesis 8(b) states that managers were focusing on group-aggregate metrics. To test it, we run the Heckman regression in two slightly different ways: (1) with covariates computed at the category level; (2) with covariates computed at the group level. We then compare which one of these two variations shows higher coefficient estimates.

2.5.3.1 Dependent Variables.

Selection part. The first step of the Heckman estimator models the probability of deviating from the DSS's recommended price. Therefore, its dependent variable is binary, and takes value 1 if the manager's confirmed price for category c in week w is different from the recommended one, i.e., $Deviated_{wc} = \mathbb{I}\left(Price_{wc} \neq PriceDSS_{wc}\right).$

Deviation part. The second step of this regression models the amount of deviation. We compute the difference between the manager's confirmed price and the DSS's recommendation and, to make this difference comparable across products, we divide it by the DSS's recommended price (otherwise expensive products would have larger price differences). We take the absolute value of this normalized price difference to be able to interpret our results in terms of price deviations respect to the DSS, and not in terms of higher or lower prices. Hence, the dependent variable in the deviation part of the Heckman model is defined as

$$AbsDeviation_{wc} = \left| \frac{Price_{wc} - PriceDSS_{wc}}{PriceDSS_{wc}} \right|.$$

2.5.3.2 Independent Variables.

Disagreement between the DSS and managers' previous heuristics. To test Hypothesis 5, we compute the binary variable $DisagreeSS_{wc}$, which takes value 1 if the DSS's recommendation was consistent with the previous rule of thumb, and 0 otherwise. To compute it, we use the continuous variable $StockoutWeek_{wc}$ defined below.

More specifically, a category is predicted to sell out before the end of the campaign if $StockoutWeek_{wc} \leq 12$, and a group is predicted to do so if $StockoutWeek_{wg} \leq 12$. The variable $DisagreeSS_{wc}$, computed at the category level, takes value 1 if either the category is predicted to sell out ($StockoutWeek_{wc} \leq 12$) but the DSS is recommending to markdown, or if c is predicted not to sell out ($StockoutWeek_{wc} > 12$) but the DSS is recommending not to markdown. Analogously, the variable $DisagreeDSS_{wg}$ takes value 1 if the group to which category c belongs is predicted to sell out ($StockoutWeek_{wg} \leq 12$) but the DSS is recommending to mark c down, or if g is predicted not to sell out ($StockoutWeek_{wg} > 12$) but the DSS is recommending not to mark c down.

We include this variable in the selection part of the Heckman regression: if there is a disagreement between the DSS and their beliefs, they will be more likely to to deviate.

Stockout period. To test Hypothesis 6, we build a variable based on one of the metrics contained in the DSS's interface, the rotation (described in Section 2.3.2). However, this metric would tend to capture a time trend (the remaining time to sell all inventory out is nonincreasing). Instead, we compute the period in which every category will be sold out if the speed of sales is the same as in the previous week. For week w and category c, this variable is defined as

$$StockoutWeek_{wc} = w + \frac{InventoryLevel_{wc}}{Sales_{w-1,c}}.$$

To study the effect of group-aggregate metrics, we use them to compute, for group g,

$$StockoutWeek_{wg} = w + \frac{InventoryLevel_{wg}}{Sales_{w-1,g}}.$$

Note that, if there have been no sales for a given category or group in a given week, these variables will not be defined. Or, if the sales have been very small compared to the inventory levels, these variables will take a value in the order of thousands. However, the clearance sales campaigns spans a limited period of time (typically 12 weeks). Therefore, we assume that managers are not sensitive to when exactly the category would be sold out if this is not going to occur even after multiple campaigns, and we cap $StockoutWeek_{wc}$ and $StockoutWeek_{wg}$ by setting any value of these variables greater than 24 to take value 24.

We include this variable in the deviation part of the Heckman estimator: we hypothesize that the exact amount of inventory that remains to be sold is related to the exact amount by which they deviate from the DSS's recommendation.

Salvage value. To make salvage values comparable across product groups and categories, we need to normalize them to the same scale. Given that managers set prices weekly, they should update beliefs on what their preferred prices are (and what the loss for salvaging products is) weekly, using the last price they set for that category as a reference point. We compute the relative salvage value of category c in week w as

$$RelSalvageValue_{wc} = \frac{SalvageValue_{wc}}{Price_{w-1,c}}.$$

We include this variable in both the selection and the deviation parts of the Heckman regression, since loss aversion can play a role in both the probability of deviating and the magnitude of price deviations, as explained in Hypothesis 7.

Number of product categories. We denote the number of different product categories in group g and week w as $NumCategs_{wg}$, and we use this variable to test Hypothesis 8(a).

Intervention stage. In Section 2.4 we have seen that Intervention 2 had a significant effect on adherence. To better understand the relationship between the covariates defined above and managers' adherence, we need to study how the interventions changed that relationship. For this reason, we define three binary variables, which represent different periods in our data: $Baseline_t$ takes value 1 if $t = W2010$, i.e., before the interventions; $Int1_t$ takes value 1 if $t \in \{S2011, W2011\}$, i.e., after Intervention 1 but before Intervention 2; $Int2_t$ takes value 1 if $t \in \{S2012, W2012, S2013, W2013\}$, after Intervention 2. We interact these three dummies with the covariates defined above to observe if the interventions changed the extent at which our covariates were related to adherence.

Control variables. We control for the number of stores in each country and campaign ($LogNumStores_{it}$, in logarithmic scale) and the number of campaigns that the manager had been using the DSS ($ExperienceWithDSS_{it}$). Additionally, we include the following dummy variables in both parts of the Heckman estimation: country type (own stores or franchises), season (spring-summer or fall-winter), week, year, product group, and country.

2.5.4 Results and Discussion

2.5.4.1 Selection Part.

Table 2.6 shows the coefficient estimates for the selection part of the Heckman regression, and Table 2.7 its corresponding average partial effects (APEs). From the APEs table, columns 1 to 3, we can see that, before the interventions, when there was a disagreement between the DSS's recommendation and the managers' previous heuristic the manager was 17.1 percentage points (or 45%) more likely to deviate; after the first intervention, they were 10.5 percentage points (or 28% of their adherence on that period); after the second intervention, this number had reduced to 6.05 percentage points (or 17%). All these APEs are all statistically significant at the 0.001 level. These numbers are higher when we run the regression using own-store countries only (19.2, 16.3 and 11.6 percentage points) than when we use only franchises. In particular, this APE is negative for franchises after Intervention 1, and insignificant after Intervention 2. When a different model specification is used (see the robustness checks in the E-Companion), this variable is not statistically significant for franchises at any stage of the interventions, but is for own-store countries. These results provide support for Hypothesis 5: managers, particularly from own-store countries, were more likely to deviate from the DSS when its recommendation was not aligned with the rule they had been following before its implementation, and the interventions mitigated this behavior. It is unclear that managers from franchise countries were following the rule of thumb as strictly as own-store country managers, given that they purchased the inventory and then used their own pricing systems, and that may be the reason why the effect of agreement between the heuristic and the DSS is weak.

For own-store country managers, the relative salvage value has a negative, statistically significant APE in all stages, and its magnitude actually increases after each intervention. Given the average relative salvage value in our data, a 10% higher salvage value of a category is related to a 2.1 percentage points (5.8%) in the probability of deviating from the DSS's recommendation before the interventions, and this effect is 2.54 percentage points (roughly

7%) after Intervention 1 and 3.05 percentage points after Intervention 2 (or 9%). This provides support for Hypothesis 7: a higher salvage value decreased their probability of deviating from the DSS's recommendations, showing that their pricing decisions were affected by loss aversion. However, the interventions did not mitigate this bias but, instead, it increased with time.

The number of categories in the group also influenced managers' willingness to adhere to the DSS's recommendations, especially for own-store country managers. Before the interventions, every additional category in a group increased by 1.08 percentage points the probability they would deviate. After Intervention 1, this number was 0.73, but after Intervention 2 it increased again to 1.2 percentage points. All these APEs are statistically significant at the 0.001 level. For franchise mangers, before the second intervention the number of categories was insignificant, but after it was 1.13 percentage points ($p < 0.001$). In the robustness checks (see the E-Companion), the number of categories is only statistically significant after Intervention 2. These results provide partial support for Hypothesis 8(a): managers, after the second intervention particularly, were trying to reduce the number of pricing decisions (categories) to make every week, and this lead them to deviate from the DSS's recommendations. When the disagreement variable is computed using group-aggregate inventory and sales (columns 4 to 6), its effect is 22.1 (before the interventions), 12.4 (after Intervention 1) and 6.89 percentage points (after Intervention 2). These coefficients are all higher than when the disagreement variable is computed using category metrics. This supports Hypothesis 8(b): when making pricing decisions for individual categories, managers were more influenced by the group-aggregate metrics than by category-level ones.

2.5.4.2 Deviation Part.

Table 2.8 contains the coefficient estimates of the deviation part of the Heckman regression. For own-store country managers, the speed of sales is significant at the 0.001 level for all intervention stages, with coefficient 0.00698 (before any intervention), 0.00636 (after Intervention 1), and 0.00676 (after Intervention 2). In other words, every additional week to sell a

	ProbDeviation					
	All countries	Own stores	Franchises	All countries	Own stores	Franchises
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline×DisagreeDSS$_c$	0.478***	0.560***	0.187***			
	(0.0187)	(0.0215)	(0.0378)			
Int1×DisagreeDSS$_c$	0.296***	0.476***	-0.0563**			
	(0.0113)	(0.0146)	(0.0185)			
Int2×DisagreeDSS$_c$	0.175***	0.349***	0.0117			
	(0.00621)	(0.00893)	(0.00882)			
Baseline×DisagreeDSS$_g$				0.621***	0.727***	0.269***
				(0.0189)	(0.0218)	(0.0384)
Int1×DisagreeDSS$_g$				0.350***	0.554***	-0.0361
				(0.0114)	(0.0147)	(0.0185)
Int2×DisagreeDSS$_g$				0.200***	0.391***	0.0186*
				(0.00625)	(0.00902)	(0.00883)
Baseline×RelSalvageValue	-2.804***	-2.627***		-2.812***	-2.581***	
	(0.0628)	(0.0825)		(0.0636)	(0.0837)	
Int1×RelSalvageValue	-2.691***	-3.175***		-2.634***	-3.110***	
	(0.0464)	(0.0630)		(0.0466)	(0.0634)	
Int2×RelSalvageValue	-2.492***	-3.813***		-2.457***	-3.746***	
	(0.0350)	(0.0441)		(0.0351)	(0.0442)	
Baseline×NumCategs	-0.0321***	0.0330***	-0.0198	-0.0383***	0.0299***	-0.0259*
	(0.00615)	(0.00791)	(0.0104)	(0.00619)	(0.00798)	(0.0104)
Int1×NumCategs	-0.00709	0.0224***	-0.000819	-0.00734	0.0202***	-0.000392
	(0.00400)	(0.00576)	(0.00571)	(0.00400)	(0.00579)	(0.00571)
Int2×NumCategs	0.0384***	0.0368***	0.0323***	0.0373***	0.0357***	0.0322***
	(0.00210)	(0.00311)	(0.00293)	(0.00211)	(0.00312)	(0.00293)
ExperienceWithDSS	-0.370**	0.346*	-0.138***	-0.381***	0.334*	-0.137***
	(0.114)	(0.138)	(0.0193)	(0.115)	(0.138)	(0.0193)
LogNumStores	0.0355	0.0372	-0.0540	0.0354	0.0339	-0.0542
	(0.0193)	(0.0271)	(0.0285)	(0.0193)	(0.0272)	(0.0285)
Franchise	-0.139***			-0.138***		
	(0.0335)			(0.0335)		
Constant	1.014***	-0.918***	0.270***	1.011***	-0.942***	0.270***
	(0.162)	(0.163)	(0.0618)	(0.162)	(0.164)	(0.0618)
N	294386	169137	125249	294386	169137	125249

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6: Coefficients of the Heckman selection part (decision to deviate from the DSS's recommended price). Its corresponding average partial effects are shown in Table 2.7. Not reported: week, season, year, group and country dummy variables.

	ProbDeviation					
	All countries	Own stores	Franchises	All countries	Own stores	Franchises
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline×DisagreeDSS$_c$	0.171***	0.192***	0.0665***			
	(0.00680)	(0.00743)	(0.0137)			
Int1×DisagreeDSS$_c$	0.105***	0.163***	-0.0196**			
	(0.00413)	(0.00511)	(0.00641)			
Int2×DisagreeDSS$_c$	0.0605***	0.116***	0.00410			
	(0.00225)	(0.00313)	(0.00326)			
Baseline×DisagreeDSS$_g$				0.221***	0.248***	0.0962***
				(0.00676)	(0.00734)	(0.0139)
Int1×DisagreeDSS$_g$				0.124***	0.189***	-0.0126
				(0.00416)	(0.00510)	(0.00645)
Int2×DisagreeDSS$_g$				0.0689***	0.129***	0.00652*
				(0.00226)	(0.00317)	(0.00327)
Baseline×RelSalvageValue	-0.965***	-0.860***		-0.966***	-0.840***	
	(0.0225)	(0.0272)		(0.0225)	(0.0269)	
Int1×RelSalvageValue	-0.926***	-1.039***		-0.904***	-1.012***	
	(0.0163)	(0.0220)		(0.0163)	(0.0219)	
Int2×RelSalvageValue	-0.858***	-1.248***		-0.843***	-1.219***	
	(0.0113)	(0.0139)		(0.0113)	(0.0139)	
Baseline×NumCategs	-0.0111***	0.0108***	-0.00694	-0.0132***	0.00973***	-0.00909*
	(0.00215)	(0.00262)	(0.00369)	(0.00216)	(0.00261)	(0.00372)
Int1×NumCategs	-0.00244	0.00732***	-0.000287	-0.00252	0.00658***	-0.000138
	(0.00138)	(0.00190)	(0.00201)	(0.00138)	(0.00188)	(0.00201)
Int2×NumCategs	0.0132***	0.0120***	0.0113***	0.0128***	0.0116***	0.0113***
	(0.000720)	(0.00101)	(0.00103)	(0.000720)	(0.00101)	(0.00103)
ExperienceWithDSS	-0.127**	0.113*	-0.0483***	-0.131***	0.109*	-0.0482***
	(0.0396)	(0.0454)	(0.00674)	(0.0395)	(0.0452)	(0.00674)
LogNumStores	0.0122	0.0122	-0.0189	0.0122	0.0110	-0.0190
	(0.00668)	(0.00892)	(0.00999)	(0.00667)	(0.00888)	(0.00999)
Franchise	-0.0474***			-0.0472***		
	(0.0111)			(0.0110)		
N	294366	169094	125249	294366	169094	125249

Delta-method standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.7: Average partial effects of the selection part of the Heckman regression on the decision to deviate from the DSS's recommended price and the magnitude of such deviation (coefficient estimates in Table 2.6). Not reported: week, season, year, group and country dummy variables.

category out, when mangers decide to deviate, is related to a 1.9% higher deviation from the DSS's recommendation for this category before the interventions, but after the interventions this decreases to 1.35%. This supports Hypothesis 6. Moreover, it seems that the interventions effectively shifted salience from inventory and speed of sales to revenue. However, in contrast to our findings in the selection part, when the stockout week is computed for the entire group, its coefficient is smaller than when it is computed for an individual category, which contradicts Hypothesis 8(b). Moreover, the stockout week variable, computed at the group level, is statistically insignificant in the robustness checks (see E-Companion).

The relative salvage value is statistically significant at the 0.001 level before Intervention 2, with coefficient estimate 0.175 (before Intervention 1) and 0.138 (after Intervention 1). Taking into account the average salvage value, a 10% increase in it is related to 1.45% higher price deviations from the DSS before the interventions, and 1.21% after Intervention 1. Even though, as we have seen, a higher salvage value made managers less likely to deviate from the DSS's recommendation, it made the size of their deviations higher, which does not support Hypothesis 7. The interventions decreased the importance that they placed on the salvage value to determine the size of their deviations, to the point of making it insignificant it after Intervention 2.

For franchise countries, the speed of sales, before any intervention, has coefficient 0.0066 (when the speed of sales is computed at the category level), and 0.00701 (when it is at the group level), with $p < 0.01$ and $p < 0.05$, respectively. After Intervention 1, these coefficients are both statistically insignificant. After Intervention 2, these coefficients are 0.00428 and 0.00125, respectively ($p < 0.001$). Given franchises' average size of their deviations, what these numbers mean is that, before any intervention, an additional week to sell a group out was related to a 4.8% larger price deviation, and this figure decreased to less than 1% after the interventions. In particular, after the interventions their attention to speed of sales was centered to category-level, and not group-level. These findings provide support for Hypothesis 6, as the speed of sales is related to the magnitude of their deviations (except during 2011), and partial support for Hypothesis 8(b), as before the interventions the coefficient of speed of sales was higher when it is computed at he group level.

A possible explanation of the seemingly odd coefficients of speed of sales and disagreement for franchises after Intervention 1 and before Intervention 2 is the following: during spring-summer and fall-winter, 2011, franchise countries purchased more inventory in each one of those campaigns than in fall-winter, 2010, but their sales dropped drastically (around 60%). Therefore, it is likely that, at the speed of sales of most products on those two campaigns, the DSS predicted that they would not be sold out during clearance sales. This explains why, on those campaigns, franchise managers did not place any importance to the exact period in which products would sell out ($StockoutWeek$ is statistically insignificant, both when computed at the category and at the group level), and disagreement between the DSS and their heuristic has a negative effect (when computed at the category level) and is statistically insignificant (when computed at the group level).

The coefficient estimates of the Mills ratio are statistically significant. This tests whether our results would be different if we did not take into account the two-step nature of this pricing process and we studied price deviations from the DSS using a linear regression. The fact that this coefficient is statistically significant proves the need for using the Heckman estimator, or any other two-step model, instead of a simpler method.

Finally, note that, although Intervention 1 did not have a significant effect on adherence, the coefficient estimates of some of our covariates are smaller after that intervention took place. In other words, Intervention 1 mitigated managers' status quo bias, as well as salience of the inventory and speed of sales. Although Intervention 1 was not sufficient to increase managers' adherence by itself, it did induce behavioral changes, paving the road for Intervention 2.

2.6. Conclusion and Managerial Insights

When companies implement advanced solutions for their operational decision making, such as optimization-based algorithms, they often do so in the form of a decision support system

	AbsDeviation					
	All countries	Own stores	Franchises	All countries	Own stores	Franchises
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline×StockoutWeek$_c$	0.00630***	0.00698***	0.00660**			
	(0.000924)	(0.000754)	(0.00208)			
Int1×StockoutWeek$_c$	0.00122***	0.00636***	0.000535			
	(0.000334)	(0.000496)	(0.000553)			
Int2×StockoutWeek$_c$	0.00464***	0.00676***	0.00428***			
	(0.000205)	(0.000258)	(0.000323)			
Baseline×StockoutWeek$_g$				0.00236	0.00269**	0.00701*
				(0.00125)	(0.00104)	(0.00288)
Int1×StockoutWeek$_g$				-0.00104**	0.00184**	-0.000376
				(0.000348)	(0.000604)	(0.000562)
Int2×StockoutWeek$_g$				0.00131***	0.00271***	0.00125***
				(0.000221)	(0.000325)	(0.000336)
Baseline×RelSalvageValue	0.0547	0.175***		0.339***	0.332***	
	(0.0355)	(0.0277)		(0.0339)	(0.0270)	
Int1×RelSalvageValue	0.110***	0.138***		0.332***	0.330***	
	(0.0305)	(0.0244)		(0.0284)	(0.0233)	
Int2×RelSalvageValue	-0.177***	0.00363		0.0722**	0.227***	
	(0.0286)	(0.0237)		(0.0264)	(0.0220)	
ExperienceWithDSS	-0.0997*	0.229***	-0.00769	-0.0816	0.212***	-0.00634
	(0.0477)	(0.0387)	(0.0117)	(0.0482)	(0.0396)	(0.0119)
LogNumStores	-0.0335***	-0.0471***	-0.00768	-0.0409***	-0.0501***	-0.00928
	(0.00819)	(0.00756)	(0.0157)	(0.00828)	(0.00774)	(0.0161)
Franchise	-0.0419**			-0.0331*		
	(0.0140)			(0.0141)		
Constant	0.612***	0.0145	0.451***	0.686***	0.132**	0.508***
	(0.0647)	(0.0457)	(0.0467)	(0.0655)	(0.0465)	(0.0476)
Mills ratio	0.0421***	0.00539	-0.126*	-0.0946***	-0.0783***	-0.222***
	(0.0123)	(0.00716)	(0.0567)	(0.0108)	(0.00639)	(0.0551)
N	294386	169137	125249	294386	169137	125249

Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.8: Coefficients of the Heckman deviation part (magnitude of price deviations). Not reported: week, season, year, group and country dummy variables.

(DSS), a tool that suggests optimal decisions to human decision makers but allows them to deviate from such recommendations. The success of a DSS's implementation, and hence its effect on the firm's performance, depends on decision makers' adherence to its suggestions. Therefore, it is critical that we understand how human decision makers interact with such tools to design them in a way that entices managers to adhere to their recommendations.

In this paper, we use data collected by Zara during seven clearance sales campaigns to study managers' adherence to price recommendations from a DSS. Although adhering to the DSS's recommendations was associated to higher revenue, mangers' initial adherence was very low. The firm performed two interventions, in the form of changes in the DSS's interface, with the objective to increase adherence. We analyze the effect of the interventions using a difference-in-differences analysis. We find that Intervention 2 had a strong, positive effect on adherence, and also significantly decreased managers' likelihood to mark a product down when the DSS recommended leaving its price unchanged. Intervention 1 had a small effect in countries with low pre-DSS adherence but not on franchises compared to own-store countries.

We investigate the drivers of managers' adherence using a Heckman regression, and relate them to cognitive biases. Managers were significantly more likely to implement the DSS's suggested prices when such recommendations were consistent with the rule of thumb they were trained to follow in the past, which can be explained by with status quo bias. When a product was selling quickly, managers' price deviations from the DSS were smaller, which can be attributed to salience of the inventory respect to a revenue forecast. When products had a large salvage value, managers were less likely to deviate from the recommended prices, but the magnitude of their deviations when they did not adhere was slightly higher, which provides partial evidence of loss aversion. The effect of some group-aggregate metrics on adherence was higher than that of individual product metrics. Moreover, managers were trying to reduce the number of distinct number of prices to assign to products. This is consistent with inattention and cognitive limitations. The interventions mitigated managers' status quo bias and effectively shifted salience from inventory to revenue, but did not eliminate managers' loss aversion and inattention. Moreover, although Intervention 1 did not have a significant effect on adherence, it did have an impact on the extent at which status quo bias and salience were related to adherence decisions, which Intervention 2 further changed. In other words, Intervention 1 induced changes in behavior that helped pave the way for Intervention 2.

More generally, we have shown empirically that the way DSS's interfaces are designed and the information they display can have a great impact on how decision makers use them, how much they adhere to their recommendations and, ultimately, on the success of their implementation and the goodness of their operational decisions.

Our findings provide a few managerial insights for any company that wants to implement a DSS, or any other analytical tool to be adopted voluntarily by its users. First, the firm should make sure decision makers' objectives and incentives are aligned with the firm's. However, this may not be enough to ensure managers' decisions are optimal. In our work, managers were incentivized to increase revenue, but they acted like they were inventory minimizers. The fact that a change in the DSS's interface significantly altered their behavior shows us that we can achieve a greater alignment between the firm's objectives and managers' behavior when the DSS makes such objectives salient. Second, the firm should provide training and explicitly teach any change in rules and policies: managers in our data were expecting the DSS to suggest decisions in line with their previous rule of thumb, but this was not the heuristic that he algorithm was following. Third, the DSS's interface should provide a reference to make sense of any quantitative information, given that human decision makers are not very good at interpreting numbers. Finally, as we have seen in our work, managers' deviations can be partially explained by the large amount of prices to set and information to process, and this behavior persisted over all the data-collection period. Firms that want to implement a DSS but leave the final decision making responsibility to managers should be mindful of humans' finite cognitive capacity, and minimize the amount of information that managers need to process, as well as the number simultaneous decisions that they need to make.

Our work has, of course, numerous limitations. Regarding internal validity, the fact that our data is observational and not a controlled experiment means that we cannot make a strong claim of causality in any of our findings. In addition, other factors could be playing a role in managers' adherence, such as trust in the accuracy of the demand and revenue forecast, or demographic characteristics of managers. Unfortunately, the data collection was performed by the company, and these variables are not available to us.

Additionally, our findings have limited external validity. The type of information displayed in the DSS's interface, the company's idiosyncratic pricing rules, the business model in which all stores in a country have the same ownership type, etc. are specific to Zara. Nevertheless, many other important fashion retailers have clearance sales periods which work similarly to those in our data, so their pricing managers could be affected by the same cognitive biases that we find. The importance of making objectives salient in the interface and providing reference points for quantitative information can be extrapolated to any context in which human managers make numerical decisions assisted by a DSS.

CHAPTER 3

Loss Aversion in Managers' Pricing Decision Making at a Fast Fashion Retailer

3.1. Introduction

In the previous chapter, we have seen that country managers at Zara constantly deviate from revenue-maximizing prices during clearance sales. Moreover, their deviations tend to be skewed towards prices that are lower than the optimal ones (see Figure 3.1).

The way in which managers at Zara deviate from revenue-maximizing prices is compatible with a number of cognitive biases. One of these is loss aversion. Last chapter's results show that Zara country managers are more likely to deviate from the optimal prices when products have a low salvage value; however, counter to what we expected to find, the size of their deviations is smaller when the salvage value is low. Thus, it is unclear if loss aversion is what best explains their pricing behavior. Moreover, loss aversion and salience of inventory are very difficult to disentangle: do managers really care about the value of unsold inventory, or only about its amount? How much of each bias do they exhibit?

In this chapter, we build a structural model of managers' pricing decision making in order to identify whether they are loss averse and, if so, how much. In our model, managers at Zara choose prices to maximize their utility over the whole clearance sales season horizon. This price optimization problem includes some constraints that reflect the firm's actual pricing rules during clearance sales, which country managers also follow. The problem's objective function, managers' utility, contains a loss aversion term which depends on a loss aversion parameter, in the style of Kőszegi and Rabin (2006).

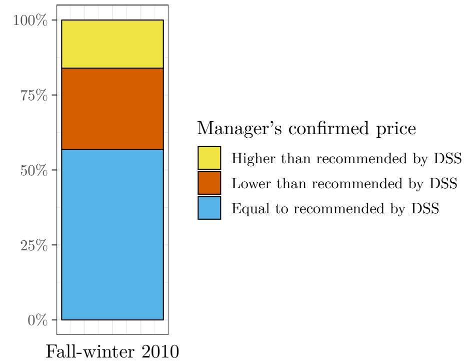


Figure 3.1: Decisions made by the managers during the Winter 2010 clearance sales campaigns with respect to the decision support system: 57% of the time, they adhered to the price recommended by the DSS; when they decided to deviate, 2/3 of the time they set prices that were lower than the recommended ones.

To calibrate our model and estimate managers' loss aversion parameter, we use data collected by Zara in two countries (Belgium and France) during 2007 and 2008 (i.e. before the decision support system that we studied in the previous chapter was implemented). At that time, managers chose clearance sales prices based entirely on the old inventory reports, the coarse heuristic we have described in the last chapter, and their intuition. Therefore, by using pre-DSS data we can analyze their pricing behavior *in the wild* and observe any cognitive biases they may have, without worrying that they might have been subject to the anchoring effect of the DSS's recommendations.

We use the data corresponding to the summer of 2007 to calibrate our demand forecast model, and then use the data from 2008 to estimate the two country managers' loss aversion parameter. For each value of this parameter, our model leads to a different set of prices that the managers would choose if they were maximizing a utility function like the one we propose. We pick the value of this parameter that provides the closest prices to the ones that they actually implemented for each product and week. We estimate a separate loss aversion parameter for each one of the two managers, to whom we refer as Manager B and Manager F, and for each product group in our data (shirts, pants, etc.).

We find that, overall, both managers are loss averse (their parameter estimates are greater than zero), and that Manager F is more loss averse than Manager B. However, we cannot find any clear pattern in their differences in loss aversion across product groups, such as being more loss averse for more fashionable products and less loss averse for basic products, or a relationship between their degree of loss aversion and products' salvage value.

Our model explains managers' pricing behavior around 11% better than if we assume that they are revenue maximizers, and this improvement is larger in product groups for which their loss aversion is higher (i.e., when the difference between the revenue function and their loss-averse utility is bigger). We also test whether managers were actually inventory minimizers, instead of revenue or utility maximizers, but this approach leads to a prediction error that is several times larger than our model's.

Given the very small sample size used in this chapter, these results are extremely preliminary and should be interpreted with caution. In §3.6 we provide a detailed list of next steps necessary for this research project to lead to a finished, publishable paper. In particular, we ultimately want to estimate the counterfactual revenue that managers at Zara would have generated if they had not been subject to loss aversion. However, we believe that our model needs to be improved and estimated with a much larger dataset and more sophisticated techniques before we can give a reliable estimate of the additional revenue that Zara is not making due to its managers' biased pricing behavior.

3.2. Literature Review

Loss aversion is a concept that was first introduced by Kahneman and Tversky (1979) as one of the components of prospect theory, which they proposed based on empirical results of laboratory experiments. Since then, this cognitive bias has been experimentally measured plenty of times (Thaler et al. 1997, Novemsky and Kahneman 2005, Gächter et al. 2007, Abdellaoui et al. 2007, 2008, Schmidt and Traub 2002), even in non-human subjects such as capuchin monkeys (Chen et al. 2006). Loss aversion "has the correct psychological foundation, it is theoretically useful, and it is a parsimonious principle that can explain many puzzles" (Camerer 2005).

Loss aversion has been observed in many real-life settings, such as consumers' choice of brand and product valuation (Tversky and Kahneman 1991, Hardie et al. 1993, Amaldoss and He 2017), labor supply in markets in which workers can choose how much or how hard to work (Camerer et al. 1997, Goette et al. 2004, Hossain and List 2012, Imas et al. 2016), and seller pricing behavior in housing markets (Genesove and Mayer 2001, Engelhardt 2003, Anenberg 2011). Loss aversion is related to the endowment effect, when individuals demand a much higher price to sell an item than what they paid for it (Thaler 1980, Kahneman et al. 1991, Ericson and Fuster 2014). Kőszegi and Rabin (2006) build on Kahneman and Tversky (1979) to provide a more general theory and model of loss aversion, and use it to explain the empirically observed consumer behavior and labor supply puzzles. For a summary of field evidence supporting prospect theory, see Camerer (1998).

These situations in which agents are loss averse, both in laboratory experiments and in the field, have one characteristic in common: subjects are not experts. For instance, in labor markets, a commonly studied group is the taxi drivers. While they are experts in driving, they are not specialists in labor economics, therefore their deviations from what economic models predict regarding how many hours they work can be attributed to lack of knowledge or ability to plan their work schedule optimally. In contrast, a context in which loss aversion has been observed both in non-experts and experts is the equity premium, or why investors hold bonds when stocks have a much higher return rate (Benartzi and Thaler 1995, Thaler et al. 1997, Gneezy and Potters 1997, Gneezy et al. 2003, Haigh and List 2005).

In this chapter, we study loss aversion in pricing behavior in subjects who are, indeed, experts in pricing. The decision makers in our data are managers at Zara, a large retailer present in tenths of countries, and these managers are responsible for all the prices during clearance sales in a whole country each one, i.e., are individuals who hold a great amount of responsibility in the company and are expected to be very sophisticated in their decision making processes. Our main contribution to the behavioral literature is that we are able to identify and quantify loss aversion in experts.

In the domain of operations management, loss aversion is most usually associated with sourcing decisions in a newsvendor setting. Experimentally, Schweitzer and Cachon (2000) find that loss aversion cannot explain decision bias in a newsvendor ordering context. However, Long and Nasiry (2014) argue that, with the proper choice of reference prices, loss aversion does explain the results in Schweitzer and Cachon (2000). Uppari and Hasija (2018) expand the model proposed by Long and Nasiry (2014). A large number of analytical papers use a newsvendor model where either the newsvendor or the consumers are loss averse, such as Su (2008), Wang (2010), Zhao and Stecke (2010), Liu et al. (2013), Baron et al. (2015), Xinsheng et al. (2015), Xu et al. (2017), among others.

Contract theory is another stream of operations management research in which loss aversion has been well studied. Ho and Zhang (2008), Katok and Wu (2009), Davis et al. (2014), Zhang et al. (2015), Davis (2015) find evidence of loss averse behavior in supply chain contracts using laboratory experiments. In contrast, in another laboratory experiment, loss aversion fails to explain the observed results (Johnsen et al. 2019). Multiple analytical papers which study supply chain contracts incorporate loss aversion in one or several agents (Wang and Webster 2007, Deng et al. 2013, Chen et al. 2014, Hu et al. 2016).

Empirically, consumers' loss aversion has been also studied in the operations management field, with applications to assortment planning and pricing (Wang 2018), and the performing arts industry (Tereyağoğlu et al. 2017).

We study loss aversion in the specific context of pricing. While this has been studied with regard to consumer behavior (see Özer and Zheng (2012)), managers' cognitive biases (loss aversion specifically) when setting prices for consumers remain understudied.

By using our unique dataset, we are able to calibrate our mode so it can capture how actual expert managers set prices in a real company. As mentioned in §3.1, our main contribution is to show that loss aversion is not unique to students who participate in laboratory experiments, or to unsophisticated individuals like taxi drivers, but that experts also suffer from this cognitive bias when making high-stakes decisions that have a large, direct impact on a firm's revenue.

3.3. Model

3.3.1 Model Formulation.

Consider a manager who is setting the price p_{nw} for product $n = 1, \dots, N$ on weeks $w = 1, \dots, W$ of clearance sales. The manager observes how much inventory of n there currently is, $\hat{I}_{n0}$, and estimates how much demand the product will have during each week in the horizon for every price under consideration, $d_{nw}(p_{nw})$. The actual demand during week w, D_{nw}, will deviate from the expected demand by an uncertain term ε_{nw}, so it will be $D_{nw} = d_{nw}(p_{nw}) + \varepsilon_{nw}$. Furthermore, the maximum number of units that can be sold corresponds to the inventory levels, as demand cannot be backlogged. Hence, the (uncertain) sales amount for product n during week w is $S_{nw} = \min\{D_{nw}, I_{n,w-1}\}$, where $I_{n,w}$ is the (uncertain) inventory level at the end of week w and, by construction, $I_{nw} = I_{n,w-1} - S_{nw}$. All inventory that remains unsold at the end of the season, I_{nW}, will be salvaged at price p_s (which takes the same value for all products within a group).

Now, if this manager was a pure revenue-maximizer, they would be searching for the price trajectory $\mathbf{p}_n = (p_{n1}, \dots, p_{nW})$ which maximized the expected revenue of this product for the whole horizon, so their utility function for product n and a given price path $\mathbf{p}$ would be

$$\mathcal{U}_n(\hat{I}_{n0}, \mathbf{p}_n) = \mathbb{E}_{\mathbf{S}}\left[\sum_{w=1}^{W} p_{nw} S_{nw} + p_s I_{nW}\right]. \tag{3.1}$$

However, the manager may be loss averse and perceive as a loss each unit that needs to be salvaged, since salvage values are significantly lower than retail prices. We model loss aversion by adding a loss aversion term to the utility function, $-\lambda\left(p_{n0} - p_s\right) I_{nW}$, where λ is the manager's loss aversion parameter[1].

[1]This form of loss aversion utility is consistent with Kőszegi and Rabin (2006) and Ericson and Fuster

Thus, the utility function for product n that the manager will maximize over $\mathbf{p}$ is

$$\mathcal{U}_n(\hat{I}_{n0}, \mathbf{p}_n) = \mathbb{E}_{\mathbf{S}} \left[\sum_{w=1}^{W} p_{nw} S_{nw} + p_s I_{nW} - \lambda \left(p_{n0} - p_s \right) I_{nW} \right]. \tag{3.2}$$

We choose, as the reference price for the loss aversion term, the product's current implemented price, p_{n0}. The choice of the most recently implemented price as a reference point is well established in the literature (Winer 1986, Popescu and Wu 2007, Bruno et al. 2012). The parameter λ will show how loss averse the manager is: the closer λ it is to 0, the closer to a revenue maximizer the manager is. If the manager is loss averse ($\lambda > 0$), we will observe larger deviations from revenue-maximizing prices when the number of units to salvage at the end of the season I_{nW} is large or when the current retail price p_{n0} is much higher than the salvage value p_s. Notice that this loss averse term of the utility function is not defined for gains, only for losses; this is because the salvage value is always strictly lower than the retail price, and so the manager can either feel neutral ($\lambda = 0$) or negatively ($\lambda > 0$) when they forecast how many units will remain unsold at the end of the season and evaluate their value compared to the reference point.

There are a number of company-specific rules that managers at Zara need to follow when setting prices during clearance sales. The first is that prices cannot increase: $p_{nv} \leq p_{nw}$ for every week $v > w$. The second is that the order of prices across products within a group is preserved over time: if $p_{nw} \leq p_{mw}$, then $p_{nv} \leq p_{mv}$ for all $v > w$. Once two product's prices coalesce, if the two products belong to the same group, their prices are coupled: if $p_{nw} = p_{mw}$, then $p_{nv} = p_{mv}$ for $v > w$. Finally, all prices need to be drawn from a discrete set that is pre-defined by the firm's marketing department: $p_{nw} \in \{p_1, \ldots, p_K\}$ for $n = 1, \ldots, N$ and $w = 1, \ldots, W$.

Note that the second and third constraints above force the manager to be thinking about the price of all the products within the same group when setting the price of each individual product. In other words, these two sets of constraints couple products together when the

(2014). In their more general framework, the loss averse utility function is $\mathcal{U} = u + \mu(u - r)$, where $\mu(x) = \eta x$ for $x > 0$ and $\mu(x) = \lambda \eta x$ for $x \leq 0$ ($\lambda > 1$), r is the reference point respect to which the decision maker evaluates the prospect u.

manager maximzes their utility. Hence, we need to write the manager's objective function as a sum of the individual product utilities,

$$\mathcal{U}(\mathbf{I}_0, \mathbf{p}|\lambda) = \sum_{n=1}^{N} \mathcal{U}_n(\hat{I}_{n0}, \mathbf{p}_n|\lambda). \tag{3.3}$$

The problem that the manager solves is:

$$\max_{\mathbf{p}} \quad \mathbb{E}_{\mathbf{S}} \left[\sum_{w=1}^{W} \sum_{n=1}^{N} p_{nw} S_{nw} + \sum_{n=1}^{N} p_s I_{nW} - \lambda \sum_{n=1}^{N} \left(p_{n0} - p_s\right) I_{nW}\right] \tag{3.4}$$

$$\text{s.t.} \quad p_{nv} \leq p_{nw} \;\; \forall n = 1, \ldots, N, \; v > w \tag{3.5}$$

$$p_{nv} \leq p_{mv} \;\; \forall n, m = 1, \ldots, N, \; v > w, \text{s.t. } \; p_{nw} \leq p_{mw} \tag{3.6}$$

$$p_{nv} = p_{mv} \;\; \forall n, m = 1, \ldots, N, \; v > w, \text{s.t. } \; p_{nw} = p_{mw} \tag{3.7}$$

$$S_{nw} = \min\{D_{nw}, I_{n,w-1}\} \;\; \forall n = 1, \ldots, N, \; w = 1, \ldots, W \tag{3.8}$$

$$I_{nw} = I_{n,w-1} - S_{nw} \;\; \forall n = 1, \ldots, N, \; w = 2, \ldots, W \tag{3.9}$$

$$I_{n1} = \hat{I}_{n0} - S_{n1} \;\; \forall n = 1, \ldots, N \tag{3.10}$$

$$p_{nw} \in \{p_1, \ldots, p_K\} \;\; \forall n = 1, \ldots, N \tag{3.11}$$

3.3.2 Certainty Equivalent Approximation.

The system described in Equations 3.4 to 3.11 is a dynamic program with a huge state space (all possible inventory levels for every product in the group) and uncertainty set (all possible demand levels for every product in the group). It cannot be solved exactly, so we use a certainty equivalent approximation to deal with it. We substitute the uncertain demand $D_{nw} = d_{nw}(p_{nw}) + \varepsilon_{nw}$ for its expected value $\mathbb{E}[D_{nw}] = d_{nw}(p_{nw})$, and so the expected sales are $s_{nw} = \mathbb{E}[S_{nw}] = \min\{d_{nw}(p_{nw}), i_{nw}\}$, where $i_{nw} = \mathbb{E}[I_{nw}] = \hat{I}_{n0} - \sum_{w=1}^{W} s_{nw}]$ is the expected inventory level for product n at the end of week w.

The program that the manager solves every week becomes, then,

$$\max_{\mathbf{p}} \quad \sum_{w=1}^{W}\sum_{n=1}^{N} p_{nw}s_{nw} + \sum_{n=1}^{N} p_s i_{nW} - \lambda \sum_{n=1}^{N} (p_{n0} - p_s)\, i_{nW} \tag{3.12}$$

$$\text{s.t.} \quad p_{nv} \le p_{nw} \;\; \forall n = 1, \ldots, N,\ v > w \tag{3.13}$$

$$p_{nv} \le p_{mv} \;\; \forall n, m = 1, \ldots, N,\ v > w, \text{s.t. } p_{nw} \le p_{mw} \tag{3.14}$$

$$p_{nv} = p_{mv} \;\; \forall n, m = 1, \ldots, N,\ v > w, \text{s.t. } p_{nw} = p_{mw} \tag{3.15}$$

$$s_{nw} = \min\{d_{nw}, i_{n,w-1}\} \;\; \forall n = 1, \ldots, N,\ w = 1, \ldots, W \tag{3.16}$$

$$i_{nw} = i_{n,w-1} - s_{nw} \;\; \forall n = 1, \ldots, N,\ w = 2, \ldots, W \tag{3.17}$$

$$i_{n1} = \hat{I}_{n0} - s_{n1} \;\; \forall n = 1, \ldots, N \tag{3.18}$$

$$p_{nw} \in \{p_1, \ldots, p_K\} \;\; \forall n = 1, \ldots, N \tag{3.19}$$

3.3.3 Discretization and Linearization.

The last constraint in the system, Equation 3.19, implies that the action set is discrete. Given that, with the certainty equivalent approximation, we are now solving a deterministic problem, we can reformulate it as a mixed integer program in the following way: we define the binary decision variables y_{nkw}, which take value 1 if product n has price p_k in week w, 0 otherwise; and the auxiliary variables $x_{nkw} = 1$ if product n has a price that is equal or lower than p_k in week w, 0 otherwise.

We now need to reformulate the constraints using this discretization. By construction, $x_{n(k-1)w} \le x_{nkw}$ and $y_{nkw} = x_{nkw} - x_{n(k-1)w}$ for all $n = 1, \ldots, N$, $k = 1, \ldots, K$, and $w = 1, \ldots, W$. Products can only be assigned one price, so $\sum_{k=0}^{K} y_{nkw} \le 1$. Constraint 3.5 above becomes $x_{nk(w-1)} \le x_{nkw}$. The ordering of products (Equation 3.14 above) is preserved with the new constraint $x_{nkw} \le x_{(n+1)kw}$. As mentioned before, one of the company's rules is that products whose prices have coalesced will remain coupled for the rest of the season; this rule (Equation 3.15) means that, if $p_{n0} = p_{n+1,0}$, then $\sum_{k=0}^{K}\sum_{w=1}^{W} x_{nkw} = \sum_{k=0}^{K}\sum_{w=1}^{W} x_{(n+1)kw}$. The expected sales for every product and week are $s_{nw} = \min\left\{\sum_{k=0}^{K} d_{nkw}y_{nkw}, i_{nw}\right\}$, and the inventory flow is formulated as $i_{nw} = i_{n,w-1} - s_{nw}$.

The manager's optimization problem becomes

$$\max_{\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{i}} \quad \sum_{w=1}^{W}\sum_{n=1}^{N}\sum_{k=1}^{K} p_k s_{nkw} + \sum_{n=1}^{N} p_s i_{nW} - \sum_{n=1}^{N} \lambda\left(p_{n0} - p_s\right) i_{nW} \tag{3.20}$$

$$\text{s.t.} \quad s_{nw} \leq d_{nkw} y_{nkw} \;\; \forall n = 1, \ldots, N,\, k = 1, \ldots, K,\, w = 1, \ldots, W \tag{3.21}$$

$$s_{nw} \leq i_{nw} \;\; \forall n = 1, \ldots, N,\, w = 1, \ldots, W \tag{3.22}$$

$$x_{n,k-1,w} \leq x_{nkw} \;\; \forall n = 1, \ldots, N,\, k = 2 \ldots, K,\, w = 1, \ldots, W \tag{3.23}$$

$$y_{nkw} = x_{nkw} - x_{n,k-1,w} \;\; \forall n = 1, \ldots, N,\, k = 2 \ldots, K,\, w = 1, \ldots, W \tag{3.24}$$

$$\sum_{k=0}^{K} y_{nkw} \leq 1 \;\; \forall n = 1, \ldots, N,\, w = 1, \ldots, W \tag{3.25}$$

$$x_{n,k,w-1} \leq x_{nkw} \;\; \forall n = 1, \ldots, N,\, w = 2, \ldots, W \tag{3.26}$$

$$x_{nkw} \leq x_{n+1,k,w} \;\; \forall n = 1, \ldots, N-1,\, w = 1, \ldots, W \tag{3.27}$$

$$\sum_{k=0}^{K}\sum_{w=1}^{W} x_{nkw} = \sum_{k=0}^{K}\sum_{w=1}^{W} x_{n+1,k,w} \;\; \forall n \text{ s.t. } p_{n0} = p_{n+1,0},\; \forall\, w \tag{3.28}$$

$$i_{nw} = i_{n,w-1} - s_{nw} \;\; \forall n = 1, \ldots, N,\, w = 2, \ldots, W \tag{3.29}$$

$$i_{n1} = \hat{I}_{n0} - s_{n1} \;\; \forall n = 1, \ldots, N \tag{3.30}$$

$$s_{nw}, i_{nw} \geq 0 \;\; \forall n = 1, \ldots, N,\, w = 1, \ldots, W \tag{3.31}$$

$$x_{n,k,w}, y_{nkw} \in \{0, 1\} \;\; \forall n = 1, \ldots, N,\, k = 1, \ldots, K,\, w = 1, \ldots, W \tag{3.32}$$

This problem is a mixed integer linear program (MILP). Note that this discretization-linearization procedure was first proposed in Caro and Gallien (2012).

3.3.4 Demand Estimation.

How does the manager forecast the expected demand of each product for the remaining weeks in the horizon? We assume that they use the following variation of a constant elasticity demand function to forecast the first week's demand:

$$d_{n1}(p_{n1}) = \alpha \left(\hat{d}_{n0}\right)^{\beta} e^{\gamma(p_{n1} - p_{n0})}. \tag{3.33}$$

In this model, the coefficient γ is customers' price elasticity, which we expect to be negative (demand increases when prices decrease, so when $p_{nw} < p_{n0}$). The factor $\left(\hat{d}_{n0}\right)^{\beta}$

captures a state dependence of the demand. In other words, when a product has been having a high demand in the previous weeks (because it is on trend, or of a higher quality), it is expected to continue having a high demand than a product with a lower past demand, even if the two products are marked down by the same amount. The coefficient α acts like a discount factor, capturing the fact that, regardless of the attractiveness of the product, and even in the absence of markdowns, prices "get old" (Caro and Gallien 2012). We do not account for the broken assortment effect, as our data is aggregated at the country level and not at the store level.

To estimate the parameters of this function we use linear regression. Given that we cannot observe the uncensored demand that products had, only their realized sales $\hat{s}_{nw}$, we use this as the dependent variable, and sales in the previous week as the first independent variable. By taking logarithms in both sides of Equation 3.33, the linear regression we estimate is

$$\log\left(\hat{s}_{nw}\right) = \log(\alpha) + \beta \log\left(\hat{s}_{n,w-1}\right) + \gamma\left(p_{nw} - p_{n,w-1}\right). \tag{3.34}$$

Once we have estimates for parameters α, β and γ, we use them to forecast the first week's demand for every possible price $p_k \in \{p_1, \ldots, p_K\}$ in the list of feasible prices, d_{nk1}. We then estimate the demand for the following weeks, for every feasible price, using a discount factor κ. The demand estimation procedure is, for the discretized problem, as follows:

$$d_{nk1} = \alpha\left(\hat{d}_{n0}\right)^{\beta} e^{\gamma(p_k - p_{n0})} \;\; \forall n = 1, \ldots, N,\; k = 1, \ldots, K \tag{3.35}$$

$$d_{nkw} = \kappa^{w} d_{nk1} \;\; \forall n = 1, \ldots, N,\; k = 1, \ldots, K,\; w = 2, \ldots, W. \tag{3.36}$$

3.3.5 Weekly Price Implementation.

Following the company's procedures, the manager solves a problem like the one described in Equations 3.20-3.32 for every product group at the beginning of each week of clearance sales and implements the chosen prices for that period, $(\hat{p}_{11}, \ldots, \hat{p}_{N1})$. A week after, after sales have been realized, the manager updates their demand forecast and solves the problem again for $w = 2, \ldots, W$ but implements only week 2's prices, $(\hat{p}_{12}, \ldots, \hat{p}_{N2})$. The process continues until the end of the clearance sales season. This process is done starting in the second week of clearance sales (which we denote by $w = 1$); for the initial markdowns ($w = 0$) of each campaign, the process is different, and done by a committee and not by the country manger alone. We denote by $\hat{\mathbf{p}}_{mg} = (\hat{p}_{11}, \ldots, \hat{p}_{N1}, \ldots, \hat{p}_{1W}, \ldots, \hat{p}_{NW})$ the set of prices that manager m implemented during the campaign for group g containing products $n = 1, \ldots, N$.

Similarly, we start by observing the realized demand for the first week of clearance sales, $(\hat{d}_{10}, \ldots, \hat{d}_{N0})$ for a group containing products $n = 1, \ldots, N$, and estimating the products' expected demands for the rest of the campaign as shown in Equations 3.35-3.36. We then solve the MILP in Equations 3.20 to 3.32 for week $w = 1, \ldots, W$ and store the first week's optimal prices, $(p^*_{11}, \ldots, p^*_{N1})$. We then use the actual realized demand (from data), $(\hat{d}_{11}, \ldots, \hat{d}_{N1})$ to re-estimate the expected demands for weeks $w = 2, \ldots, W$, solve the MILP again for this shorter set of weeks, and store $(p^*_{12}, \ldots, p^*_{N2})$. We repeat this method to obtain the vector of optimal prices $\mathbf{p}^*_{mg} = (p^*_{11}, \ldots, p^*_{N1}, \ldots, p^*_{1W}, \ldots, p^*_{NW})$, which contains the prices of group g's products that our model predicts to be implemented every week by manager m.

3.3.6 Loss Aversion Parameter Estimation.

Note that the MILP in Equations 3.20 to 3.32 depends on the manager's loss aversion parameter, λ, so the vector of optimal prices is actually of the form $\mathbf{p}^*_{mg}(\lambda)$. We now need to estimate the value of this parameter. To allow for variation in degree of loss aversion not only between managers, but also between product types (for instance, a manager might be more loss averse when setting prices for fashion products that are not as easy to sell as basic products), we estimate a value of λ for every manager and group, λ_{mg}.

We do so by line search over λ to minimize the mean absolute percentage error[2],

$$\lambda^*_{mg} = \arg\min \sum_{w=1}^{W} \sum_{n=1}^{N_g} 100 \cdot \left| \frac{p^*_{nw}(\lambda) - \hat{p}_{nw}}{\hat{p}_{nw}} \right|. \tag{3.37}$$

[2]When we minimized the sum of squared errors, $\lambda^*_{mg} = \arg\min \sum_{w=1}^{W} \sum_{n=1}^{N_g} \left[p^*_{nw}(\lambda) - \hat{p}_{nw}\right]^2$, we obtained very similar values of λ^*_{mg}.

3.4. Data

To be able to fit our model to best capture managers' behavior, we used a dataset which was collected by Zara before the implementation of the decision support system that we analyzed in the previous chapter. This way, we can be sure that the observed prices that managers implemented are a result of their own internal decision making process, and not subject to any effect from the DSS such as anchoring. Our pre-DSS dataset contains two clearance sales campaigns (summer of 2007 and summer of 2008), in two countries, Belgium and France. For the remainder of this paper, we will refer to those countries' managers as Manager B and Manager F, respectively.

Our dataset contains observations for 9 different product groups (all of them are women's apparel). In Zara's inner jargon, a group is a product type which is targeted towards a specific consumer group. For instance, there are three groups corresponding to blazers for women: "blazer - basic", "blazer - woman", and "blazer - TRF". "Basic" products are those which are made usually in solid, neutral colors, and are less risky and less fashionable. In the opposite extreme, "TRF" is Zara's apparel line for younger women. Products in this collection are meant to be trendy, usually have unique designs, colors or prints, and are more risky for the company. Between basic and TRF products, "woman" is the product line targeted to adult women. This collection contains products with non-basic colors, prints or designs, but not too risky to be worn in a work setting. For example, in last winter's assortment we found that basic blazers were black or dark gray, women's blazers had a discreet print, and the only TRF blazer was all glittery and sparkly.

Table 3.1 shows how many observations there are (and what percent of the data that amount accounts for) for every country manager, every campaign, and every product group.

Recall that one of Zara's pricing rules is that all products (SKUs) in the same group which have the same price shall be priced together for the remainder of the season. A set of SKUs with the same price during the regular season is called a *cluster*. A week-cluster pair is the smallest unit of observation in our data, as country managers make pricing decisions

	Num. Observations	% Observations
Country manager		
Manager B (Belgium)	1,068	47.59
Manager F (France)	1,176	52.41
Clearance sales campaign		
Summer 2007	1,024	45.63
Summer 2008	1,220	54.37
Product group		
Knitwear	272	12.12
T-shirt	438	19.52
Skirt - TRF	199	8.87
Pants - TRF	324	14.44
Pants - woman	226	10.07
Pants - basic	258	11.50
Blazer - TRF	132	5.88
Blazer - woman	196	8.73
Blazer - basic	199	8.87
Total	2244	100

Table 3.1: Summary statistics of the categorical variables in our data.

at the cluster level (and not each SKU separately)[3].

An observation in our dataset is shown in Table 3.2. It corresponds to a pricing decision made by Manager B (the one setting prices for Belgium) in the Summer 2008 campaign. This is the 4990 cluster of TRF blazers, i.e, blazers in the TRF collection which were priced at 49.90€ during the regular season. This is the pricing decision made at the beginning of week 2 of clearance sales; during week 1, this cluster was priced at 29.95€. In week 2, the manager decides to price it at 19.95€. At the beginning of week 2 there are 724 units of inventory of this cluster in Belgium, of which 177 are sold during the week. Any units that remain unsold by the end of clearance sales will be salvaged at 4.74€.

Table 3.3 contains summary statistics of the numerical variables in our dataset.

[3]In §3.3, index $n = 1, \ldots, N$ actually refers to clusters.

Country	Year	Season	Group	Cluster	Week
Belgium	2008	Summer	Blazer - TRF	4990	2
Initial inventory	**Quantity sold**	**Regular price**	**Previous price**	**Implemented price**	**Salvage value**
724	177	49.90€	29.95€	19.95€	4.74€

Table 3.2: One observation in our dataset.

	Mean	St. Dev.	Median	Min.	Max.
Inventory per cluster at the beginning of the clearance sales campaign	10,986.59	21,380.55	3,057.50	0	149,487
Weekly sales per cluster	1,058.10	2,707.72	184.50	0	39,686
Inventory per cluster at the end of the clearance sales campaign	1,758.99	3,078.85	580.50	0	17,065
Number of weeks that a cluster is on sale	8.976	1.232	9	7	10
Number of cluster per product group	6.667	2.029	6	2	10
% markdown chosen by the country manager (unconditional)	12.90	17.02	0	0	86.79
% markdown chosen by the country manager (given that he/she decided to markdown)	31.12	11.43	30.15	12.03	86.79

Table 3.3: Summary statistics of the numerical variables in our data. Note that 41.49% of the time products were marked down, while the remaining 58.51% of the time country managers decided to keep the previous week's price.

3.5. Results

3.5.1 Demand Forecast Function Parameters.

We first estimate the demand forecast function parameters using the linear regression described in Equation 3.34.

The coefficient estimates of this regression are shown in Table 3.4. From these coefficient estimates, we compute the parameters of the demand function as $\alpha = e^{-0.099} = 0.906$, $\beta = 0.939$, and $\gamma = -0.049$. For the demand in weeks $w > 1$, we use the discount factor $\kappa = 0.84$ as suggested in Caro and Gallien (2012). The demand forecast model is, then,

$$d_{nk1} = 0.906 \cdot \left(\hat{d}_{n0}\right)^{0.939} \cdot e^{-0.049(p_k - p_{n0})} \;\; \forall n = 1, \ldots, N,\; k = 1, \ldots, K \tag{3.38}$$

$$d_{nkw} = 0.84^{w} \cdot d_{nk1} \;\; \forall n = 1, \ldots, N,\; k = 1, \ldots, K,\; w = 2, \ldots, W. \tag{3.39}$$

As mentioned in §3.4, we used the observations from 2007 to calibrate the parameters of the demand forecast function, and the observations from 2008 to estimate managers' loss

	$\log(\text{Sales}_{nw})$
$\log(\text{Sales}_{n,w-1})$	0.939^{***}
	(0.012)
$\text{Price}_{nw} - \text{Price}_{n,w-1}$	-0.049^{***}
	(0.007)
Intercept	-0.099
	(0.073)
Observations	840
R^2	0.877
Adjusted R^2	0.877
Residual Std. Error	0.737 (df = 837)
F Statistic	$2{,}984.965^{***}$ (df = 2; 837)
Standard errors in parentheses	*p<0.1; **p<0.05; ***p<0.01

Table 3.4: Coefficient estimates of the linear regression (Equation 3.34) of the demand forecast model.

aversion parameters. We discard the first week of 2007 to be able to use that week's observed sales as the lagged observed sales independent variable in the regression (we cannot observe the pre-clearance sales information in our data). Our final data set with which we estimate the coefficients of the demand function contains 840 observations. Attempts at estimating separate demand models for each one of the managers, or product groups, etc. led to spurious results, such as non-significant coefficients or positive price elasticities, most likely due to the small sample size.

3.5.2 Loss Aversion Parameter Estimation.

We used the data from 2008 to estimate the value of λ_{mg}, the loss aversion parameter of each manager for every product group. As a preliminary step, we computed the mean absolute percentage error,

$$MAPE_{mg}(\lambda) = \sum_{w=1}^{W} \sum_{n=1}^{N_g} 100 \cdot \left| \frac{p^*_{nw}(\lambda) - \hat{p}_{nw}}{\hat{p}_{nw}} \right|, \tag{3.40}$$

for different values of λ (by varying this parameter in small increments). We then plotted

	Knitwear	T-shirt	Skirt TRF	Pants TRF	Pants woman	Pants basic	Blazer TRF	Blazer woman	Blazer basic
Manager B	0.613	1.908	0.896	0.083	0.370	0.056	0.138	0.768	0.005
Manager F	1.760	2.305	0.813	1.376	1.201	0.162	1.829	0	2.292

Table 3.5: Estimated value of the loss aversion parameter λ_{mg} for every manager and product group.

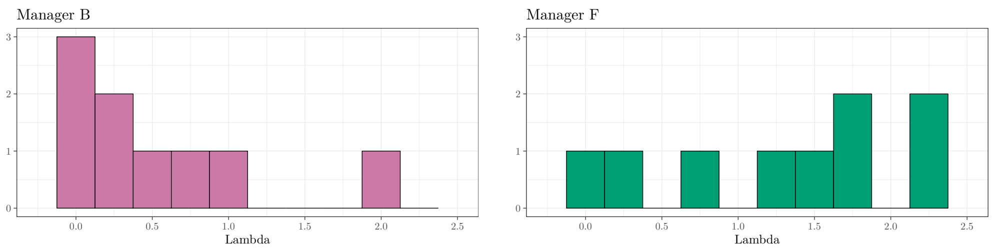


Figure 3.2: Histogram of the loss aversion parameter λ_{mg} estimates, for Manager B (left) and Manager F (right), and all product groups.

$MAPE_{mg}(\lambda)$ to visually inspect if it had only one minimum, and we identified an interval in which this minimum was contained (for each manager and group). We used this interval as the initial bounds for the line search of the minimum of $MAPE_{mg}$ on λ.

Table 3.5 contains our λ estimates. These values are also shown in Figure 3.2. The estimated values of this parameter vary widely (from 0 to 2.305) between country managers and between product groups. However, from the histograms in Figure 3.2 we can see that Manager F seems to be more loss averse than Manager B. In particular, Manager B's estimated λ's take values below 1 in all groups except two, whereas Manager F's estimates are above 1 in six of the nine product groups.

One might suspect that the variation in λ within each country manager is driven by differences between product groups. However, as we can see in Figure 3.3, there does not appear to be any clear pattern. Both managers seem to be very loss averse when they are setting prices for t-shirts (with λ estimates equal to 1.908 for Manager B and 2.305 for Manager F), moderately loss averse when they set prices for TRF skirts (with parameter

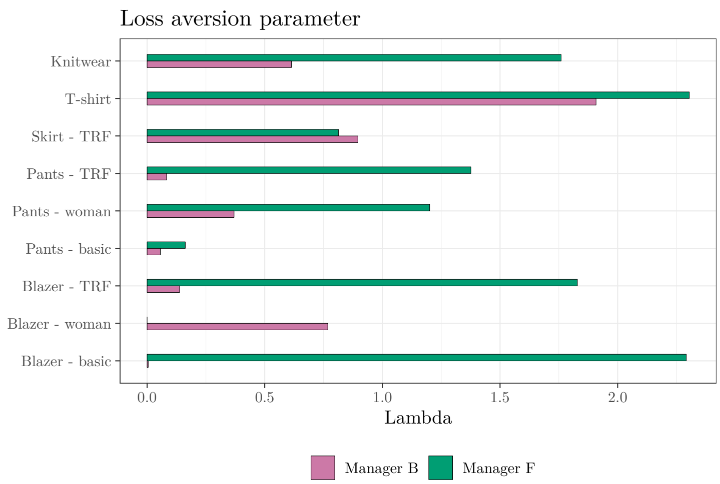


Figure 3.3: Estimates of the loss aversion parameter λ_{mg}, for each manager and product group. The upper bar (green) of every group corresponds to Manager F, and the lower bar (pink) corresponds to manager B.

estimates 0.896 and 0.813, respectively), and have very small λ estimates for basic pants (0.056 and 0.162). These are the only three product groups where both managers seem to have a similar behavior. An extreme example of this are the three types of blazers: for TRF blazers, Manager F is very loss averse ($\lambda = 1.829$) and Manager B is very little loss averse ($\lambda = 0.138$); for basic blazers, Manager F is even more loss averse ($\lambda = 2.292$) and Manager B is almost not loss averse at all ($\lambda = 0.005$); however, we find the opposite order for women's blazers, where Manager B's estimated λ is 0.768, and Manager F's is 0.

Therefore, from this visual inspection, we cannot draw any clear conclusion on what product group's characteristics are driving the differences in loss aversion parameter estimates within each manager. In particular, it is unclear if the trendiness or basicness of a product group has any relationship with managers' degree of loss aversion respect to that group, as the loss aversion parameter estimates for these two broad types of products do not seem to be distinctively clustered.

A possible explanation of the variation in loss aversion between products may be the difference in salvage value across groups. However, many different (but related) product

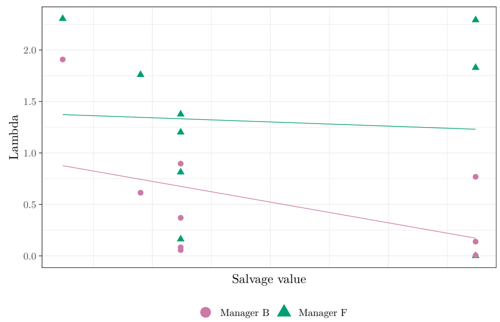


Figure 3.4: Relationship between the loss aversion parameter λ_{mg} and the salvage value s_{mg}.

groups share salvage value: for instance, all blazer groups in our data have the same salvage value, and so do all bottoms (skirts and pants). But, as we have seen, both managers have different λ estimates for each type of blazer, skirt, etc. We can confirm that the estimates of the loss aversion parameter have very little correlation with the salvage value of different product groups by looking at Figure 3.4.

3.5.3 Model Comparison.

Our goal is to model manager's pricing decisions in a way that explains their actual behavior better than just assuming that they are perfect revenue maximizers. To test how successful we were, we repeat the price optimization in Equations 3.20-3.32 for $\lambda = 0$. In other words, we replace the objective function in Equation 3.20 for the following discretization of the expected revenue:

$$\sum_{w=1}^{W} \sum_{n=1}^{N} \sum_{k=1}^{K} p_k s_{nkw} + \sum_{n=1}^{N} p_s i_{nW}. \tag{3.41}$$

Like in the original problem, we rerun the MILP with this objective function and Constraints 3.21-3.32 for every week, updating the demand forecast using the previous week's realized sales, and store the first week's prices (the ones that the manager would implement). We obtain a vector of revenue-maximizing prices $\mathbf{p}^R_{mg} = (p^R_{11}, \ldots, p^R_{N1}, \ldots, p^R_{1W}, \ldots, p^R_{NW})$ for every manager m and froup g. We then compute the mean absolute percentage error for this price path as

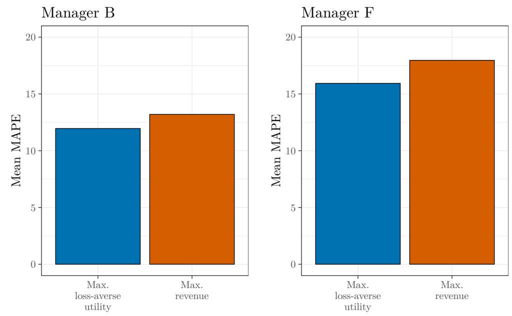


Figure 3.5: Comparison of the average over groups of $MAPE_{mg}$, for Manager B (left) and Manager F (right). In each plot, the fist bar (blue) corresponds to the error in our loss-averse utility maximization model, $MAPE^*_{mg}$, while the second bar (orange) corresponds to the error in the revenue maximization model, $MAPE^R_{mg}$.

$$MAPE^R_{mg} = \sum_{w=1}^{W} \sum_{n=1}^{N_g} 100 \cdot \left| \frac{p^R_{nw} - \hat{p}_{nw}}{\hat{p}_{nw}} \right|, \tag{3.42}$$

where $\hat{p}_{nw}$ is the price that manager m set for product n in week w, as observed in our data.

Figure 3.5 shows a comparison between the average over all groups of $MAPE^R_{mg}$ and the average over groups of $MAPE^*_{mg}$, our loss-averse utility maximization model's error (Equation 3.37). As we can see, our model's error is lower than that of revenue maximization: for Manager B, the average $MAPE^R$ is 13.20, while the average $MAPE^*$ is 11.96; for Manager F, the average $MAPE^R$ is 17.96, while the average $MAPE^*$ is 15.93. If we take the average over all the groups and country managers of the ratio $MAPE^*_{ig}/MAPE^R_{ig}$, we find that our model explains managers' behavior 11.04% better, on average, than assuming that managers are revenue maximizers.

The exact values of the mean absolute percentage error, for each group and manager, are shown in Table 3.6. Figure 3.6 shows a histogram of our model's percentage improvement respect to revenue maximization. In six manager-group pairs, our model improves revenue maximization by an amount that is lower than 5%; in four cases, the improvement is between

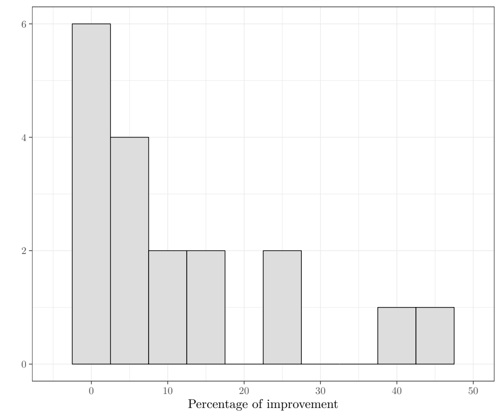


Figure 3.6: Histogram of the percentage improvement (from modeling managers' behavior as a revenue maximization to doing so as a loss-averse utility maximization), for all country managers and product groups.

5% and 10%; in the remaining eight cases, our model explains manager's behavior more than 10% better than revenue maximization. In two of these eight cases, the improvement is higher than 40%.

What makes our model better at explaining managers' price decision making than revenue maximization in some cases, and not so different in others? A possible explanation is that, for groups in which the loss aversion coefficient estimate λ is higher, the improvement is also higher, as the objective function is further from revenue (compared to when λ is small). Indeed, in Figure 3.7 we can see that there is a positive correlation between the loss aversion parmeter estimate and the percentage improvement that our model offers respect to revenue maximization.

Our hypotheses that managers were loss averse, and not revenue maximizers, arose from observing that they were setting overly aggressive markdowns, as if they were trying to minimize the amount of unsold inventory at the end of the season. What if their actual objective was as simple as minimizing inventory, and not maximizing their loss-averse utility?

We repeat the price optimization procedure given by the MILP in Equations 3.20-3.32

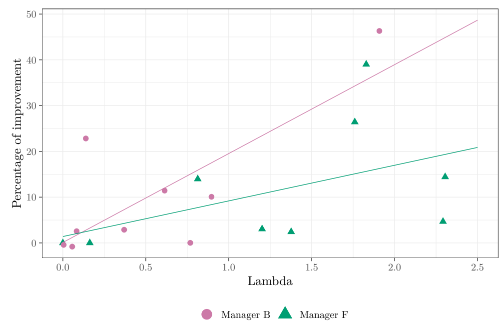


Figure 3.7: Relationship between the percentage improvement from using a loss-averse utility maximization model (respect to a revenue maximization model) and the loss aversion parameter λ_{mg}, for all country managers and product groups.

but we replace the problem's objective, Equation 3.20, for

$$\min_{\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{i}} \quad \textstyle\sum_{n=1}^{N} i_{nW}. \tag{3.43}$$

Like before, we store the first week's optimal prices for every weekly run of the algorithm, and obtain the vector of inventory-minimizing prices $\mathbf{p}^I_{mg} = (p^I_{11}, \ldots, p^I_{N1}, \ldots, p^I_{1W}, \ldots, p^I_{NW})$ for manager m and group g. The mean absolute percentage error between this vector of prices and the actually implemented prices is $MAPE^I_{mg}$.

As we can see in Figure 3.8, this model explains managers' behavior visibly worse than both loss-averse utility maximization and revenue maximization, with average across groups equal to 54.2 (Manager B) and 53.1 (Manger F). Table 3.6 shows the exact $MAPE^I_{mg}$ for every manager and group.

3.6. Conclusion and Next Steps

As we had seen in the previous chapter, country managers at Zara choose prices during clearance sales in a way that deviates from the revenue-maximizing price path, and more than 60% of the time they do so by setting overly aggressive markdowns, i.e., prices that are below the optimal ones. As we saw, this behavior is consistent with loss aversion: their

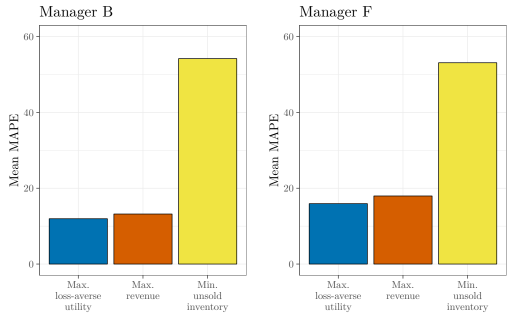


Figure 3.8: Comparison of the mean across groups of $MAPE_{mg}$, for Manager B (left) and Manager F (right). In each plot, the fist bar (blue) corresponds to our loss-averse utility maximization model, the second bar (orange) corresponds to a revenue maximization model, and the third bar (purple) corresponds to an inventory minimization model.

		Knitwear	T-shirt	Skirt TRF	Pants TRF	Pants woman	Pants basic	Blazer TRF	Blazer woman	Blazer basic
Manager B	Utility Max.	15.58	5.50	12.79	8.89	16.81	12.43	8.14	17.69	9.78
	Revenue Max.	17.59	10.25	14.23	9.13	17.31	12.33	10.54	17.69	9.74
	Inventory Min.	69.98	74.85	65.06	62.54	48.13	66.65	50.36	26.04	24.10
Manager F	Utility Max.	13.93	19.93	16.60	11.30	16.09	13.32	8.73	26.61	16.88
	Revenue Max.	18.93	23.29	19.29	11.58	16.59	13.32	14.31	26.61	17.71
	Inventory Min.	57.53	58.30	61.58	51.68	41.15	62.79	57.62	35.67	51.64

Table 3.6: Comparison of the mean absolute percentage error $MAPE_{mg}$ of the utility maximization model we propose in §3.3 and the revenue maximization and inventory minimization models suggested above, for every manager and product group.

probability of deviating from the revenue-maximizing prices is higher when the products' salvage value is low, which suggests that they are concerned not only about the amount of unsold inventory at the end of the season, but more specifically about the value of this unsold inventory.

Given that salvage values are significantly lower than retail prices or even clearance sales prices, managers may perceive having to salvage every unit of inventory as a loss, compared to how much more revenue that item would have made if it had been sold during clearance sales. Given that the decision support system that was suggesting revenue-maximizing prices to mangers already took into account what the salvage value of each product was, deviating from such prices in relation to the salvage value suggests that managers were overestimating the importance of the loss from salvaging products, i.e., they were loss averse.

To test whether their deviations from revenue-maximizing prices were driven by loss aversion, we build a structural model of managers' pricing decision making. In our model, country managers at Zara who set prices during clearance sales optimize their utility function, which includes a loss aversion term. The model's constraints reflect the company's actual pricing rules. We use a certainty equivalent approximation to deal with the very high dimensionality of the problem, and a discretization based on the discrete list of prices that managers are allowed to use. We compute the loss aversion parameter λ_{mg} that leads to the price path which is the closest to the prices that the managers actually chose.

We estimate the model's parameters using a small dataset collected by the company before the decision support system was implemented, so that we can observe managers' preferred prices without the anchoring effect that the DSS may be introducing. Our dataset contains nine product groups, sold during two separate clearance sales campaigns (spring summer of 2007 and 2008), in two countries, Belgium and France. Because of the small size of this dataset, all the results presented in the paper are extremely preliminary and need to be interpreted with caution.

Our initial results suggest that the two country managers (to whom we refer as Manager B and Manager F, respectively) were, indeed, loss averse, to a degree that varies widely by product group. Manager B's loss aversion parameters are generally lower than Manager F's, which suggests that they were less loss averse than their French counterpart. Our model's mean absolute percentage error, averaged over all groups and managers, is approximately 11% lower than that of a revenue maximization model, and several times lower than if we assume managers were inventory minimizers. To summarize, our model is promising in that it can explain managers' pricing behavior more accurately than a revenue maximization model can.

As mentioned above, the dataset we used to calibrate our model is very small. Our most immediate next step would be to use a larger dataset and observe whether our preliminary results still hold. There are two possible options for extending the size of our sample: the first one is to use additional pre-DSS data from a few other countries (which were the control groups when the DSS's pilot test was performed). This dataset still contains prices that managers set without the influence of the DSS's recommendation. However, this is not a large dataset either. Moreover, it would require a cleaning, organizing and consolidating effort, as the different variables are contained in several different spreadsheets with little information to link them together.

The second option is to use the same dataset we used in the previous chapter, and assume that, when managers deviated, the prices they chose reflect their inner price choices (with a small anchoring effect from the DSS), but when they adhered we cannot observe their true price preferences. In other words, we could use this dataset by ignoring the observations in which they adhered to the DSS's recommendations, and use only the instances in which they deviated. This approach would allow us to build counterfactuals for what prices they would have set if they had not adhered to the DSS's recommendations, and estimate the gain in revenue for Zara over all the countries (not only the two participating in the pilot test) from having implemented the DSS.

Our model greatly relies on the ability to forecast demand, so the demand estimation method needs to be improved. The parameters of the current function need to be estimated using a larger amount of data, which would possibly allow us to use different price elasticities for different product groups or for different countries. In addition, other functional forms of the demand forecast should be tested. In the most immediate future, we will estimate the current function's parameters with the current dataset using nonlinear least squares instead of linear regression and ordinary least squares on the demand's logarithm.

Similarly, we will use nonlinear least squares to be able to not only estimate the values of λ but also to obtain standard errors of our estimates. However, the small sample sizes (we are estimating one value of λ for each manager and group) might still lead to large standard errors even when a parameter is significant, so this should be done after expanding our dataset as explained above.

The choice of the last implemented price as a reference price for the loss aversion part of the utility function is rooted in the current literature (Winer 1986, Popescu and Wu 2007, Bruno et al. 2012). However, there is some evidence that consumers use a more complex reference point for prices, namely a combination of the most recent and the lowest price (Nasiry and Popescu 2011). In the future, we need to explore alternative reference prices like a combination of the most recent and the highest price (the best one from the manager's perspective). Other structural estimation experts have suggested that we could estimate the reference prices as model parameters.

In previous discussions of this project, the utility function contained a loss aversion term for every week in the horizon. In other words, we assumed managers do not only perceive as a loss having to salvage inventory, but also having to implement markdowns. However, we found that the current version of the utility function did a better job at explaining managers' behavior. In addition, the computation times for the old, more complex mode were significantly longer than for the one we have presented here.

Ultimately, the goal of structural estimation is to allow us to build counterfactuals in non-experimental settings. In our case, as the final next step, we aim to estimate the revenue that could have been made if managers were not biased or, more specifically, not loss averse. An initial attempt at doing so yields the following result: the two managers in our dataset were generating a revenue that was 1.8% lower than what they would have done had they been perfect revenue maximizers. Nevertheless, we think that, before trusting this estimate, we need to ensure that the demand is forecasted more accurately, the loss aversion parameter is estimated more carefully, and that we use data from more countries (i.e., more managers) to get a more representative picture of the situation we are analyzing.

APPENDIX A

Robustness Checks for Chapter 1

A.1. Robustness Checks for Section 1.5

A.1.1 Collinearity.

In Table A.1 we can see the correlation matrix for the order-level variables of our model and, in Table A.2, the variation inflation factors (VIF) for the regression without category/time/buyer dummies. The largest VIF corresponds to the order size control and is 4.31, which is much lower than the usual rule-of-thumb threshold of 10 (Wooldridge 2015). In the case of the full-dummies model, as expected, the VIFs (not reported here) are much higher for the dummy variables, as there is a high correlation between the different outcomes of the same categorical variable, but also between buyers and categories (for instance, some buyers only ordered shirts, or pants, etc.).

	Correlation matrix			
	(1)	(2)	(3)	(4)
(1) RelPriceFC	1			
(2) PropThisCategory	0.0102^{+}	1		
(3) BuyerIsMajorBrand	0.0543^{***}	0.110^{***}	1	
(4) LogOrderSize	-0.0718^{***}	-0.0104^{+}	0.0167^{**}	1

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Correlation matrix			
	(1)	(2)	(3)	(4)
(1) RelPriceFC	1			
(2) PropThisCategory	0.0207^{**}	1		
(3) BuyerIsMajorBrand	0.0835^{***}	0.0593^{***}	1	
(4) LogOrderSize	-0.0813^{***}	-0.0928^{***}	0.0165^{*}	1

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.1: Correlation matrices of the model's variables. All sample (left), type B factories (right).

VIF		
All sample		
Variable	VIF	1/VIF
LogOrderSize	4.31	0.231908
PropThisCategory	3.05	0.327687
$\text{USC}_{i,j-1}$	2.07	0.483542
BuyerIsMajorBrand	1.04	0.961219
RelPriceFC	1.01	0.986433
Mean VIF	2.29	

VIF		
Subsample		
Variable	VIF	1/VIF
LogOrderSize	4.31	0.232132
PropThisCategory	3.50	0.285977
$\text{USC}_{i,j-1}$	1.59	0.629142
BuyerIsMajorBrand	1.04	0.957988
RelPriceFC	1.01	0.992310
Mean VIF	2.29	

Table A.2: Variation inflation factors of the Arellano-Bond regression for all the data set (left) and the type B factories' subsample (right).

A.1.2 Autocorrelation.

The Arellano-Bond estimator assumes that the differenced errors are correlated at most with order 1, i.e., $\rho(\Delta\varepsilon_{it}, \Delta\varepsilon_{i,t-1})$ will not be zero, as both differences contain $\varepsilon_{i,t-1}$, but it is necessary that $\rho(\Delta\varepsilon_{it}, \Delta\varepsilon_{is}) = 0$, for $t - s > 1$. We run the Arellano-Bond autocorrelation test with order up to 5 for both the all-data and the type B factories' subsample estimates. The z-statistics and its corresponding p-values are shown in Table A.3. We reject H_0 (no autocorrelation) at order 1, but fail to reject it at higher orders. Therefore, the no-autocorrelation assumption of the model is satisfied.

H_0: **no autocorrelation**		
All sample		
Order	z	p
1	-3.0464	0.0023
2	1.4372	0.1507
3	-1.1044	0.2694
4	0.95331	0.3404
5	-1.3021	0.1929

H_0: **no autocorrelation**		
Subsample		
Order	z	p
1	-2.9779	0.0029
2	1.4067	0.1595
3	-1.1162	0.2643
4	0.95473	0.3397
5	-1.2944	0.1955

Table A.3: Arellano-Bond test for serially correlated errors, for the model using all the data set (left) and the subsample of type B factories (right).

A.1.3 Joint Significance of the Order-level Variables.

A distinctive aspect of our study is that we have order-level information. Though some of the order-level variables are not always statistically significant on their own, when considered jointly they become highly significant, which is testament to the order-by-order discretion followed by some factories.

We test the joint significance of the order-level variables by considering nested subsamples of the data based on the average level of unauthorized subcontracting $\overline{USC}_i$. First, we consider the full dataset; then, type B factories, i.e., $0 < \overline{USC}_i < 1$; after that, factories that subcontracted more than 10% of the orders but less than 90%; finally, factories that subcontracted more than 20% but less than 80% of their orders. We report the χ^2 statistics and the corresponding p-values in Table A.4. We can see that, for each subset, the order variables gain in significance, suggesting that factories that have a less defined subcontracting behavior (less polarized towards 0% or 100%) are more sensitive to order-specific characteristics when making a subcontracting decision.

Joint significance tests H_0: order-level variables are jointly insignificant				
Factories	All	Type B	$10\% < \overline{USC}_i < 90\%$	$20\% < \overline{USC}_i < 80\%$
		Model with order-level variables only		
χ^2	124.09	600.06	617.82	614.21
p	0.0000	0.0000	0.0000	0.0000
N	32028	18646	11813	10879
$\text{USC}_{i,j-1}$	0.297***	0.299***	0.312***	0.305***
		Model with order-level and dummy variables		
χ^2	22.46	30.53	58.62	78.85
p	0.0001	0.0000	0.0000	0.0000
N	32028	18646	11813	10879
$\text{USC}_{i,j-1}$	0.295***	0.298***	0.310***	0.305***

Table A.4: Joint significance of the order-level variables for the Arellano-Bond estimates. The top panel is for the model with order-level variables only (*RelPriceFC*, *PropThisCategory*, *BuyerIsMajorBrand*) and the bottom panel is for the model with *RelPriceFC*, *PropThisCategory*, and dummies.

A.1.4 Alternative Model Specifications.

We test our hypotheses using two different regression types. First, the Hsiao estimator, which is constructed in the same way as the Arellano-Bond, but is estimating using 2SLS instead of GMM. The coefficients in this estimation and their significance are very similar to those of the Arellano-Bond regression (Table A.5). To account for the fact that the dependent variable is binary, we also use a binary outcome model: the random effects dynamic probit (Wooldridge 2010). The coefficient and APE estimates, shown in Table A.6, are fairly robust to those of the Arellano-Bond estimation. Unfortunately, this model does not allow for an easy IV approach, so the coefficient estimates are likely affected by some degree of endogeneity. It does not allow for the inclusion of category, buyer and time dummy variables.

	Unauthorized subcontracting			
	Hsiao			
	(1)	(2)	(3)	(4)
$USC_{i,j-1}$	0.271***	0.270***	0.276***	0.272***
	(0.0652)	(0.0649)	(0.0630)	(0.0619)
RelPriceFC	-0.0729*	-0.0732*	-0.0732*	-0.0719*
	(0.0361)	(0.0360)	(0.0353)	(0.0352)
PropThisCategory	0.0145	-0.00294	-0.000219	-0.00135
	(0.0329)	(0.0321)	(0.0322)	(0.0345)
BuyerIsMajorBrand	-0.0497*	-0.0550*	-0.0601**	
	(0.0204)	(0.0217)	(0.0233)	
LogOrderSize	-0.00215*	-0.00211*	-0.00245**	-0.00231**
	(0.000843)	(0.000833)	(0.000906)	(0.000860)
Category dummies	No	Yes	Yes	Yes
Month dummies	No	No	Yes	Yes
Buyer dummies	No	No	No	Yes
N	32028	32028	32028	32028

Robust standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Unauthorized subcontracting			
	Hsiao			
	(1)	(2)	(3)	(4)
$USC_{i,j-1}$	0.271***	0.268***	0.279***	0.275***
	(0.0653)	(0.0649)	(0.0624)	(0.0613)
RelPriceFC	-0.110*	-0.114*	-0.112*	-0.106*
	(0.0496)	(0.0491)	(0.0481)	(0.0482)
PropThisCategory	0.0296	-0.0104	-0.0149	-0.0183
	(0.0696)	(0.0702)	(0.0695)	(0.0701)
BuyerIsMajorBrand	-0.0847*	-0.105+	-0.121*	
	(0.0431)	(0.0551)	(0.0609)	
LogOrderSize	-0.00337*	-0.00337*	-0.00344*	-0.00337*
	(0.00140)	(0.00139)	(0.00147)	(0.00138)
Category dummies	No	Yes	Yes	Yes
Month dummies	No	No	Yes	Yes
Buyer dummies	No	No	No	Yes
N	18646	18646	18646	18646

Robust standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Hsiao regression using all the data set (left) and subsampling for type B factories (right).

	Unauthorized subcontracting			
	Dynamic random effects probit			
	All sample		Subsample	
	Coefficients	APEs	Coefficients	APEs
	(1)	(2)	(3)	(4)
$USC_{i,j-1}$	1.780***	0.152***	1.778***	0.403***
	(0.0281)	(0.0116)	(0.0281)	(0.0182)
RelPriceFC	-0.220***	-0.0189***	-0.222***	-0.0504***
	(0.0513)	(0.00460)	(0.0514)	(0.0118)
PropThisCategory	0.249***	0.0213***	0.253***	0.0575***
	(0.0628)	(0.00559)	(0.0629)	(0.0145)
BuyerIsMajorBrand	-0.235*	-0.0201*	-0.234*	-0.0532*
	(0.105)	(0.00907)	(0.105)	(0.0239)
LogOrderSize	-0.00379	-0.000324	-0.00405	-0.000920
	(0.00896)	(0.000767)	(0.00898)	(0.00204)
USC_{i1}	3.445***	0.295***	0.654***	0.148***
	(0.296)	(0.0123)	(0.168)	(0.0357)
Mean RelPriceFC	-1.353	-0.116	0.616	0.140
	(1.171)	(0.100)	(1.207)	(0.274)
Mean PropThisCategory	-0.862+	-0.0738+	-0.528	-0.120+
	(0.449)	(0.0391)	(0.322)	(0.0727)
Mean BuyerIsMajorBrand	-3.631	-0.311	-0.176	-0.0400
	(3.794)	(0.321)	(5.013)	(1.136)
Mean LogOrderSize	0.0871	0.00746	0.0284	0.00644
	(0.132)	(0.0113)	(0.124)	(0.0281)
Constant	-3.253***		-1.081	
	(0.910)		(0.860)	
$\ln(\sigma_u^2)$	0.780***		-0.917***	
	(0.177)		(0.190)	
N	32028	32028	18646	18646

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

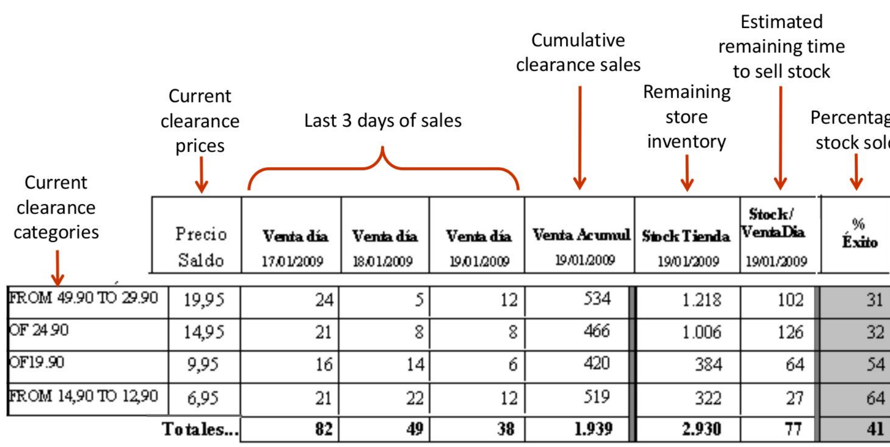
Table A.6: Dynamic probit regression using all the data set (columns 1 and 2) and the subsample of type B factories (columns 3 and 4). Odd columns contain coefficients, even columns contain APEs.

APPENDIX B

Robustness Checks for Chapter 2

B.1. Screenshots of the DSS's Interface

Figure B.1 shows one of the legacy inventory reports on which managers used to base their decisions before the DSS's implementation. Figure B.2 shows the DSS's interface after the interventions.



	Precio Saldo	Venta día 17/01/2009	Venta día 18/01/2009	Venta día 19/01/2009	Venta Acumul 19/01/2009	Stock Tienda 19/01/2009	Stock/ VentaDia 19/01/2009	% Éxito
FROM 49.90 TO 29.90	19,95	24	5	12	534	1.218	102	31
OF 24.90	14,95	21	8	8	466	1.006	126	32
OF19.90	9,95	16	14	6	420	384	64	54
FROM 14,90 TO 12,90	6,95	21	22	12	519	322	27	64
	Totales...	**82**	**49**	**38**	**1.939**	**2.930**	**77**	**41**

Figure B.1: One of the weekly inventory and sales reports on which managers based their pricing decisions before the DSS's implementation. The last two columns correspond to the rotation and success (see Section 2.3.2).

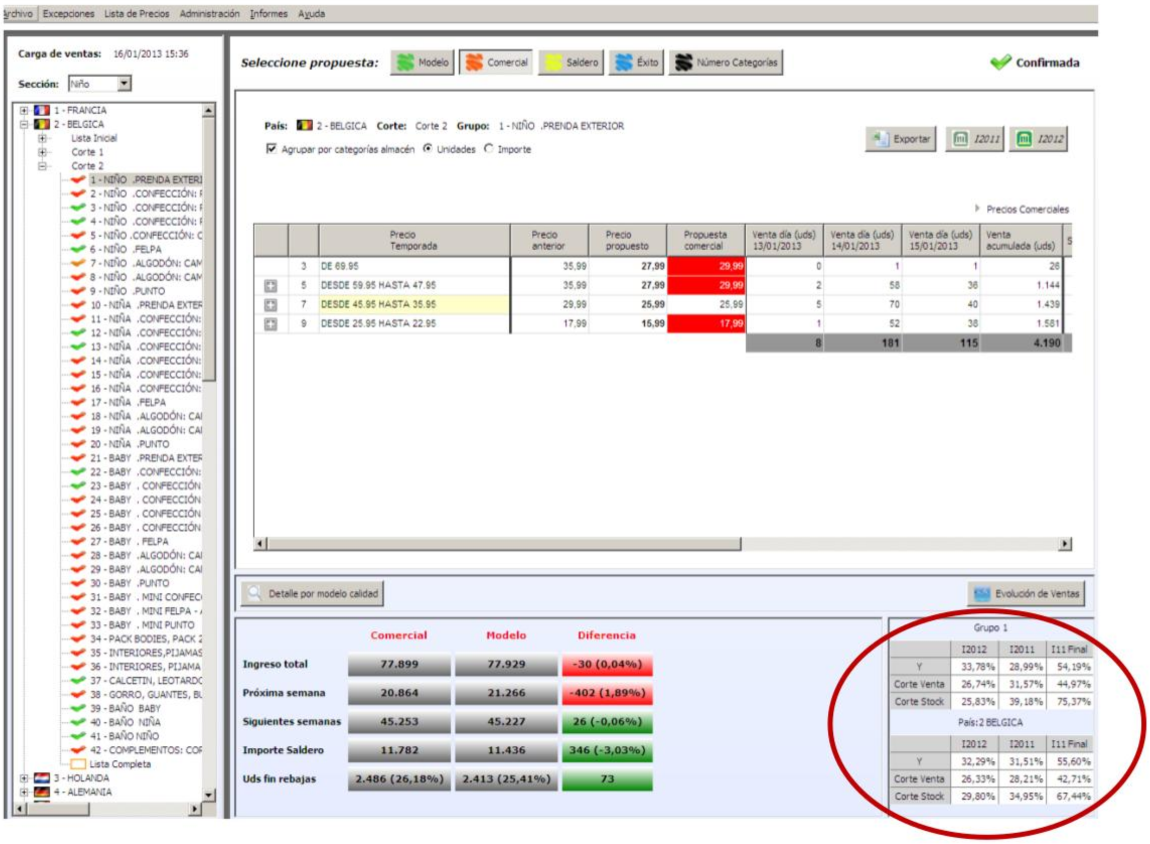


Figure B.2: The DSS's interface after the interventions. The top area contains the inventory and sales reports. The bottom left area contains the price recommendation and the managers' confirmed price, and their respective revenue and sales forecasts (for different horizons). The bottom right area contains what was added during the interventions: the Y metric for that group and country (first column of the tables), plus the Y metric for that group and country in the same week of the previous year as a reference point (second column), and at the end of the season corresponding to the reference point (third column).

B.2. Robustness Checks for Section 2.4

B.2.1 Difference-in-Differences Regression for a No-Intervention Period

The two interventions performed by Zara took place in a spring-summer campaign. Therefore, all our DiD analyses of the effect of the interventions contain one fall-winter season as pre-intervention period, and one spring-summer season as post-intervention. Potentially, our results could be capturing only a seasonality trend in adherence behavior. In other words, it could happen that managers consistently adhere more in spring-summer than in fall-winter, for reasons that are unrelated to the interventions, such as reacting to different purchasing behavior from consumers in different times of the year, or others. Note that we find a smaller effect of Intervention 1 than of Intervention 2, and no effect at all of Intervention 1 on franchises compared to own-store countries, which shows that the time at which Intervention 2 occurred was different from Intervention 1's.

To strengthen our analysis and to rule out the possibility of a consistent seasonality effect, we run the previous DiD analyses using data from W2012 and S2013, exactly a year after the two campaigns we use to study Intervention 2, but this time no intervention took place. If there is no seasonality effect, and the previous results can be attributed to the interventions, then here we expect to see no significant difference in adherence, and in probability of marking down conditional on the DSS recommending keep, because there was no intervention. In this regression, the three different splits between control and treated are identical to those in Intervention 2's regression, and the time variable (equivalent to $Int2_t$ in Intervention 2's analysis), called $Time_t$, takes value 0 in W2012, 1 in S2013.

The results of this robustness check are in Table B.1. Indeed, all the DiD coefficients of interest are statistically insignificant. This shows that there is no seasonality effect such that managers' adherence behavior is intrinsically different in spring-summer and in fall-winter.

Notice, too, that the $Franchise$ binary variable's coefficient is also statistically insignificant, which was not the case when we studied the interventions. This shows that, after the interventions, franchise managers' adherence behavior was similar to that of own-store country managers.

B.2.2 Fixed-Effects Linear Regression with Intervention Indicators

As an additional robustness check, we run two fixed effects (within-country) linear regressions: one has $Adherence_{itg}$ as dependent variable, and the other one has the probability

	No intervention		
	Adherence		
	(1)	(2)	(3)
Time	0.000644	0.00517	-0.00927
	(0.00848)	(0.00997)	(0.0133)
Franchise	-0.0279		
	(0.0155)		
Time×Franchise	-0.0243		
	(0.0169)		
LowAdherence75		-0.0665***	
		(0.0154)	
Time×LowAdherence75		-0.0191	
		(0.0141)	
LowAdherence90			-0.0733***
			(0.0198)
Time×LowAdherence90			-0.0000169
			(0.0160)
Constant	0.630***	0.667***	0.683***
	(0.00939)	(0.0129)	(0.0182)
N	2996	2996	2996

Robust standard errors in parentheses

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

	No intervention		
	CMarkdown		
	(1)	(2)	(3)
Time	-0.00595	0.0103	0.0141
	(0.00955)	(0.0106)	(0.0154)
Franchise	-0.0177		
	(0.0166)		
Time×Franchise	0.0231		
	(0.0154)		
LowAdherence75		0.0242	
		(0.0192)	
Time×LowAdherence75		-0.00741	
		(0.0137)	
LowAdherence90			0.0316
			(0.0286)
Time×LowAdherence90			-0.0103
			(0.0172)
Constant	0.264***	0.237***	0.227***
	(0.0112)	(0.0166)	(0.0273)
N	2996	2996	2996

Robust standard errors in parentheses

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

Table B.1: Change in managers' adherence (left) and in managers' probability of marking a product down when the DSS recommended keeping its price unchanged (right), as a DiD regression, using data from fall-winter 2012 and spring-summer 2013, when the company did not perform any intervention. In each regression, the first column corresponds to the DiD estimator when own-store countries are the control and franchises are the treated; the second column, when managers in the top quartile of pre-intervention adherence are control; the third column, when the top decile of pre-intervention adherence are control.

of marking a product down when the DSS recommended keeping its price unchanged, or $CMarkdown_{itg}$. Both contain indicator variables which take value 1 after each intervention had taken place, and a number of control variables: the number of stores in that country ($LogNumStores_{it}$), the number of campaigns the manager has been using the DSS ($ExperienceWithDSS_{it}$), the salvage value ($SalvageValue_{itg}$ in euros), and season (fall-winter or spring-summer) and product group dummies. The data used in this regression is, like in the DiD case, aggregated at the group-campaign level, i.e., the adherence and probability of marking down are averaged for all weeks in a campaign and all product categories in a group, and contains all campaigns from W2010 to W2013.

The results of this pair of regressions, shown in Table B.2, are robust to those of our DiD analysis: Intervention 1 had no significant, robust effect, but Intervention 2 did, and it was all driven by franchises. Intervention 2 increased franchises' adherence and decreased their likelihood of marking down a product against the DSS's recommendations.

B.3. Robustness Checks for Section 2.5

B.3.1 Limitations of the Heckman Estimator

The Heckman estimator models a two-step decision like the one that managers at Zara faced: first, whether to deviate from the DSS's recommendation or not; second, if they deviated, by how much. Moreover, it is a useful estimator for any setting in which the dependent variable follows a continuous distribution but has a few points with non-zero mass (Wooldridge 2010). In our case, the magnitude of deviations has a large mass at $AbsDeviation_{wc} = 0$, which makes this a well-suited model for our setting. The fact that the Mills ratio has a significant coefficient estimate in our results shows the need to account for this type of two-step process.

However, this estimator does not take into account the panel nature of our data: relatively few (84) country managers made multiple decisions over time (around 3,500 on average). The covariance matrix correction proposed by Heckman (1977) does take into account some sources of heteroskedasticity, but not serial correlation between each cross-section. In this

	Adherence		
	All countries	Own stores	Franchises
	(1)	(2)	(3)
Int1	-0.00217	-0.0142	0.0257
	(0.0122)	(0.00858)	(0.0257)
Int2	0.0565***	0.00158	0.120***
	(0.0140)	(0.0122)	(0.0242)
ExperienceWithDSS	0.00523	0.00853**	-0.000182
	(0.00284)	(0.00270)	(0.00512)
LogNumStores	-0.00666	-0.000260	0.0465
	(0.0267)	(0.0254)	(0.0372)
SalvageValue	-0.00586**	-0.000918	
	(0.00178)	(0.00493)	
Franchise	0.0586***		
	(0.0153)		
Constant	0.509***	0.541***	0.382***
	(0.0507)	(0.0782)	(0.0457)
N	10431	5639	4792

Robust standard errors in parentheses

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

	CMarkdown		
	All countries	Own stores	Franchises
	(1)	(2)	(3)
Int1	-0.00610	0.0186	-0.0797*
	(0.0154)	(0.0115)	(0.0306)
Int2	-0.0949***	0.0140	-0.223***
	(0.0210)	(0.0113)	(0.0326)
ExperienceWithDSS	-0.00881**	-0.0115***	-0.00209
	(0.00278)	(0.00297)	(0.00429)
LogNumStores	0.0462	0.00372	-0.0333
	(0.0425)	(0.0288)	(0.0424)
SalvageValue	0.00521*	-0.00115	
	(0.00221)	(0.00368)	
Franchise	-0.0458		
	(0.0437)		
Constant	0.379***	0.398***	0.639***
	(0.0781)	(0.0763)	(0.0498)
N	10275	5598	4677

Robust standard errors in parentheses

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

Table B.2: Regression of managers' adherence (left) and of managers' probability of marking a product down when the DSS recommended keeping its price unchanged (right), as a fixed effects (within-country) linear regression with intervention indicator variables plus controls. In each regression, the first column corresponds to all country managers; the second column, to managers in from countries in which Zara owns the stores; the third column, to franchise countries. Not reported: season and group dummy variables.

section, we repeat each part of the Heckman estimator using panel data-specific methods which correct for serial correlation.

B.3.2 Probability of Deviating as a Fixed Effects Linear Probability Model

The first part that we test is the selection part. In the Heckman estimator, this is modeled like a pooled probit, where the binary $Deviated_{wc}$ is the dependent variable. The panel data-specific equivalent of this would be a random effects probit model, or a fixed effects or random effects logit model, to account for the binary nature of the dependent variable. However, given the large size of the data, and the number of dummy variables to include as controls (group, week, season, year), this becomes computationally infeasible.

Instead, we run a simple linear probability model, with $Deviated_{wc}$ as dependent variable, within-country fixed effects, and all covariates and controls that we included in the selection part of the Heckman estimator (except country dummies, given that now the dependent variable has its country average subtracted). The errors are clustered at the country level, correcting for any possible serial correlation.

Table B.3 shows the results of this regression. We observe that $DisagreeDSS_{wc}$ and $DisagreeDSS_{wg}$, for franchises, are statistically significant, and that the number of categories in the group is only statistically significant after Intervention 2. For own-store countries, $DisagreeDSS_{wc}$, $DisagreeDSS_{wg}$ and $RelSalvageValue$ are statistically significant. Moreover, all the statistically significant coefficients in this table have values that are very similar to their equivalent APE in the selection part of the Heckman regression. These results provide robustness to our coefficient and APE estimates in the Heckman model's selection part.

B.3.3 Magnitude of Price Deviations as a Fixed Effects Linear Model

Analogously to the previous section, we run a within-country fixed effects linear regression using the continuous variable $AbsDeviation_{wc}$ using the same covariates and controls that

	Deviated					
	All countries	Own stores	Franchises	All countries	Own stores	Franchises
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline*DisagreeDSS$_c$	0.177***	0.206***	0.0715			
	(0.0201)	(0.0198)	(0.0529)			
Int1*DisagreeDSS$_c$	0.113***	0.174***	-0.0226			
	(0.0194)	(0.0115)	(0.0364)			
Int2*DisagreeDSS$_c$	0.0745***	0.135***	0.0146			
	(0.0121)	(0.0118)	(0.0154)			
Baseline*DisagreeDSS$_g$				0.228***	0.266***	0.102
				(0.0219)	(0.0203)	(0.0583)
Int1*DisagreeDSS$_g$				0.132***	0.202***	-0.0146
				(0.0205)	(0.0120)	(0.0377)
Int2*DisagreeDSS$_g$				0.0847***	0.151***	0.0175
				(0.0147)	(0.0138)	(0.0194)
Baseline*RelSalvageValue	-0.838***	-0.776***		-0.817***	-0.731***	
	(0.0765)	(0.0742)		(0.0747)	(0.0688)	
Int1*RelSalvageValue	-0.805***	-0.926***		-0.777***	-0.884***	
	(0.0571)	(0.0655)		(0.0553)	(0.0633)	
Int2*RelSalvageValue	-0.681***	-1.043***		-0.665***	-1.010***	
	(0.0504)	(0.0394)		(0.0494)	(0.0393)	
Baseline*NumCategs	-0.0113	0.0102	-0.00865	-0.0129*	0.00965	-0.0109
	(0.00573)	(0.00509)	(0.0138)	(0.00589)	(0.00509)	(0.0139)
Int1*NumCategs	-0.00270	0.00796	-0.000910	-0.00265	0.00758	-0.000736
	(0.00291)	(0.00399)	(0.00382)	(0.00288)	(0.00395)	(0.00375)
Int2*NumCategs	0.0134***	0.0134***	0.0112***	0.0129***	0.0129**	0.0111***
	(0.00185)	(0.00378)	(0.00209)	(0.00187)	(0.00370)	(0.00215)
ExperienceWithDSS	-0.0265***	-0.000890	-0.0249*	-0.0249***	0.00225	-0.0249*
	(0.00525)	(0.00504)	(0.00945)	(0.00532)	(0.00479)	(0.00943)
LogNumStores	0.00849	0.00434	-0.0186	0.00876	0.00347	-0.0187
	(0.0192)	(0.0207)	(0.0281)	(0.0193)	(0.0217)	(0.0281)
Franchise	-0.0469			-0.0467		
	(0.0244)			(0.0243)		
Constant	0.673***	0.528***	0.657***	0.657***	0.496***	0.657***
	(0.0523)	(0.0633)	(0.0688)	(0.0529)	(0.0639)	(0.0688)
N	294386	169137	125249	294386	169137	125249

Robust standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: Coefficients of a fixed effects (within-country) linear regression of the decision to deviate from the DSS's recommended price. Not reported: week, season, year, and group dummy variables.

we used for the deviation part of the Heckman estimator. Given that the decision not to deviate would bias our results (because in the Heckman regression the Mills ratio has a statistically significant coefficient), we need to take into account only the times in which managers decided to deviate, i.e., we run this regression using only the subsample of data in which $Deviated_{wc} = 1$.

Table B.4 shows the coefficient estimates. We can see that the variable $StockoutWeek_{wg}$, computed at the group-aggregate level, is almost never statistically significant. Other than this one, the rest of covariates ($StockoutWeek_{wc}$ and $RelSalvageValue_{wc}$) show the same significance and similar values to their equivalents in the Heckman model's deviation part, which strengthens our previous estimates' robustness.

	AbsDeviation					
	All countries	Own stores	Franchises	All countries	Own stores	Franchises
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline*StockoutWeek$_c$	0.00616***	0.00699**	0.00698***			
	(0.00160)	(0.00222)	(0.00158)			
Int1*StockoutWeek$_c$	0.00113	0.00637***	0.000499			
	(0.000574)	(0.00140)	(0.000656)			
Int2*StockoutWeek$_c$	0.00464***	0.00676***	0.00435***			
	(0.000779)	(0.00113)	(0.00103)			
Baseline*StockoutWeek$_g$				0.00286	0.00248	0.00793**
				(0.00212)	(0.00279)	(0.00274)
Int1*StockoutWeek$_g$				-0.000793	0.00186	-0.000410
				(0.000531)	(0.00101)	(0.000627)
Int2*StockoutWeek$_g$				0.00129*	0.00274*	0.00139
				(0.000641)	(0.00117)	(0.000762)
Baseline*RelSalvageValue	0.127**	0.184***		0.188***	0.212***	
	(0.0451)	(0.0479)		(0.0457)	(0.0509)	
Int1*RelSalvageValue	0.183***	0.149***		0.175***	0.176***	
	(0.0415)	(0.0400)		(0.0419)	(0.0412)	
Int2*RelSalvageValue	-0.102	0.0178		-0.0920	0.0275	
	(0.0608)	(0.0668)		(0.0626)	(0.0695)	
ExperienceWithDSS	0.00590	0.00126	0.00836	0.00830*	0.00188	0.0146**
	(0.00320)	(0.00431)	(0.00415)	(0.00344)	(0.00483)	(0.00436)
LogNumStores	-0.0347	-0.0473	-0.0127	-0.0379	-0.0468	-0.0181
	(0.0251)	(0.0380)	(0.0220)	(0.0243)	(0.0394)	(0.0201)
Franchise	-0.0385			-0.0405*		
	(0.0203)			(0.0195)		
Constant	0.391***	0.384***	0.366***	0.410***	0.408***	0.361***
	(0.0558)	(0.104)	(0.0378)	(0.0562)	(0.111)	(0.0383)
N	109619	59844	49775	109619	59844	49775

Robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4: Coefficients of a fixed effects (within-country) linear regression of the magnitude of price deviations, conditional on having decided to deviate (i.e., using only the subsample of observations in which the manager did not adhere to the DSS's recommendation). Not reported: week, season, year, and group dummy variables.

BIBLIOGRAPHY

Abdellaoui, M., H. Bleichrodt, and O. lHaridon. 2008. A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and uncertainty* 36 (3): 245.

Abdellaoui, M., H. Bleichrodt, and C. Paraschiv. 2007. Loss aversion under prospect theory: A parameter-free measurement. *Management Science* 53 (10): 1659–1674.

Agrawal, V., and D. Lee. 2016. The effect of sourcing policies on a supplier's sustainable practices. Working paper, available at SSRN 2601974.

Alan, Y., and M. A. Lapré. 2018. Investigating operational predictors of future financial distress in the US airline industry. *Production and Operations Management* 27 (4): 734–755.

Amaldoss, W., and C. He. 2017. Reference-dependent utility, product variety, and price competition. *Management Science* 64 (9): 4302–4316.

Anenberg, E. 2011. Loss aversion, equity constraints and seller behavior in the real estate market. *Regional Science and Urban Economics* 41 (1): 67–76.

Angist, J. D., and J.-S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Ashida, M., and E. Plinke. 2004. Made in China: Is this a sustainable label? Risks and opportunities of the China boom from a socially responsible investment perspective. Technical report, Association for Sustainable and Responsible Investment in Asia.

Babich, V., and C. S. Tang. 2012. Managing opportunistic supplier product adulteration: Deferred payments, inspection, and combined mechanisms. *Manufacturing Service Oper. Management* 14 (2): 301–314.

Baron, O., M. Hu, S. Najafi-Asadolahi, and Q. Qian. 2015. Newsvendor selling to loss-averse consumers with stochastic reference points. *Manufacturing & Service Operations Management* 17 (4): 456–469.

Barrio, P. J., D. G. Goldstein, and J. M. Hofman. 2016. Improving comprehension of numbers in the news. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2729–2739. ACM.

Bazerman, M. H., and D. A. Moore. 2008. *Judgment in managerial decision making*. Wiley.

Bearden, J. N., R. O. Murphy, and A. Rapoport. 2008. Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management* 10 (4): 625–636.

Benartzi, S., and R. H. Thaler. 1995. Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics* 110 (1): 73–92.

Bendoly, E., K. Donohue, and K. L. Schultz. 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of operations management* 24 (6): 737–752.

Bolton, G. E., and E. Katok. 2008. Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management* 10 (3): 519–538.

Bordalo, P., N. Gennaioli, and A. Shleifer. 2013. Salience and consumer choice. *Journal of Political Economy* 121 (5): 803–843.

Bruno, H. A., H. Che, and S. Dutta. 2012. Role of reference price on price and quantity: insights from business-to-business markets. *Journal of Marketing Research* 49 (5): 640–654.

Butler, S. 2014. Primark to pay £6m more to victims of Rana Plaza factory in Bangladesh. *The Guardian* March 16.

Camerer, C. 2005. Three cheerspsychological, theoretical, empiricalfor loss aversion. *Journal of Marketing Research* 42 (2): 129–133.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler. 1997. Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics* 112 (2): 407–441.

Camerer, C. F. 1998. Prospect theory in the wild: Evidence from the field. *California Institute of Technology Social Science Working Papers* 1037.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics using Stata*. Stata press College Station, TX.

Caplin, A., and M. Dean. 2015. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review* 105 (7): 2183–2203.

Caro, F. 2012. Zara: Staying fast and fresh. Technical report, The Case Center. Reference number 612-006-1.

Caro, F., P. Chintapalli, K. Rajaram, and C. S. Tang. 2016. Improving supplier compliance through joint and shared audits. *Manufacturing & Service Operations Management (forthcoming)*.

Caro, F., and J. Gallien. 2012. Clearance pricing optimization for a fast-fashion retailer. *Operations Research* 60 (6): 1404–1422.

Chen, L., and H. L. Lee. 2016. Sourcing under supplier responsibility risk: The effects of certification, audit, and contingency payment. *Management Science* 63 (9): 2795–2812.

Chen, M. K., V. Lakshminarayanan, and L. R. Santos. 2006. How basic are behavioral biases? evidence from capuchin monkey trading behavior. *Journal of Political Economy* 114 (3): 517–537.

Chen, X., G. Hao, and L. Li. 2014. Channel coordination with a loss-averse retailer and option contracts. *International Journal of Production Economics* 150:52–57.

Cheng, T. E., J. N. Gupta, and G. Wang. 2000. A review of flowshop scheduling research with setup times. *Production and operations management* 9 (3): 262–282.

Cheremukhin, A., A. Popova, A. Tutino et al. 2011. Experimental evidence on rational inattention. *Federal Reserve Bank of Dallas Working Paper* 1112.

Chetty, R., A. Looney, and K. Kroft. 2009. Salience and taxation: Theory and evidence. *American economic review* 99 (4): 1145–77.

Clifford, S., and S. Greenhouse. 2013. Fast and flawed inspections of factories abroad. *The New York Times* September 1.

Cui, R., G. Allon, A. Bassamboo, and J. A. Van Mieghem. 2015. Information sharing in supply chains: An empirical and theoretical valuation. *Management Science* 61 (11): 2803–2824.

Davis, A. M. 2015. An experimental investigation of pull contracts in supply chains. *Production and Operations Management* 24 (2): 325–340.

Davis, A. M., E. Katok, and N. Santamaría. 2014. Push, pull, or both? a behavioral study of how the allocation of inventory risk affects channel efficiency. *Management Science* 60 (11): 2666–2683.

Delmas, M., and I. Montiel. 2009. Greening the supply chain: When is customer pressure effective? *Journal of Economics & Management Strategy* 18 (1): 171–201.

Deng, X., J. Xie, and H. Xiong. 2013. Manufacturer–retailer contracting with asymmetric information on retailer's degree of loss aversion. *International Journal of Production Economics* 142 (2): 372–380.

Distelhorst, G., J. Hainmueller, and R. M. Locke. 2016. Does lean improve labor standards? management and social performance in the nike supply chain. *Management Science* 63 (3): 707–728.

Donaldson, T. 2015. Ethical sourcing forum: Subcontracting more prevalent than brands believe. *The Sourcing Journal* March 27 online.

Drake, D., and S. Spinler. 2013. OM Forum - sustainable operations management: An enduring stream or a passing fancy? *Manufacturing Service Oper. Management* 15 (4): 689–700.

Drennan, J. 2015. Laundering the global garment industry's dirty business. *Foreign Policy* April 24 online.

Dyer, J. H. 1996. Specialized supplier networks as a source of competitive advantage: Evidence from the auto industry. *Strategic management journal*:271–291.

Eisen, M. Amazon's $23,698,655.93 book about flies. In *Michael Eisen's Blog*: http://www.michaeleisen.org/blog/?p=358.

Elmaghraby, W., W. Jank, S. Zhang, and I. Z. Karaesmen. 2015. Sales force behavior, pricing information, and pricing decisions. *Manufacturing & Service Operations Management* 17 (4): 495–510.

Engelhardt, G. V. 2003. Nominal loss aversion, housing equity constraints, and household mobility: evidence from the united states. *Journal of urban Economics* 53 (1): 171–195.

Ericson, K. M. M., and A. Fuster. 2014. The endowment effect.

Flicker, B. 2018. Managerial insight and "optimal" algorithms. *Working Paper*.

Gächter, S., E. J. Johnson, and A. Herrmann. 2007. Individual-level loss aversion in riskless and risky choices. *SSRN*.

Genesove, D., and C. Mayer. 2001. Loss aversion and seller behavior: Evidence from the housing market. *The quarterly journal of economics* 116 (4): 1233–1260.

Gilbert, B., and J. G. Zivin. 2014. Dynamic salience with intermittent billing: Evidence from smart electricity meters. *Journal of Economic Behavior & Organization* 107:176–190.

Gneezy, U., A. Kapteyn, and J. Potters. 2003. Evaluation periods and asset prices in a market experiment. *The Journal of Finance* 58 (2): 821–837.

Gneezy, U., and J. Potters. 1997. An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics* 112 (2): 631–645.

Goette, L., D. Huffman, and E. Fehr. 2004. Loss aversion and labor supply. *Journal of the European Economic Association* 2 (2-3): 216–228.

Goodman, P. 2013. Bangladesh tragedy proves global garment trade too big to supervise. *Huffington Post* May 9.

Greene, W. H. 2003. *Econometric analysis.* Pearson Education India.

Greene, W. H. 2007. *Econometric analysis.* Prentice Hall.

Greenhouse, S. 2013. Europeans fault American safey effort in Bangladesh. *The New York Times* Nov 18.

Guo, R., H. Lee, and R. Swinney. 2016. Responsible sourcing in supply chains. *Management Science* 62 (9): 2722–2744.

Haigh, M. S., and J. A. List. 2005. Do professional traders exhibit myopic loss aversion? an experimental analysis. *The Journal of Finance* 60 (1): 523–534.

Hardie, B. G., E. J. Johnson, and P. S. Fader. 1993. Modeling loss aversion and reference dependence effects on brand choice. *Marketing science* 12 (4): 378–394.

Hastie, R., and R. M. Dawes. 2010. *Rational choice in an uncertain world: The psychology of judgment and decision making.* Sage.

Heckman, J. 1981. Heterogeneity and state dependence. In *Studies in Labor Markets*, ed. S. Rosen: Univeristy of Chicago Press.

Heckman, J. J. 1977. Sample selection bias as a specification error (with an application to the estimation of labor supply functions). *National Bureau of Economic Research Cambridge, Mass., USA.*

Ho, T.-H., and J. Zhang. 2008. Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science* 54 (4): 686–700.

Hoch, S. J., and D. A. Schkade. 1996. A psychological approach to decision support systems. *Management Science* 42 (1): 51–64.

Hogarth, R. M. 1981. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin* 90 (2): 197.

Hossain, T., and J. A. List. 2012. The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science* 58 (12): 2151–2167.

Hu, B., C. Meng, D. Xu, and Y.-J. Son. 2016. Three-echelon supply chain coordination with a loss-averse retailer and revenue sharing contracts. *International Journal of Production Economics* 179:192–202.

Huang, L., J. Song, and R. Swinney. 2017. Managing social responsibility in multi-tier supply chains. Working paper, available at SSRN 2837332.

Hurst, R., H. Murdoch, and D. Gould. 2005. Changing over time: Tackling supply chain labour issues through business practice. Technical report, The Impact Overtime Project.

Ibanez, M. R., J. R. Clark, R. S. Huckman, and B. R. Staats. 2017. Discretionary task ordering: Queue management in radiological services. *Management Science*.

Imas, A., S. Sadoff, and A. Samek. 2016. Do people anticipate loss aversion? *Management Science* 63 (5): 1271–1284.

Jiang, B. 2009. The effects of interorganizational governance on supplier's compliance with SCC: An empirical examination of compliant and non-compliant suppliers. *J. Oper. Manag.* 27 (4): 267–280.

Johnsen, L. C., G. Voigt, and C. J. Corbett. 2019. Behavioral contract design under asymmetric forecast information. *Decision Sciences*.

Kahneman, D. 1973. *Attention and effort*, Volume 1063. Citeseer.

Kahneman, D. 2003a. Maps of bounded rationality: Psychology for behavioral economics. *American economic review* 93 (5): 1449–1475.

Kahneman, D. 2003b. A perspective on judgment and choice: mapping bounded rationality. *American psychologist* 58 (9): 697.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives* 5 (1): 193–206.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263–291.

Katok, E., and D. Y. Wu. 2009. Contracting in supply chains: A laboratory investigation. *Management Science* 55 (12): 1953–1968.

Kerppola, M., R. Moody, L. Zheng, and A. Liu. 2014. H&M's global supply chain management sustainability: Factories and fast fashion. Technical report, William Davidson Institute (WDI) at the University of Michigan.

Kim, S.-H. 2015. Time to come clean? Disclosure and inspection policies for green production. *Operations Research* 63 (1): 1–20.

Kleinmuntz, D. N. 1985. Cognitive heuristics and feedback in a dynamic decision environment. *Management Science* 31 (6): 680–702.

Kőszegi, B., and M. Rabin. 2006. A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121 (4): 1133–1165.

Kraft, T., L. Valdes, and Y. Zheng. 2017. Improving supplier social responsibility under incomplete visibility. Working paper, available at SSRN 3028117.

Labowitz, S., and D. Baumann-Pauly. 2014. Business as usual is not an option. Technical report, New York University Center for Business and Human Rights.

Labowitz, S., and D. Baumann-Pauly. 2015. Beyond the tip of the iceberg: Bangladesh's forgotten apparel workers. Technical report, New York University Center for Business and Human Rights.

Lee, H. L., and E. Plambeck. 2009. Ma Jun and IPE: Using information to improve China's environment. Technical report, Stanford Graduate School of Business, Teaching Case SI-115.

Liu, W., S. Song, and C. Wu. 2013. Impact of loss aversion on the newsvendor game with product substitution. *International Journal of Production Economics* 141 (1): 352–359.

Locke, R., M. Amengual, and A. Mangla. 2009. Virtue out of necessity? Compliance, commitment, and the improvement of labor conditions in global supply chains. *Politics & Society* 37 (3): 319–351.

Locke, R., F. Qin, and A. Brause. 2007. Does monitoring improve labor standards? Lessons from Nike. *Industrial & Labor Relations Review* 61 (1): 3–31.

Long, X., and J. Nasiry. 2014. Prospect theory explains newsvendor behavior: The role of reference points. *Management Science* 61 (12): 3009–3012.

McIntosh, A., B. J. Reys, and R. E. Reys. 1992. A proposed framework for examining basic number sense. *For the learning of mathematics* 12 (3): 2–8.

McKenzie, N., and R. Baker. 2016. Surf clothing label Rip Curl using 'slave labour' to manufacture clothes in North Korea. *The Sydney Morning Herald* February 21 online.

Meeker, D., J. A. Linder, C. R. Fox, M. W. Friedberg, S. D. Persell, N. J. Goldstein, T. K. Knight, J. W. Hay, and J. N. Doctor. 2016. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: A randomized clinical trial. *Journal of the American Medical Association* 315:562–570.

Moore, B., T. Hunt, and S. Ditty. 2016. The fashion transparency index. Technical report, Ethical Consumer and Fashion Revolution.

Nasiry, J., and I. Popescu. 2011. Dynamic pricing with loss-averse consumers and peak-end anchoring. *Operations research* 59 (6): 1361–1368.

Neate, R. 2014. Amazon sellers hit by nightmare before christmas as glitch cuts prices to 1p. *The Guardian*.

Novemsky, N., and D. Kahneman. 2005. The boundaries of loss aversion. *Journal of Marketing research* 42 (2): 119–128.

Nuttall, C. 2012. Apple agrees to China pollution audit. *Financial Times* April 2012 online.

Orsdemir, A., B. Hu, and V. Deshpande. 2015. Responsible sourcing via vertical integration and supply chain partnership. Working paper, available at SSRN 2630733.

Özer, Ö., and Y. Zheng. 2012. Behavioral issues in pricing management.

Peters, E. 2012. Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science* 21 (1): 31–35.

Peters, E., D. Västfjäll, P. Slovic, C. Mertz, K. Mazzocco, and S. Dickert. 2006. Numeracy and decision making. *Psychological science* 17 (5): 407–413.

Plambeck, E., and T. Taylor. 2016. Supplier evasion of a buyer's audit: Implications for motivating compliance with labor and environmental standards. *Manufacturing Service Oper. Management* 18 (2): 184–197.

Popescu, I., and Y. Wu. 2007. Dynamic pricing strategies with reference effects. *Operations Research* 55 (3): 413–429.

Porteous, A., S. Rammohan, and H. Lee. 2015. Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties. *Production Oper. Management* 24 (9): 1402–1413.

Robertson, H., and S. Heng. 2014. Brands dodge blame over substandard factory. *The Cambodian Daily* July 22 online.

Rosa-Hatko, W., and E. A. Gunn. 1997. Queues with switchover: A review and critique. *Annals of Operations Research* 69:299–322.

Samuelson, W., and R. Zeckhauser. 1988. Status quo bias in decision making. *Journal of risk and uncertainty* 1 (1): 7–59.

Sarkar, S., and R. S. Sriram. 2001. Bayesian models for early warning of bank failures. *Management Science* 47 (11): 1457–1475.

Schmidt, U., and S. Traub. 2002. An experimental test of loss aversion. *Journal of risk and Uncertainty* 25 (3): 233–249.

Schweitzer, M. E., and G. P. Cachon. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science* 46 (3): 404–420.

Sharda, R., S. H. Barr, and J. C. MCDonnell. 1988. Decision support system effectiveness: a review and an empirical test. *Management science* 34 (2): 139–159.

Sims, C. A. 2003. Implications of rational inattention. *Journal of monetary Economics* 50 (3): 665–690.

Smithers, R. 2015. Rana Plaza disaster: Benetton donates $1.1m to victims' fund. *The Guardian* April 17.

Su, X. 2008. Bounded rationality in newsvendor models. *Manufacturing & Service Operations Management* 10 (4): 566–589.

Tan, T., and B. R. Staats. 2016. Behavioral drivers of routing decisions: Evidence from restaurant table assignment. *Working Paper*.

Tang, C. S., and V. Babich. 2014. Using social and economic incentives to discourage Chinese suppliers from product adulteration. *Business Horizons* 57 (4): 497–508.

Tereyağoğlu, N., P. S. Fader, and S. Veeraraghavan. 2017. Multiattribute loss aversion and reference dependence: Evidence from the performing arts industry. *Management Science* 64 (1): 421–436.

Thaler, R. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization* 1 (1): 39–60.

Thaler, R. H., A. Tversky, D. Kahneman, and A. Schwartz. 1997. The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics* 112 (2): 647–661.

Theuws, M. 2015. Hidden subcontracting in the garment industry. Technical report, Centre for Research on Multinational Corporations.

Thorlakson, T., J. F. de Zegher, and E. F. Lambin. 2018. Companies contribution to sustainability through global supply chains. *Proceedings of the National Academy of Sciences*:201716695.

Tiefenbeck, V., L. Goette, K. Degen, V. Tasic, E. Fleisch, R. Lalive, and T. Staake. 2016. Overcoming salience bias: how real-time feedback fosters resource conservation. *Management Science*.

Toffel, M. W., J. Short, and M. Ouellet. 2015. Codes in context: How states, markets, and civil society shape adherence to global labor standards. *Regulation & Governance* 9 (3): 205–223.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *science* 185 (4157): 1124–1131.

Tversky, A., and D. Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics* 106 (4): 1039–1061.

Uppari, B. S., and S. Hasija. 2018. Modeling newsvendor behavior: A prospect theory approach. *Manufacturing & Service Operations Management*.

Vairaktarakis, G. L. 2013. Noncooperative games for subcontracting operations. *Manufacturing Service Oper. Management* 15 (1): 148–158.

Van Donselaar, K. H., V. Gaur, T. Van Woensel, R. A. Broekmeulen, and J. C. Fransoo. 2010. Ordering behavior in retail stores and implications for automated replenishment. *Management Science* 56 (5): 766–784.

Van Mieghem, J. 2008. *Operations strategy*. Belmont, MA: Dynamic Ideas.

Vaughan-Whitehead, D., and L. Pinedo Caro. 2017. Purchasing practices and working conditions in global supply chains: Global survey results. Technical report, International Labour Organization, INWORK Issue Brief No. 10.

Wang, C. X. 2010. The loss-averse newsvendor game. *International Journal of Production Economics* 124 (2): 448–452.

Wang, C. X., and S. Webster. 2007. Channel coordination for a supply chain with a risk-neutral manufacturer and a loss-averse retailer. *Decision Sciences* 38 (3): 361–389.

Wang, R. 2018. When prospect theory meets consumer choice models: Assortment and pricing management with reference prices. *Manufacturing & Service Operations Management* 20 (3): 583–600.

Winer, R. S. 1986. A reference price model of brand choice for frequently purchased products. *Journal of consumer research* 13 (2): 250–256.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data.* MIT press.

Wooldridge, J. M. 2015. *Introductory econometrics: A modern approach.* Nelson Education.

Xinsheng, X., M. Zhiqing, S. Rui, J. Min, and J. Ping. 2015. Optimal decisions for the loss-averse newsvendor problem under cvar. *International Journal of Production Economics* 164:146–159.

Xu, X., H. Wang, C. Dang, and P. Ji. 2017. The loss-averse newsvendor model with backordering. *International Journal of Production Economics* 188:1–10.

Zhang, Y., K. Donohue, and T. H. Cui. 2015. Contract preferences and performance for the loss-averse supplier: Buyback vs. revenue sharing. *Management Science* 62 (6): 1734–1754.

Zhao, X., and K. E. Stecke. 2010. Pre-orders for new to-be-released products considering consumer loss aversion. *Production and Operations Management* 19 (2): 198–215.

Zheng, Y., and Ö. Özer. 2012. Behavioral issues in pricing management. *The Oxford handbook of pricing management*:415–460.