Operations Management in Apparel Retailing: Processes, Frameworks and Optimization

Felipe Caro
UCLA Anderson School of Management
110 Westwood Plaza, Suite B420
Los Angeles, CA 90095, USA
fcaro@anderson.ucla.edu

Victor Martínez-de-Albéniz
IESE Business School
University of Navarra
Av. Pearson 21 08034 Barcelona, Spain
vs valbeniz@iese.edu

Abstract

Apparel retailing is a large industry that requires managing simple yet highly uncertain and time-sensitive processes. Operations management frameworks can help in taking good design, production and distribution decisions quickly, using all the available information. Indeed, the success of “fast-fashion” retailers such as Zara is built on the use of simple principles that allow these companies to react to rapid changes in market conditions. In this article, we review the decision models at the basis of the current best practices.

Keywords: Assortment planning, Sourcing, Inventory management, Distribution policies.
AMS Subject classifications: 90B05, 90B30.

1. Introduction

The purpose of this article is to provide an overview of apparel retailing from an operational perspective. We start by providing an overview of the industry in Section 2 and the existing practices in Section 3. This allows us to identify three types of problems that can be analyzed using Operations Research techniques. We start from the last decision to be made: how to distribute inventory over the network of stores. The description of this problem and several solution
approaches are discussed in Section 4. We then consider the previous decision, prior to distribution: how much and when to buy of one particular product. We discuss the different models in Section 5. We finally tackle the first decision: how many and which products to design, and when to introduce them into the stores. This is presented in Section 6. We conclude in Section 7 with a brief overview of future research opportunities.

2. The Apparel Retailing Industry

Apparel is defined as “something that covers or adorns” and refers to outer garments or clothing. As such, it includes a number of subcategories such as clothing, footwear or accessories. Studies suggest that clothing has existed for more than 100,000 years (Kittler et al. 2003). The industry to make and sell apparel has a long history too, with changes in manufacturing processes (e.g., development of denim in the 1600s, mechanized textile looms in the 1760-70s, sewing machines in 1840s, etc.), distribution/retail processes (e.g., department stores in the 1830s, mail ordering in the 1880s, store chains after World War II, online sales in the 1990s) and in customer needs (fashion trends).

Given that apparel is a basic good, consumed by everyone, and easy to produce (e.g., this is a sector that developing countries focus on at the early stages of industrialization), the influence of this industry in the economy is quite significant. The annual size of the industry was USD 2.6 trillion globally in 2010 (Treehugger 2012). In terms of household consumption, €366 billion was spent in clothing in the EU in 2010 (Eurostat 2013), for USD 348 billion in the US (OECD 2013). Production is nowadays strong in low-cost countries, e.g., in Bangladesh, where the textile industry contributes 81.6% of export earnings (World Trade Organization 2012).

Given the relatively low barriers to entry in the industry, it is not surprising that there is a very large number of retailers involved. For example, there are more than 1,400 brands with significant market share reported in Euromonitor International (2013). Unfortunately, there are no standard rankings of apparel retailers, because general merchandise retailers also sell apparel (e.g., Wal-Mart, Marks and Spencer) and luxury goods are usually separated from mass-produced goods. We provide in Figure 1 the evolution of sales of some of the largest retailers according to Forbes (2012). We can see that The Gap, a long time leader, has been surpassed by H&M and Inditex over the last years.

3. Industry Practices

In order to bring a product to market, companies have traditionally organized work around collections: the Spring-Summer collection (typically released in January-February) and the Fall-Winter collection (typically released in July-
August). Thus, twice a year, firms get rid of the old collection in a discount period, which in countries like Spain was regulated by law; they in parallel introduce the new collection. As a result, the set of products during the season is fixed and renewed periodically. Although some of the products in the new collection may be the same as in the old one (called carry-over or repetition products), most of them have a limited life on the shelf by design.

The traditional process from design to market is relatively simple and we describe it for a collection to be released in January 2013. The process starts with the design phase, where creative designers create a product concept, both digitally (through computer-aided design CAD software) and physically (through the making of samples of fabric, prints, etc.). The duration of this phase greatly varies across firms: some firms start designing very early (in November 2011) so as to have the full collection designed on time (in May 2012); while others can postpone the design to a few months before the next phase takes place. One of the key determinants of when the collection needs to be designed is the fact that some firms have a wholesale channel, selling to multi-brand stores, department stores, etc. They thus need to have the complete offer to show these customers so as to take early orders from them. Once design is completed, the firm contacts suppliers (external or in-house) and places production orders with them. The production phase then begins. First, the fabric is made from thread of raw material such as cotton, linen, etc. or procured for certain materials such as leather. Second, the fabric undergoes a treatment to fix its touch and feel, including color dying, washing or printing (although for certain items this can be done later, e.g., printing T-shirts after cutting or washing jeans for wrinkles after sewing). Third, the fabric is cut into pieces for the different products.
Fourth, the product is “assembled”, through the sewing of the pieces together into the finished product. Fifth and final, the product is packed and shipped to the retailer’s warehouse. This part of the process requires specific equipment and labor. Economies of scale are significant due to the fact that most of the work is independent of the quantity of product to be produced, and that the labor-intensive step has a steep learning curve due to quality issues, and thus requires sufficient volume to be cost-competitive. The timing of this phase varies depending on supplier lead-time. For example, for a firm selling in Europe, for flat-knit fabrics made in China and transported by sea, lead-time may be 4-6 months, which would require the production order to be finalized by July 2012; for knitwear made in Turkey (thus shipped by truck), lead-time may be 1 month, which would allow the production order to be postponed until November 2012. Finally, once the product is in the retailer’s warehouse, it needs to be distributed to the stores. This final part of the process is very quick typically (days if transportation is by road within Europe or by air). There are two types of shipments involved. Initially, stores are loaded with large quantities of inventory at the beginning of the season (in January 2013). They are then restocked during the season in small quantities, which is called replenishment.

In recent years, a new breed of retailers has challenged this traditional process. These are players such as the Inditex group (including Zara, Bershka, Massimo Dutti, Pull and Bear or Stradivarius), H&M or Topshop, who are called fast-fashion retailers, or pronto moda in Spanish. Their practices are well documented, see Ferdows et al. (2002), Ghemawat and Nueno (2003), McAfee et al. (2004), Lewis et al. (2004) or Caro (2012). They have undertaken a radical change to the approach so as to provide fashion almost on demand: “When Madonna gave a series of concerts in Spain, teenage girls were able to sport at her last performance the outfit she wore for her first concert, thanks to Zara” (The Economist 2005). Specifically, they have chosen to work at the item level, rather than using collections. They can do this because they typically do not have a wholesale channel that is demanding a full collection, and they control the retail point of sales. This move allows them to avoid batching thousands of products together, and to accelerate lead times across the board. For example, it is no longer needed to design together products with quick and slow supplier lead times. Moreover, this also gives the freedom to introduce products in the store continuously, not only twice a year. This implies that the utilization of all resources (designers, factories, distribution) can be balanced better over time, avoiding unnecessary peaks twice a year. Costs and response times can thus be reduced. Furthermore, fast-fashion retailers have typically used quick-response production to reach stores as soon as possible, thereby allowing them to respond to nascent demand trends first, so as to provide and capture more value from the consumers. This requires them to accelerate not only the production phase, by using nearshore suppliers close to market, in countries such as Portu-
gal, Morocco, Bulgaria or even Turkey, but also the design phase, by directing the creative aspects towards a commercial need to reduce design iterations, and by using standard methods and materials to reduce efforts on samples. As a result, the total design-to-market time for an item to be launched in January 2013 can be reduced to a mere 6 weeks if the appropriate fabric is used and the go decisions (authorizations to move from sample to industrialization) are not delayed. In a way, they are like a surfer that is able to catch a wave before any other notices it. Both systems are compared in Figure 2.

![Figure 2: Traditional vs. fast-fashion design-to-sales processes.](image)

Regardless of whether we are considering traditional or fast-fashion retailers, there are mainly three types of operational decisions that companies face. These are taken sequentially. First, design needs to decide a strategy of how many and which products to make, and when to introduce them over the season (although traditional firms push them all at the same time at the beginning of the season). Second, production/purchasing needs to decide a strategy of how much and when to buy of each of the items that have been designed. Third, distribution needs to decide when inventory should be transferred between the central warehouse(s) and the stores. Since the former decisions should incorporate the impact they have on the latter ones, we analyze them in reverse order.

4. Distribution Decisions

During the selling season, the retailer’s ability to influence the sales of a product is limited. Indeed, design and production have been finalized, and the retailer can only place the inventory in the right place at the right time at the
right price so as to maximize the sell-out, i.e., the ratio of quantity sold divided by the quantity purchased. In the clearance sales period (e.g., January and July, or more generally when the product is phased out), price reductions are typically used to push up the sell-out (Caro and Gallien 2012). In contrast, during the full-price selling season, the retailer’s focus is on distributing the inventory correctly over the network, and in particular in deciding how much to keep in the central warehouse vs. in the stores. This becomes especially relevant when the product is successful, i.e., when stock-outs are common, because a proper centralization of inventory allows the retailer to ship inventory to the stores where it is needed.

To model the distribution decision, two-echelon inventory models have been used, e.g., Federgruen and Zipkin (1984), Robinson (1990), Nahmias and Smith (1994), Graves (1996), Axsäter et al. (2002) among others. Most of the work allows for warehouse replenishment from a production facility, although in apparel there is typically either no replenishment opportunity or at most one. The problem is usually described via dynamic programming. There are $T$ periods, $t = 1, \ldots, T$ and $N$ stores $n = 1, \ldots, N$. In period $t$, the starting inventory in the warehouse is denoted $x_{0t}$, and the inventory in store $n$ is $x_{nt}$. Shipments from warehouse to store $n$ are denoted $q_{nt} \geq 0$ and increase the inventory at each store to $y_{nt} = x_{nt} + q_{nt}$, without any delay (which is a realistic approximation because in apparel lead-times take 1-3 days, while periods are weeks), while the inventory at the warehouse becomes $y_{nt} = x_{nt} - \sum_{n=1}^{N} q_{nt} \geq 0$. Demand in each point of sales is then realized. It is denoted $D_{nt}$ and is assumed to be stochastic; if it exceeds the available inventory, sales are lost.

The dynamic program can be expressed as follows, using the profit-to-go function $J_t(\bar{x}_t) = J_{T+1}(\bar{x}_{T+1}) = 0$ and

$$J_t(x_{0t}, \bar{x}_t) = \max_{\bar{q} \mid \sum_{n=1}^{N} q_{nt} \leq x_{0t}} \mathbb{E} \left\{ - \sum_{n=1}^{N} c_n q_{nt} + r_n \min\{D_{nt}, y_{nt}\} + J_{t+1}(y_{nt} - \bar{D}_t) \right\}.$$  

(4.1)

It is easy to show that $J_t$ is jointly concave in $(x_{0t}, \bar{x}_t)$ for all $t$, and as a result, it is optimal to follow a state-dependent base-stock policy in each period: there exists $y_{nt}$ (function of $x_{nt} + \sum_{n=1}^{N} x_{nt}$) such that if $x_{nt} \leq y_{nt}$ for all $n = 1, \ldots, N$, it is optimal to set $y_{nt} = y_{nt}$.

Heuristics exist to approximate the solution of this dynamic programming. The most common one assumes that inventory is balanced, see e.g., Zipkin (1984). The problem can then be decomposed into single-dimensional dynamic programs, which are weakly coupled (Adelman and Mersereau 2008).

Alternative solution approaches, specifically tailored to the apparel industry, have been proposed too. For instance, Caro and Gallien (2010) and Caro et al. (2010) describe an integer program developed to optimize shipment decisions at
Zara. One of the main features of the model is incorporating Zara’s store level display policy by which a product is removed from the shop floor to the backroom when it is missing the major sizes. Hence, distribution must be done carefully to make sure that scarce inventory is shipped to stores where the combination of sizes available will be put on display. Through a controlled field experiment it was shown that using the optimization model increased sales by 3-4%.

Notice that one of the key inputs of the distribution problem is the initial quantity in the warehouse at the beginning of the season. If it is too high, stores can essentially be managed independently, and the problem becomes quite simple. If it is too low, all the stores will stock out early, so again there is little room for optimization. We examine next how the purchase quantity should be determined.

5. Sourcing Decisions

Sourcing decisions are critical to good distribution and sales performance. It has therefore been extensively studied in the literature. To focus on the quantity decision, most work ignores the fact that sales occur over a network, and rather considers a single aggregate stochastic demand that the company faces. There are two main streams of work: single-purchase and multiple-purchase models.

Single-purchase models are usually variations of the well-known newsvendor or newsboy model in inventory management (Zipkin 2000, Axsäter 2006). The firm faces a stochastic demand \( D \), with c.d.f. \( F_D \) and to serve it, buys a quantity \( q \), at a cost \( c \) per unit. If an item is sold, it generates a revenue of \( r > c \) and, if not, it can be salvaged at a value \( v < c \). The retailer thus maximizes expected profits:

\[
\max_q \mathbb{E} \left[ r \min\{D, q\} + v \max\{q - D, 0\} - cq \right].
\]  

(5.1)

The solution is to buy the critical fractile quantity \( q \), so that

\[
F_D(q^\star) = \frac{c - v}{r - v}.
\]  

(5.2)

This type of model is appropriate when the production lead-times are so large that it is impossible to obtain more production to reach the stores before the end of the season, once the season has started. In contrast, when lead-times are shorter, it is possible to use early demand information to refine the forecast and produce more when demand turns out to be high. This requires a model with multiple purchase opportunities.

Multi-purchase models have been studied in the inventory management literature. For retail applications with finite horizons, it is important to consider


how demand forecasts can be improved over time. Thus, it is necessary to consider correlation of demands $D_t$ over time. Consider for simplicity that there are two purchase opportunities, as in Martínez-de-Albéniz (2011). There is an initial demand forecast and, after early sales $d_1$ are observed, it is updated into a new demand forecast with c.d.f. $F_{D|d_1}$. As a result, the total demand is $D$, where $D$ is dependent on the realization of $D_1$. Inventory can be purchased at cost $c_1$ before $D_1$ is realized, or at cost $c_2 > c_1$ after the forecast has improved. We denote $q_1, q_2$ the quantities purchased in each case, respectively. Keeping the same cost structure $r, v$ for full-price revenue and salvage value, the problem can be solved by backward induction: given $d_1, q_1$, $q_2^*(d_1, q_1)$ satisfies

$$F_{D|d_1}(q_1 + q_2^*) = \frac{c_2 - v}{r - v}.$$  

(5.3)

The retailer can solve the initial problem, so as to maximize

$$E_{D_1}\left[\Pi(D_1, q_1) - c_1 q_1\right],$$

where $\Pi(D_1, q_1) = \max_{q_2} E_D[r \min\{D, q_1 + q_2\} + v \max\{q_1 + q_2 - D, 0\} - c_2 q_2 | D_1]$.

Since $\partial \Pi / \partial q_1 = \min\{c_2, v + (r - v) F_{D|D_1}(q_1)\}$, we can determine the optimal $q_1$:

$$E_D \min\left\{c_2 - v, v : F_{D|D_1}(q_1)\right\} = \frac{c_1 - v}{r - v}.$$  

(5.4)

This type of result can be extended to more than two periods, e.g., Wang et al. (2012), and to incorporate fixed ordering costs, e.g., Song and Zipkin (1993). It has been used in various settings. Fisher and Raman (1996) include production capacities and minimum batches. Iyer and Bergen (1997) compare the systems with single purchase, either before or after forecasts are updated, and discuss when quick response is Pareto-improving for retailer and manufacturer. Agrawal et al. (2002) describe an integrative decision model to determine volume commitments to suppliers. Fisher et al. (2001) and Li et al. (2009) incorporate a replenishment lead-time for the second order. Including lead-times usually makes the problem intractable, as the optimal policy loses the base-stock structure, see Fukuda (1964) or Whittmore and Saunders (1977). Heuristics are developed instead, e.g., Veeraraghavan and Scheller-Wolf (2008) or Allon and van Mieghem (2010). The effect on supply chain incentives has also been studied: Anand et al. (2008), Erhun et al. (2008), Martínez-de-Albéniz and Simchi-Levi (2012) and Calvo and Martínez-de-Albéniz (2012) discuss the effect of multiple purchasing opportunities on supplier pricing, and Krishnan et al. (2010) study the consequences of quick response on retailer efforts.
6. Design Decisions

Product decisions are the first step in planning an apparel season. Here the main challenge is to select a combination of products that conform an attractive assortment. From an optimization perspective, a major source of complexity is how to deal with substitution effects. Several consumer choice models have been proposed in the literature, with the multinomial logit (MNL) being the most recurrent due to its tractability (Anderson et al. 1992). If the “universe” of products available is \( P = \{1, \ldots, |P|\} \), then the assortment problem can be recast as find the subset \( S \subseteq P \) that maximizes expected profits. Given an assortment \( S \), a quantity of interest is the probability \( q_j \) that consumers will choose product \( j \in S \). Since the market size can always be normalized to one, \( q_j \) represents the demand for product \( j \) and it depends on the other products available in \( S \).

In a single-period or single shot problem, the assortment \( S \) is chosen once and for all. The problem can be simplified by assuming that products have the same retail price \( r \) and procurement cost \( c \), which could be reasonable when products are horizontally differentiated (e.g., shirts with different colors). A further simplification is to assume that consumers do not substitute based on the items they see in stock but rather based on whether they were included or not in the initial selection \( S \) (the latter is called assortment-based substitution whereas the former corresponds to stockout-based substitution). In this setting, finding the optimal assortment \( S \) can be formulated as the following optimization problem:

\[
\max_{S \subseteq P} \sum_{j \in S} [(r - c)q_j - C(q_j)],
\]

(6.1)

where \( C(q_j) \) is an increasing and concave function that represents any operational costs associated with offering product \( j \). There are \( 2^{|P|} \) possible assortments, so in principle, solving the assortment problem (6.1) is hard when \( |P| \) is large. Fortunately, the solution can be characterized in some important special cases.

Under the MNL model, the chance that a consumer will choose product \( j \in S \) is given by \( q_j = \frac{v_j}{\sum_{i \in S} v_i + v_0} \), where \( v_j \) is the preference or attractiveness of product \( j \) and \( v_0 \) is the weight of the no-purchase option (it is convenient to sort the products in decreasing order of attractiveness, so \( v_1 \geq v_2 \geq \ldots \geq v_n \)). For this case, van Ryzin and Mahajan (1999) show that the optimal assortment \( S^* \) that solves problem (6.1) is always a popular set containing a serial sequence of the most attractive products (formally, a popular set is any subset of the kind \( \{\}, \{1\}, \{1, 2\}, \ldots, \{1, 2, \ldots, |P|\} \)). This result simplifies problem (6.1) to just evaluating the expected profit of the \( n + 1 \) popular sets. Kök and Xu (2011) extend this fundamental result by allowing endogenous prices and by considering
a nested MNL model, which allows for product groups or subcategories such that products are homogeneous within each group but heterogenous across groups. It is worth pointing out that there is a revenue management variation of this problem in which all costs are sunk and products have different prices \( r_j, j \in P \). In that context, the optimal assortment is \textit{revenue-ordered} in the sense that it is an ordered sequence of the most expensive products (Talluri and van Ryzin 2004, Rusmevichientong and Topaloglu 2011).

Besides the MNL case, the assortment problem (6.1) has been studied under other choice models. For instance, Smith and Agrawal (2000) look at a model in which consumers make a single substitution attempt while Gaur and Honhon (2006) consider a Hotelling-Lancaster locational model. From an implementation standpoint, Fisher and Vaidyanathan (2009) develop a procedure to estimate the probabilities \( q_j \) from sales data. All these papers assume assortment-based substitution. The problem with stockout-based substitution adds an additional layer of complexity, which is studied in Mahajan and van Ryzin (2001) and Honhon et al. (2010). For further references on the (single-period) assortment problem, see the surveys by Kök et al. (2008) and Mantrala et al. (2009).

The fast-fashion context does not fit the single-period problem because products are introduced regularly. Moreover, fast-fashion retailers like Zara can use this capability to do some trial-and-error to learn what styles are selling. Hence, finding the optimal assortment becomes a dynamic problem with demand learning. Caro and Gallien (2007) tackle this problem in which the main trade-off is between \textit{exploration} and \textit{exploitation} (exploration consumes resources that otherwise could be used to sell “safe bets”). When products are independent, Caro and Gallien (2007) show that an index policy is near-optimal (an index policy dictates that the assortment in each period should be composed by the products with the highest indices). The indices are based on the unknown demand rates and have a mean-variance form, where the variance is weighted by a factor that decays as the season expires. The model is extended to allow for single substitution attempts (as in Smith and Agrawal 2000), which requires solving a quadratic knapsack problem. Ulus et al. (2012) study the same exploration versus exploitation trade-off but under the Hotelling-Lancaster locational model.

7. Conclusions and Research Opportunities

We have identified here three main decisions that are key determinants of success in fashion apparel retailing. First, design decisions require a good understanding of how consumers choose among the products within the assortment. Choosing the right assortment requires balancing the attractiveness for the consumer, the costs associated with an increased variety and the possible learning that can be achieved if assortments can be changed over the season. Second, purchasing decisions require managing the risks of over-ordering and under-ordering
compared to the demand. There, it is important to improve the forecasts by observing early sales, and order more if needed. Third, once the assortment and the purchase quantities are fixed, distribution decisions need to properly locate inventory across a network of stores. The main objective is to stock enough for the stores to provide adequate service and replenish them as sales are realized.

In this article we have summarized most of the existing literature related to apparel retailing. However, there are many research opportunities in this field. For example, there is ample room for more work on analytical models of retail processes. Changing practices indeed require integrating new approaches to the existing models, such as how to manage dynamic assortments and product introductions, how to manage store space, how to guide in-season purchasing and design, etc. These problems require complex dynamic optimization techniques. Furthermore, the changing nature of market trends calls for an advanced use of stochastic demand models, continuous and discrete. Moreover, game theoretical analyses can be useful to better understand the strategic interactions between retailers. For example, while dynamic assortments can only be beneficial for a retailer taken in isolation, under competition they have the risk of triggering a “product war” where all retailers launch many new products, thereby increasing their operating costs, without expanding their market shares. Finally, it is worth mentioning that the field is ready for more empirical work too: data is abundant and statistics can help determine the true value of the existing practices.

Acknowledgements

Research supported in part by the Spanish Ministry of Economics and Competitiveness (Ministerio de Economía y Competitividad, formerly Ministerio de Ciencia e Innovación) – ref. ECO2011-29536.

References


B. Dankworth, and D. Shah. 2009. Why is assortment planning so difficult
71–83.

[37] Martínez-de-Albéniz, V. 2011. Using Supplier Portfolios to Manage Demand
Risk. In Handbook of Integrated Risk Management in Global Supply Chains,
Sons Inc.

[38] Martínez-de-Albéniz, V., and D. Simchi-Levi. 2012. Supplier–Buyer Nego-
tiation Games: Equilibrium Conditions and Supply Chain Efficiency. Production
and Operations Management forthcoming.


[40] Nahmias, S., and S. A. Smith. 1994. Optimizing inventory levels in a two-
echelon retailer system with partial lost sales. Management Science 40 (5):
582–596.


[42] Robinson, L. W. 1990. Optimal and approximate policies in multiperiod,
multiplocation inventory models with transshipments. Operations Research 38 (2):
278–295.

[43] Rusmevichientong, P., and H. Topaloglu. 2011. Robust Assortment Opti-
mization in Revenue Management Under the Multinomial Logit Choice Model.
USC Marshall School of Business, working paper.

Inventories Systems with Demand Substitution. Operations Research 48:60–
64.


general discrete choice model of consumer behavior. Management Science 50
(1): 15–33.


About the Authors

Felipe Caro is an Associate Professor of Decisions, Operations, and Technology Management at the UCLA Anderson School of Management. He holds a Ph.D. in Operations Management from the Massachusetts Institute of Technology and an Industrial Engineering degree from the University of Chile. His research interests span retail operations, supply chain management, carbon footprinting and natural resources, with a strong emphasis on practical applications.

Victor Martínez-de-Albéniz is an Associate Professor of Production, Technology and Operations Management at IESE Business School - University of Navarra. He holds a Ph.D. from the Massachusetts Institute of Technology and an engineering degree from École Polytechnique in France. His research focuses on supply chain management, where procurement, production and distribution decisions can help companies compete more successfully in the global arena. More recently, he has worked on retail topics, especially in the fashion industry.