

Clearance Pricing Optimization for a Fast-Fashion Retailer

Felipe Caro

UCLA Anderson School of Management, Los Angeles, California 90095, fcaro@anderson.ucla.edu

Jérémie Gallien

London Business School, London NW1 4SA, United Kingdom, jgallien@london.edu

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 since designed and implemented 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.

Subject classifications: retailing; markdown pricing; clearance sales; fast-fashion; revenue management; forecasting; field test; model implementation.

Area of review: OR Practice.

History: Received June 2011; revision received May 2012; accepted June 2012.

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 (Caro 2012). 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 are 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 et al. 1998, Heching et al. 2002, Smith 2009), which all consider each article independently. Indeed, the high 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 later).

The present paper describes the development, implementation, and evaluation in the field of the markdown optimization system that Zara has since deployed worldwide to address these challenges. To the best of our knowledge, this constitutes the first large-scale application of a multiproduct 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) and then describe its legacy clearance pricing process (in §1.2). Some of the data presented in this paper have 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.

1.1. Clearance Sales at Zara

Like 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 clearance sales for any retailer is to maximize the revenue derived from merchandise that is still unsold when a new collection is about to be introduced, so it works as a final correction mechanism for products that did not sell as expected during the regular season. At Zara, merchandise is often deliberately withdrawn from the store display area during the season 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 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 to aggregate merchandise in fewer locations and improve their display quality (e.g., complete missing sizes). Finally, the sales period ends with the step known as *liquidation*, when all remaining clearance inventory is collected from Zara stores and sold (sometimes by weight) 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 rearranged: 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 rearrange display areas from the backroom, 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 uses prices from only a discrete and finite set of so-called *commercial prices* all ending with "0.99" 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 30 to 700 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 constrained 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 it 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.

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

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”).

Current clearance categories	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/ Venta Día 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
OF 19.90	9.95	16	14	6	420	384	64	54
FROM 14.90 to 12.90	6.95	21	22	12	519	322	27	64
Totale ...		82	49	38	1.939	2.930	77	41

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.

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 multiproduct pricing problem (see Maglaras and Meissner 2006, 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 et al. (1998), Mantrala and Rao (2001), and Heching et al. (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 contain only 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 a multiproduct markdown optimization model, whereas all the existing application-oriented pricing papers cited discuss only 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.” While our model assumes independent demand across products, it does capture the many dependencies between pricing decisions for different products that are introduced by store-level implementation constraints. 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 endeavor, both because the fast-fashion retail model might provide some important competitive advantages (Caro 2012) and because clearance pricing is arguably more challenging in that specific context (see §1).

3. Demand Prediction Model

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 data set (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 Data Set Construction

Let \mathcal{R} and \mathcal{F} 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 $I_r^w := \sum_{s \in \mathcal{S}(r), j \in \mathcal{F}} I_{rsj}^w$ 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 the e-companion, and it presents some challenges due to the presence of seasonality effects (e.g., Christmas or weekends), which cause 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 data set 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 data set 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 describe shortly. For many different model specifications, we applied this procedure to the training data set defined above and computed predictions for the sales realized in the clearance period of 2008 (testing data set). The final model selection was based on a combination of managerial judgment, the t -statistics and overall goodness-of-fit in the two-stage procedure, and most importantly, the aggregate forecasting error for the testing data set. The final validation took place during the live pilot described in §5, when forecast

accuracy was measured as part of a field implementation (see §5.2.2 for a discussion of these results). Although our forecasting model was thus derived through extensive experimentation, we describe here only 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:

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

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. Note that the purchase quantity is decided several weeks or months before clearance sales start so it acts as a predetermined variable at the time of the forecast estimation. This contrasts with studies that use panel data across several seasons and treat aggregate purchasing decisions endogenously (see Kesavan et al. 2010).

- 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 ($\min\{1, I_r^w/f\}$): 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 because, when inventory is low, the remaining items are usually those that are less attractive to customers (see Smith and Achabal 1998, Fisher and Raman 2010, §A-3).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. Although this parameter can be defined for an individual store, we calibrated it for the entire country to keep the regression aggregated at that level.

- Price discount (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 $\beta_{0r}, \beta_1, \dots, \beta_5$ 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 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 are 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

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

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

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

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 nonupdated regressors from Stage 1 are removed to obtain the residuals that contain only the effects of the updated regressors. Note that β_4^w was included in the Stage 1 regression to improve the fit, but it is disregarded in Equation (3) because 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:

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

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

To leverage data from the past and current seasons, the aforementioned two-stage procedure 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 $\{\tilde{\beta}_4^{P,w}, \tilde{\beta}_5^{P,w}\}_{w \geq 1}$, where we use the superscript P to denote that it comes from past season data. Of course, from Stage 1 we also obtained the set of parameters $\tilde{\beta}_{0r}^P, \tilde{\beta}_1^P, \tilde{\beta}_2^P, \tilde{\beta}_3^P$, but these played a role only 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 $\tilde{\beta}_{0r}^C, \tilde{\beta}_1^C, \tilde{\beta}_2^C, \tilde{\beta}_3^C$, and from Stage 2 we obtained $\{\tilde{\beta}_4^{C,w}, \tilde{\beta}_5^{C,w}\}_{1 \leq w < c}$, where we use the superscript C to denote that it comes from current season data. Finally, the parameters $\hat{\beta}_4^w$ and $\hat{\beta}_5^w$ used in the forecast for the current period ($w = c$) were computed with the following recursive equation:

$$\hat{\beta}_i^1 = \tilde{\beta}_i^{P,1}, \quad (5)$$

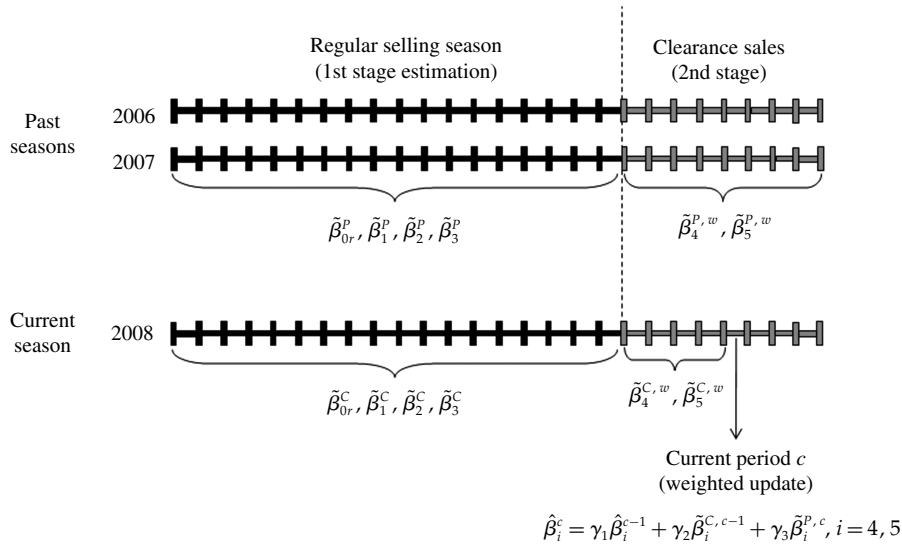
$$\hat{\beta}_i^w = \gamma_1 \hat{\beta}_i^{w-1} + \gamma_2 \tilde{\beta}_i^{C,w-1} + \gamma_3 \tilde{\beta}_i^{P,w} \quad w > 1, \quad (6)$$

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 $\hat{\beta}_4^w$ and $\hat{\beta}_5^w$ obtained from Equations (5)–(6) we forecasted the demand rate for the current period ($w = c$) according to the formula

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

Figure 2. Two-stage estimation procedure using past (P) and current (C) season data.



Note. The current period is denoted by $w = c$.

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 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):

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

where $\tilde{\epsilon}_r^w$ corresponds to the estimated errors in Equation (4). The bias in the retransformation 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 e-companion 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. Note first 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) depends only on the price and inventory of article r . Therefore, the model does not capture substitution or complementarity effects between products. Although there exists recent literature on how to estimate primary demand for substitutable products under the presence of stockouts (e.g., Vulcano et al. 2012 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 judgment, 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. Overall, our demand model as described in Equation (1) is a particular case of the general demand model introduced in Smith and Achabal (1998), and we also borrow from that paper the specific formulation for the broken assortment effect. However, in contrast with the specific model definition estimated for application purposes in Smith and Achabal (1998) and featuring exponential price sensitivity—see Equation (30) in that paper—our specification is a log–log model with a constant price elasticity, which also captures additional idiosyncratic explanatory variables and temporal auto-correlation. We refer the reader to §5.2.2 for a discussion of the actual forecasting accuracy performance achieved by our demand model when implemented in the field, which constitutes in our view an important 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 contains 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 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 the e-companion, 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). Although 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, are 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). For small instances with a single article we performed a numerical study, and the suboptimality gap was below 0.5% on average and less than 3% in the worst case (see the e-companion for details).

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 $\alpha_{rsj}\tilde{\lambda}_r^w$, where $\tilde{\lambda}_r^w$ is given by the forecast formula (7) and α_{rsj} is the sales weight of SKU rs at store j (see the e-companion for details on the computation of this last parameter). Let $k \in \mathcal{K} := \{1, \dots, K\}$ be the clearance price index. The set of clearance prices is $\{p_k, k \in \mathcal{K}\}$, which by convention is increasing, i.e., $p_0 \leq p_1 \leq p_2 \leq \dots \leq p_K$, and p_0 is the final unit salvage value or liquidation price. Let $n \in \mathcal{N} := \{1, \dots, N\}$ 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 $\mathcal{W} := \{w \mid c \leq w < W\}$ 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 are data available.

For the decision variables, $x_{nk}^w \in \{0, 1\}$ indicates whether cluster n should be sold at clearance price p_k or lower during pricing period $w \in \mathcal{W}$, with $x_{n0}^w = 0$ for all $(n, w) \in \mathcal{N} \times \mathcal{W}$. The prices from the previous period are given by x_{nk}^{c-1} , which is input data in the current period. In particular, recall that $w = 1$ corresponds to the first period, therefore one can use x_{nk}^0 to impose a minimum markdown at the beginning of clearance sales. The auxiliary variable $y_{nk}^w \in \{0, 1\}$ indicates whether cluster n should be sold at clearance price p_k during period w ; $\lambda_{nk}^w \geq 0$ represents the expected sales for cluster n in period $w \in \mathcal{W}$ if sold at price p_k ; $z_k^w \in \{0, 1\}$ indicates whether clearance price p_k is used for any cluster ($z_k^w = 1$) or not at all ($z_k^w = 0$) during period $w \in \mathcal{W}$; and $I_n^c := \sum_{r \in \mathcal{R}_n, s \in \mathcal{S}(r), j \in \mathcal{J}} I_{rsj}^c$ represents the inventory of cluster n available at the beginning of period w . Note that the inventory level for the current period I_n^c is input data. The uncertainty in the problem is given by $Sales_{rsj}^w$, 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:

$$\max \sum_{w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}} p_k \lambda_{nk}^w + \sum_{n \in \mathcal{N}} p_0 I_n^w, \quad (9)$$

$$\text{s.t. } \lambda_{nk}^w = y_{nk}^w \sum_{r \in \mathcal{R}_n} \sum_{s \in \mathcal{S}(r)} \sum_{j \in \mathcal{J}} \mathbb{E}[Sales_{rsj}^w \mid p_k, I_{rsj}^w], \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (10)$$

$$y_{nk}^w = x_{nk}^w - x_{nk-1}^w, \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (11)$$

$$x_{nk-1}^w \leq x_{nk}^w, \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (12)$$

$$x_{nk}^w \leq x_{n+1k}^w, \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (13)$$

$$\sum_{w \in \mathcal{W}, k \in \mathcal{K}} x_{nk}^w = \sum_{w \in \mathcal{W}, k \in \mathcal{K}} x_{n+1k}^w, \quad \forall n \text{ such that} \quad (14)$$

$$\sum_{k \in \mathcal{K}} x_{nk}^{c-1} = \sum_{k \in \mathcal{K}} x_{n+1k}^{c-1}, \quad (14)$$

$$x_{nk}^{w-1} \leq x_{nk}^w, \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (15)$$

$$I_n^{w+1} = I_n^w - \sum_{k \in \mathcal{K}} \lambda_{nk}^w, \quad \forall w \in \mathcal{W}, n \in \mathcal{N} \quad (16)$$

$$y_{nk}^w \leq z_k^w, \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (17)$$

$$\sum_{k \in \mathcal{K}} z_k^w \leq N^w, \quad \forall w \in \mathcal{W}, \quad (18)$$

$$\sum_{n \in \mathcal{N}} I_n^w y_{nk}^w \geq Q^w z_k^w, \quad \forall w \in \mathcal{W}, k \in \mathcal{K}, \quad (19)$$

$$I_n^w, \lambda_{nk}^w \geq 0, x_{nk}^w \in \{0, 1\}, z_k^w \in [0, 1], \quad \forall w \in \mathcal{W}, n \in \mathcal{N}, k \in \mathcal{K}. \quad (20)$$

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_{nk}^w and y_{nk}^w 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 because pricing decisions are made at that level. Constraint (17) implements the definition of z_k^w . Constraint (18) ensures that the number of distinct price categories in period w does not exceed 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 nonnegative or binary requirements for the decision variables. Observe that variables y_{nk}^w and z_k^w do not need to be defined as binary provided that x_{nk}^w is. Also, the nonnegativity of I_n^{w+1} together with the inventory balance constraint (16) ensure that the expected sales λ_{nk}^w never exceed the available inventory I_n^w .

The formulation above is still hard to solve in practice due to the nonlinearity of Constraint (10). Therefore, we linearize these equations. This requires some attention because the random variables $Sales_{rsj}^w$ 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 \hat{I}_{rsj}^w be an upper bound for the inventory level of SKU rs at store j in period w . Then, we have that

$$\begin{aligned} \lambda_{rk}^w &:= \sum_{s \in \mathcal{S}(r)} \sum_{j \in \mathcal{J}} \mathbb{E}[Sales_{rsj}^w | p_k, I_{rsj}^w] \\ &= \tilde{\lambda}_r^w(p_k, I_r^w) \sum_{s \in \mathcal{S}(r)} \sum_{j \in \mathcal{J}} \alpha_{rsj} \mathbb{E}[\tau_{rsj}^w | p_k, I_{rsj}^w] \\ &\leq \tilde{\lambda}_r^w(p_k, I_r^w) \sum_{s \in \mathcal{S}(r)} \sum_{j \in \mathcal{J}} \alpha_{rsj} \mathbb{E}[\tau_{rsj}^w | p_k, \hat{I}_{rsj}^w] \\ &= \frac{(\min\{1, I_r^w/f\})^{\tilde{\beta}_4}}{(\min\{1, \hat{I}_r^w/f\})^{\tilde{\beta}_4}} \sum_{s \in \mathcal{S}(r)} \sum_{j \in \mathcal{J}} \mathbb{E}[Sales_{rsj}^w | p_k, \hat{I}_{rsj}^w] \\ &= \left(\min\left\{1, \frac{I_r^w}{f}\right\} \right)^{\tilde{\beta}_4} \frac{E_r^w(p_k)}{(\min\{1, \hat{I}_r^w/f\})^{\tilde{\beta}_4}}, \end{aligned} \quad (21)$$

where $E_r^w(p_k) := \sum_{s \in \mathcal{S}(r), j \in \mathcal{J}} \mathbb{E}[Sales_{rsj}^w | p_k, \hat{I}_{rsj}^w]$ and $\tilde{\lambda}_r^w(p_k, I_r^w)$ 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 $I_{rsj}^w \leq \hat{I}_{rsj}^w$ for all stores and SKUs. The fourth step uses the definition of $\tilde{\lambda}_r^w(p_k, I_r^w)$ 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 the e-companion, the resulting expression is

$$\begin{aligned} &\left(\min\left\{1, \frac{I_r^w}{f}\right\} \right)^{\tilde{\beta}_4} \frac{E_r^w(p_k)}{(\min\{1, \hat{I}_r^w/f\})^{\tilde{\beta}_4}} \\ &\approx \min\left\{1, 1 - \mu + \mu \frac{I_r^w}{f}\right\} \frac{E_r^w(p_k)}{(\min\{1, \hat{I}_r^w/f\})^{\tilde{\beta}_4}}, \end{aligned} \quad (22)$$

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

$$\lambda_{rk}^w \leq E_r^w(p_k) y_{nk}^w, \quad \forall w \in \mathcal{W} \setminus \{c\}, r \in \mathcal{R}_n, \\ n \in \mathcal{N}, k \in \mathcal{K}, \quad (23)$$

$$\lambda_{rk}^w \leq \left(1 - \mu + \mu \frac{I_r^w}{f}\right) E_r^w(p_k), \quad \forall w \in \mathcal{W} \setminus \{c\}, \\ r \in \mathcal{R}, k \in \mathcal{K}, \quad (24)$$

where $F_r^w(p_k) := E_r^w(p_k)/(\min\{1, \hat{I}_r^w/f\})^{\tilde{\beta}_4}$. Constraint (23) is the relevant bound when $1 - \mu + \mu(I_r^w/f) \geq 1$ 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 $I_r^w \leq \hat{I}_r^w$). Note that Constraints (23) and (24) do not need to be defined for $w = 1$ because 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 the e-companion 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 because 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 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 fewer 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 of 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 because 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 because 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 (later the model was also migrated to Java). 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. Third, quantify the model's specific impact on markdown decisions, which is the main issue discussed in this 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 §5.2.2), pricing behavior (in §5.2.3), and financial impact (in §5.2.5).

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 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 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. Although 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 it in §5.2.3.

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

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

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 because 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

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

Because 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

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

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 purposely designed to measure the model's impact on that metric by using a difference-in-differences statistical procedure. Difference-in-differences is a standard technique to evaluate the impact of a change/treatment in an experimental setting in which one group of subjects has experienced the treatment, whereas another, comparable group has not (Kennedy 2003). The estimated impact corresponds to the difference between the changes experienced by the two groups.⁶ Specifically, for each store in Belgium we computed the aggregate realized income for groups 1–12 and 13–20 (denoted Y_{1-12} and Y_{13-20} , respectively). Then, we took first differences and computed the average across all stores (denoted $Y_{1-12}^{\text{BEL}} - Y_{13-20}^{\text{BEL}}$). 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

$$\Delta^{\text{BEL}} := (Y_{1-12}^{\text{BEL}} - Y_{13-20}^{\text{BEL}}) - (Y_{1-12}^{\text{RWE}} - Y_{13-20}^{\text{RWE}}), \quad (28)$$

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

As mentioned in §5.1.1, 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 $Y_{1-12}^{\text{BEL}} - Y_{13-20}^{\text{BEL}}$ and $Y_{1-12}^{\text{RWE}} - Y_{13-20}^{\text{RWE}}$, 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. Summary Statistics. To forecast sales, we followed the two-stage estimation procedure described in §3.2.

Table 1. Range and average value for Stage 1 parameters $\tilde{\beta}_{0r}^C, \tilde{\beta}_1^C, \tilde{\beta}_2^C, \tilde{\beta}_3^C$.

Regressor	Parameter	Parameter range	Avg. parameter value
Intercept	$\tilde{\beta}_{0r}^C$	[-3.84, -0.08]	-2.73
$\ln(C_r)$	$\tilde{\beta}_1^C$	[0.05, 0.42]	0.28
A_r^w	$\tilde{\beta}_2^C$	[-0.01, 0.00]	-0.006
$\ln(\lambda_r^{w-1})$	$\tilde{\beta}_3^C$	[0.32, 0.81]	0.66

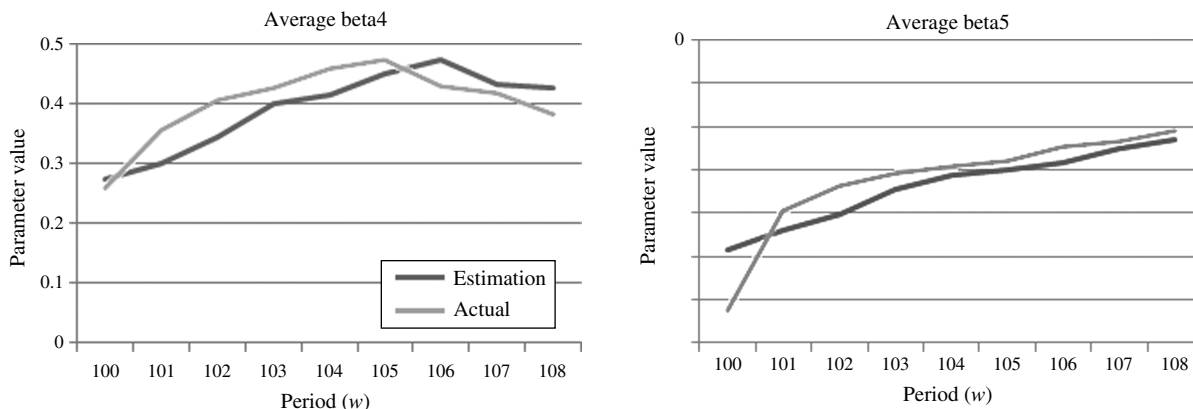
The Stage 1 regression given by Equation (2) had an average R^2 of 0.72. Table 1 provides summary statistics for the estimated parameters $\tilde{\beta}_{0r}^C, \tilde{\beta}_1^C, \tilde{\beta}_2^C, \tilde{\beta}_3^C$, that were used in the forecasting formula (7). As expected, $\tilde{\beta}_1^C$ and $\tilde{\beta}_3^C$ were positive and $\tilde{\beta}_2^C$ was negative, although this last parameter was rather small so the age explanatory variable A_r^w was relevant only for articles that were introduced early in the season. The Stage 2 regression given by Equation (4) had an average R^2 of 0.40. In Figure 3 we compare the estimated and actual mean values of β_4^w and β_5^w during clearance sales (in the figures we use Zara’s convention and denote the first period of clearance sales by 100). The estimated value for $\beta_i^w, i = 4, 5$, corresponds to $\hat{\beta}_i^w$ from the recursive formulas (5)–(6), whereas the actual value corresponds to $\tilde{\beta}_i^{C,w}$, which was computed after sales were observed in each period w . Overall, the estimated values closely followed the actual values. The bigger differences occurred in the initial periods, which should be expected because in those periods the estimated values relied more on the elasticities computed from historical data. The broken assortment parameter β_4^w was positive, but in the inelastic range: at its peak, a 1% decrease in inventory decreased demand by less than 0.5% on average. The nonmonotone pattern of β_4^w is interesting because it means that once the inventory reached a certain level (around week six), further percentage decreases had a smaller impact on demand. The price elasticity parameter β_5^w was negative but increasing, meaning that on average consumers became less sensitive to price as clearance sales advanced.

5.2.2. 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.⁷ The results are shown in Table 2, 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 six-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.⁸ 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).

Table 2 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 toward the end of clearance sales. Although this seems a plausible explanation, it did not have a major impact because the last periods accounted for a small fraction of total clearance sales.

The model formulation in §4 assumes Poisson demand. We validated this assumption by performing the statistical test described in Bitran et al. (1998) for each cluster in each group.⁹ In the e-companion we provide the details of the test, but the main results stratified by groups’ sales volume are summarized in Table 3. The Poisson hypothesis was

Figure 3. Estimation vs. actual values for β_4^w (left) and β_5^w (right).



Notes. Period 100 is the first period of clearance sales. The elasticities in the right-hand graph are negative (the scale is blank for confidentiality reasons).

Table 2. Forecast accuracy measured in MAD per period at the country level.

Country	Clearance sales period w (%)									Sales-weighted average (%)
	100	101	102	103	104	105	106	107	108	
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

not rejected in 61.8% of the cases with a significance of 10% (p -value ≥ 0.1). Although this represents the majority of the product clusters, Table 3 shows that the assumption was less prevalent in high-volume groups. A likely explanation is the Poisson property that requires the coefficient of variation to decrease with mean demand, which was usually not observed in practice. Despite the limited Poisson validation, Table 3 shows that the forecast was more accurate for high-volume groups. A forecast error of 23.1% is very good by industry standards. However, it could possibly have been better if the Poisson assumption had been confirmed. In fact, there was one high-volume group (basic trousers and shorts) that had a perfect Poisson fit, and this group achieved the lowest forecast error, 15.2%, across all groups (see the e-companion for details).

In contrast to the groups with ample sales, the low-volume groups did better in the Poisson test but had a less accurate forecast. We believe this is because Zara's trendier products—the T.R.F. collection, composed mostly by low-volume groups—have more volatile sales. Moreover, a lower volume translates into fewer observations to run the regressions. In future work, one could test other demand specifications—like the negative binomial (see Agrawal and Smith 1996)—but any improvement in forecast accuracy would have to be contrasted with the loss in modeling tractability.

5.2.3. Adherence to Model Recommendations. 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). As shown in Figure 4, however, the overall adherence to the model prices was very high, with a few exceptions in Ireland that we discuss next.

In the cases of G16-P100 and G20-P102 (we use “G” and “P” to abbreviate Group and Period, respectively), the

Table 3. Poisson validity (% of nonrejected hypotheses) and forecast accuracy (measured in MAD) stratified by groups' sales volume.

Groups	# groups	Poisson validity (%)	Forecast error (%)
Top 80% in sales volume	6	43.6	23.1
Bottom 20% in sales volume	14	75.3	26.8
All groups	20	61.8	23.9

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

5.2.4. Analysis of Model Prices. It is also clear from the number of dots shown in Figure 4 that the prices suggested by the model often differed from the manual ones. We therefore set out to determine whether these differences in price recommendations followed any systematic pattern. To that end, we show the average price P^w and the fraction sold X^w in each period w of clearance sales for the treatment and control groups in Figure 5 and Table 4, 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. As shown in Table 4, these differences in prices are fully consistent with the observed fraction of inventory sold over time. Indeed, groups 1–12 in Belgium, which were priced with the model in a fairly aggressive manner initially, ended with a higher fraction sold than the same (control) groups in Ireland. Furthermore, the higher model prices applied to groups 13–20 in Ireland resulted in a lower fraction of inventory sold than the lower manual prices applied to the same groups in Belgium. 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 mark down sooner the more expensive items to increase sales volume, and it chose to collect more revenue out of 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 the pricing strategies on revenue can be seen in Figure 6, in which we plot the trajectory of the first difference $Y_{1-12}^w - Y_{13-20}^w$ in Belgium and Ireland. The curve for the rest of western Europe (RWE) represents

Figure 4. Adherence to model prices in intervention groups.

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

Note. Shaded cells indicate when model prices were not followed; dots indicate when they differed from manual prices.

the baseline. Although the final distance with respect to the RWE curve was almost the same in both countries (see §5.2.5 for a statistical validation of this observation), it is interesting to see that in Ireland the distance was generated early, whereas in Belgium it happened toward the end of the clearance horizon. This is consistent with the pricing strategies the model followed in each country: initially conservative in Ireland, whereas more aggressive in Belgium.

Another important component of the model output is the recommended number of price categories in each period, which we now discuss. To investigate this we considered

$\sum_{k \in \mathcal{K}} z_k^w / N$ for each group and clearance period w , i.e., the ratio of the number of different price categories used in period w to the initial number of price clusters for that group during the regular selling season. From the averages of these ratios across groups shown in Table 5, it is clear that the model had a systematic tendency to recommend a larger number of price categories than was used by the manual pricing policy.

Inspection of the data reveals that the model also had a systematic tendency to use a larger number of different prices over the entire course of the clearance period for

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

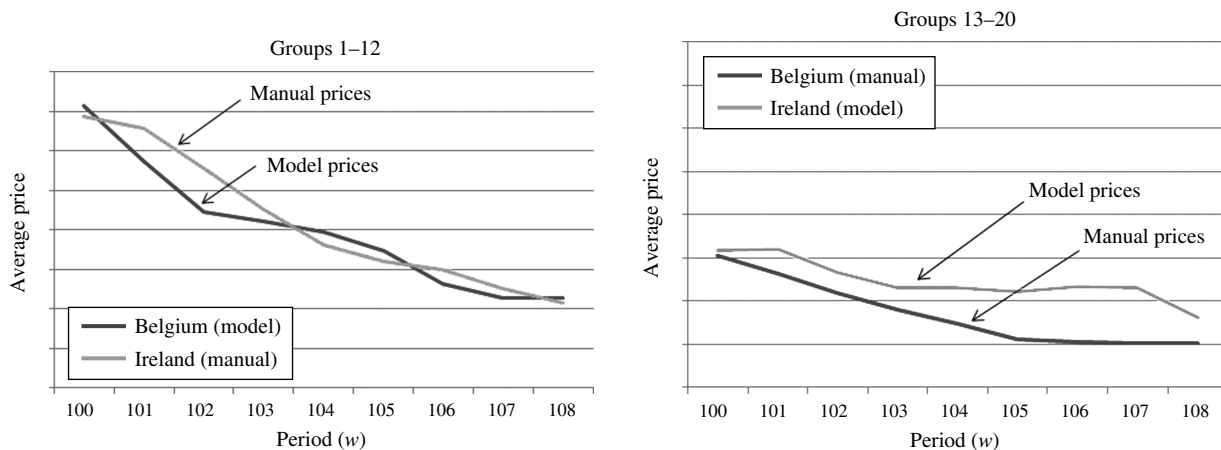


Table 4. Fraction sold (X^w) across periods for groups 1–12 and 13–20 in Belgium and Ireland.

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

Note. The last column reports the fraction sold two weeks after the last markdown.

a given group. Our interpretation is that the decisions by the pricing committee to aggregate price clusters and to select specific prices were occasionally driven by a desire to simplify the problem structure because of time constraints and perhaps cognitive limitations. This behavior sometimes led to ignoring of opportunities for maximizing revenues through finer segmentations of the offer across different products or different price points, which the model was able to recognize. We turn next to the financial impact of all these differences in pricing behavior.

5.2.5. Financial Impact Assessment. The overall impact of the model is based on the difference-in-differences metric Δ^q , $q = \text{BEL, IRL}$, defined in Equation (28). The results are summarized in Table 6. In particular, the first row corresponds to the figures for the live pilot in 2008. The mean and median of the first difference $Y_{1-12} - Y_{13-20}$ 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

–4.8 pp 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 2.2 pp, which provided a baseline value for these average differences. The estimated impact of the model on the realized income was therefore an increase of 2.7 pp in Belgium and an increase of 2.6 pp 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. Finally, we compared the 95% confidence intervals and verified that they did not overlap and had opposite signs as desired.

To validate our methodology, in rows three and four of Table 6 we report the difference-in-differences calculation for 2006 and 2007. Because 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 shown in the last two columns of Table 6 were consistently smaller in absolute terms than their equivalent in 2008. None of them was statistically significant at the 5% level, except for the 2007 difference in Belgium. We looked at this result carefully and considered it to be irrelevant. Indeed, it had a negative sign, which if anything indicated that there was even more merit to the positive result observed in 2008, and the latter was obtained with a significance level that was two orders of magnitude better than the value in 2007. Moreover, when the t -test was performed with the Welch correction, the result in 2008 remained unchanged, while in 2007 it became insignificant at the 5% level. Finally, the 95% confidence intervals for Belgium and Ireland overlapped completely in 2006 and 2007, which did not happen in 2008. We took this analysis as evidence that supported the overall methodology.

The remarks drawn from Table 6 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

Figure 6. Trajectory of the first difference $Y_{1-12}^w - Y_{13-20}^w$ in Belgium (BEL), Ireland (IRL), and the rest of western Europe (RWE).

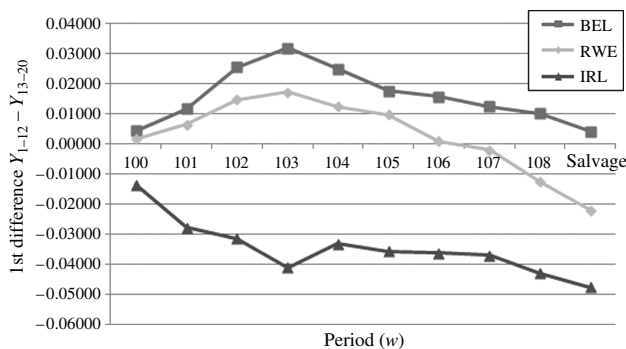


Table 5. Averages for groups 1–12 and 13–20 in Belgium and Ireland of the ratios of the number of different price categories used in each period to the initial number of price clusters.

Groups	Country	Clearance sales period w									Time averages
		100	101	102	103	104	105	106	107	108	
1–12	Belgium (model)	0.65	0.50	0.45	0.43	0.41	0.41	0.37	0.30	0.30	0.42
	Ireland (manual)	0.49	0.46	0.43	0.33	0.27	0.26	0.25	0.24	0.22	0.33
13–20	Ireland (model)	0.61	0.61	0.54	0.46	0.45	0.43	0.40	0.29	0.17	0.44
	Belgium (manual)	0.52	0.39	0.35	0.28	0.24	0.19	0.18	0.18	0.18	0.28

Note. The last column reports the averages of these ratios across time periods.

in Ireland. However, recall that the model recommendations were not implemented as closely for group 20 in Ireland (c.f. §5.2.3). Hence, it is important to observe that when group 20 was excluded from the calculations in Ireland, (see Table 7) the impact estimation on the realized income increased from 2.6 pp to 3.1 pp, 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 6 and for Ireland in Table 7, we concluded that the price optimization model increased the realized income Y by approximately 2.9 pp during the pilot experiment.

A rough historic average for the realized income Y at the end of clearance sales is 50%. Therefore, an increase in 2.9 percentage points in Y means a 5.8% increase in revenues. To see this, let v^{legacy} and v^{model} be the clearance revenue under the legacy and model-based markdown policies, respectively. Let V be the clearance inventory valued at regular season prices, so $Y^q = v^q/V$, with $q = legacy, model$. A 2.9-pp increase in realized income means that $Y^{model} = Y^{legacy} + 0.029$, which implies that $v^{model} = v^{legacy}(1 + 0.029/Y^{legacy})$. Replacing Y^{legacy} in this last equation by its historical value (50%) shows that v^{model} represents a 5.8% increase with respect to v^{legacy} . For a monetary value of this impact, we considered 2006 when Zara reported \$7,194 M

in revenues. Following Ghemawat and Nueno (2003), we assumed that 17.5% of sales were generated at markdown prices (see §1), which resulted in a clearance sales income of \$1,259 M (because Zara avoids markdowns during the regular season, we assumed that discounted sales prior to the clearance period were negligible). Therefore, the impact of the model in 2006 represented an increase of \$73 M—i.e., 5.8% of \$1,259 M—in clearance revenue. In 2007 this meant \$83 M in additional sales, and \$90 M in 2008.¹⁰

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) is the only other markdown system implementation in which the observed financial impact is reported. Specifically, the most successful case described in Smith and Achabal (1998) increased the realized income by 4% with respect to the previous year, which is lower than the 5.8% estimated in the present application. This difference in impact could be driven by the particular features of the systems implemented but might also be explained by the relative performance of the legacy markdown policy that was replaced, as well as the estimation methodology—for instance, the impact estimations in Smith and Achabal (1998) do not involve controls for external factors.

Table 6. 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.

Year	1st difference: $Y_{1-12} - Y_{13-20}$			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)
95% confidence interval				(1.2, 4.2)	(−5.1, −0.2)
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)
95% confidence interval				(−3.2, −0.3)	(−4.6, 0.0)
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)
95% confidence interval				(−2.4, 0.8)	(−4.1, 1.7)

Notes. Statistical significance from two-tailed test: * $p < 5\%$, ** $p < 1\%$, *** $p < 0.2\%$, **** $p < 0.04\%$. Figures are percentage points.

Table 7. Model impact assessment in Ireland without G20, i.e., the first difference is only between groups 1–12 (control) and 13–19 (intervention).

Year	1st difference: $Y_{1-12} - Y_{13-19}$		2nd difference
	IRL	RWE	IRL – RWE
2008			
Mean (median)	–6.9 (–6.2)	–3.8 (–3.8)	–3.1** (–2.5***)
<i>t</i> -stat (<i>z</i> -stat)			2.65 (3.16)
95% confidence interval			(–5.4, –0.8)

Notes. Statistical significance from two-tailed test: * $p < 5\%$, ** $p < 1\%$, *** $p < 0.2\%$, **** $p < 0.04\%$. Figures are percentage points.

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 clearance sales for all product groups (see the e-companion 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 and Bear, have shown interest in adapting this tool for their own stores.

In conclusion, our work is the first documented application of a complete multiproduct 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 that 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 clearance sales revenue by about 6%, corresponding for example to \$90 M in 2008. This financial impact is explained by the model's ability, relative to the legacy process, of maximizing revenue rather than liquidating stock. The model 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 it 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.

Electronic Companion

An electronic companion to this paper is available as part of the online version at <http://dx.doi.org/10.1287/opre.1120.1102>.

Endnotes

1. The implemented data collection module involves standard database access queries and is not discussed here.
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 $\max\{1, I_r^w/f\}$ in the regression model. However, we chose not to use it as an explanatory variable because 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).

3. In our implementation, for the second period we set $\gamma_2 = \gamma_3 = -\gamma_1 = 1$, and then for the third period onward we used $\gamma_1 = 0.15$, $\gamma_2 = 0.85$, $\gamma_3 = 0$.
4. This constraint was usually dropped in the final periods.
5. For this experiment, RWE consisted of Spain, Portugal, France, Italy, Austria, Holland, and the United Kingdom.
6. In most applications, the same subject is measured before and after receiving treatment. That is not possible at Zara because a particular article is sold only once, so instead we used the rest of western Europe as the baseline without treatment. This distinction does not affect the statistical procedure; see Stock and Watson (2003).
7. 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.
8. It should also be noted that the forecast in Fisher and Vaidyanathan (2009) is for an assortment problem instead of markdown pricing.
9. In Bitran et al. (1998) the test is preformed for each article. In contrast, we tested the Poisson assumption per cluster because our optimization model is formulated at that level.
10. The financial impact estimations provided here assume a 1.3 \$/Euro exchange rate. They were performed independently by the authors and do not engage the responsibility of the Inditex group.

Acknowledgments

This paper owes much to Rodolfo Carboni, who performed a substantial part of the implementation work at Zara. The authors 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 paper 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. Finally, the authors thank Steve Smith, Ming Hu, and Richard Lai for their useful remarks on a preliminary version of the paper, and they thank Izak Duenyas, the associate editor, and two anonymous referees for their comments that helped improve the paper.

References

Agrawal N, Smith SA (1996) Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Res. Logist.* 43:839–861.

Anderson E, Simester D (2003) Effects of \$9 price endings on retail sales: Evidence from field experiments. *Quant. Marketing Econom.* 1(1):93–110.

Araman V, Caldentey R (2009) Dynamic pricing for nonperishable products with demand learning. *Oper. Res.* 57(5):1169–1188.

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

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

Besbes O, Phillips R, Zeevi A (2010) Testing the validity of a demand model: An operations perspective. *Manufacturing Service Oper. Management* 12(1):162–183.

Bitran G, Caldentey R, Mondschein S (1998) Coordinating clearance markdown sales of seasonal products in retail chains. *Oper. Res.* 46(5):609–624.

Cachon G, Swinney R (2009) Purchasing, pricing, and quick response in the presence of strategic customers. *Management Sci.* 55(3):497–511.

Carboni R (2009) Clearance pricing optimization at Zara. M.S. thesis, LFM, Massachusetts Institute of Technology, Cambridge, MA.

Caro F (2012) Zara: Staying fast and fresh. The European Case Clearing House, ECCH Case 612-006-1, Anderson School of Management, University of California, Los Angeles, Los Angeles.

Caro F, Gallien J (2010) Inventory management of a fast-fashion retail network. *Oper. Res.* 58(2):257–273.

Caro F, Martinez de Albeniz V (2010) The impact of quick response in inventory-based competition. *Manufacturing Service Oper. Management* 12(3):409–429.

Dana JD, Petrucci NC (2001) Note: The newsvendor model with endogenous demand. *Management Sci.* 47(11):1488–1497.

DeHoratius N, Mersereau AJ, Schrage L (2008) Retail inventory management when records are inaccurate. *Manufacturing Service Oper. Management* 10(2):257–277.

Dong L, Kouvelis P, Tian Z (2009) Dynamic pricing and inventory control of substitute products. *Manufacturing Service Oper. Management* 11(2):317–339.

Duan N (1983) A nonparametric retransformation method. *J. Amer. Statist. Assoc.* 78(383):605–610.

Elmaghraby W, Keskinocak P (2003) Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Sci.* 49(10):1287–1309.

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

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

Fisher M, Vaidyanathan R (2009) An algorithm and demand estimation procedure for retail assortment optimization. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

Gallego G, Hu M (2007) Dynamic pricing of perishable assets under competition. Working paper, Rotman School of Management, University of Toronto, Toronto, Ontario.

Gallego G, van Ryzin G (1997) A multiproduct dynamic pricing problem and its application to network yield management. *Oper. Res.* 45(1):24–41.

Gaur V, Fisher M (2005) In-store experiments to determine the impact of price on sales. *Production Oper. Management* 14(4):377–387.

Ghemawat P, Nueno JL (2003) ZARA: Fast fashion. Harvard Business School Multimedia Case 9–703–416, Harvard Business School, Boston.

Heching A, Gallego G, van Ryzin G (2002) An empirical analysis of policies and revenue potential at one apparel retailer. *J. Revenue and Pricing Management* 1(2):139–160.

Jasin S, Kumar S (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, Stanford, CA.

- Kennedy P (2003) *A Guide to Econometrics*, 5th ed. (MIT Press, Cambridge, MA).
- Kesavan S, Gaur V, Raman A (2010) Do inventory and gross margin data improve sales forecasts for U.S. public retailers? *Management Sci.* 56(9):1519–1533.
- Little J (1966) A model of adaptive control of promotional spending. *Oper. Res.* 14(6):1075–1097.
- Maglaras C, Meissner J (2006) Dynamic pricing strategies for multi-product revenue management problems. *Manufacturing Service Oper. Management* 8(2):136–148.
- Manning W, Mullahy J (2001) Estimating log models: To transform or not to transform? *J. Health Econom.* 20:461–494.
- Mantrala MK, Rao S (2001) A decision-support system that helps retailers decide order quantities and markdowns for fashion goods. *Interfaces* 31(3):146–165.
- Musalem A, Olivares M, Bradlow ET, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. *Management Sci.* 56(7):1180–1197.
- Sigurdsson V, Foxall G, Saevarsson H (2010) In-store experimental approach to pricing and consumer behavior. *J. Organ. Behav. Management* 30:234–246.
- Smith S (2009) Clearance pricing in retail chains. Agrawal N, Smith S, eds. *Retail Supply Chain Management*, Chapter 11 (Springer Science + Business Media, New York), 271–291.
- Smith S (2011) Personal communication.
- Smith S, Achabal D (1998) Clearance pricing and inventory policies for retail chains. *Management Sci.* 44(3):285–300.
- Smith S, McIntyre S, Achabal D (1994) A two-stage sales forecasting procedure using discounted least squares. *J. Marketing Res.* 31(1):44–56.
- Soon W (2011) A review of multi-product pricing models. *Appl. Math. Comput.* 217(21):8149–8165.
- Stock JH, Watson MW (2003) *Introduction to Econometrics* (Pearson Education, Inc., Boston).
- Talluri K, van Ryzin G (2004) *The Theory and Practice of Revenue Management* (Springer, New York).
- Valkov T (2006) From theory to practice: Real-world applications of scientific pricing across different industries. *J. Revenue and Pricing Management* 5(2):143–151.
- Verdugo O (2010) Coordination of inventory distribution and price markdowns for clearance sales at Zara. Master's thesis, LGO, Massachusetts Institute of Technology, Cambridge, MA.
- Vulcano G, van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Oper. Res.* 60(2):313–334.

Felipe Caro is an assistant professor of decisions, operations, and technology management at the UCLA Anderson School of Management. His research interests span retail operations, supply chain management, carbon footprint, and natural resources, with a strong emphasis on practical applications.

Jérémie Gallien is an associate professor of management science and operations at London Business School. His research focuses on supply chain management, global health delivery and logistics, hospital and healthcare operations, mathematical modeling, and simulation.