Clearance Pricing Optimization for a Fast-Fashion Retailer

Felipe Caro [*] Jérémie Gallien [†]

June 10, 2011

Abstract

Fast-fashion retailers such as Zara offer continuously changing assortments and use minimal in-season promotions. Their clearance pricing problem is thus challenging because it involves comparatively more different articles of unsold inventory with less historical price data points. Until 2007, Zara used a manual and informal decision-making process for determining price markdowns. In collaboration with their pricing team, we designed and implemented since an alternative process relying on a formal forecasting model feeding a price optimization model. As part of a controlled field experiment conducted in all Belgian and Irish stores during the 2008 Fall-Winter season, this new process increased clearance revenues by approximately 6%. Zara is currently using this process worldwide for its markdown decisions during clearance sales.

1. Introduction

Markdown pricing is an important activity for many retailers of seasonal goods (Talluri and van Ryzin 2004), and with more than a billion euros generated through clearance sales in 2008 this is certainly true of Spanish apparel retailer Zara. However, clearance pricing is arguably more challenging for Zara than for many of its competitors. This is because its innovative fast-fashion model (also adopted by Sweden-based H&M, Japan-based World Co., and Spain-based Mango) involves selling many more articles with shorter life cycles that are almost never discounted during the regular selling season (Ghemawat and Nueno 2003). As a result, when Zara decided in 2007 to develop internally a markdown optimization system, it quickly realized that its needs were at the forefront of revenue management practice, for two main reasons. First, almost no historical price response data is available for its articles at the beginning of each clearance period. As a result, it had to devise a methodology for estimating price elasticity based exclusively on features common with articles sold in previous seasons, and updating this estimation based on actual sales information once the clearance period starts. Second and perhaps most importantly, it could not rely on any of the published price optimization models known to have been implemented or tested with real data (e.g., Smith and Achabal 1998, Bitran, Caldentey and Mondschein 1998, Heching, Gallego and van Ryzin 2002 or Smith 2009), which all consider each article independently. Indeed, the high

[*]UCLA Anderson School of Management, Los Angeles, CA 90095, fcaro@anderson.ucla.edu
[†]London Business School, London NW1 4SA, United Kingdom, jgallien@london.edu

number of different articles available during clearance sales makes it impractical for its stores to implement pricing policies determined at the article level. For example, stores have to ensure that groups of similar articles (e.g., men's shirts) do not include too many different price points, that a minimum amount of inventory across different articles is associated with any advertised price point, that two groups of articles merged under a common price point will not be subsequently separated again, etc. (a complete discussion of these and other store-level markdown implementation issues associated with the fast-fashion model is provided further).

The present paper describes the development, implementation and evaluation in the field of the markdown optimization system that Zara has deployed worldwide since to address these challenges. To the best of our knowledge, this constitutes the first large-scale application of a multi-product price optimization model by a firm for which all relevant technical and implementation details as well as related impact estimation results are made public. In the remainder of this introduction, we provide additional background on Zara's clearance sales period (in §1.1), then describe its legacy clearance pricing process (in §1.2) and the structure of the new process developed along with the paper organization (in §1.3).

1.1 Clearance Sales at Zara

As many other apparel retailers, Zara holds its clearance sale periods for about two months following each biannual selling season, with country-specific starting dates at the beginning of January (Fall/Winter season) and late June or early July (Spring/Summer season). The generic goal of these clearance sales is to maximize the revenue derived from merchandise that is still unsold when a new collection is about to be introduced. At Zara, merchandise is often deliberately withdrawn from the store display area during the season in order to make room for more recent incoming articles. As a result, clearance sales offer a biannual opportunity to reduce the cost of not only the traditional assortment/collection transitions between seasons mentioned above, but also the frequent in-season assortment transitions associated with the fast-fashion model. While some of that inventory withdrawn during the season is sometimes kept in the store backroom, the majority of it is returned to a warehouse in order to subsequently enable a more efficient re-distribution before the clearance sales. Because clearance selling rates exceed the replenishment capacity of Zara's distribution system however, this re-distribution of clearance inventory to stores takes place over a period of several weeks preceding the sales period. As the sales period unfolds, Zara dedicates an increasing portion of its store space to the new collection; this overlap strategy is designed to promote the new collection and induce upsales. Some transfers of remaining clearance inventory may also be organized between nearby stores in order to aggregate merchandise in fewer locations and improve their display quality (e.g., complete missing sizes). Finally, the sales period end with the step know as *liquidation*, when all remaining clearance inventory is collected from Zara stores and sold (sometimes by weigth) to wholesale buyers working for low-price channels, typically in developing countries.

At the store level, clearance sales induce many substantial changes. The price reductions consented then, which are visibly advertised, attract substantially more visitors and the volume of merchandise sold drastically increases. As a result, the store workload increases both before and during clearance sales. In the weeks before the sales, the clearance inventory must be received, labeled and stored in the backroom. The day (and night) before the sale begins, the store display is entirely re-arranged: while garments are displayed with a relatively low density according to matching colors and styles during the regular season, during the sales period they are displayed with a very high density according to their type (e.g., men's shirts) and selling price. During the sales, store associates are solicited by many more customers, must re-fold and re-place many more tried garments, replenish and/or re-arrange display areas from the backroom and update price labels and signs more frequently, process more customers at the register, etc... Consequently, store staffing must be substantially increased during clearance sales, often with temporary employees.

For legal, marketing and organizational reasons, Zara's pricing policy during both the regular season and the clearance period is country-specific and to date Zara is not considering segmenting large countries into smaller pricing regions (e.g., U.S. east and west coasts). During clearance sales, another barrier to price segmentation shared with many other apparel retailers is the strong appeal to visitors of a limited number of signs displayed within the store to signal specific price points or markdowns (e.g., "everything at €4.99", or "-20%"). In addition, during both clearance sales and the regular season Zara only uses prices from a discrete and finite set of so-called *commercial prices* all ending with ".99" which, as discussed in the marketing literature (Anderson and Simester 2003) tend to be more appealing to customers. Another key feature more specific to fast-fashion is the substantial increase of the number of different articles present in the starting clearance sale inventory compared to that found in stores at any time during the regular season (which results from the short season life-cycles). As a result, it would be neither easy nor desirable for stores to implement a different pricing policy for each article during the clearance period, as is done during the regular season. In particular, because the current price display technology is paper-based, the implementation of markdowns requires store associates to locate and retrieve articles from the store backroom or display area and attach a price update sticker on each article tag. In this environment, Zara learned that the retrieving workload and probability of costly mislabeling errors associated with an independent pricing policy for each individual article are particularly high.

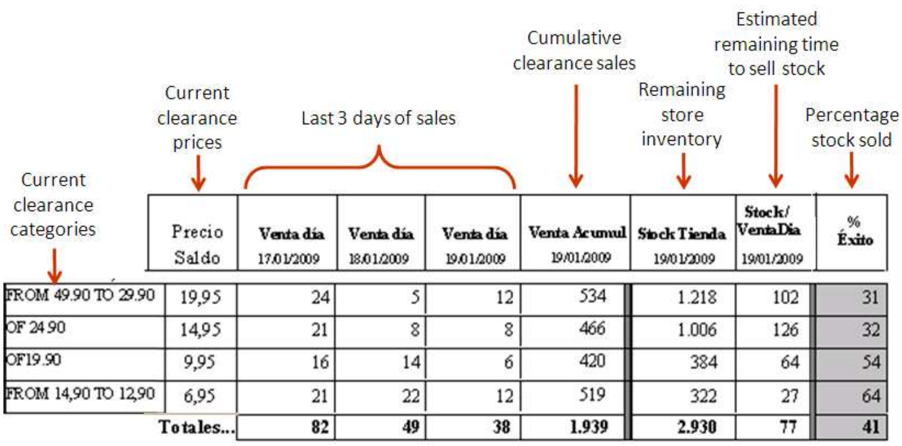
To address these challenges, Zara makes clearance pricing decisions in each country instead at the level of a set of articles called a *group*, which corresponds to a relatively high-level descriptor encompassing anywhere from 20 to 350 different articles; examples of groups include “woman blazers”, “man knitwear”, “basic skirts”, etc. Each group is partitioned into *clusters*, which are subsets of articles that were sold at the same price during the regular season; each group typically includes between 4 and 12 clusters. Finally, clearance pricing decisions are implemented by aggregating one or several clusters into a *category*, and assigning a different clearance sales price to each category. A important feature is that categories are almost always defined as an interval of regular season prices. For example, the group “basic skirts” could include 4 clusters of 9, 15, 25 and 12 articles sold respectively at €19.99, €24.99, €29.99 and €35.99 during the regular season; at a specific time during the clearance period Zara might decide to form a first category comprising the €19.99 cluster and assign it a marked-down price of €9.99, and form a second category comprising all the articles with a regular season selling price between €24.99 and €35.99 (i.e., the three remaining clusters) and assign it a clearance price of €19.99.

A key rationale for the clearance pricing methodology just described is that it is particularly easy to communicate and implement at the store level. This is because during clearance sales articles are both displayed (in stores) and stored (in their backrooms) by group. In addition, by design the regular season selling price is the feature of each article which is most prominently displayed on its tag. This methodology also makes it easy to ensure that clearance prices are always lower than regular season prices by a sufficient amount (a legal requirement in many countries), and that the total number of different price points for each group (which is often contrained by store display space limitations) remains sufficiently low. Finally, with this method clearance pricing decisions can be updated for each country and group on a weekly basis. In each such update the clearance price of a given cluster is never allowed to increase, in part to avoid arbitrage and returns. In addition, to simplify store execution and progressively reduce the store space dedicated to clearance sales, categories are allowed to merge but not split over time. In the next subsection, we describe the legacy process used by Zara to make the clearance pricing decisions just described.

1.2 Legacy Clearance Pricing Decision Process

Until 2007, the decision process used by Zara for clearance pricing consisted of two steps. The first consisted of determining initial categories and markdowns (see §1.1) for the very first week of the sales, and took place over a period of about a month preceding the beginning of the clearance period. This started with a systematic review of the unsold inventory and sales performance during the selling season for all the product groups in Spain (Zara’s top sales country to date), performed by the *pricing committee* –a small group of 4-5 key executives combining financial, commercial and distribution expertise that included Zara’s CFO, in conjunction with the two sales managers for that country. This team then determined all the categories and markdowns for Spain, based on an exchange of views and experience by its members. These decisions were then transposed to all the other countries using standard conversion tables taking into account differences in national factors (e.g., overall pricing strategy, income level, competitive positioning), and communicated to all the country managers for review. Finally, the initial markdown lists were finalized as part of subsequent meetings where country managers would discuss possible modifications to the default list for their country with one or several members of the pricing committee, based on their experience and judgment. An important challenge for this first step was the lack of data, as for nearly all articles no price markdowns are performed during the season, which made it difficult to predict the response to a given price cut in the first week of sales.

The second step focused on updating the categories and markdowns after the clearance sale had started. This was performed independently by each country manager, typically on a weekly basis, in consultation with one or several members of the pricing committee. The main source of information used to make these update decisions was the weekly country clearance sales report generated at least every week for each country and each group of articles, as shown in Figure 1.



	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 1: Example of a weekly country clearance sales report (second week of the Winter 2009 clearance sales in Italy for the group "Basic Skirts").

Specifically, the country manager and pricing committee representative would typically review then the estimated time to sell the remaining stock of each category at the current price (calculated based on the average sales rate over the last three days) and compare it with the time remaining in the clearance period. When these time comparisons indicated a substantial risk of unsold inventory at the end of clearance sales, they would further markdown the category, but otherwise leave the current price unchanged. In other words, the primary heuristic pursued qualitatively consisted in minimizing the amount of inventory sold through liquidation, but keeping prices as high as possible when doing so. Additionally, markdowns of individual categories were sometimes determined so that two separate adjacent categories would merge into a single one, in particular near the end of the clearance period when the inventory remaining in one category would not be deemed sufficient to justify a separate price point in stores. An important observation however is that no formal or explicit guidelines were followed when making these decisions. Rather, country managers and members of the pricing committee would rely then on their experience and an exchange of views with their colleagues. Finally, the price update decisions for all categories in each country had to be made under significant time pressure. This was due to both the short delay between the availability of the country clearance sales report and the desirable time for communicating price update decisions to stores, and to the labor-intensive nature of this process for the members of the pricing committee.

1.3 Paper Organization

The high-level structure of the markdown optimization system we developed with Zara to improve the clearance pricing process just described matches fairly closely that of a typical Revenue Management solution as described in Chapter 1 of Talluri and van Ryzin (2004). Specifically, it involves a data collection module, a demand prediction model and a price optimization model. After a literature review and summary of our contributions in Section 2, Sections 3 and 4 discuss both the development and the formulation of this demand prediction model and price optimization model, respectively. Section 5 then describes a pilot implementation study we conducted with Zara in order to assess the impact of the resulting new clearance pricing decision system, and we offer concluding remarks in Section 6. Some of the data presented in this paper has been disguised to protect its confidentiality, and we emphasize that the views presented in this paper do not necessarily represent those of the Inditex Group.

2. Literature Review

Within the relatively vast literature on markdown optimization and dynamic pricing (see the survey by Elmaghraby and Keskinocak 2003 and the monograph by Talluri and van Ryzin 2004), this paper is characterized by its consideration of a multi-product pricing problem (see Maglaras and Meissner 2006 and Gallego and van Ryzin 1997 for seminal theoretical models, and Soon 2011 for a recent survey). Its main distinguishing feature however is a focus on the development, implementation and use of a novel operational markdown optimization model by an actual firm. That is, we seek to shed light on the frontier of clearance pricing optimization practice through a rigorous case study of a pricing system development and implementation in a challenging environment.

With a similar concern for application, Bitran, Caldentey and Mondschein (1998), Mantrala and Rao (2001) and Heching, Gallego and van Ryzin (2002) analyze historical demand and pricing data from various firms in order to generate useful insights on the likely additional revenue and qualitative pricing policy differences associated with the potential implementation of a markdown optimization model. An important additional step is taken in Smith and Achabal (1998), Smith (2009) and Valkov (2006), which describe the implementation and use of markdown optimization systems by various companies, and report some related results. However, these last three references only contain limited example data and do not discuss the methodology used for calculating the impact estimates provided. In addition, Valkov (2006) does not contain a detailed description of the pricing optimization models used as part of the implementations reported. In contrast, our paper contains complete descriptions of the clearance pricing process of the firm under study and the technical details of the pricing system developed, and provides an extensive discussion of its implementation. A second critical difference is our focus on the implementation of multi-product markdown optimization model, whereas all the existing application-oriented pricing papers just cited only discuss single-product models. That difference is significant and positions our work at the forefront of OR practice because, as stated in Talluri and van Ryzin (2004), "[...] many commercial applications of dynamic-pricing models make the simplifying assumption of unrelated products and independent demand and solve a collection of single-product models as an approximation." Finally, our paper relies on a rigorously designed controlled field experiment spanning several countries to estimate the resulting impact on both prices and revenue. This also seems significant, because we are aware of no other markdown optimization paper reporting an estimation of impact involving a control for external factors.

Because of the impact estimation methodology just mentioned, our work is also related to the set of papers discussing empirical tests in retail networks. Those tests are typically designed to estimate the effects of many possible marketing interventions such as packaging, shelf placement or price on sales, or to estimate network-wide season demand based on preliminary sales observations from a limited set of stores – see Fisher and Rajaram (2000) and Gaur and Fisher (2005) for discussions on experimental design methodologies, application examples and references. In particular, several such studies show how valuable insights on customers' price response behavior can be generated by testing empirically the impact of various price points on sales (e.g., Gaur and Fisher 2005, Sigurdsson et al. 2010). As previously mentioned however, our paper seems to be the first one among this group to describe the test of a markdown optimization system (as opposed to specific price points) as part of a controlled field experiment.

Finally, this paper is relevant to the literature investigating the operational problems that are specific to fast-fashion retailing, including the studies of assortment by Caro and Gallien (2007), of distribution by Caro and Gallien (2010), and of operations strategy by Cachon and Swinney (2009) and Caro and Martinez de Albeniz (2010). Among this group, our paper is the first to investigate the clearance pricing problem faced by a fast-fashion retailer. This seems an important endeavour, both because the fast-fashion retail model may provide some important competitive advantages (Ghemawat and Nueno 2003), and because clearance pricing is arguably more challenging in that specific context (see §1).

3. Demand Prediction Model

A key input data to a markdown optimization system are the predictions of demand for the various clearance prices considered. Given the pricing process described in §1.1, for Zara these predictions (and the ensuing price optimization to be described in §4) must be performed independently for each group of articles in each country. We discuss next the two steps we followed in order to develop these forecasts, namely the construction of a historical demand dataset (in §3.1) and the specification and fitting of a prediction model (in §3.2).[1] We then conclude this section in §3.3 with a brief discussion of the resulting model's underlying assumptions. It is worth noting that here we describe the development of the forecasting method that gave the best results in our application. A comprehensive theoretical analysis of the forecasting problem itself is beyond the scope of this paper but would be an interesting avenue for future research.

3.1 Demand Dataset Construction

Let $\mathcal{R}$ and $\mathcal{J}$ denote all the articles and stores in a given group and country respectively, and let $s \in \mathcal{S}(r)$ denote the size-color combinations available for each article $r \in \mathcal{R}$. An SKU (fully specified article) corresponds then to a pair $(r, s) \in \mathcal{R} \times \mathcal{S}(r)$, which for brevity we write rs. Let $w \in \mathbb{Z}$ denote a clearance sales period which is usually a week. By convention, we write $w = 1$ and $w = c$ to denote the first and the current period of clearance sales respectively, and we write $w < 1$ to represent the weeks during the regular selling season, i.e., prior to clearance sales. For article r, let $I_r^w := \sum_{s \in \mathcal{S}(r),\, j \in \mathcal{J}} I_{rsj}^w$ be the inventory position of article r available in the entire country at the beginning of period w, where I_{rsj}^w is the inventory position of SKU rs at store j at the beginning of that period. Let λ_r^w denote the demand rate in period w for article r which is roughly computed as the sales observed in period w divided by the number of days the article was on display. The exact computation of λ_r^w is described in Appendix A and it presents some challenges due to the presence of seasonality effects (e.g., Christmas or weekends), which cause

[1]The implemented data collection module involves standard database access queries and is not discussed here.

variations in demand that are unrelated to prices, and because of stockouts and inventory display policies, which cause demand censoring.

Historical sales and inventory data for each SKU at the store level are readily available at Zara. Using this information we constructed a dataset of historical weekly demand λ_r^w and inventory I_r^w spanning four representative product groups and the three years from 2006 to 2008. This dataset was further split into a training set (all data from 2006 and 2007 and regular selling season of 2008) and a testing set (2008 clearance sales).

3.2 Forecasting Model Specification

The main challenge when forecasting the demand rate λ_r^w for future periods is the initial lack of price sensitivity data, due to the fixed-price policy that Zara applies to most of its articles in the regular season. To overcome this, we relied on a two-stage estimation procedure, which we will describe shortly. For many different model specifications, we applied this procedure to the training dataset defined above, and computed predictions for the sales realized in the clearance period of 2008 (testing dataset). The final model selection was based on a combination of managerial judgement, the t-statistics and overall goodness-of-fit in the two-stage procedure, and most importantly, the aggregate forecasting error for the testing dataset. The final validation took place during the live pilot described in §5, when forecast accuracy was measured as part of a field implementation (see §5.2.1 for a discussion of these results). Although our forecasting model was thus derived through extensive experimentation, we only describe here the final implemented result, and refer the reader to Carboni (2009) for more details on that development process.

The process just described resulted in the selection of the following functional form

$$\begin{aligned}\lambda_r^w &= F\big(C_r, A_r^w, \lambda_r^{w-1}, I_r^w, p_r^w\big)\\ &= \exp\left(\beta_{0r} + \beta_1 \ln(C_r) + \beta_2 A_r^w + \beta_3 \ln(\lambda_r^{w-1}) + \beta_4^w \ln(\min\{1, \frac{I_r^w}{f}\}) + \beta_5^w \ln\left(\frac{p_r^w}{p_r^T}\right)\right), \quad (1)\end{aligned}$$

where the dependent variable is the demand rate of article r in period w and the regressors are:

- Purchase quantity (C_r): Size of the purchase made for article r (measured in number of units). We explain the selection of this variable by its correlation with the "fashion" component of an article. Usually, articles with low fashion content (also known as "basic") are purchased in large quantities, whereas more trendy items are deliberately purchased in small amounts. Because each article purchase covers Zara's entire store network, this variable is the same across countries.

- Age of an article (A_r^w): Number of days since article r was introduced at the stores. The selection of this variable is intuitive because sales typically peak shortly after a product is introduced then gradually decrease as weeks go by. This variable can be country-dependent.

- Previous period demand (λ_r^{w-1}): The demand rate showed some degree of autocorrelation. We considered a first-order autoregressive term because it gave a good fit and kept the model simple. We used the Dickey-Fuller test to discard the presence of a unit root.

- Broken assortment effect ($\min\{1, \frac{I_r^w}{f}\}$): In retailing it has been well-documented that the demand rate of an article usually declines when the inventory goes below a certain level. This fact is known as the *broken assortment effect* and it is especially prevalent in apparel since, when inventory is low, the remaining items are usually those that are less attractive to customers (see Smith and Achabal 1998 and Section A-3 in Fisher and Raman 2010).[2] To incorporate this in our model, we define the threshold f, which can be article-dependent, and represents the minimum on-hand inventory required for an adequate in-store presentation of the product. Though this parameter can be defined for an individual store, we calibrated it for the entire country in order to keep the regression aggregated at that level.

- Price discount ($\frac{p_r^w}{p_r^T}$): Price is obviously a key sales driver in the clearance period. The selection of this specific variable reflects however that customers are more sensitive to the relative markdown than to the absolute price cut. Indeed, a common practice at Zara and other apparel retailers is to advertise specific markdowns (expressed as negative percentages) using signs posted in various areas of the store. In addition, Zara deliberately shows the current price p_r^w together with the regular season price p_r^T on the article's price tag, so customers immediately know how big is the markdown. The selected regressor captures these features as the ratio between the two prices.

The parameters $\beta_{0r}, \beta_1, \ldots, \beta_5$ in Equation (1) are regression coefficients. In particular, β_5 represents the price elasticity, which in this model is constant and identical for all articles in the given product group for each country. This gave better results than alternative specifications with price- or article-dependent elasticities. Similarly, the multiplicative/exponential functional form in (1) provided a better fit for price response than a linear model, as is also noted by other studies in the literature (Smith *et al.* 1994). To linearize the regression model, we took logarithms in Equation (1). Note that the error term in the linearized model becomes a multiplicative error factor in

[2] Note that the literature also documents cases where large quantities of inventory correlate positively with demand. This could be incorporated by including the complement of the broken assortment effect $\max\{1, \frac{I_r^w}{f}\}$ in the regression model. However, we chose not to use it as an explanatory variable since the causality was less clear. Indeed, the inventory might be high because Zara anticipated higher demand. In contrast, Zara avoids holding incomplete assortments even if demand is low (see Caro and Gallien 2010).

the original model – this apparently innocuous transformation has important consequences when forecasting, as will be seen later.

An important observation is that when applying Equation (1) to predict demand during clearance sales, the desirable frequencies at which the regression coefficients should be updated vary for different regressors. Specifically, while $\beta_{0r}, \beta_1, \beta_2$ and β_3 may be estimated once using regular season data, it is desirable to update the estimation of β_4 and β_5 more frequently, in part because very little price response data is initially available for most articles of the current season. This motivates the two-stage estimation procedure we apply: in Stage 1 coefficients for some regressors are estimated with regular season data, while in Stage 2 the coefficient of the other explanatory variables such as price elasticity are estimated and periodically updated as clearance sales data becomes available. Note that this approach closely resembles the two-stage method developed by Smith *et al.* (1994) in the context of temporary in-season promotions.

More specifically, in Stage 1 we used regular selling season data ($w < 1$) and ran the regression

$$\ln(\lambda_r^w) = \beta_{0r} + \beta_1 \ln(C_r) + \beta_2 A_r^w + \beta_3 \ln(\lambda_r^{w-1}) + \beta_4 \ln(\min\{1, \frac{I_r^w}{f}\}) + u_r^w, \quad \forall\, r \in \mathcal{R}, w < 1, \quad (2)$$

with error term u_r^w, from where we obtained the set of parameters $\widetilde{\beta}_{0r}, \widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3$. In Stage 2, we used clearance sales data ($w \geq 1$) to compute the residuals

$$\phi_r^w = \ln(\lambda_r^w) - \widetilde{\beta}_{0r} - \widetilde{\beta}_1 \ln(C_r) - \widetilde{\beta}_2 A_r^w - \widetilde{\beta}_3 \ln(\lambda_r^{w-1}), \quad \forall r \in \mathcal{R}, w \geq 1. \quad (3)$$

As in Smith *et al.* (1994), this two-stage procedure is based on the pragmatic assumption that season-wide effects–i.e., those estimated in the first stage–are more stable than the key parameters that are updated in the second stage. Therefore, in Equation (3) the effects of the non-updated regressors from Stage 1 are removed to obtain the residuals that contain only the effects of the updated regressors. Note that β_4^t was included in the Stage 1 regression to improve the fit, but it is disregarded in Equation (3) since the broken assortment effect is particularly relevant during clearance sales, and therefore, the parameter should be updated. Then, Stage 2 is accomplished by regressing the residuals with respect to the broken assortment effect and the price markdowns:

$$\phi_r^w = \beta_4^w \ln(\min\{1, \frac{I_r^w}{f}\}) + \beta_5^w \ln\left(\frac{p_r^w}{p_r^T}\right) + \epsilon_r^w, \quad \forall r \in \mathcal{R}, w \geq 1, \quad (4)$$

where ϵ_r^w is the regression error term. This yields the estimated parameters $\widetilde{\beta}_4$ and $\widetilde{\beta}_5$.

The leverage date from the past and current seasons, the two-stage procedure aforementioned was executed twice. This is depicted in Figure 2. First, the procedure was executed pooling the data from past seasons, which generated the set of estimated parameters $\{\widetilde{\beta}_4^{P,w}, \widetilde{\beta}_5^{P,w}\}_{w \geq 1}$, where we use the superscript P to denote that it comes from past season data. Of course, from Stage 1 we

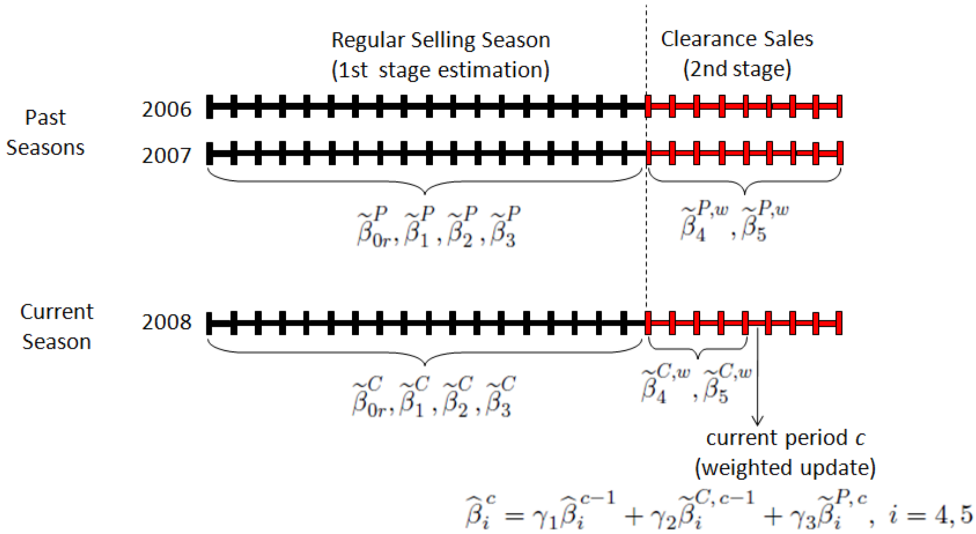


Figure 2: Two-stage estimation procedure using past (P) and current (C) season data. The current period is denoted by $w = c$.

also obtained the set of parameters $\widetilde{\beta}_{0r}^P, \widetilde{\beta}_1^P, \widetilde{\beta}_2^P, \widetilde{\beta}_3^P$, but these only played a role in the computation of the residuals. The two-stage procedure was then executed using current season data up to the most recent period available ($w = c-1$). From Stage 1 we obtained the set of estimated parameters $\widetilde{\beta}_{0r}^C, \widetilde{\beta}_1^C, \widetilde{\beta}_2^C, \widetilde{\beta}_3^C$, and from Stage 2 we obtained $\{\widetilde{\beta}_4^{C,w}, \widetilde{\beta}_5^{C,w}\}_{1\le w<c}$, where we use the superscript C to denote that it comes from current season data. Finally, the parameters $\widehat{\beta}_4^w$ and $\widehat{\beta}_5^w$ used in the forecast for the current period ($w = c$) were computed with the following recursive equation:

$$\widehat{\beta}_i^1 = \widetilde{\beta}_i^{P,1} \tag{5}$$

$$\widehat{\beta}_i^w = \gamma_1 \widehat{\beta}_i^{w-1} + \gamma_2 \widetilde{\beta}_i^{C,w-1} + \gamma_3 \widetilde{\beta}_i^{P,w} \quad w > 1, \tag{6}$$

where $i = 4, 5$ and $\gamma_1 + \gamma_2 + \gamma_3 = 1$.

The recursion (6) is a direct generalization of the one-dimensional exponential smoothing, which is a common procedure used to update parameters in adaptive systems (see Little 1966). In the first period of clearance sales, all the weight was given to the historical parameters due to the lack of current pricing information. For subsequent periods, the value of γ_2 determined the weight given to the most recent data.[3] Note that as clearance sales progressed, only the residual regression for the most recent period had to be run. With the parameters $\widehat{\beta}_4^w$ and $\widehat{\beta}_5^w$ obtained from Equations (5)-(6) we forecasted the demand rate for the current period ($w = c$) according to the formula

$$\widetilde{\lambda}_r^w \approx \exp\left(\widetilde{\beta}_{0r}^C + \widetilde{\beta}_1^C \ln(C_r) + \widetilde{\beta}_2^C A_r^w + \widetilde{\beta}_3^C \ln(\lambda_r^{w-1}) + \widehat{\beta}_4^w \ln(\min\{1, \frac{I_r^w}{f}\}) + \widehat{\beta}_5^w \ln\left(\frac{p_r^w}{p_r^T}\right)\right). \tag{7}$$

There is one caveat to the forecasting formula (7): it ignores the fact that the expectation of the regression error term (ϵ_r^w) is usually greater than one when the logarithmic transformation is

[3] In our implementation, for the second period we set $\gamma_2 = \gamma_3 = -\gamma_1 = 1$, and then for the third period onwards we used $\gamma_1 = 0.15, \gamma_2 = 0.85, \gamma_3 = 0$.

reversed. Indeed, by Jensen's inequality, $\mathbb{E}[e^{\epsilon}] \geq 1$ when $\mathbb{E}[\epsilon] = 0$, and in many cases the inequality is strict, which introduces a systematic downwards bias if ignored. In the test runs we observed a tendency to underestimate demand. Therefore, we multiplied the forecast (7) by a correction factor, and specifically used the *smearing* factor H^w introduced by Duan (1983):

$$H^w := \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \exp\left(\widetilde{\epsilon}_r^w\right), \quad w \geq 1. \tag{8}$$

where $\widetilde{\epsilon}_r^w$ corresponds to the estimated errors in Equation (4). The bias in the re-transformation can be even more significant in the presence of heteroscedasticity, see Manning and Mullahy (2001). To avoid this issue, we tested systematically for heteroscedasticity, and if present, we modified the correction factor (8) accordingly (see the Appendix D for details). In general, the correction factor took values ranging from 1.2 to 1.6.

3.3 Demand Model Discussion

We now provide a brief discussion on the assumptions underlying the demand model presented in §§3.1 and 3.2. To begin, note from Equation (31) that the estimation of the demand rate λ_r^w for article r only uses sales data of that article. Similarly, the functional form in Equation (1) only depends on the price and inventory of article r. Therefore, the model does not capture substitution or complementarity effects between products. Though there exists recent literature on how to estimate primary demand for substitutable products under the presence of stock-outs (e.g., Vulcano *et al.* 2009 and Musalem *et al.* 2010), the data and computational requirements make them more suitable for in-season promotions rather than markdown sales. Moreover, finding the dynamic pricing solution becomes very challenging even in stylized settings where the substitution structure is known (e.g., Dong *et al.* 2009). Given the additional complexity that cross-product dependencies impose, we chose not to model them explicitly. For the same reason, we did not consider directly the impact of competition (see Gallego and Hu 2007) or strategic customers (see Cachon and Swinney 2009). While some of these effects may be incorporated indirectly by restricting the range of feasible prices based on Zara's informed judgement, these observations certainly constitute limitations of our model. Note however that these limitations also apply to the legacy pricing process, which constitutes our benchmark in this particular practical setting. Furthermore, we refer the reader to §5.2.1 for a discussion of the actual forecasting accuracy performance achieved by our demand model when implemented in the field, which constitutes in our view the ultimate validation of this model and its assumptions.

4. Price Optimization Model

This section describes the development and formulation of the markdown optimization model we implemented (in §4.1) and a brief discussion of its underlying assumptions (in §4.2).

4.1 Model Development

The decision problem considered arises at least every week during the clearance sales period for every combination of country (e.g., Belgium) and product group (Woman Blazer or T-Shirts). It consists of partitioning each product group in each country into price categories (e.g., Woman Blazer from €99 to €55), and assign to each subset a clearance price (e.g. €29.99) at which all the articles in that subset will be sold. An important general constraint is that different articles with the same regular season price or the same price at some point during the clearance period are typically always sold subsequently at the same clearance price. That is, price categories aggregate but do not separate over time. As a result, instead of considering individual articles we can use the concept of a cluster (see §1.1), which are articles that were sold for the same price during the regular selling season. Also, all clearance prices for a given product group must be chosen within a discrete feasible price set (e.g. {€9.99, €14.99, €19.99, €24.99, €29.99, €34.99}).

A natural approach to model clearance pricing is dynamic programming (DP) (Talluri and van Ryzin 2004). Such formulation for our problem is given in Appendix , but as with most DPs it is subject to the curse of dimensionality (Bertsekas 1995) and therefore difficult to implement in a practical setting. This leads us to consider approximate formulations. Specifically, for the inventory dynamics we use a certainty equivalent approximation by which future sales are replaced by their expected values (see Bertsimas and Popescu 2003). Though each period the problem is solved for the entire horizon, only the actions suggested for the current period are implemented. After sales are observed, the input data including the forecast is updated and the model is solved again. Besides its advantages from an implementation standpoint, the certainty equivalent controller also has a good theoretical performance as discussed in Jasin and Kumar (2010).

In what follows, we assume that customers demanding SKU rs at store j in period w arrive according to a Poisson process with arrival rate $\alpha_{rsj}\,\widetilde{\lambda}_r^w$, where $\widetilde{\lambda}_r^w$ is given by the forecast formula (7) and α_{rsj} is the sales weight of SKU rs at store j (see Appendix A for details on the computation of this last parameter). Let $k \in \mathcal{K} := \{1, ..., K\}$ be the clearance price index. The set of clearance prices is $\{p_k, k \in \mathcal{K}\}$ which by convention is increasing, i.e., $p_0 \leq p_1 \leq p_2 \leq \ldots \leq p_K$, and p_0 is the final unit salvage value or liquidation price. Let $n \in \mathcal{N} := \{1, ..., N\}$ denote clusters and let $\mathcal{R}_n$ be the set of products in cluster n, so the entire product group is $\mathcal{R} = \bigcup_{n\in\mathcal{N}} \mathcal{R}_n$. Clusters are ordered in reverse order of regular season prices. That is, $\mathcal{R}_1$ contains the most expensive items, and $\mathcal{R}_N$ the cheapest. Let $\mathcal{W} := \{w|\, c \leq w < W\}$ be the remaining pricing periods, where W is the last period when all the remaining inventory will be liquidated at price p_0. As before, $w = c - 1$ represents the most recent period for which there is data available.

For the decision variables, $x^w_{nk} \in \{0, 1\}$ indicates whether cluster n should be sold at clearance price p_k or lower during pricing period $w \in \mathcal{W}$, with $x^w_{n0} = 0$ for all $(n, w) \in \mathcal{N} \times \mathcal{W}$. The prices from the previous period are given by x^{c-1}_{nk}, which is input data in the current period. In particular, recall that $w = 1$ corresponds to the first period, therefore one can use x^0_{nk} to impose a minimum markdown at the beginning of clearance sales. The auxiliary variable $y^w_{nk} \in \{0, 1\}$ indicates whether cluster n should be sold at clearance price p_k during period w; $\lambda^w_{nk} \geq 0$ represents the expected sales for cluster n in period $w \in \mathcal{W}$ if sold at price p_k; $z^w_k \in \{0, 1\}$ indicates whether clearance price p_k is used for any cluster ($z^w_k = 1$) or not at all ($z^w_k = 0$) during period $w \in \mathcal{W}$; and $I^w_n := \sum_{r\in\mathcal{R}_n,\, s\in\mathcal{S}(r),\, j\in\mathcal{J}} I^w_{rsj}$ represents the inventory of cluster n available at the beginning of period w. Note that the inventory level for the current period I^c_n is input data. The uncertainty in the problem is given by $Sales^w_{rsj}$, the sales of SKU rs at store j in period w, which is a random variable for $w \in \mathcal{W}$ that depends on price and the inventory position. Finally, let Q^w and N^w be input parameters that represent the minimum inventory per category and the maximum number of distinct prices in period w respectively. We formulate the pricing optimization model as follows:

$$\max \quad \sum_{w\in\mathcal{W}, n\in\mathcal{N},\, k\in\mathcal{K}} p_k \lambda^w_{nk} + \sum_{n\in\mathcal{N}} p_0 I^W_n \tag{9}$$

$$s.t.$$

$$\lambda^w_{nk} = y^w_{nk} \sum_{r\in\mathcal{R}_n} \sum_{s\in\mathcal{S}(r)} \sum_{j\in\mathcal{J}} \mathbb{E}\big[Sales^w_{rsj} \big|\, p_k, I^w_{rsj}\big], \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K} \tag{10}$$

$$y^w_{nk} = x^w_{nk} - x^w_{nk-1}, \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \tag{11}$$

$$x^w_{nk-1} \leq x^w_{nk}, \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \tag{12}$$

$$x^w_{nk} \leq x^w_{n+1k}, \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \tag{13}$$

$$\sum_{w\in\mathcal{W}, k\in\mathcal{K}} x^w_{nk} = \sum_{w\in\mathcal{W}, k\in\mathcal{K}} x^w_{n+1k}, \quad \forall\, n \text{ such that } \sum_{k\in\mathcal{K}} x^{c-1}_{nk} = \sum_{k\in\mathcal{K}} x^{c-1}_{n+1k}, \tag{14}$$

$$x^{w-1}_{nk} \leq x^w_{nk}, \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \tag{15}$$

$$I^{w+1}_n = I^w_n - \sum_{k\in\mathcal{K}} \lambda^w_{nk}, \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N} \tag{16}$$

$$y^w_{nk} \leq z^w_k, \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \tag{17}$$

$$\sum_{k\in\mathcal{K}} z^w_k \leq N^w, \quad \forall\, w \in \mathcal{W}, \tag{18}$$

$$\sum_{n\in\mathcal{N}} I^w_n y^w_{nk} \geq Q^w z^w_k, \quad \forall\, w \in \mathcal{W}, k \in \mathcal{K}, \tag{19}$$

$$I^w_n, \lambda^w_{nk} \geq 0,\ x^w_{nk} \in \{0, 1\}, z^w_k \in [0, 1], \quad \forall\, w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}. \tag{20}$$

The objective (9) is the sum of the revenue from all clusters up until the last week, and the revenue from liquidation at price p_0 after clearance sales. Constraint (10), which we soon discuss further, links predicted sales volume with prices, i.e., it represents the underlying price response model. Constraints (11) and (12) follow from the definition of the x^w_{nk} and y^w_{nk} variables. Constraint (13) ensures that the initial ordering of clusters by prices is maintained throughout the clearance period. Constraints (14) make sure that clusters that were priced together, remain together. Constraint (15) ensures that the clearance sales price for any cluster decreases over time. Constraint (16) implements the inventory dynamics as a function of the pricing decisions. Note that inventory is aggregated by cluster since pricing decisions are made at that level. Constraint (17) implements the definition of z^w_k. Constraint (18) ensures that the number of distinct price categories in period w does not exceeds N^w. Constraint (19) ensures that the amount of inventory available at the beginning of period w is at least Q^w for each category.[4] Finally, constraint (20) imposes the non-negative or binary requirements for the decision variables – observe that variables y^w_{nk} and z^w_k do not need to be defined as binary provided that x^w_{nk} is. Also, the non-negativity of I^{w+1}_n together with the inventory balance constraint (16) ensure that the expected sales λ^w_{nk} never exceed the available inventory I^w_n.

The formulation above is still hard to solve in practice due to the non-linearity of constraint (10). Therefore, we linearize these equations. This requires some attention since the random variables $Sales^w_{rsj}$ depend on inventory in two ways. First, sales are bounded by the inventory available (as in the usual newsvendor), and second, the demand rate of the Poisson process is affected by inventory level through the broken assortment effect (see §3). Moreover, the latter takes place at the article level, while our problem is formulated at the cluster level. Taking this into account, our approximation is based on the following observation: Let $\widehat{I}^w_{rsj}$ be an upper bound for the inventory level of SKU rs at store j in period w. Then, we have that

$$
\begin{aligned}
\lambda^w_{rk} \quad &:= \quad \sum_{s\in\mathcal{S}(r)}\sum_{j\in\mathcal{J}} \mathbb{E}\big[Sales^w_{rsj}\,\big|\,p_k, I^w_{rsj}\big] \\
&= \quad \widetilde{\lambda}^w_r(p_k, I^w_r) \sum_{s\in\mathcal{S}(r)}\sum_{j\in\mathcal{J}} \alpha_{rsj}\,\mathbb{E}\big[\tau^w_{rsj}\,\big|\,p_k, I^w_{rsj}\big] \\
&\leq \quad \widetilde{\lambda}^w_r(p_k, I^w_r) \sum_{s\in\mathcal{S}(r)}\sum_{j\in\mathcal{J}} \alpha_{rsj}\,\mathbb{E}\big[\tau^w_{rsj}\,\big|\,p_k, \widehat{I}^w_{rsj}\big] \\
&= \quad \frac{\big(\min\{1, \frac{I^w_r}{f}\}\big)^{\widetilde{\beta}_4}}{\big(\min\{1, \frac{\widehat{I}^w_r}{f}\}\big)^{\widetilde{\beta}_4}} \sum_{s\in\mathcal{S}(r)}\sum_{j\in\mathcal{J}} \mathbb{E}\big[Sales^w_{rsj}\,\big|\,p_k, \widehat{I}^w_{rsj}\big] \\
&= \quad \big(\min\{1, \frac{I^w_r}{f}\}\big)^{\widetilde{\beta}_4} \frac{E^w_r(p_k)}{\big(\min\{1, \frac{\widehat{I}^w_r}{f}\}\big)^{\widetilde{\beta}_4}},
\end{aligned}
\tag{21}
$$

[4] This constraint was usually dropped in the final periods.

where $E_r^w(p_k) := \sum_{s\in\mathcal{S}(r),\, j\in\mathcal{J}} \mathbb{E}\big[Sales_{rsj}^w \,\big|\, p_k, \widehat{I}_{rsj}^w\big]$ and $\widetilde{\lambda}_r^w(p_k, I_r^w)$ is the arrival rate given by the forecast (7) evaluated at $p_r^w = p_k$. The first equality is the definition of λ_{rk}^w. The second equality follows from the Poisson process where τ_{rsj}^w is the stopping time until when SKU rs is on display at store j in period w (the same property is used in Caro and Gallien 2010). The third step follows from $I_{rsj}^w \le \widehat{I}_{rsj}^w$ for all stores and SKUs. The fourth step uses the definition of $\widetilde{\lambda}_r^w(p_k, I_r^w)$ and again the Poisson property, and the last step is just the definition of $E_r^w(p_k)$.

We linearize the broken assortment term in equation (21) by fitting the linear form used in Smith and Achabal (1998). While details are given in Appendix C, the resulting expression is

$$\big(\min\{1, \frac{I_r^w}{f}\}\big)^{\widetilde{\beta}_4} \frac{E_r^w(p_k)}{\big(\min\{1, \frac{\widehat{I}_r^w}{f}\}\big)^{\widetilde{\beta}_4}} \approx \min\left\{1, 1-\mu+\mu\frac{I_r^w}{f}\right\} \frac{E_r^w(p_k)}{\big(\min\{1, \frac{\widehat{I}_r^w}{f}\}\big)^{\widetilde{\beta}_4}}, \tag{22}$$

where $\mu := \dfrac{3\rho^2+9\rho}{2\rho^2+6\rho+4}$ and $\rho := \widetilde{\beta}_4$. We can use Equation (22) to write the following (approximate) linear constraints for λ_{rk}^w:

$$\begin{aligned}
\lambda_{rk}^w \;&\le\; E_r^w(p_k) y_{nk}^w, && \forall\, w\in\mathcal{W}\backslash\{c\}, r\in\mathcal{R}_n, n\in\mathcal{N}, k\in\mathcal{K}, && (23)\\
\lambda_{rk}^w \;&\le\; \Big(1-\mu+\mu\frac{I_r^w}{f}\Big) F_r^w(p_k), && \forall\, w\in\mathcal{W}\backslash\{c\}, r\in\mathcal{R}, k\in\mathcal{K}, && (24)
\end{aligned}$$

where $F_r^w(p_k) := \dfrac{E_r^w(p_k)}{\big(\min\{1, \frac{\widehat{I}_r^w}{f}\}\big)^{\widetilde{\beta}_4}}$. Constraint (23) is the relevant bound when $1-\mu+\mu\frac{I_r^w}{f} \ge 1$ or $y_{nk}^w = 0$, and constraint (24) captures the complementary case. Note that in the first case the denominator in the right hand side of equation (22) is equal to one (here we are using again the fact that $I_r^w \le \widehat{I}_r^w$). Note that constraints (23) and (24) do not need to be defined for $w = 1$ since the inventory levels for the current period are known so $E_r^w(p_k)$ and $F_r^w(p_k)$ can be computed exactly.

Constraints (23) and (24) linearize constraint (10) in our pricing optimization model. Note that constraints (23) and (24) are defined at the article level r while the constraint they replace is defined at cluster level n. Therefore, a complete formulation would also have to include the identity $\lambda_{nk}^w := \sum_{r\in\mathcal{R}_n} \lambda_{rk}^w$ and an inventory balance equation at the article level. An alternative is to aggregate constraints (23) and (24) across articles of the same cluster. This reduces the size of the model and is also consistent with the fact that pricing decisions are made at the cluster level. In Appendix C we provide the details on how to aggregate these constraints and how to compute the parameters $E_r^w(p_k)$ and $F_r^w(p_k)$ for $w > 1$.

4.2 Optimization Model Discussion

The premise in the pricing optimization model described in §4.1 is that the firm's objective is to maximize the total revenue across the entire duration of clearance sales. This seems natural since at the time of clearance sales, inventory is a sunk cost. However, in practice many retailers make pricing decision with the *de facto* objective of liquidating inventory in order to open up space for the upcoming season. The consequences of this is discussed later in §5, but for the model formulation the question is how to account for the fact that markdown items hold up valuable retail display area. One approach is to include an opportunity cost that captures the relative value between old and new items (see Araman and Caldentey 2009), but estimating this parameter can be a difficult task. An alternative approach preferred by Zara is to deliberately reduce the number of posted prices as clearance sales goes by. That is the purpose of constraint (18), and the rationale is that having less prices allows the store manager to consolidate the inventory on display so it uses less store space. In some situations, it is imperious that most of the inventory is sold, e.g., when there is little opportunity to salvage stock and disposing it would imply a cost. For those cases, we included an optional constraint that explicitly limits the amount of inventory left over.

The Poisson assumption is another central premise in the model formulation, which is needed for tractability reasons as in many revenue management problems. We specifically take advantage of it to approximate the price response where the demand rate also depends on the inventory level (see Equation (21)). We believe this feature constitutes a novel feature of the model since most of the literature on inventory-dependent demand is for the single-period newsvendor (e.g., Dana and Petruzzi 2001). Finally, it must be noted that we rely on the certainty equivalent approximation to solve a math program instead of a DP. This is driven by the need of efficient run times since there is a small time window to make the pricing decisions.

However restrictive the assumptions and approximations discussed here may seem, we note in closing that they are supported and partially justified from a practical standpoint by the implementation results to be presented next.

5. Pilot Implementation Study

A working prototype of the entire new pricing system was completed in 2008. The forecast described in §3 was implemented in Java and the optimization model from §4 was coded in AMPL and solved with CPLEX. Pulling the data from Zara's databases to feed the forecast was the most time-consuming task and was usually done over the weekend. Solving the optimization problem was done overnight and the usual instance for a group in a given country would have up to 12 prices, 15 clusters and 8 periods. Therefore, the number of binary variable in the model rarely exceeded 1,500, and each instance was typically solved in a few minutes.

We tested the resulting model-based pricing process in a controlled field experiment that took place from January to March 2009, corresponding to clearance sales of the 2008 Fall-Winter season. The objective of the live pilot was threefold: First, establish the applicability of the model in the field. Second, refine the solution based on user feedback. And third, quantify the model's specific impact on markdown decisions, which is the main issue discussed in the present section. Specifically, the details of our methodology are given in §5.1, where we present the experimental design (in §5.1.1) and the performance metrics used (in §5.1.2). The results of the pilot are then reported in §5.2, specifically the observed forecast accuracy (in §5.2.1), pricing behavior (in §5.2.2) and financial impact (in §5.2.3).

5.1 Methodology

5.1.1 Experimental Design

Zara's assortment in the Women section consists of 20 product groups, not including accessories. For the pilot, we divided the assortment in two large sets of product groups: groups 1-12, that include relatively more classic designs targeted to women in their late twenties and above; and groups 13-20, that include more fashionable items targeted to younger women. Articles in the latter groups tend to have lower prices than those in the former. Zara gave our team the entire countries of Belgium (BEL) and Ireland (IRL) to run the experiment. We carefully designed the pilot in the following way: In Belgium, the optimization model was used to suggest prices for groups 1-12, whereas the manual legacy process was used to price groups 13-20. Conversely, in Ireland, the manual process was used for groups 1-12, and the model suggested prices for groups 13-20. Finally, in the rest of Western Europe (RWE), the manual process was used to price all the product groups.[5]

The groups subject to the model, i.e., groups 1-12 in Belgium and 13-20 in Ireland, represented the treatment set, while the rest served as the control set. On the one hand, this partition allowed to rule out country and store specific factors, i.e., whether the conditions in a given country or store were intrinsically more (or less) favorable for clearance sales. On the other hand, the fact that the treatment and control groups were inverted between Belgium and Ireland made sure that the particular selection of the treatment set was not driving the results. Finally, any base difference between groups 1-12 and 13-20 was captured by using the rest of Western Europe as a reference point, against which Belgium and Ireland were compared. This design was also constructed to minimize potential demand substitution across test and control groups. Specifically, Zara managers felt that demand substitutions would be plausible between any two articles within either groups 1-12 or groups 13-20, but unlikely across those two sets of groups because they appealed to fairly different customer types. These considerations ruled out designs involving (say) even-numbered groups for the intervention and odd-numbered group for the control. Finally, the inventory available at the

[5]For this experiment, RWE consisted of Spain, Portugal, France, Italy, Austria, Holland, and the United Kingdom.

beginning of clearance sales, while not explicitly controlled for, was fairly uniform across groups and countries (i.e., the amount of initial inventory relative to past season demand was about the same everywhere). This resulted from the use by Zara of the same pre-clearance inventory positioning process worldwide (Verdugo 2010).

Note that for the treatment groups, the model was used to suggest markdowns, but the manual prices were still generated in parallel. In other words, at each price revision, two lists of suggested prices were available for the treatment set: one list from the model, and the other from the legacy pricing process (followed as usual without knowledge of the model pricing recommendations). The actual decision to follow the model prices required the approval from Zara's pricing committee together with the country managers of Belgium and Ireland. Though in an ideal experiment we would have liked the model to dictate the prices for the treatment groups, this was not allowed given that there was still uncertainty, and even some skepticism, on whether the model would perform well. Moreover, the model had always been envisioned as a support tool rather than an automated decision maker. Hence, letting the pricing committee and country manager have the final say, not only was pivotal for the experiment to happen, but also was closer to the actual use intended for the model. We did keep track of the adherence to the model's suggestion and report this later in §5.2.

5.1.2 Metrics

The primary financial metric used by Zara to evaluate its clearance sales performance is the *realized income* Y defined as

$$Y := \frac{\text{clearance period income } + \text{ liquidation income}}{\text{value of clearance inventory at regular season prices}}, \tag{25}$$

which can be calculated for a store, country or the entire chain. If only the inventory sold until period w is considered in Equation (25), then the metric is denoted Y^w. The realized income measures the ratio of the actual revenue from clearance sales to the maximum revenue achievable by selling the inventory at regular season prices. A higher realized income is better as it reflects more revenue generated out of a given amount of initial stock, valued at the prices prior to markdowns. Note that, if the numerator and denominator in Equation (25) are divided by the inventory expressed in units, then the metric Y can be seen as the ratio of the average price in clearance sales to the average price in the regular season. Therefore, $1 - Y$ is the average markdown, or the average *price cut*, as it is known internally at Zara.

As discussed before, even though the objective of clearance sales is to maximize revenue, an indirect goal is also to liquidate stock. Therefore, a secondary metric that is very relevant to Zara is the percentage sold X^w, internally known as the *fraction sold*, which is defined as

$$X^w := \frac{\text{units sold up to period } w}{\text{initial clearance sales inventory (in units)}}. \tag{26}$$

Since the fraction sold can always be improved by introducing more aggressive markdowns, it is complemented by the average price P^w in period w, defined as

$$P^w := \frac{\text{period } w \text{ initial inventory valued at period } w \text{ prices}}{\text{period } w \text{ initial inventory (in units)}}. \tag{27}$$

Note that P^w is computed at the beginning of period w and reflects the pricing decisions, while X^w and Y^w are trailing metrics that are computed at the end of period w once sales have been observed.

Historically, the fraction sold X^w, the average price P^w and the (trailing) average markdown $1 - Y^w$ had been the metrics most closely monitored by managers at Zara, usually comparing them across different countries. For the pilot, we were also interested in the trajectory followed by these metrics. However, for our purpose of comparing two pricing methods, the key metric was the realized income Y at the end of clearance sales. The live experiment was specifically designed to measure the model's impact on that metric by using a difference-in-differences statistical procedure – see Stock and Watson (2003) for a textbook discussion and see Imbens and Wooldridge (2009) for a survey on the use of difference-in-differences for program evaluation. Indeed, for each store in Belgium we computed the aggregate realized income for groups 1-12 and 13-20 (denoted $Y_{\text{1-12}}$ and $Y_{\text{13-20}}$, respectively). Then, we took first differences and computed the average across all stores (denoted $\overline{Y}^{\text{BEL}}_{\text{1-12}} - \overline{Y}^{\text{BEL}}_{\text{13-20}}$). We did the same for the stores in Ireland and in RWE. Finally, we computed the (second) difference between the averages in Belgium and RWE to obtain

$$\Delta^{\text{BEL}} := \left(\overline{Y}^{\text{BEL}}_{\text{1-12}} - \overline{Y}^{\text{BEL}}_{\text{13-20}}\right) - \left(\overline{Y}^{\text{RWE}}_{\text{1-12}} - \overline{Y}^{\text{RWE}}_{\text{13-20}}\right), \tag{28}$$

and we did the same between Ireland and RWE to obtain Δ^{IRL}.

As mentioned in §5.1, the first difference in Equation (28) removed any country or store specific factors, while the second difference removed any intrinsic performance differential between groups 1-12 and 13-20. Averaging across stores removed any random noise. Moreover, we used the store sample to perform a t-test comparing the means $\overline{Y}^{\text{BEL}}_{\text{1-12}} - \overline{Y}^{\text{BEL}}_{\text{13-20}}$ and $\overline{Y}^{\text{RWE}}_{\text{1-12}} - \overline{Y}^{\text{RWE}}_{\text{13-20}}$, which determined whether the expression in Equation (28) was significantly different from zero. Similarly, we used the Mann-Whitney test to determine whether the medians of the first differences across stores in Belgium and RWE were significantly apart. The same calculations were repeated for Ireland and RWE.

5.2 Results

5.2.1 Forecast Accuracy

We begin this section by looking at the quality of the forecast. In general, the forecast was computed overnight after the weekend and the pricing decisions were made the following day. For each price decision implemented in Belgium and Ireland, we computed the forecast error, i.e., the difference between the actual and predicted sales. Then, we computed the mean absolute deviation (MAD) for each period (each period corresponds approximately to a week) and each country.[6] The results are shown in Table 1, where the last column provides the sales-weighted average across all periods. Overall, we found the forecast accuracy to be reasonable and within range of other studies. For instance, Fisher and Vaidyanathan (2009) report an out-of-sample MAD of 25.8% at the chain-SKU level over a 6-month period. Our forecast had a similar MAD at a slightly more aggregate product level (price categories) but for a shorter time window and a single country rather than the entire chain.[7] Experience at other retailers confirms that it is very difficult to get better than 25% MAD for weekly sales of specific retail products, especially when prices are changing each week (Smith 2011).

Country	Clearance Sales Period (Week)									Sales-weighted
	100	101	102	103	104	105	106	107	108	Average
Belgium	24.7%	28.7%	19.7%	17.8%	28.4%	21.3%	22.7%	23.3%	26.8%	23.8%
Ireland	20.1%	19.4%	27.1%	22.9%	27.0%	23.4%	27.6%	31.9%	41.7%	24.1%

Table 1: Forecast accuracy measured in MAD per period at the country level.

Table 1 shows that the quality of the forecast was quite consistent across time. Only in the last two periods in Ireland the MAD was above 30%. We believe that this was in part due to differences between the actual stock at the stores and the inventory levels in the database. In fact, record inaccuracy is a well-known issue for retailers (DeHoratius *et al.* 2008), and it becomes more prevalent as stock is depleted towards the end of clearance sales. Although this seems a plausible explanation, it did not have a major impact since the last periods accounted for a small fraction of total clearance sales.

For validation purposes, we checked the quality of the forecast aggregated across all groups. We expected the accuracy to improve significantly, which was indeed the case. Figure 3 shows the weekly aggregate forecast versus actual sales for Belgium and Ireland (we report it as percentages of

[6]Here we used the definition of MAD given in Fisher and Raman (2010), p. 65, which is equivalent to the sales-weighted mean absolute percentage error, or MAPE, commonly used by practitioners.

[7]It should also be noted that the forecast in Fisher and Vaidyanathan (2009) is for an assortment problem instead of markdown pricing.

total sales since the absolute values cannot be disclosed). In both countries, the aggregate forecast closely followed sales, with a small underestimation mostly in the first periods. Part of this could be due to systematic errors of the kind that motivated the correction factor in Equation (8). However, we believe it is also explained by the fact that the forecast in the initial periods relied more on the the elasticity computed from historical data, see Equation (5)-(6). As clearance sales progressed, the elasticity was updated with current data so the later periods in Figure 3 could reflect some degree of learning by the model over time.

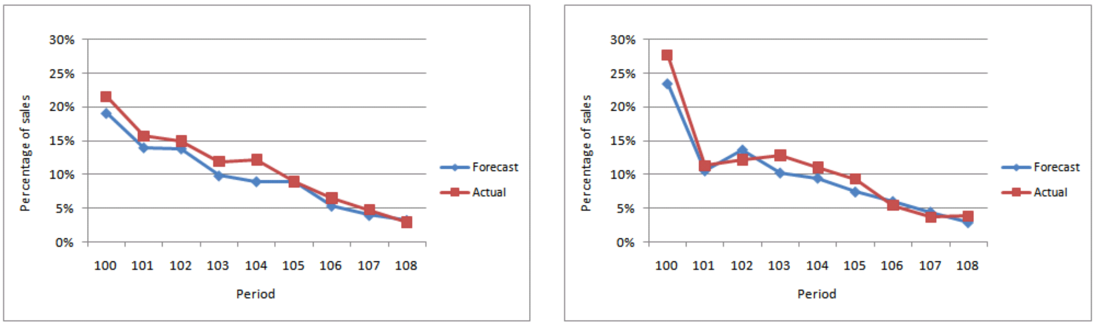


Figure 3: Weekly aggregate forecast versus actual sales for Belgium (left) and Ireland (right).

5.2.2 Pricing Behavior

As described in §5.1.1, the model prices were evaluated by the pricing committee and the country managers. Whenever the model and the manual prices differed, there would be a discussion, and occasionally the team chose to implement the manual ones (recall that the legacy process was still performed in parallel as a back-up for the treatment groups). Figure 4 shows the adherence to the model prices. A solid dot indicates a group-period where the model and the manual prices differed, and a shaded box indicates that the latter was followed. From this figure it is clear that the two pricing methods differed quite often. However, in many cases the discrepancy was only in a couple of price categories, and frequently the model's suggestion was the price above or below in the discrete set $\mathcal{K}$.

Overall, the adherence to the model prices was very high, with a few exceptions in Ireland. In the cases of G16-P100 and G20-P102 (we use "G" and "P" to abbreviate Group and Period, respectively), the choice of the manual prices were judgement calls following the perception that the model prices were too conservative. In the other cases, the manual prices were selected due to a lack of confidence in the model's suggestion, mostly induced by a large forecast error in the previous period. The case of G20 deserves particular attention. In period 102 the manual prices were implemented, and for those prices the forecast overestimated sales by 35%. This led to believe

Country – Group		100	101	102	103	104	105	106	107	108
		Period								
Belgium	1	●	●		●	●	●	●		
	2	●	●	●		●		●	●	
	3	●		●			●	●		
	4		●			●		●	●	
	5	●	●	●			●	●	●	
	6	●	●	●	●					
	7	●	●	●	●		●	●	●	
	8	●	●	●	●	●	●		●	
	9	●	●	●			●	●	●	
	10	●	●	●			●		●	
	11	●	●		●	●	●	●		
	12	●	●		●	●	●		●	
Ireland	13	●	●	●	●			●		●
	14	●		●	●		●	●	●	●
	15	●		●	●	●	●	●	●	●
	16	●	●	●	●		●	●	●	●
	17	●		●	●	●	●	●	●	●
	18	●		●	●	●	●	●		●
	19	●		●	●	●	●	●	●	●
	20	●	●	●	●		●			

Figure 4: Adherence to model prices. Shaded cells indicate when they were not followed.

that the model had suggested higher prices for P102 because it was overestimating demand. The latter created some skepticism towards the model prices for G20, and therefore, in the following weeks when the price suggestions differed, the manual ones were implemented.

The average price P^w and the fraction sold X^w in each period w of clearance sales for the treatment and control sets are shown in Figure 5 and Table 2, respectively. The trajectories in Figure 5 show that in Belgium the model discounted prices a bit more aggressively in the first couple of weeks but then came fairly close to the manual prices after that. In Ireland, however, the model markdowns remained less aggressive than the manual markdowns for the entire sales period. In terms of inventory, Table 2 shows that groups 1-12 in Belgium ended with a higher fraction sold that the same groups in Ireland, while the opposite happened with groups 13-20. Recall that groups 1-12 had articles with higher initial prices than groups 13-20. Hence, compared to the manual process, in the experiment the model chose to markdown sooner the more expensive items to increase sales volume, and it chose to collect more revenue out the cheaper items at the expense of selling lower quantities. Put simply, for the expensive items the model's strategy called for volume, whereas for the cheaper articles it called for price. The impact of these decisions in terms of revenue is described in the next section.

5.2.3 Financial Impact Assessment

The overall impact of the model is based on the on the difference-in-differences metric $\Delta^q, q =$ BEL, IRL, defined in Equation (28). The results are summarized in Table 3. In particular, the

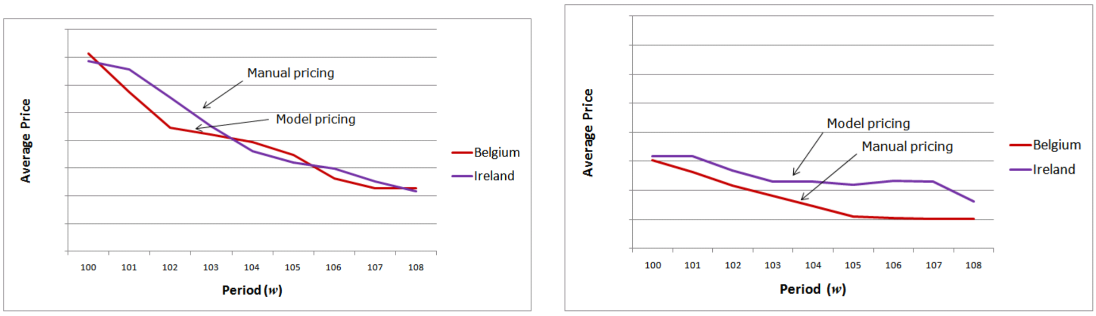


Figure 5: Average price (P^w) across periods for groups 1-12 (left) and groups 13-20 (right) in Belgium and Ireland.

Groups – Country		Period (w)									Final
		100	101	102	103	104	105	106	107	108	
Groups 1-12	Belgium	19.1%	34.0%	51.4%	63.1%	71.0%	75.2%	79.6%	82.7%	84.4%	85.1%
	Ireland	20.5%	28.6%	38.0%	47.4%	59.3%	67.0%	72.5%	76.7%	79.8%	82.2%
Groups 13-20	Belgium	19.0%	32.3%	43.5%	53.3%	65.5%	75.2%	81.5%	86.3%	89.2%	89.9%
	Ireland	23.4%	33.1%	43.1%	53.9%	60.7%	68.1%	71.6%	73.7%	76.8%	80.3%

Table 2: Fraction sold (X^w) across periods for groups 1-12 (intervention in Belgium, control in Ireland) and 13-20 (control in Belgium, intervention in Ireland). The last column reports the fraction sold two weeks after the last markdown

first row corresponds to the figures for the live pilot in 2008. The mean and median of the first difference $Y_{1\text{-}12} - Y_{13\text{-}20}$ across stores in Belgium, Ireland, and RWE are reported in columns two, three, and four, respectively. The second difference between the averages observed in Belgium (Ireland) and in RWE is reported in column five (six) and corresponds to the empirical value of Δ^{BEL} (Δ^{IRL}). We also report the difference between the medians. We provide the t-statistics to assess the significance of the difference between the means, and for the medians we provide a z-statistic that corresponds to the usual Normal approximation of the Mann-Whitney U-statistic. The significance of the statistics is reported conservatively by considering the two-tailed versions of the tests.

The actual average first difference observed during the pilot was 0.5 percentage points (pp) in Belgium and -4.8pp in Ireland. However, one of the two sets of product groups could have been intrinsically harder to sell during clearance sales than the other, and indeed the numbers for RWE show that groups 13-20 had a higher realized income than groups 1-12 by about 2.2pp, which provided a baseline value for these average differences. The estimated impact of the model on the realized income was therefore an increase of 2.7pp in Belgium and an increase of 2.6pp in Ireland. These results were not driven by outliers because the mean and median changes were consistently alike and all had the same sign. The t-test comparing the means showed that this estimation of impact on Y was significant at the 5% level in Ireland, and at the 0.04% level in Belgium. This observation was confirmed by the Mann-Whitney test comparing the medians. For completeness, we performed the t-test with the Welch correction for unequal variances. The significance level was the same in Belgium but actually improved in Ireland, which showed that our results were robust and most likely conservative.

Year		1st Difference: $Y_{1\text{-}12} - Y_{13\text{-}20}$			2nd Difference	
		BEL	IRL	RWE	BEL − RWE	IRL − RWE
2008	mean (median)	0.5(0.4)	−4.8(−4.6)	−2.2(−2.1)	**$2.7^{****}(2.5^{****})$**	**$-2.6^{*}(-2.5^{**})$**
	t-stat (z-stat)				3.59(3.73)	2.12(2.93)
2007	mean (median)	−0.2(0.2)	−0.8(−0.5)	1.5(1.7)	$-1.7^{*}(-1.5^{*})$	−2.3(−2.2)
	t-stat (z-stat)				2.37(2.17)	1.93(1.94)
2006	mean (median)	−3.2(−2.8)	−3.7(−3.5)	−2.4(−2.7)	−0.8(−0.1)	−1.2(−0.8)
	t-stat (z-stat)				0.97(0.89)	0.82(1.05)

Note: Statistical significance from two-tailed test: $^{*}p < 5\%$, $^{**}p < 1\%$, $^{***}p < 0.2\%$, $^{****}p < 0.04\%$.

Table 3: Model impact assessment using difference-in-differences, where the first difference is between groups 1-12 (intervention in Belgium, control in Ireland) and 13-20 (intervention in Ireland, control in Belgium), and the second difference is between Belgium/Ireland and RWE. Figures are percentage points.

In order to validate our methodology, in rows three and four of Table 3 we report the difference-in-differences calculation for 2006 and 2007. Since in those years no pilot took place, we expected any difference between Belgium (Ireland) and RWE to be statistically insignificant and closer to zero. In fact, the values for 2006 and 2007 in the last two columns of Table 3 are consistently smaller in absolute terms than their equivalent in 2008. None of them is statistically significant at the the 5% level, except for the 2007 difference in Belgium. We looked at this result carefully and we believe that it is irrelevant. Indeed, it has a negative sign, which if anything, would indicate that there is even more merit to the positive result observed in 2008, and the latter was obtained with a significance level that is two orders of magnitude better than the value in 2007. Moreover, if the t-test is performed with the Welch correction, the result in 2008 remains unchanged, while in 2007 it becomes insignificant at the 5% level. We took this analysis as evidence that supported the overall methodology.

The remarks drawn from Table 3 rely on the significance of the results given by the statistical tests. In that sense, the impact of the model in Belgium could seem stronger than in Ireland. However, recall that the model recommendations were not implemented as closely for group 20 in Ireland (c.f. §5.2.2). Hence, it is important to observe that when group 20 was excluded from the calculations in Ireland, see Table 4, the impact estimation on the realized income increased from 2.6pp to 3.1pp, and more importantly, the significance level increased from 5% to 1%, and from 1% to 0.2%, for the difference of the means and medians, respectively. Averaging the result for Belgium in Table 3 and for Ireland in Table 4, we can therefore conclude that the price optimization model increased the realized income Y by approximately 2.9pp during the pilot experiment.

Year		1st Diff: $Y_{1\text{-}12} - Y_{13\text{-}19}$		2nd Difference
		IRL	RWE	IRL − RWE
2008	mean (median)	$-6.9(-6.2)$	$-3.8(-3.8)$	$\mathbf{-3.1^{**}(-2.5^{***})}$
	t-stat (z-stat)			2.65(3.16)

Note: Statistical significance from two-tailed test: $^{*}p < 5\%$, $^{**}p < 1\%$, $^{***}p < 0.2\%$, $^{****}p < 0.04\%$.

Table 4: Model impact assessment in Ireland without G20, i.e., the first difference is only between groups 1-12 and 13-19. Figures are percentage points.

A rough historic average for the realized income at Zara during clearance sales is 50%. Therefore, an additional 2.9pp in Y means a 5.8% increase in revenues. For a monetary value of this impact estimate, consider 2006 when Zara reported $7,194M in revenues. Following Ghemawat and Nueno (2003), we assume that 17.5% of sales were generated at markdown prices (see §1), which would result in a clearance sales income of $1,259M (since Zara avoids markdowns during the regular season we assume that discounted sales prior to the clearance period are negligible). Increasing the realized income by 2.9pp therefore corresponds to an increase of annual clearance revenue by about $73M (i.e., 5.8% of $1,259M). In 2007 this would mean $83M in additional sales, and $90M in 2008.[8] Given that the use of the model does not have a major impact on Zara's costs, the increase in revenues due to this new pricing process is likely to translate directly into additional net profits. To the best of our knowledge, Smith and Achabal (1998) report the only other observed relative financial impact of a markdown optimization system implementation in the literature. Specifically, for the most successful of three implementations described in that paper it was observed that the realized income increased 4% over the previous year, which is consistent with but lower than our increase estimate of 5.8% for the present application. This difference in impact could be driven by the relative features of the systems implemented, but also the relative performance of the legacy markdown policies they replaced as well as the estimation methodology (the impact estimations in Smith and Achabal 1998 do not involve controls for external factors).

6. Pilot Aftermath and Conclusion

Following the pilot test, Zara's IT group completed a distributed software application allowing managers of all countries to use the model continuously through clearances sales for all product

[8]The financial impact estimations provided here assume a 1.3 $/Euro exchange rate. They were performed independently by the paper authors and do not engage the responsibility of the Inditex group.

groups (see Appendix E for snapshots of the user interface). This application provides the complete pricing recommendations and corresponding sales and revenue predictions for all clusters and all remaining weeks, and it also enables what-if scenario analysis relative to a baseline of specified pricing decisions, as well as a visualization of the expected revenue and quantity sold corresponding to different possible selling prices in a given situation. At the time of writing, the model and its user interface have become the standard markdown pricing tool within the company. Country managers have been trained and can now access the application independently from their desktop computer. The model is used to make pricing decisions in all the countries where Zara has company-managed stores and the commercial regulations allow for discretionary markdowns. This represents about 80% of the entire store network. The remainder corresponds to either countries with franchised stores, about 12% of the store network in 2009, or countries where clearance sales must adhere to specific markdown regulations, preventing the use of the model or even the legacy pricing process. Given the successful results, other brands within Inditex, such as Stradivarius and Pull & Bear, have shown interest in adapting this tool for their own stores.

In conclusion, our work is the first documented application of a complete multi-product markdown optimization solution to the setting of fast-fashion retailing. It involves a rigorous impact assessment through a pilot experiment designed to provide a control-adjusted estimation, which contrasts with many other applications where the specific impact is either not estimated at all or estimated through a before versus after methodology which completely ignores than many other factors can affect the difference between before and after. Finally, we believe this to be the first large-scale application of a pricing optimization solution by a global firm for which all relevant technical and implementation details as well as related impact estimation results are made public. By exposing important aspects of how pricing is performed in practice, this paper opens the field for more theoretical research.

In terms of impact, we showed that the solution implemented increases clearances sales revenue by about 6%, corresponding for example to $90M in 2008. This financial impact is explained by the model's ability, relative to the legacy process, of maximizing revenue rather than liquidating stock. Indeed, the model usually made price suggestions that were slightly more aggressive at the beginning and more conservative towards the end. It also showed its ability to correctly identify the appropriate markdown strategy depending on the type of article (e.g., classic vs. fashion) considered.

Beyond the financial aspect, this project also had a cultural impact on Zara. First, it changed Zara's approach to markdowns from intuition-based to model-based, and demonstrated that pricing decisions can be improved by a scientific approach. Second, the project created consensus on the objective of clearance sales, which in turn provided a basis for discussion and pushed the country managers to find stronger arguments to justify their intuition. Third, a slightly more subtle cultural impact resulted from the introduction of a formal forecasting method. Initially, the forecast error received most of the attention at the pricing meetings, and the model-based process was evaluated based on the accuracy of its sales predictions. It took a concerted communicational effort to shift the discussion to the impact of the suggested prices on revenue, which was the actual purpose of the model. Keeping track of the forecast error is relevant, and there is always room to improve it, but it was important to anchor the debate on what really mattered, and that despite some level of forecast inaccuracy, the model could still generate better pricing decisions. We considered this to be a key learning, especially given the fact that many optimization projects do not materialize because the performance is measured exclusively based on the forecast rather than the realized profits (a similar observation motivated the work by Besbes *et al.* 2010).

From a process standpoint, the pricing solution we implemented enables more consistency, scalability and organizational distribution of pricing decisions. In other words, it provides a yardstick that unifies the pricing criteria across a diverse pool of country managers. This is particularly relevant for Zara in light of that firm's growth aspirations. Finally, we believe that the public dissemination of this successful and fully documented application of revenue management in a global company with a visible brand should also generate a substantial impact beyond Zara.

Acknowledgments. This paper owes much to Rodolfo Carboni, who performed a substantial part of the related implementation work at Zara. We are also most grateful to Miguel Díaz, Javier García, Lorena Rodríguez, Christian Montejo, Francisco Babio, José Manuel Corredoira, Felipe Peña, César Suárez and Jesús Echevarria from Inditex for their support and many contributions to this project. The authors benefited from feedback at the 2009 Conference of the Consortium for Operational Excellence in Retailing, in particular from the organizers Ananth Raman and Marshall Fisher, and from the judges of the 2010 Practice Competition of the Revenue Management and Pricing Section of INFORMS including Robert Shumsky, Warren Lieberman, Maarten Oosten and Robert Phillips. The authors also want to thank Steve Smith, Ming Hu and Richard Lai for useful comments on a preliminary version of the paper.

References

Anderson, E., D. Simester. 2003. Effects of $9 Price Endings on Retail Sales: Evidence from Field Experiments. *Quantitative Marketing and Economics* **1**(1) 93–110.

Araman, V., R. Caldentey. 2009. Dynamic Pricing for Nonperishable Products with Demand Learning. *Operations Research* **57**(5) 1169–1188.

Bertsimas, D., I. Popescu. 2003. Revenue management in a dynamic network environment. *Transportation Science* **37**(3) 257–277.

Bertsekas, D. 1995. *Dynamic Programming and Optimal Control.* Athena Scientific, Cambridge, MA.

Besbes, O., R. Phillips, A. Zeevi. 2010. Testing the Validity of a Demand Model: An Operations Perspective. *Manufacturing and Service Operations Management* **12**(1) 162–183.

Besbes, O., A. Muharremoglu. 2010. On the Price of Demand Censoring in the Newsvendor Problem. Working Paper. Columbia University, NY.

Bitran, G., R. Caldentey and S. Mondschein. 1998. Coordinating Clearance Markdown Sales of Seasonal Products in Retail Chains. *Operations Research* **46**(5) 609–624.

Cachon, G., R. Swinney. 2009. Purchasing, Pricing, and Quick Response in the Presence of Strategic Customers." *Management Sci.* **55**(3) 497–511.

Cachon, G., R. Swinney. 2010. The Value of Fast Fashion: Rapid Production, Enhanced Design, and Strategic Consumer Behavior. Working Paper, The Wharton School.

Carboni, R. 2009. *Clearance Pricing Optimization at Zara.* M.S. Thesis. LFM, MIT.

Caro, F., J. Gallien. 2010. Inventory Management of a Fast-Fashion Retail Network. *Operations Research* **58**(2) 257–273.

Caro, F., V. Martinez de Albeniz. 2010. The Impact of Quick Response in Inventory-Based Competition. *Manufacturing and Service Operations Management* **12**(3) 409–429.

Dana, J. D., N. C. Petruzzi. 2001. Note: The Newsvendor Model with Endogenous Demand. *Management Science* **47**(11) 1488–1497.

DeHoratius, N., A. J. Mersereau, L. Schrage. 2008. Retail Inventory Management When Records Are Inaccurate. *Manufacturing and Service Operations Management* **10**(2) 257–277.

Dong, L., P. Kouvelis, Z. Tian. 2009. Dynamic Pricing and Inventory Control of Substitute Products. *Manufacturing and Service Operations Management* **11**(2) 317–339.

Duan, N. 1983. A Nonparametric Retransformation Method. *Journal of the American Statistical Association* **78**(383) 605–610.

Elmaghraby, W. and P. Keskinocak. 2003. Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions. *Management Science*

49(10) 1287-1309.

Ferdows, K., J. AD Machuca and M. Lewis. 2002. Zara. The European Case Clearing House. Case 603-002-1.

Fisher, M. and K. Rajaram. 2000. Accurate Retail Testing of Fashion Merchandise: Methodology and Application. *Marketing Science* **19**(3) 266–278.

Fisher, M., A. Raman. 2010. *The New Science of Retailing.* Harvard Business Press, Massachusetts.

Fisher, M., R. Vaidyanathan. 2009. An Algorithm and Demand Estimation Procedure for Retail Assortment Optimization. Working Paper, The Wharton School, UPenn, PA.

Fraiman, N., M. Singh L. Arrington and C. Paris. 2002. Zara. Columbia Business School Case.

Gallego, G., G. van Ryzin. 1994. Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science* **40**(8) 999–1020.

Gallego, G., G. van Ryzin. 1997. A Multi-Product Dynamic Pricing Problem and its Application to Network Yield Management. *Operations Research* **45**(1) 24–41.

Gallego, G., M. Hu. 2007. Dynamic Pricing of Perishable Assets under Competition. Working Paper, Rotman School of Management, University of Toronto.

Gaur, V. and M. Fisher. 2005. In-Store Experiments to Determine the Impact of Price on Sales. *Production and Operations Management*, **14**(4) 377–387.

Ghemawat, P., J. L. Nueno. 2003. ZARA: Fast Fashion. Harvard Business School Multimedia Case 9–703-416.

Green. W. 2003. *Econometric Analysis.* 5-th Edition, Pearson Education, New Jersey.

Heching, A., G. Gallego and G. van Ryzin. 2002. An Empirical Analysis of Policies and Revenue Potential at One Apparel Retailer. *Journal of Revenue and Pricing Management* **1**(2) 139–160.

Imbens, G. W., J. M. Wooldridge. 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* **47**(1) 5–86.

Jasin, S., S. Kumar. 2010. Re-solving Heuristic with Bounded Revenue Loss for Network Revenue Management with Customer Choice. Working paper, Institute for Computational and Mathematical Engineering, Stanford University.

Little, J. 1966. A Model of Adaptive Control of Promotional Spending. *Operations Research* **14**(6) 1075–1097.

Maglaras, C. and J. Meissner. 2006. Dynamic Pricing Strategies for Multi-Product Revenue

Management Problems. *Manufacturing and Service Operations Management* **8**(2) 136–148.

Manning, W., J. Mullahy. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* **20** 461–494.

Mantrala, M. K., and S. Rao. 2001. A Decision-Support System that Helps Retailers Decide Order Quantities and Markdowns for Fashion Goods. *Interfaces* **31**(3) 146–165.

Musalem, A., M. Olivares, E. T. Bradlow, C. Terwiesch, D. Corsten. 2010. Structural Estimation of the Effect of Out-of-Stocks. *Management Science* **56**(7) 1180–1197.

Sigurdsson, V., G. Foxall and H. Saevarsson. 2010. In-Store Experimental Approach to Pricing and Consumer Behavior. *Journal of Organizational Behavior Management* **30** 234–246.

Smith. S., S. McIntyre, D. Achabal. 1994. A Two-Stage Sales Forecasting Procedure Using Discounted Least Squares. *Journal of Marketing Research* **31**(1) 44–56.

Smith, S., D. Achabal. 1998. Clearance Pricing and Inventory Policies for Retail Chains. *Management Science* **44**(3) 285–300.

Smith, S. 2009. Clearance Pricing in Retail Chains. Chapter 11 in *Retail Supply Chain Management*, N. Agrawal and S. Smith (eds.). Springer Science+Business Media, NY.

Smith, S. 2011. Personal communication.

Soon, W. 2011. A Review of Multi-Product Pricing Models. *Applied Mathematics and Computation* **217**(21) 8149–8165.

Stock, J. H., M. W. Watson. 2003. *Introduction to Econometrics.* Pearson Education, Inc., Boston, MA.

Talluri, K., G. van Ryzin. 2004. *The Theory and Practice of Revenue Management.* Springer, New York, NY.

Valkov, T. 2006. From theory to practice: Real-world applications of scientific pricing across different industries. *Journal of Revenue and Pricing Management* **5**(2), 143-151.

Verdugo, O. 2010. *Coordination of inventory distribution and price markdowns for clearance sales at Zara.* M.S. Thesis. LGO, MIT.

Vulcano, G., G. van Ryzin, R. Ratliff. 2009. Estimating Primary Demand for Substitutable Products from Sales Transaction Data. Working Paper, New York University.

Appendix (to be posted online)

A. Demand Rate Estimation

In the following, let $Sales^d_{rsj}$ denote sales of SKU rs at store $j \in \mathcal{J}$ in day $d \in \mathcal{D}$ and let I^d_{rsj} be the inventory position of SKU rs at store j at the beginning of day d. We want estimate the demand rate λ^w_r, which can differ from sales due to demand censoring caused by stockouts. We must also take into account seasonality effects that affect demand but are not related to price changes.

The first step in the demand estimation is thus to de-seasonalize the data, using daily seasonality factors defined as

$$\delta_d := \left(\frac{\sum_{r\in\mathcal{R}} Sales^{w(d)}_r}{\overline{Sales}}\right) \cdot \left(7\frac{M_{m(d)}}{\sum_{m\in\mathcal{M}} M_m}\right), \quad \forall\, d \in \mathcal{D}, \tag{29}$$

where $\overline{Sales}$ is the average weekly sales in the given country, $w(d)$ is the week in which day d falls, $Sales^w_r := \sum_{s\in\mathcal{S}(r),\, j\in\mathcal{J}, d\in\mathcal{D}(w)} Sales^d_{rsj}$ are the sales for article r in period w aggregated across sizes, colors and stores, $\mathcal{D}(w)$ is the set of days in period w, m denotes a day of the week, $m(d) \in \mathcal{M} :=$ $\{\text{Mon}, \ldots, \text{Sun}\}$ is the day of the week for d, $M_m := \dfrac{\sum_{d\in\mathcal{D},\, r\in\mathcal{R},\, s\in\mathcal{S}(r),\, j\in\mathcal{J}} 1_{\{m(d)=m\}}\, Sales^d_{rsj}}{\sum_{d\in\mathcal{D}} 1_{\{m(d)=m\}}}$ is the average sales that occur on a day $m \in \mathcal{M}$, and 1_E is the indicator function associated with event E. In words, the factor δ_d captures the expected daily sales variations for the brand at the country level. Note that the definition in Equation (29) has two components which normalize sales data with respect to inter-week as well as intra-week variations. Although these components should be updated every season, we have observed that they have remained in fact quite constant over the years (see Carboni 2009), which shows the formula's robustness and validates its use.

Correcting for demand censoring is more complicated, even for a single SKU (see Besbes and Muharremoglu 2010). Moreover, due to Zara's inventory display policy, an article r is usually moved from display when some of its key sizes or colors are missing, even if there is stock available for other sizes/colors, which exacerbates the censoring effect. To capture this, we followed Caro and Gallien (2010) and defined the indicator $1_{DND^d_{rsj}}$ that equals one if SKU rs was not on display at store j in day d. Here DND^d_{rsj} stands for <u>D</u>ays <u>N</u>ot on <u>D</u>isplay, and formally the event is defined as

$$DND^d_{rsj} := \left\{\{I^d_{rsj} = 0\} \text{ or } \{\min_{\tilde{s}\in\mathcal{S}^+(r)} I^d_{r\tilde{s}j} = 0 \text{ and } \max_{\tilde{s}\in\mathcal{S}(r)} Sales^d_{r\tilde{s}j} = 0\}\right\}, \tag{30}$$

where $\mathcal{S}^+(r)$ has the key sizes/colors for article r.

We finally estimated the demand rate for article r in period w as

$$\lambda^w_r = \frac{Sales^w_r}{Time^w_r}, \tag{31}$$

with

$$Time_r^w := \sum_{s\in\mathcal{S}(r),\, j\in\mathcal{J}} \alpha_{rsj} \cdot \sum_{d\in\mathcal{D}(w)} \delta_d (1 - 1_{DND^d_{rsj}}), \tag{32}$$

where $\alpha_{rsj} := \sum_{w<1} Sales^w_{rsj} \big/ \sum_{w<1} Sales^w_r$ denote the relative sales weight of store j for SKU rs and $Sales^w_{rsj} := \sum_{d\in\mathcal{D}(w)} Sales^d_{rsj}$ are the sales of SKU rs at store j in period w. Note that Equation (32) is a weighted average of the deseasonalized "days on display" in period w. Therefore, the ratio in Equation (31) corresponds to a normalized demand rate. The sales weight parameters α_{rsj} are also used to disaggregate the forecast in Equation (7). Note that Zara frequently reallocates inventory prior to clearance sales. When that happened, we used

$$\alpha_{rsj} = \frac{\sum_{w<1,\, j\in\mathcal{J}} Sales^w_{rsj}}{\sum_{w<1} Sales^w_r} \cdot \frac{\sum_{w<1,\, r\in\mathcal{R},\, s\in\mathcal{S}(r)} Sales^w_{rsj}}{\sum_{w<1,\, r\in\mathcal{R}} Sales^w_r}$$

to disaggregate the demand rate to the SKU and store level.

B. Dynamic Programming Formulation

In the dynamic programming formulation, the state of the system is the current inventory level (I^w) and the previous price decisions (x^w). Note that here we use the vector notation $I^w := \{I^w_{rsj}\}_{r\in\mathcal{R},\, s\in\mathcal{S}(r),\, j\in\mathcal{J}}$, $Sales^w := \{Sales^w_{rsj}\}_{r\in\mathcal{R},\, s\in\mathcal{S}(r),\, j\in\mathcal{J}}$, $x^w := \{x^w_{nk}\}_{n\in\mathcal{N},k\in\mathcal{K}}$, and $p^w := \{p^w_n\}_{n\in\mathcal{N}}$, where the variable p^w_n represents the price assigned to cluster n in period w (it should not be confused with the input parameters $p_k, k\in\mathcal{K}$, which are the feasible prices). After clearance sales, the remaining inventory is salvaged at price p_0. Hence, for any final state, the revenue from the liquidating the inventory is $J^W(I^W, x^{W-1}) = \sum_{r\in\mathcal{R},\, s\in\mathcal{S}(r),\, j\in\mathcal{J}} p_0 I^W_{rsj}$. The Bellman equation in period $w\in\mathcal{W}$ of clearance sales is the following:

$$J^w(I^w, x^{w-1}) \quad = \quad \max \sum_{n\in\mathcal{N}} p^w_n \lambda^w_n + \mathbb{E}\big[J^{w+1}(I^w - Sales^w, x^w)\big|\, p^w, I^w\big] \tag{33}$$

$$s.t.$$

$$\lambda^w_n = \sum_{r\in\mathcal{R}_n} \sum_{s\in\mathcal{S}(r)} \sum_{j\in\mathcal{J}} \mathbb{E}\big[Sales^w_{rsj}\big|\, p^w_n, I^w_{rsj}\big], \quad \forall\, n\in\mathcal{N}, \tag{34}$$

$$p^w_n = \sum_{k\in\mathcal{K}} p_k\, y^w_{kn}, \quad \forall\, n\in\mathcal{N}, \tag{35}$$

$$y^w_{nk} = x^w_{nk} - x^w_{nk-1}, \quad \forall\, n\in\mathcal{N}, k\in\mathcal{K}, \tag{36}$$

$$x^w_{nk-1} \le x^w_{nk}, \quad \forall\, n\in\mathcal{N}, k\in\mathcal{K}, \tag{37}$$

$$x^w_{nk} \le x^w_{n+1k}, \quad \forall\, n\in\mathcal{N}, k\in\mathcal{K}, \tag{38}$$

$$\sum_{k\in\mathcal{K}} x^w_{nk} = \sum_{k\in\mathcal{K}} x^w_{n+1k}, \quad \forall\, n \text{ such that } \sum_{k\in\mathcal{K}} x^{w-1}_{nk} = \sum_{k\in\mathcal{K}} x^{w-1}_{n+1k}, \tag{39}$$

$$x^{w-1}_{nk} \le x^w_{nk}, \quad \forall\, n\in\mathcal{N}, k\in\mathcal{K}, \tag{40}$$

$$y^w_{nk} \le z^w_k, \quad \forall\, n\in\mathcal{N}, k\in\mathcal{K}, \tag{41}$$

$$\sum_{k\in\mathcal{K}} z_k^w \leq N, \tag{42}$$

$$\sum_{n\in\mathcal{N}}\sum_{r\in\mathcal{R}_n}\sum_{j\in\mathcal{J}} I_{rj}^w\, y_{nk}^w \geq Q z_k^w, \quad \forall\; k\in\mathcal{K}, \tag{43}$$

$$x_{nk}^w \in \{0,1\}, z_k^w \in [0,1], \quad \forall\; n\in\mathcal{N}, k\in\mathcal{K}. \tag{44}$$

C. Model Linearization

Let $\rho = \widetilde{\beta}_4$ be the parameter associated to the broken assortment term $\ln(\min\{1, \frac{I_r^w}{f}\})$ in the demand regression. We assume that $0 \leq \rho \leq 1$, which is consistent with our empirical results. We would like to fit the following linear form suggested by Smith and Achabal (1998) to capture the broken assortment effect:

$$\min\left\{1, 1-\mu+\mu\frac{I_r^w}{f}\right\}$$

For $\mu \geq 0$, this expression is relevant when it is less than one, which only occurs when $\frac{I_r^w}{f} \leq 1$. We can chose μ to be the value that minimizes the quadratic distance between the two functional forms. This amounts to minimize $\int_0^1 (x^\rho - 1 + \mu - \mu x)^2\, dx$ with respect to μ. The first-order conditions provide the closed-form formula $\mu = \dfrac{3\rho^2+9\rho}{2\rho^2+6\rho+4}$.

The linearization of the broken assortment effect yields constraints (23) and (24) which are defined at the article level. Constraint (23) can be easily aggregated for all articles r that belong to a given cluster n to obtain

$$\lambda_{nk}^w \leq E_n^w(p_k) y_{nk}^w, \quad \forall\; w\in\mathcal{W}\backslash\{c\}, n\in\mathcal{N}, k\in\mathcal{K}, \tag{45}$$

where $E_n^w(p_k) := \sum_{r\in\mathcal{R}_n} E_r^w(p_k)$. For constraint (24), by adding the right hand side over $r\in\mathcal{R}_n$ we have

$$\begin{aligned}
\sum_{r\in\mathcal{R}_n}\Big(1-\mu+\mu\frac{I_r^w}{f}\Big)F_r^w(p_k) &= (1-\mu)\sum_{r\in\mathcal{R}_n} F_r^w(p_k) + \mu\Big(\sum_{r\in\mathcal{R}_n}\frac{I_r^w}{I_n^w}F_r^w(p_k)\Big)\frac{I_n^w}{f} \\
&\approx (1-\mu)\sum_{r\in\mathcal{R}_n} F_r^w(p_k) + \mu\Big(\sum_{r\in\mathcal{R}_n}\frac{\widehat{I}_r^w}{\widehat{I}_n^w}F_r^w(p_k)\Big)\frac{I_n^w}{f},
\end{aligned}$$

where in the last parenthesis we have replaced I_r^w by its estimated upper bound $\widehat{I}_r^w$. We use this last expression to formulate a linear constraint for the broken assortment effect at the cluster level:

$$\lambda_{nk}^w \leq (1-\mu)F_n^w(p_k) + \mu\overline{F}_n^w(p_k)\frac{I_n^w}{f}, \quad \forall\; w\in\mathcal{W}\backslash\{c\}, n\in\mathcal{N}, k\in\mathcal{K}, \tag{46}$$

where $F_n^w(p_k) := \sum_{r\in\mathcal{R}_n} F_r^w(p_k)$, and $\overline{F}_n^w(p_k) := \sum_{r\in\mathcal{R}_n}\frac{\widehat{I}_r^w}{\widehat{I}_n^w}F_r^w(p_k)$ is a weighted average of the parameters $F_r^w(p_k)$. Replacing constraint (10) by Equations (45) and (46) we finally obtain a mixed integer programming (MIP) formulation for the clearance pricing problem.

In order to compute $E_r^w(p_k)$ and $F_r^w(p_k)$ we would need an upper bound $\widehat{I}_r^w$ for the inventory level in period $w > 1$. One possibility is to use the current inventory level $\widehat{I}_r^1$, but that can be a loose upper bound for periods beyond $w = 1$ which could lead to a significant overestimation of sales as indicated by the derivation of Equation (21). Instead, we introduce a factor $0 < \kappa < 1$ such that $E_r^w(p_k) = \kappa^{w-1}E_r^1(p_k)$. The parameter κ can be seen as a discount factor that captures how prices age, regardless of the inventory level. Indeed, from our conversation with Zara managers it appears that consumers' perception of novelty during clearance sales seems to decay very quickly. That is, one expects to sell considerably more today than a week later, everything else remaining constant. Note also that κ is different from the A_r^w variable used in the demand estimation model since the latter captures how the product ages (not the price) and it is computed during the regular selling season when prices are kept fixed.

In order to estimate κ, we use actual sales data from periods that had the same price in past clearance sales. In particular, let $\mathcal{V}$ be the set of triplets (r, w, w') from past years such that $w < w'$, $p_r^{w-1} > p_r^w$, and $p_r^{w'} = p_r^w$. We want to make sure that the "aging prices" effect is not confounded with the broken assortment effect, since the latter is captured explicitly by constraints (46). For that reason, we only consider triplets (r, w, w') where there is plenty of inventory (i.e., the broken assortment effect is not present). Then, we perform a minimization with respect to κ of the squared differences

$$\sum_{(r,w,w')\in\mathcal{V}} \left(\kappa^{w'-w} - \frac{E_r^{w'}(p_r^{w'})}{E_r^w(p_r^w)}\right)^2, \tag{47}$$

where $E_r^w(p_r^w)$ is substituted by $Sales_r^w$. Note that $\kappa < 1$ guarantees that in future periods less sales are expected at the same price, regardless of the stock level. Finally, the parameter $F_r^w(p_k)$ can be approximated by $\kappa^{w-1}F_r^1(p_k)$.

D. Bias Correction with Heteroscedasticity

An (implicit) assumption in OLS is that the variance of the error in the regression does not depend on the the explanatory variables. However, during clearance sales it could be that for some groups/countries the average magnitude of the error (i.e., its standard deviation) depends on the price level or the broken assortment effect. If the latter occurs, then we have a case of *heteroscedasticity* and the bias correction factor (due to the logarithmic transformation) will depend on the explanatory variables. To deal with this, we propose the following steps that are based on White's test, which is commonly used to detect heteroscedasticity (see Green 2003).

1. As before, run the regression of the residuals and compute the errors as:

$$\widetilde{\epsilon}_r^w = \ln(\lambda_r^w) - \widetilde{\beta}_{0r} - \widetilde{\beta}_1 \ln(C_r) - \widetilde{\beta}_2 A_r^w - \widetilde{\beta}_3 \ln(\lambda_r^T) - \widetilde{\beta}_4^w \ln(\min\{1, \frac{I_r^w}{f}\}) - \widetilde{\beta}_5^w \ln\left(\frac{p_r^w}{p_r^T}\right).$$

2. To detect heteroscedasticity, we can regress the square of the errors $(\widetilde{\epsilon}_r^w)^2$ with respect to the explanatory variables, their squares, and cross-products. In other words, let $BA(I_r^w) := \ln(\min\{1, \frac{I_r^w}{f}\})$ and $EL(p_r^w) := \ln\left(\frac{p_r^w}{p_r^T}\right)$. Then, run the following regression:

$$(\widetilde{\epsilon}_r^w)^2 = \nu_0 + \nu_1 BA(I_r^w) + \nu_2 EL(p_r^w) + \nu_3 BA(I_r^w)^2 + \nu_4 EL(p_r^w)^2 + \nu_5 BA(I_r^w) \cdot EL(p_r^w).$$

3. If none of the estimated parameters $\widetilde{\nu}_1$ to $\widetilde{\nu}_5$ are significant, i.e., the (absolute) t-statistics are less than 1.96, then nothing changes and the (logarithmic) bias correction factor is computed as before using the formula in Equation (8).

4. If any of the estimated parameters $\widetilde{\nu}_1$ to $\widetilde{\nu}_5$ is significant, then we should adjust the bias correction factor for heteroscedasticity. For simplicity, assume that only $\widetilde{\nu}_2$ is significant. Then, for price p, the bias correction factor is given by

$$H^w(p) = \frac{1}{|\mathcal{R}(t_c)|} \sum_{r \in \mathcal{R}(t_c)} \exp\left(\widetilde{u}_r^w \sqrt{\widetilde{\nu}_0 + \widetilde{\nu}_2 EL(p)}\right), \quad w \geq 1, \tag{48}$$

where $\widetilde{u}_r^w = \dfrac{\widetilde{\epsilon}_r^w}{\sqrt{\widetilde{\nu}_0 + \widetilde{\nu}_2 EL(p_r^w)}}$. If more parameters $\widetilde{\nu}_i$ are significant, then we must include the respective term under the square root in Equation (48) and in the denominator of $\widetilde{u}_r^w$. It is easy to see that when the parameters $\widetilde{\nu}_1$ to $\widetilde{\nu}_5$ are not significant (i.e., they are equal to zero), then Equation (48) reduces to Equation (8).

E. User Interface Snapshots

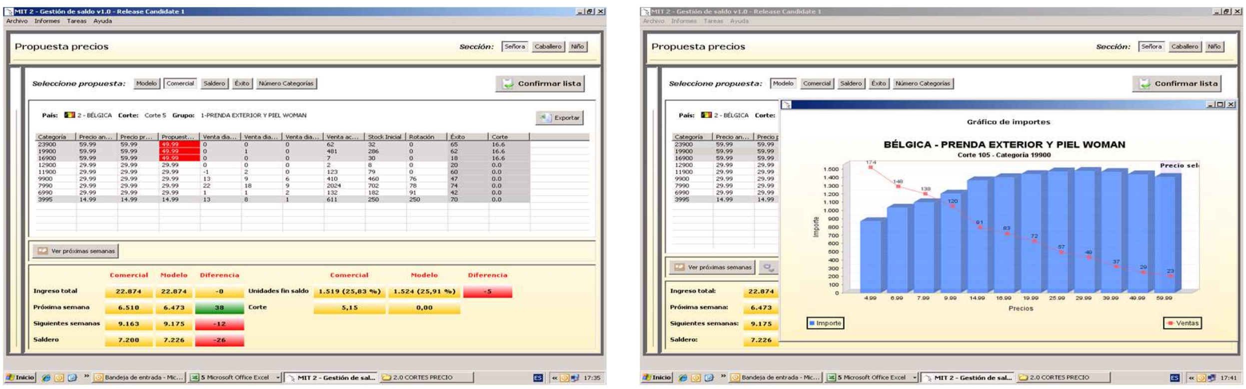


Figure 6: Two snapshots of the distributed software application that implements the model-based pricing process.