Chapter 4
The Role of Quick Response in Accelerating Sales of Fashion Goods

Alejandro Lago, Victor Martínez-de-Albéniz, Philip Moscoso and Andreu Vall

Abstract Quick response has been proposed as an appropriate operational strategy to serve volatile markets. In fashion, postponing design, production, and distribution as much as possible may indeed reduce the uncertainty related to product success. In this paper, we provide an empirical study of the influence of lead time and sourcing origin on product success, based on data provided by a European fast fashion retailer. We provide a model of sales diffusion over time where product success is characterized by the *speed of sales*. We then evaluate how the speed of sales is influenced by the design time and the time-to-market of each particular product. We find that delaying the time of design is very beneficial, because it allows the firm to learn about fashion trends. The effect of time-to-market is more subtle. For a shorter time-to-market, speed of sales is considerably higher, but there is limited learning obtained by postponing design. In contrast, for longer time-to-market, speed of sales is lower, but the learning is higher, so for products designed late in the season, the speed of sales is similar to that of items with short time-to-market.

Keywords Dynamic demand model · Demand estimation · Lead time management

---

A. Lago (✉) · V. Martínez-de-Albéniz · P. Moscoso · A. Vall
IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain
e-mail: alago@iese.edu

V. Martínez-de-Albéniz
e-mail: valbeniz@iese.edu

P. Moscoso
e-mail: pmoscoso@iese.edu

A. Vall
e-mail: andreu.vall@jku.at

T.-M. Choi (ed.), *Analytical Modeling Research in Fashion Business*,
Springer Series in Fashion Business, DOI 10.1007/978-981-10-1014-9_4

4.1 Introduction

Reducing the lead time required to design, manufacture, and distribute the products is nowadays a powerful source of competitive advantage for fast fashion companies. Indeed, in apparel retailing, supply chain management has undergone a major shift of focus in the last decades: While historically supply efficiency was the primary concern, since the 1980s, the need for agility and speed to respond to demand trends has received increased attention both from industry and academia (Christopher et al. 2004).

Following introduction of pull production concepts in the early 1980s, which linked production decisions to evidence of downstream demand, a major milestone in supply chain management was quick response (QR), an apparel manufacturing initiative initiated in the USA in the 1980s (Hammond and Kelly 1990). Essentially, QR is a set of practices aiming at a better match between supply and demand. Through the years, QR practices have evolved and transferred to other industries and eventually have become a prevailing principle in supply chain management (Fisher 1997).

The primary objective of QR is to postpone ordering decisions as much as possible to benefit from better demand information and consequently reduce the risk of demand–supply mismatches. As such, QR allows a company (1) to make comparatively smaller investments in inventory and therefore mitigates the cost of stock clearances, as well as (2) reduces the opportunity costs of lost sales with to the possibility of faster replenishment. To achieve these goals, QR practices target the reduction of lead times and setup costs. Companies try to reduce lead times sufficiently so that production can be done in response to initial sales information (Fisher and Raman 1996).

Firms like Benetton first attempted to reduce lead time by postponing distribution through expedited logistics methods. Later on, QR practices were extended to postponement of production as well, typically achieved by nearshoring through local suppliers, and in recent years, of product design as well. This has enabled companies to respond even quicker to market trends and introduce new products very frequently, which is a key characteristic of fast fashion players such as Zara (Caro and Martínez-de-Albéniz 2015).

QR and fast fashion practices have been discussed extensively in the operations management literature. However, to the best of our knowledge, very limited empirical analysis exists on the practical impact on industry. There are unresolved fundamental questions such as whether a company that reduces its lead time sells more. The objective of this paper is thus to empirically establish whether lead time reduction is valuable, and if so, when and how much. We present results from an original data set, provided by a European fast fashion retailer for the fall–winter 2008 season. This single environment provides an experimental setting where product success is solely driven by variations in lead times and supply conditions. Specifically, the same type of product may be sourced from different types of suppliers, ranging from offshore low-cost suppliers to nearshore suppliers with significantly shorter lead times. This variability in the lead times for the same category of goods allows us to discriminate the value of QR.

To study the data, we develop a model of product success which is characterized by the *speed of sales*. This metric is defined as the percentage of total inventory available at the firm that is sold on average per unit time. It is a proxy for product success in our fast fashion context, for three main reasons. First, the faster a certain portion of a product is sold at full price, the less likely markdowns are needed at the end of the season because of excess inventory. Hence, a higher speed of sales is associated with a reduction in the excess-inventory cost present in traditional QR models. Furthermore, because new products are introduced continuously all along the season, including some that may resemble the current product, a higher speed of sales is associated with additional sales of similar items, similar to the reduction of lost-sales cost in traditional QR models. Second, retailers can introduce more frequently new products in their stores, so in a given period (e.g., a year), more units will be sold per square meter; hence, a higher speed of sales is linked to a lower opportunity cost of space per unit time that the end-of-season demand–supply mismatch approach in the traditional QR literature disregards. Finally, there is an indirect effect associated with higher speed of sales: The higher frequency of introductions will often lead to customer visiting the stores more often and typically buying more products (Bernstein and Martínez-de-Albéniz 2014). This is related to making strategic customers experience product scarcity (Cachon and Swinney 2011), which leads them to procrastinate less, i.e., postpone less often a purchase in the hope of later markdowns.

Our empirical analysis tests how this metric is influenced by the sourcing choices made on the different products, controlling for all other observable product features such as color, product type, average sales price, or cost. In particular, we focus on the role of two main variables: the time of design of the item, which can be delayed if an item is nearshored and will be important if fashion trends can be learned during the season; and the lead time or time-to-market itself, which becomes relevant if trends drift randomly, in which case a shorter time between design and sales should increase the success of the product.

We uncover several important findings. First, we find that the time of design strongly influences product success: Products designed later in the season tend to be more successful and sell faster. This confirms that fashion trends can be learned over the season and that products designed later will be more attractive to the customers. Second, we analyze the effect of time-to-market and find that short time-to-market makes products more successful, by penalizing items made offshore, which on average will sell slower. We also observe that the uncertainty in success rate is higher for offshore items, which suggests that there are more ‘flops’ in offshore products compared to nearshore ones. Third, we study the interaction between design time and time-to-market: We find that the learning obtained by delaying design is faster when time-to-market is higher, for offshore origins. In other words, for items with long time-to-market, being able to delay design time is very valuable, because the improvement of information about fashion trends can be translated into better products; in contrast, for items with short time-to-market, information cannot be improved by much when design time is delayed. As a result, speeds of sales of items designed early in the season are much higher for nearshore products compared to offshore ones, while they are indistinguishable when design happens later in the season. Hence, our work provides a first empirical evidence that QR, through delayed design and short time-to-market, can create significantly better products, even though they may be more expensive.

The remainder of this paper is organized as follows. Section 4.2 reviews the relevant literature. We present in Sect. 4.3 a sales diffusion model that we use to identify the parameters that determine product success. We then discuss in Sect. 4.4 the data and the empirical findings. Finally, Sect. 4.5 summarizes the conclusions, discusses some implications of this work, and lists some topics for further research.

4.2 Literature Review

Quick response (QR) is 'a set of practices designed to reduce the cost of mismatches between supply and demand' (Cachon and Terwiesch 2009). These practices include, for example, lead time reduction, postponement, product differentiation, and better forecasting. There is a broad literature on QR and fast fashion related to the present paper. On the one hand, starting from the mid-1990s, there are numerous papers presenting analytical models of QR practices and analyzing their implications. These papers suggest that the closer one gets to actual sales realization, the smaller the forecast error, and this is why QR leads to lower inventory costs and lower lost sales. We confirm this with our data, although we do not follow the traditional demand–supply mismatch approach. On the other hand, there is also some work of empirical nature focusing on the relationship between inventories and sales. We complement this stream of literature by providing an empirical estimate of the benefits of QR on the speed of sales.

4.2.1 Analytical Models

As companies aim at increasing their sales and margin through the introduction of new products in increasingly shorter cycles, they realize that these sales also become more unpredictable. Such innovative products are thus challenging in terms of inventory and/or stock-out costs, and companies will need for a responsive supply chain, rather than an efficient supply chain adapted to functional, stable, and lower-margin products (Fisher 1997). QR practices are the operational levers to enable such responsiveness. QR has been extensively studied and modeled in literature as a mean to reduce the risks of supply–demand mismatches.

Fisher and Raman (1996) and Iyer and Bergen (1997) proposed the first models that capture the improvement in the demand forecast between an initial ordering opportunity and a later one. The key trade-off is that in the later order, forecast is more reliable, but cost is higher. In centralized settings, QR combined with optimized decision making can significantly increase profits for a firm facing stochastic demand (Fisher et al. 2001). Fisher (2009) shows how forecasts can be improved by using initial sales data. More sophisticated demand models have also been proposed so that dynamic forecast updates can be incorporated, e.g., Heath and Jackson (1994). Song and Zipkin (2012) analyze such inventory planning problems with uncertain demand as a general case of the newsvendor problem. They observe that the dynamic programming formulation is closely related to a serial inventory system with stochastic demand over an infinite horizon. As a result, one can characterize the optimal policy and assess the value of additional demand information.

The postponement of supply to take advantage of better information can take different forms. First, one can consider the postponement of the distribution of products to stores, where the central pooling of inventory can provide higher sales with the same inventory levels. Caro and Gallien (2010) dynamically optimize shipments from a central warehouse to a network of stores and present how the proposed system was implemented at Zara. The implementation is further detailed in Caro et al. (2010). Second, it is possible to postpone manufacturing too. This practice is related to the decision on where the point of product differentiation should be located (Lee and Tang 1997). In particular, it may improve the service level and reduce inventories when dealing with product proliferation, as it will typically reduce manufacturing process complexity, and leverage better pooling effects. However, it may also require investments in the redesign of the products and processes (Feitzinger and Lee 1997). Third, product design is the last process that can be postponed. In other words, retailers can carry dynamic assortments to offer the most attractive product in each moment of the season. Caro and Gallien (2007) discuss how to optimally introduce new products based on sales expectations and demand learning opportunities and provide a closed-form policy that captures the key trade-offs between exploration and exploitation. Gaur and Honhon (2006) study product assortment and inventory decisions of a retailer under a locational choice model. They find that in some circumstances, static substitutions solutions can be a good proxy for the dynamic case problem and that in these cases, products of an optimal assortment should be equally spaced out at intervals to avoid substitution risks. Caro et al. (2014) and Çınar and Martínez-de-Albéniz (2013) study how to dynamically release products into the stores to keep assortments sufficiently attractive despite decaying interest from the part of the consumers.

In all the papers above, QR is shown to be useful in reducing the costs from demand–supply mismatches, through reduced inventory costs and lost sales. In contrast, we associate QR to higher speed of sales, which means that not only inventory costs are lower, but overall store sales are also higher (as new products can be introduced and sold), thereby reducing the opportunity cost of lost sales, and the costs of the store (rent, labor) will be better utilized. This effect is similar to the improvement in sales described in the dynamic assortment literature.

Finally, our dynamic model of sales borrows some elements from the work that links sales to inventory levels. Curhan (1973) relates availability of shelf space (which ultimately affects the amount of in-store inventory) to retail sales. Given that the amount of new products has grown faster that retail space, optimal space allocation has become a central issue for retailers. Corstjens and Doyle (1981) present an optimization model that takes into account elasticities to shelf space. Smith and Achabal (1998) and Caro et al. (2010) also provide retail models where sales are increasing concave functions of inventory levels. Balakrishnan et al. (2004) and Baron et al. (2011) study how to manage inventory in such setting. Finally, supply chain coordination issues have also been studied in this setting. Wang and Gerchak (2001) show that in order to coordinate a manufacturer and a retailer, the manufacturer should not just push the wholesale price lever, but offer as well an inventory holding cost subsidy to the retailer. Martínez-de-Albéniz and Roels (2011) study the competitive dynamics that result from such a shelf scarcity and demonstrate that, with wholesale price contracts, the inefficiencies from decentralized decision making are limited and can be fully solved through pay-to-stay fee contracts.

4.2.2 Empirical Models

In contrast with all the analytical papers cited so far, to the best of our knowledge, there is scarce empirical work on QR practices in fashion retailing besides estimates on the value of QR based on data in Fisher and Raman (1996), and business school cases (e.g., Ghemawat and Nueno 2003; Ferdows et al. 2002; McAfee et al. 2004; Caro 2011).

Most papers related to our work examine the effect of inventory (lower under QR) on sales. Chen et al. (2005) examine across different industries the relationship between inventory management practices and financial performance during 1981–2000. They find that the companies that have very high inventories show poor stock returns in the long term but that this relationship does not hold in the opposite case, as the firms with the lowest inventories show average stock returns. Gaur et al. (2005) empirically study inventory turnover at US retailers between 1987 and 2000. As turnover varies widely across retailers and over time, they conclude that the metric can be improved by adjusting for changes in gross margins, capital intensity, and sales surprise (which they define as the ratio of actual to expected sales for the year). These parameters condition significantly effective inventory productivity of a firm. Similarly, Kesavan et al. (2010) conclude that sales forecast of US retailers can be improved by incorporating exogenous variables into the estimation, such as inventory and gross margin. They test their model with historical data of inventory and gross margin and show that the model generates more accurate sales forecast than the forecasts of equity analysts.

The drivers for inventory have also been studied. Rumyantsev and Netessine (2007) use aggregate inventory data to measure the relationship between variables as demand uncertainty, lead times, gross margins, and firm size on the level of inventory. They find evidence that companies facing a higher demand uncertainty, longer lead times, and higher margins carry larger inventories. They also confirm that companies leverage economies of scale in inventories, as larger firms have relatively less inventory. Olivares and Cachon (2009) test empirically the key assumption that an increase in sales will lead to a less than proportional increase in inventory, with car dealer data. However, their results indicate that dealers carry more inventory when they face stronger competition, because a higher service level will compensate the potential sales reduction due to more competition. This is in contrast with Amihud and Medenelson (1989), where public data on manufacturing firms is used to estimate the effect of market power (which they proxy by margins and market shares) on inventory levels and variability. It is found that when market power decreases (i.e., competition intensifies), firms tend to lower their inventories.

Finally, the relationship to consumer choice has also been explored. Musalem et al. (2010) present a structural demand model that can capture the effect of stock-outs on customer choice, which they test empirically for consumer products. The model illustrates how the amount of lost sales induced by stock-out can be quantified. Ton and Raman (2010) analyze store data to study the effects of product variety and inventory levels on store sales in electronics retailing. They show that an increase of product variety and inventories leads to more misplaced products, which again decreases store sales.

In our paper, we take a more direct approach: We consider an alternative metric of product success and the speed of sales, and then study how it is driven by lead time choices.

4.3 The Model

4.3.1 A Continuous Inventory-Dependent Sales Model

Consider a fast fashion retailer that sells fashion goods over a network of stores. To sell a given product i, the firm follows a process that starts with the design of the product, continues with a sourcing decision where a manufacturer (in-house or external) is selected together with a production quantity Q_i and ends with the reception of the merchandise at a central warehouse from which it is distributed to the stores. After these steps, the product is displayed in the stores, seen by the customers, and purchased if they find it good enough.

We focus on fashion products, where there is significant uncertainty on the success of the product. Traditionally, due to the relative short life cycle of these goods compared to the design-to-store lead time, the literature has used the newsvendor model to capture such risks. Namely, a stochastic demand is realized; either excess inventory is salvaged or inventory is sold out; and some sales are lost. We claim that this model is inadequate in fast fashion as it overlooks some important elements. First, this is intrinsically a multi-period model, and sales do occur over many weeks. In fact, generally it is possible to sell all the inventory of a given product if the retailer keeps it in the store long enough. Second, demand is not the same when the product is introduced or when the product has been in the store long enough, because, among other things, inventory availability is not the same, which influences how the product is displayed and how much consumer attention it captures (Caro et al. 2014; Çınar and Martínez-de-Albéniz 2013). Third, retailers do not keep all the products in the store for the same amount of time, which implies that two products that have the same amount of sales may be very different if one has only stayed in the store for a few weeks while the other has been there for months. As a result, we build a more detailed sales model that integrates these facts. Our data will then be used to calibrate the model, and we will be able to test some hypotheses on what makes a product successful.

For product i, we let $t = 0$ be the time of introduction of the product in the stores. In every time interval $[t, t + \mathrm{d}t]$ after the product introduction, sales are materialized. We hence consider a continuous-time sales model, as in Balakrishnan et al. (2004), which provides tractability for the analysis and can then be easily extended to discrete sales periods (weeks in our data). Furthermore, this is a realistic choice when the lot sizes are sufficiently large, as is the case in our data (hundreds to thousands of units per product).

The amount of infinitesimal sales is a function of the inventory available in each of the stores, together with a random component that essentially depends on how much the visitors of the store like the product's design, quality, and price. Interestingly, the random element in a product's success turns out to be highly correlated across time. We observe this phenomenon in our data, and most of the existing literature suggests a similar behavior. For instance, Caro and Gallien (2007) assume that a product has a certain 'true' sales rate common to the store network that needs to be learned over time; Fisher (2009) observes that early sales are highly predictive of future sales. This implies that, even though sales may be unpredictable before the introduction of a product, once the product is introduced into the stores and some early demand is revealed, most of the demand uncertainty is resolved.

Hence, we include a random element $\tilde{\epsilon}_i$ that is unknown at the product design-sourcing stage, but is revealed by the time the product is in the store. For simplicity, we assume that sales become deterministic for $t > 0$, although we show below how to include a random component after the product is in the store.

Furthermore, we see in our data that the sales of any given product are also highly correlated across stores. In other words, even though the success of skirts may be higher in France than in Austria, a given skirt will have either relatively high sales in France and Austria or relatively low sales in France and Austria. This allows us to consider network sales s_{it} rather than focusing on individual store sales. Specifically, we use the following modeling of the infinitesimal sales quantity

$$s_{it} = \alpha_i q_{it}^{\beta_i} \tag{4.1}$$

where q_{it} is the network inventory level and $\alpha_i > 0$, $0 \leq \beta_i \leq 1$ product-dependent parameters that may depend on $\tilde{\epsilon}_i$. This formulation that links sales to inventory is common in the literature, e.g., Corstjens and Doyle (1981), Wang and Gerchak (2001), Balakrishnan et al. (2004), or Martínez-de-Albéniz and

Roels (2011). It captures the fact that better displays through higher inventory quantities lead customers to pay more attention to the product and hence increase the purchase probability. It also suggests that there are decreasing returns from carrying more inventory: Doubling inventory only increases expected sales by a factor $2^{\beta_i} \leq 2$. Although the functional model is different, the same characteristics are present in the piecewise-linear relationship between sales and inventory of Smith and Achabal (1998).

Furthermore, it is worth pointing out that this formulation is robust to aggregation, i.e., valid both at the store and the network level. Indeed, if the sales-inventory sensitivity β_i is common to all the stores and the retailer distributes inventory in stores so as to maximize total sales over the network, then the sales over all stores k are as follows:

$$\max_{q_{itk} \geq 0 | \sum_k q_{itk} = q_{it}} \sum_k \alpha_{ik} q_{itk}^{\beta} = \alpha_i q_{it}^{\beta}$$

with $\alpha_i = \left(\sum_k \alpha_{ik}^{1/(1-\beta)}\right)^{1-\beta}$. This means that the retailer can distribute the inventory appropriately such that the marginal value of placing an additional item in a store is the same across all the stores. Then, the relationship (1) between inventory and sales exists both at the store level and the network level, which makes our focus on the network robust. For simplicity, we focus on the case $\beta_i = 1$ in this paper.

Now that we have defined sales as a function of inventory, we can use the inventory balance equation

$$\mathrm{d}q_{it} = -s_{it}\mathrm{d}t \tag{4.2}$$

to establish that

$$q_{it} = Q_i e^{-\alpha_i t}. \tag{4.3}$$

Note that the analysis can be easily extended to include an uncertain element in the sales trajectory: $\mathrm{d}q_{it} = -q_{it}(\alpha_i \mathrm{d}t + \sigma_i \mathrm{d}W_t)$, where W_t is a Wiener process. This would allow us to use our model for random demand processes too.

We shall refer to α_i as the *speed of sales* for product i. This metric provides a relative-to-inventory measure of how fast sales are materializing. This is appropriate because in our fast fashion context the key constraint is store space, that is, consumed proportionally to the amount of inventory displayed, as pointed out in the introduction. As a result, the speed of sales α_i indirectly measures how much each square meter of store can turn per unit time when carrying product i. Since for fast fashion retailers new products are being introduced continuously, it is always preferable to carry products with higher speeds of sales, so sales and margins are maximized, while inventory and store costs remain constant and discounts are minimized.

4.3.2 Relationship Between Sales Parameters and Operational Choices

Thus far, we have characterized the inventory trajectory after a product is introduced, which is exponentially decreasing over time. We turn here to the question of how the speed of sales for product i, α_i, is formed. Recall that it depends on some observable product $FEATURES_i$ but also on a random component $\tilde{\epsilon}_i$ that describes how much consumers like the product, as determined by some unobservable design characteristics, which is independent of observable product characteristics. Specifically, the variables $FEATURES_i$ include product category, color, average sales price, and cost.

In addition, we are particularly interested in understanding how the retailer's operational choices affect the realization of the speed of sales parameter, specifically the choice of suppliers and sourcing location. This drives the two variables on which we focus: the effective design-to-store *time-to-market*, T_i; and the design time D_i, defined as the first time the product was introduced in the store minus the time-to-market.

In other words, we pose the following relationship:

$$\alpha_i = f(D_i, T_i, FEATURES_i, \tilde{\epsilon_i}). \tag{4.4}$$

Since we do not know the time-to-market T_i exactly, we shall use the region of origin as a proxy for it. We shall consider the following regions (the average T_i for each region is provided): North Africa (4 weeks), East Asia (16 weeks), South Asia (12 weeks), East Europe (4 weeks), and West Europe (2 weeks). Furthermore, we shall also assume that the random element $\tilde{\epsilon}_i$ is exogenously generated and that it is uncorrelated with T_i, D_i. This is reasonable, since as we explain later in Sect. 4.4, we do not expect that sourcing areas are self-selected based on an a priori evaluation of the product sales forecasts.

In a fashion context, we expect that a shorter time-to-market allows the company to react faster to revealed consumer tastes. As discussed earlier, fashion is characterized by a high degree of uncertainty and volatility of consumer preferences. A long time-to-market forces companies to commit to design choices well before any real consumer preferences about colors, styles, or specific designs are revealed. On the other hand, a shorter time-to-market may enable, once some consumer preferences about certain product characteristics are known, to design and manufacture the product according to these preferences quickly enough to bring it to market before changes in preferences occur. Hence, a shorter time-to-market would result on a reduction on the uncertainties with customer tastes and will then lead to more successful products: In other words, products that are procured from closer locations and have a shorter T_i and a later D_i, should have a higher (i.e., faster) speed of sales α_i. This argument can be elaborated further. It is generally agreed in the literature that fashion trends can be only roughly guessed at the beginning of the season by a company, and they are only gradually revealed as products are being sold and the season advances. One could imagine two extreme ways in which these season trends evolve and are revealed.

In one extreme, trends are relatively stable and converge during the season (e.g., gooseneck shirts with Chinese motives start to be sold and as they are seen they become more successful and by the end of the season most people wear them), as in Eppen and Iyer (1997) or Fisher and Raman (1996). In this case, the company will gradually observe the trend as the season advances. Products designed later in the season will take advantage of this learning and hence will likely be more successful on average. In this case, a later design time will be correlated with a higher speed of sales. Of course, products designed later in the season will increasingly be produced in nearshore origins so that they can arrive on time to the stores. Hence, one would expect that products from nearshore origins are more successful, but this is exclusively due to the fact that they are designed later, i.e., with higher D_i.

On the other extreme, however, one could imagine that season trends are short-lived and that they are continuously evolving on a unpredictable manner during the season length (i.e., Chinese motives are being fancy this week, but next month Madonna will promote Scottish prints that substitute Chinese motives), as in Caldentey and Caro (2010). In this case, when some trend is spotted, one would like to design, produce, and bring the product to market as soon as possible in nearshore origins. However, there will be no convergent learning about the season trends. In this case, one can imagine preferences as moving in a random walk manner with no clear convergence. Products such that the lead time between design and introduction is short are able to take advantage of the trend before it dies and hence will be more successful. But a later design time will need not be influencing the success per se, as trends evolve during the season. Hence, in this case, products from nearshore origins would be more successful because of their shorter lead time, independently of the time of design or introduction, i.e., with lower T_i.

According to industry experts, reality may well be a combination of these depending on the season and product categories. Hence, our model considers both the time of design to capture season learning, and the origin location (as a proxy for design lead time) to capture design adaptation to short-lived trends. Being this the first empirical paper of this type, our scope will be restricted to show that nearshoring has a positive effect via a possible combination of each of these effects. This leads to the formulation of several empirical questions that we can test through the data. One may additionally try to estimate the relative importance of one of the effects against the other, but unfortunately our data do not allow to do that, because T_i and D_i exhibit high correlation.

4.3.3 Hypotheses

We are especially interested in evaluating the impact of the D_i and T_i on the speed of sales α_i. This leads to three hypotheses.

H1. *A later time of design $D_i$has a positive impact on speed of sales α_i, across all origins*.

According to H1, we expect to find that items designed later in the season sell faster, because the retailer can learn about customer tastes and thereby design more attractive products.

H2. *A shorter time-to-market T_i, given the same time of design D_i, has a positive impact on speed of sales α_i.*

According to H2, we expect to find that items with shorter time-to-market tend to be more successful. The proximity to market allows the firm to follow the market trends more accurately and therefore to offer products that sell faster. In contrast, a longer time-to-market means that there will be more time between design and distribution for customers to change their preferences. Hence, H2 would confirm that there is some uncertainty associated with a trend 'drift,' independent of the time of design and thus not associated with in-season learning of market trends (which is the focus of H1).

Furthermore, we shall consider one additional hypothesis on the interaction between D_i and T_i.

H3. *The marginal increase of speed of sales α_i due to a shorter time of design T_i is reduced when the time of design D_i increases.*

H3 means that the advantage of shorter time-to-markets is reduced as the season progresses. In other words, products designed in long lead time origins benefit more from the learning accumulated during the season; hence, for products introduced at the end of the season, the advantage of a short time-to-market is not as important.

4.4 Empirical Analysis

To validate the model presented in Sect. 4.3, we use data from a fast fashion retailer. We describe next the practices of the company and the details of the data set, and then present the empirical analysis to test our hypotheses.

4.4.1 Industry Setting

We use data from a fast fashion retailer based in Spain that wished to remain anonymous. We will call the firm F. The company's business model is similar to that of the Inditex group (holding of many brands, including Zara), the company that coined the term *fast fashion*, and is described in Caro and Martínez-de-Albéniz (2013, 2015). Fast fashion retailers have not only become market leaders (Keeley and Clark 2008), but have also dramatically changed the competitive dynamics in the industry. These companies are able to offer very fashionable products at affordable prices. This requires carrying a very dynamic store assortment that is able to adjust the products in the store to emerging demand trends. Such strategy allows fast fashion retailers to differentiate from traditional fashion retailers that introduce a new collection every six months only but also from newly emerged low-cost apparel retailers as Uniqlo or Kiabi, which focus their offering on basic, less fashionable items (e.g., plain-color T-shirts). The fast fashion practices have been described extensively in the business press, for example in Foroohar (2006), El País (2008, 2011), Butler (2013), or Berfield and Baigorri (2013). In addition, many case studies have been written, in particular on Zara, considered the current best practice (e.g., Ferdows et al. 2002; Ghemawat and Nueno 2003; McAfee et al. 2004; Caro 2011).

Fast fashion retailers have made supply chain management a centerpiece of their business. They have developed a much tighter control of their supply chains, structurally aiming for a vertical integration of most of the sourcing, production, distribution, and retailing tasks, and even assuming economic ownership of large portions of the value chain (e.g., warehouses and stores). The priority is very often put on execution speed (i.e., minimizing the time from design to store), rather than on cost. For instance, when choosing sourcing locations, nearshoring may be preferred. Furthermore, to manage their supply chains, they have typically implemented end-to-end information systems that provide real-time visibility of sales at the item level, which can be very useful for distribution, sourcing, and design decisions. This does not mean that these systems are complex: Zara is known for comparatively basic IT solutions where business needs are always put first before technological possibilities (McAfee et al. 2004).

It is important to note that, while the fast fashion model makes these companies different, it is not always the most appropriate strategy for any type of product being sold (Fisher 1997). Specifically, it is applied almost exclusively for fashion items, i.e., those with a design with high fashion content, that is, truly new compared to past designs. For more basic items where no real novelty is present in the design, a traditional, efficient supply chain is used, even at Inditex. As a result, to cope with the distinctive requirements of basic and fashion items, two distinct supply chain models are operated in parallel: A very efficient supply chain with a primary focus on cost optimization delivers their basic items that have comparatively constant demand patterns and high volumes; and an agile supply chain produces the trendier products, with the priority being speed and responsiveness to demand.

F also operates a similar dual-supply chain structure. This setting is ideal for the study of quick response, because it provides a controlled experiment where different products are produced with a long or a short time-to-market. Of course, the products may have different characteristics, but these will be controlled for in the empirical analysis. To be able to extract insights, we nevertheless require a certain amount of variability: to discriminate between long and short time-to-market, we need that products with the same characteristics (category, price, features, etc.) are supplied by both supply chains. In other words, we need that F sources in Asia (long time-to-market) and Europe or North Africa (short time-to-market) the same type of products. Fortunately, this is so for most of the product categories. However, for certain categories, only one origin is used, due to country specializations (e.g., all belts at F are made in China), which means that we cannot discriminate the value of QR from the data, and as a result, we do not include these in our analysis.

4.4.2 Company Background

F started operations in the year 2000 as a multi-brand store. While initially, F had decided to offer around 50 % of high-street brands, 25 % of low-cost own brands, and another 25 % of newly developed fashion brands, F decided in 2005 to focus on its own brands completely. These brands, grouped in 5 key brand families, are controlled by F from the design to the store. Our data are from the fall–winter 2008 (FW08) season, from April 2008 to January 2009, a season representative of the new strategy of the firm. We describe F's situation in 2008 next.

As of 2008/2009, the retail footprint of F is made of 141 stores with both owned and franchised stores, which are treated identically (same distribution and retail processes). Furthermore, the company complements its store network through corners, wholesalers, and shop-in-shops. Owned stores represent more than half of total sales, franchises slightly more than a third, with rest being corners, wholesalers, and shops-in-shops. Average revenue per point of sale and year is about €200,000, although owned stores and franchises may have significantly higher sales that the other points of sales. F has retail presence in 7 countries: The majority of sales originate in Spain, but it also has stores in France, Portugal, Italy, Greece, Romania, and Russia. Target customers are young people, primarily women, 15–30 years old, who are value-oriented but looking for highly fashionable products.

The company sells a wide range of fashion products. During a typical season (spring–summer and fall–winter), including FW08, F offers between 500 and 1000 different stock-keeping units (SKUs) , defined as a unique combination of model and color. The products are grouped in product families, such as T-shirts, dresses, bags, and shirts. T-shirts are the main product categories representing around 15 % of sales and 25 % of the SKUs. The remaining revenues and SKUs are relatively equally distributed among the other product families, with an average product family representing around 6 % of SKUs and 8 % of sales. The average price of a product is in the range of 10–20, and the company sells in the range of 1 million units per year in total.

The products are sourced worldwide from 14 different countries, such as Bangladesh or China, but also some production is allocated to quick-response suppliers in Spain, Portugal, Bulgaria, or Morocco. Production quantities are determined in one shot, after a design has been completed, and no additional production of the same product is ever made. However, new products are developed in season that can be similar to successful existing products, but they are treated as different SKUs. The firm in fact learns about customer preferences during the season: Designers permanently track fashion trends and customer preferences. There is a single, centralized design department that is organized around the main product families. Thus, the firm may be able to make these newly designed products more attractive, which should result in a higher speed of sales. Different types of products are sourced from different types of suppliers, ranging from offshore suppliers to nearshore suppliers with significantly shorter lead times. The cost of an item, which is a control variable in our study, may of course depend on the location

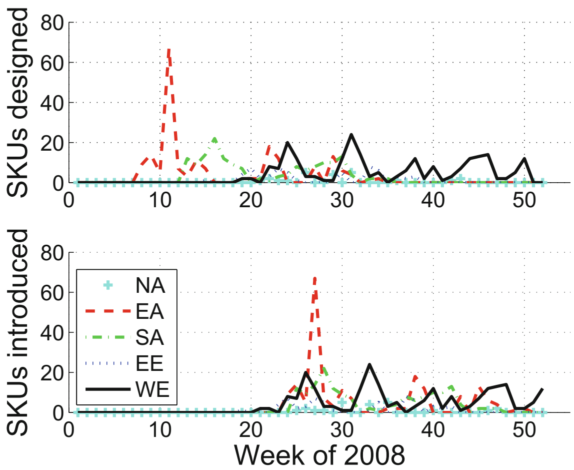

Fig. 4.1 Number of SKUs designed (*top*) and introduced (*bottom*) by week and region

of the supplier, but nearshore goods are not systematically more expensive: For instance, the average cost of a shirt or a pair of trousers is higher in Europe than in Asia, although the opposite is true for dresses. In both cases, the differences are minimal.

The process of deciding which supplier makes which product works as follows. The firm uses a master calendar for their product category introductions over the season, with the aim of having a balanced assortment in the stores. In the calendar, the decision of using an offshore versus nearshore supplier is specified, so as to be able to plan design, production, and distribution activities (including early communication of capacity requirements to suppliers, before products have been designed). These introductions are depicted in Fig. 4.1. It is worth highlighting that the amount of products being introduced at the beginning versus in the middle of the season has remained more or less stable over several seasons. As a result, when F decides to introduce a product at a certain date, the choice of offshore versus nearshore has already been fixed upfront. Thus, it should not create a self-selection problem in our analysis, although we discuss the impact of potential endogeneity in Sect. 4.4.7.

Hence, when the design for a specific product starts, its lead time requirement is already decided. The moment at which this occurs is what we call design time. It takes place with the latest available information about trends. After the main design phase is complete, procurement managers are able to quickly make initial cost estimates of the new products and track available capacity at the suppliers. Once the managers select a prototype for production, designers make final adjustments in colors and materials and coordinate specifications with the chosen supplier, that is, only eligible if it can meet the required lead time. The order is then placed. Finally, after the prespecified lead time, the items are introduced in the stores.

4.4.3 Data Set Description

We use a company database with information at the SKU level. Every SKU is uniquely defined by the model (design) and the color. That implies that different sizes are not considered different SKUs. Since, the distribution of sizes is fairly homogenous within each product, no bias is introduced with this aggregation. Data were properly filtered to remove inconsistencies. Only the ten product families with higher percentage of revenue over the total annual revenue and with at least three different origin areas are included, namely *scarf*, *shirt*, *T-shirt*, *knitwear jacket*, *cardigan sweater*, *jersey*, *trousers*, *shorts*, *top,* and *dress*. Furthermore, the products with very small purchase volumes (Q less than 5 units) were removed. These products are not representative of the company usual batch size (see Table 4.1 for relevant statistics of the products). Similarly, a few items had total sales higher than the purchase quantity, which we removed as well. After these filters, the final number of SKUs is 653.

The database includes on the one hand, sales volumes in units, per SKU, day and point of sales; on the other hand, the features of each SKU include family, color, average sales price per week, and purchasing cost per unit as well as the total purchased volume in units (inbound quantity from supplier into F), and the day of arrival of the product at F's unique warehouse. As we mention above, we take this day of arrival to the warehouse as the time of introduction of the product since cross-docking operations make the available the product at the store in the next few days.

Note that family *t-shirt* shows the highest variety of SKUs, while the families' *shorts* and *top* show little variety. For all the families, the average batch size is far

Table 4.1 Descriptive statistics of the key variables by product family

Product family	Number of SKUs	% overall SKUs	Average (std. dev.) of batch size	% of SKUs sourced in				
				EA	SA	NA	EE	WE
Cardigan sweater	40	6	690 (242)	50	15	0	0	35
Dress	84	13	701 (290)	31	12	2	6	49
Jersey	83	13	752 (402)	30	47	0	0	23
Knitwear jacket	86	13	900 (694)	65	12	0	0	23
Scarf	41	6	286 (258)	76	17	0	0	7
Shirt	45	7	841 (168)	4	7	9	58	22
Shorts	7	1	738 (138)	0	43	14	29	14
Top	12	2	615 (438)	33	0	0	33	33
Trousers	60	9	1223 (1038)	2	20	27	18	33
T-shirt	195	30	903 (726)	14	36	0	5	45
All categories	653	100	824 (633)	30	25	3	9	33

from the threshold of 5 units below which the product was eliminated. Generally, there is variation in the batch size, as shown by the standard deviations, although there was no particular association between batch size and other indicators such as product family or origin. Furthermore, each SKU has a unique sourcing origin. Countries of origin are classified into regions that share similar practices and delivery lead times: East Asia (EA: China, Vietnam), South Asia (SA: Pakistan, India, and Bangladesh), North Africa (NA: Morocco), East Europe (EE: Bulgaria, Romania, Turkey), and West Europe (WE: Spain, Portugal, France, and Italy). Since exact lead times are not available at the SKU level, we use these regions as proxies for supplier lead times and hence time-to-market. As stated in Sect. 4.3.2, the representative lead times for each region are 2 weeks for WE, 4 weeks for EE and NA, 12 weeks for SA, and 16 weeks for EA, respectively. As shown in Table 4.1 for each product category, nearshore and offshore origins are represented by a minimum amount of SKUs (at least 20 %), with the exception of *scarfs*.

The previous Figure 4.1 depicts the number of SKU designed and introduced as the season progresses, per week. The first introduction for the fall–winter season takes place in April, on week 17 (earlier than one may think), which means our timeline spans a total of 38 weeks. One can observe that product introductions tend to be lumped around some specific weeks for each specific origin (as shown by the spikes in the figure). Nevertheless, each origin has products introduced regularly through most of the season. In addition, one can observe that there is a strong collinearity between design time and time-to-market: For items which have late design times, the time-to-market must be shortened to launch them in timely manner. Furthermore, the retailer also tends to introduce earlier in the season a larger share of items made in Asia, with larger time-to-market; hence, products from these origins tend to be designed earlier as well. These observations imply that when incorporating both D_i and T_i in our regressions, we will have to work with cross-effects.

Unfortunately, for regions North Africa and East Europe, the number of new SKUs in any week is small. This does affect some of our empirical conclusions regarding these regions. Namely, when analyzing the effects of both design time and time-to-market on sales, we will need to consider aggregating our regions into broader categories to avoid spurious effects: nearshore (North Africa, West Europe, East Europe) and offshore origins (East Asia and South Asia). With this broader aggregation, we will have enough products which are introduced from offshore origins in weeks 20–30, which have been designed and produced starting between weeks 4 and 18, that is, when seasonal customers' preferences have already been revealed at the beginning of the season. Thus, this allows us to test whether there is a learning effect even for offshore origins. Hence, despite the loss of granularity, this approach will allow us to make robust inferences about the impact of time-to-markets (defined as either nearshore or offshore procurement) on sales as the season progresses and learning occurs.

4.4.4 Computation of α_i

In a fast fashion context, few measures of sales performance have been developed or tested. In our data, total production quantities are small, and there is no product replenishment. Hence, absolute measures such as total season sales volume are not adequate to measure performance since sales volumes are censored due to small quantities produced. Sell-through (i.e., the ratio of sales to purchase volumes) is also problematic because it is difficult to define an adequate selling period where the sell-through can be observed. Indeed, a large number of products tend to sell out eventually and replaced by new similar models, while, for others selling poorly, prices may be reduced at some point to accelerate sales. As a consequence, depending on the time frame used, products may eventually show different behavior. To avoid these problems, we shall consider as a measure of sales performance the speed of sales α_i, described in Sect. 4.3. We will nevertheless explore sell-through directly in Sect. 4.4.6.

To compute the value of α_i for product i, we first define the sell-through as the ratio

$$ST_{it} = \frac{SALES_{i[0,t]}}{Q_i} \tag{4.5}$$

where t is the time, measured in weeks, elapsed since the initial date when the product arrived to the warehouse, $SALES_{i[0,t]}$, is the cumulative sales of product i up to time t in all stores, and Q_i the total inventory purchased of product i. Of course, $0 \leq \mathrm{ST}_{it} \leq 1$.

From Eq. (4.3), $\ln(1 - \mathrm{ST}_{it}) = -\alpha_i t$. Furthermore, since by construction $\mathrm{ST}_{i0} = 0$ and hence $\ln(1 - \mathrm{ST}_{i0}) = 0$, we define α_i as

$$\alpha_i = -\frac{\sum_{t=1}^{W} t\mathrm{ln}(1 - \mathrm{ST}_{it})}{\sum_{t=1}^{W} t^2}. \tag{4.6}$$

We choose a window W to be the minimum of 17 weeks (4 months) and the time where the discount season starts, i.e., at the end of December (week 52 is the last week to be included). The reason we limit the number of weeks is that, after a long time, inventory starts running low and so is the sales quantity, and it may add noise to our estimator. Instead, we focus on the initial period when the product is introduced. Note however that our results are very robust to all other time windows that were tried. Furthermore, we limit our attention to the weeks where items are typically sold at full price, see detailed explanation below.

The definition of α_i in Eq. (4.6) deserves some explanation. One can see that it is a weighted average of $\ln(1 - \mathrm{ST}_{it})/t$, i.e., the rate at which sell-through decreases. The weights are proportional to t^2, which means that data from the later weeks are more important in determining α_i, although at these times $\ln(1 - \mathrm{ST}_{it})/t$ is also much smaller. The formula in (4.6) is in fact the parameter that best fits an exponential sell-through evolution over time. Specifically, α_i minimizes the squared deviations between $\ln(1-\mathrm{ST}_{it})$ and $-\alpha_i t$, as in the ordinary least squares approach:

$$\min_{\hat{\alpha}_i \geq 0} \sum_{t=1}^{W} (\ln(1-\mathrm{ST}_{it}) + \hat{\alpha}_i t)^2.$$

To ensure that the sell-through evolution is sufficiently well fitted through an exponential pattern, we calculated for each SKU the value $r_i^2 = 1 - \left[\sum_{t=1}^{W} (\ln(1-\mathrm{ST}_{it}) + \alpha_i t)^2\right] / \left[\sum_{t=1}^{W} (\ln(1-\mathrm{ST}_{it}) - \mu)^2\right]$, where μ is the average of $\ln(1-\mathrm{ST}_{it})$ over the window. For a vast majority products, characterizing the sales diffusion as an exponential function of t closely fits the actual diffusion path: on average $r_i^2 = 0.93$.

Furthermore, it is worth highlighting that to make Eq. (4.6) consistent with the model of Sect. 4.3, it is necessary that α_i remains constant over the weeks in which the parameter is calculated. This is the case. Clearly, product characteristics (origin, family, color, cost, etc.) do not change. The only potential problem may come from changes in the pricing of the item. Fortunately, firm F has a policy of not running markdowns in season. To ensure that this policy was being followed in our data, we examined whether the same product had been sold at different prices in the same store (because different countries use different prices): We found that 67 % of store–SKU pairs had the same price recorded in the company's ticket database during the window over which we did the estimation; 89 % had a coefficient of variation of less than 1 %; and only 8 % of the pairs had a variation of more than 10 %. Even in those cases, there was no systematic markdown being applied: It seemed to be due to occasional discounting, e.g., through vouchers to the best customers or volume discounts, as prices seemed to vary erratically. This indicates that the price information includes factors unrelated to the scope of this paper that should not affect the computation of the speed of sales. As a result, when we include price effects later on, we only include the average sales price, instead of more granular price information.

Figure 4.2 shows the distribution of α_i, by region. For each region, we depict the empirical cumulative distribution function (c.d.f.) of speed of sales for all products being sourced for the same origin. We can already observe that regions with a shorter time-to-market have speed of sales distributions with higher averages and less dispersion. Of course, to perform a robust test of our hypothesis, we shall control for all the other factors that may influence α_i, which we do next.

4.4.5 Testing the Hypotheses

To validate our hypotheses, we test several models. We start with a one-factor model: Model I only includes the effect of design time but neglects time-to-market effects, and it can thus be used to evaluate the learning that takes place during the

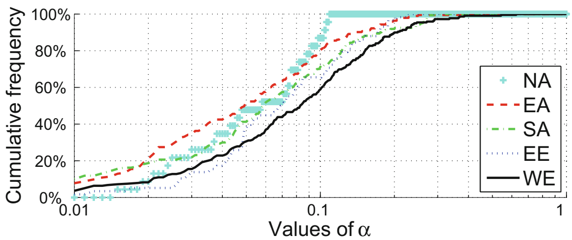

Fig. 4.2 C.d.f. of α_i for each origin

season, i.e., hypothesis H1. We then proceed to Model II, which considers design time and time-to-market variables. As pointed out in Sect. 4.4.3, these two variables exhibit strong collinearity, so the model also considers the interaction between design time and time-to-market. Furthermore, in all models, we control for all the observable product features, such as product family or color to take into account observable seasonal preferences. We also control for average sales price and cost, to take into account differences in product value. These two variables are somewhat correlated, and we observe that there is a relatively stable multiplier, i.e., the ratio between average price and cost, across all products. Finally, we also control by product purchase quantity to take into account possible effects of batch size on the speed of sales. Note nevertheless that our analytical model from Eq. (4.3) suggests that the sell-through is independent of the volume. Hence, although there is no conceptual support to consider that the speed of sales α_i may depend on the purchasing volume, we include it for completeness, and we find that it is not significant.

Our two models have a multiplicative form, i.e., we regress $\ln(\alpha_i)$ with the independent variables, rather than α_i. Indeed, from the tails of the distributions in Fig. 4.2, it seems more reasonable to use a multiplicative error model. In addition, we know that α_i must be positive, and a multiplicative form will avoid providing negative estimates. Finally, we should point out that additive models were also tested, but the fits are much better with multiplicative ones.

4.4.5.1 Model I

Model I is described by:

$$\begin{aligned} \ln(\alpha_i) = \beta_0 + \beta_D D_i + \beta_{\text{PRICE}} \cdot PRICE_i + \beta_{\text{COST}} \cdot COST_i \\ + \beta_Q \cdot Q_i + \sum_{c \in \text{Color}} \beta_c \cdot \chi_{c_i} + \sum_{c \in \text{Family}} \beta_f \cdot \chi_{f_i} + \epsilon_i \end{aligned} \tag{4.7}$$

D_i is the week in which the product was designed, as described in Sect. 4.3.2. Additionally, *Color* and *Family* are categorical variables for product color and

Table 4.2 OLS estimation for Model I (7)

	Estimate	Std. error	p value
Intercept	−3.58800	0.25410	2.0e−16***
D	0.02593	0.00323	5.3e−15***
Q	0.00002	0.00006	7.5e−01
PRICE	−0.03878	0.00782	9.2e−07***
COST	0.02884	0.01655	8.2e−02.
R^2	0.334		

Significance levels: '***': 0.001 and '.': 0.1

family, $PRICE_i$ is the average price recorded over all sales for the item in the store network, $COST_i$ is the product cost, and Q_i contains the product batch size.

The model parameters are estimated using OLS. For this model, there was no apparent relationship between D_i and the residual of the regression, so OLS was appropriate. In contrast, in later models, there is some heterogeneity across time-to-market and origins, so weighted least square (WLS) methods are used in those cases. Table 4.2 shows the results (color and family coefficients have been omitted).

The key insight from this model is that β_D is positive and significant, at the 0.001 level. This supports H1 and means that items designed later in the season will be more attractive and sell faster. The magnitude of this coefficient suggests that an item will sell 2.5 % faster than some other item designed 1 week before. Still, the low R^2 suggests that despite the significant average effect on sales, other product aspects may also be important to determine the speed of sales.

Moreover, one may wonder whether the learning takes place at the same pace during the season, e.g., whether most of the learning takes place in the beginning. This would suggest a nonlinear effect of design time on speed of sales. For this purpose, we tested a variation of this model with $\ln(D_i)$ and D_i^2, but the coefficients for these additional variables turned out to be non-significant.

Finally, it is interesting to report (from the complete model in the left) that items with lower price and higher cost sell faster, although the cost relationship is significant at the 10 % level only. On the one hand, this reflects the price sensitivity of customers. On the other, it suggests that customers also react to product cost: In other words, products that cost more, i.e., require more expensive fabrics or more labor, offer higher value to consumers.

4.4.5.2 Model II

Model I only included the effect of design time on sales. This implies that there is no difference in treatment between products with different time-to-market. For instance, a product designed in week 25 and made in West Europe is introduced into the stores on week 27, while another made in South Asia arrives in week 37. We thus include in Model II.a both variables D_i and T_i. In addition, the interaction term $D_i \times T_i$ is also incorporated, to identify whether the learning effect during the

Table 4.3 WLS estimation for Model II.a (8) on the left and Model II.b (9) on the right

	Estimate	Std. error	p value	Estimate	Std. error	p value
Intercept	−3.21100	0.35830	2.0e−16***	−2.82300	0.33190	2.0e−16***
D	0.01552	0.00729	3.4e−02*	0.00626	0.00622	3.2e−01
T	−0.03451	0.01988	8.3e−02.			
D × T	0.00104	0.00074	1.6e−01			
OFFSHORE				−0.94390	0.25600	2.5e−04***
D × OFFSHORE				0.02771	0.00937	3.2e−03**
Q	0.00002	0.00006	7.1e−01	−0.00001	0.00006	8.4e−01
PRICE	−0.03875	0.00796	1.4e−06***	−0.03832	0.00776	1.0e−06***
COST	0.02991	0.01671	7.4e−02.	0.02829	0.01628	8.3e−02.
R^2	0.344			0.354		

Significance levels: '***': 0.001; '**': 0.01; '*': 0.05; and '.': 0.1

season (seen in Model I) is stronger or weaker depending on the speed of execution. It is worth pointing out that there was some heteroskedasticity across regions, so we used WLS to estimate the parameters, with the weights being equals to the variance of the residuals obtained from OLS estimation. Interestingly, the variance of the residuals is higher for higher time-to-market, i.e., there is more uncertainty in the success rate of products made off shore.

$$\begin{aligned}\ln(\alpha_i) = {} & \beta_0 + \beta_D D_i + \beta_T T_i + \beta_{D\times T} D_i T_i + \beta_{\text{PRICE}} \cdot \text{PRICE}_i + \beta_{\text{COST}} \cdot \text{COST}_i \\ & + \beta_Q \cdot Q_i + \sum_{c\in\text{Color}} \beta_c \cdot \chi_{c_i} + \sum_{c\in\text{Family}} \beta_f \cdot \chi_{f_i} + \epsilon_i \end{aligned} \tag{4.8}$$

The results are shown in the left of Table 4.3. Since D_i, T_i and $D_i \times T_i$ are correlated, we see that the significance of all these variables is reduced compared to Model I. In particular, part of the learning effect identified in Model I is taken now by variable T_i (this effect is negative because items made in regions with short time-to-market are designed later). In addition, the goodness of fit is not improved compared to Model I, as R^2 is only marginally higher. As a result, the model in (4.8) does not seem to capture correctly the joint effects of D_i and T_i. Hence, we develop Model II.b, a simpler alternative to (4.8) in (4.9), by including the variable $OFFSHORE_i$ instead of T_i (which was numerical but limited to values of 2, 4, 12, or 16 weeks). Recall that $OFFSHORE_i$ is equal to one when the item is sourced in Asia (East and South) and zero otherwise:

$$\begin{aligned}\ln(\alpha_i) = {} & \beta_0 + \beta_D D_i + \beta_{\text{OFFSHORE}} \text{OFFSHORE}_i + \beta_{D\times\text{OFFSHORE}} D_i \text{OFFSHORE}_i \\ & + \beta_{\text{PRICE}} \cdot \text{PRICE}_i + \beta_{\text{COST}} \cdot \text{COST}_i + \beta_Q \cdot Q_i \\ & + \sum_{c\in\text{Color}} \beta_c \cdot \chi_{c_i} + \sum_{c\in\text{Family}} \beta_f \cdot \chi_{f_i} + \epsilon_i \end{aligned} \tag{4.9}$$

The right of Table 4.3 contains the results of this alternative Model II.b. First, we see that the model offers a slightly higher goodness of fit, because T_i and D_i are correlated so the informativeness of T_i beyond D_i is not that high. To obtain a higher R^2, one could consider the categorical variables $ORIGIN_i$ instead of T_i, but then there would be few observations in East Europe and North Africa, so we focused on this more robust specification.

The learning effect during the season is now weakened, and most of this effect is taken by $OFFSHORE_i$: Items made in Asia have lower speed of sales. But this does not mean that there is no learning. Indeed, the positive interaction term implies that the learning effect is different depending on the origin. It is significantly larger for offshore origins compared to nearshore ones. This indicates that as season progresses, the success of products from offshore countries improves faster than those from nearshore ones. In other words, offshore products benefit a lot by delaying the design time, which suggests that 'blind' designs made very early carry high design risk and hence will on average be less successful, while 'informed' designs made after the beginning of the season perform much better. In contrast, the learning for nearshore products is small. We interpret that this is due to having most of these products designed in season, so that most of the learning from early sales is already present. This suggests that learning is much higher in the first half of the season (weeks 20–35), while it is limited in the second half (weeks 35–50).

To evaluate the absolute effect of short time-to-market on speed of sales, consider for instance a product designed in week 20: The difference between the 'predicted' average log speed of sales for offshore products ($-2.79 - 0.94 + 0.027 \times 20$ weeks) and nearshore ones (-2.79) is quite important, about 49 % higher for nearshore items. This difference is reduced as design time increases: By week 25, it is 30 % higher; by week 30, only 14 %. For a product designed in week 35, the difference disappears, but this is already too late in the season to order items from Asia. As a result, although short time-to-market remains valuable, within the range of the data (T_i being smaller than 35), H2 is not supported because by the middle of the season the difference between offshore and nearshore becomes insignificant.

In contrast, we find support for H1 and H3. As the season progresses, the average success rates increase, so postponing design is valuable; and the difference of success rates across offshore and nearshore origins tends to vanish gradually as design time increases. In other words, the learning is stronger for products made in Asia, which are designed very early. Indeed, products from faraway origins are mostly 'wild guesses' of season trends prior to the season, and one would expect much lower average success rates, as we observe. As the season advances and customer preferences start to be observed, designs from offshore origins are much closer to what the market is expecting but, still, long time-to-market makes these products less successful compared to those sourced near shore. Eventually, designs made well into the season are able to have designs closer to what the market expects and thus have success rates comparable to nearshored items. As discussed in Sect. 4.3.2, this seems to indicate that trends can fashion trends are persistent, as opposed to short-lived. They are revealed by early sales, and by week 35, there is no significant advantage of a short

Table 4.4 WLS estimation to explain the sell-through after 4 weeks (ST_4, top) and 8 weeks (ST_8, bottom)

	Estimate	Std. error	p value	Estimate	Std. error	p value
$\ln(ST_4)$						
Intercept	−2.55500	0.48320	1.7e−07***	−2.18600	0.43710	7.5e–07***
D	0.02773	0.01089	1.1e−02*	0.02044	0.00905	2.4e–02*
T	−0.14910	0.02849	2.3e−07***			
D × T	0.00625	0.00104	3.6e−09***			
OFFSHORE				−2.66500	0.35070	1.1e–13***
D × OFFSHORE				0.10050	0.01269	1.2e–14***
Q	−0.00002	0.00007	7.4e−01	−0.00008	0.00008	2.9e–01
PRICE	−0.03871	0.01083	3.8e−04***	−0.03209	0.01007	1.5e–03**
COST	0.04648	0.02225	3.7e−02*	0.03333	0.02084	1.1e−01
R^2	0.443			0.480		
$\ln(ST_8)$						
Intercept	−1.48100	0.45670	1.3e−03**	−1.03200	0.42100	1.5e–02*
D	0.02539	0.01169	3.0e−02*	0.01639	0.01018	1.1e−01
T	−0.10010	0.02650	1.8e−04***			
D × T	0.00409	0.00099	3.9e−05***			
OFFSHORE				−1.92200	0.34350	3.5e–08***
D × OFFSHORE				0.07059	0.01256	3.0e–08***
Q	−0.00002	0.00006	7.9e−01	−0.00003	0.00006	5.9e–01
Price	−0.03962	0.00904	1.4e−05***	−0.03995	0.00869	5.3e–06***
Cost	0.04483	0.01825	1.4e−02*	0.04385	0.01760	1.3e−02*
R^2	0.468			0.496		

Significance levels: '***': 0.001; '**': 0.01; '*': 0.05; and '.': 0.1

time-to-market. Hence, the value of QR resides in being able to defer the design time so as to take advantage of trend information.

4.4.6 Alternative Measures of Success

So far we have focused our attention on speed of sales. To ensure the robustness of our conclusions, we have tested our explanatory models with other measures of product success. Namely, we have regressed the traditional metric of sell-through after a fixed number of weeks, with all the variables present in our study. Table 4.4 contains the results of Models II.a and II.b applied to the logarithm of sell-through rates after 4 and 8 weeks. The conclusions remain unchanged: being able to design later and having a short time-to-market both increase the success rate, and the learning effect is more pronounced for offshore sources.

4.4.7 Sourcing Choices

Based on the process that F uses to allocate products to origins and suppliers, we have assumed throughout our analysis that the sourcing choice is exogenous. One potential concern is that such choice may be endogenous which would clearly impact our results. In fact, endogeneity would actually strengthen our conclusions, although the actual estimates would be different.

Specifically, one could imagine that the firm chooses where to produce a product based on its prospects. Unfortunately, we have no way of identifying such behavior with our data, because product forecasts or other qualitative assessments are not available. But if product-source allocation was indeed endogenous, Fisher (1997) describes how to so: Products with higher a priori uncertainty should be sent nearshore. That is, the company may source riskier products from nearshore origins and more basic risk-free products from offshore origins, and this should endogenously affect speed of sales. As a result, one should expect a lower or more uncertain success rate of products with short time-to-market. We find the opposite, as discussed before Eq. (4.8). Thus, our results are robust in sign and significance to this potential bias, although the absolute value of the estimator may be biased.

4.5 Discussion

Our paper presents an empirical study of the value of quick response. We first build an analytical model where sales are related to inventory availability. We use this model to explain the dynamics of sales in our data and to construct the success indicator of a product: We define speed of sales as the share of inventory that is sold per time unit. Thus, more successful products sell faster, generally more but not necessarily (because they may stay in the store for a shorter period). We then compute the speed of sales of all the products in the sample and study how it depends on production choices, specifically the design time and the time-to-market (long if made off shore, short near shore).

Several conclusions can be extracted, which are new to the literature. First, our results confirm that QR practices can help a fashion company to increase the speed of sales. In our case, the ability of postponing design provides a significantly higher speed of sales (H1). Second, we observe that shorter time-to-market in itself is also valuable, although only significant early in the season (H2). Third, the improvement obtained by postponing design is more pronounced when time-to-market is longer, hence for offshore origins compared to nearshore ones (H3). This suggests that nearshore production is generally able to produce faster-selling products, but offshore production after early seasonal trends are observed can perform well too. In terms of magnitude, from our Model II, we can state that products made in West Europe, East Europe, and North Africa in the middle of the season (week 25) have a speed of sales that is about 30 % higher ($e^{0.94-0.027\times 25}$) than those made in East Asia and South Asia.

Our work has important implications. Namely, our results suggest that choosing a supplier in a particular region does have an impact on the speed at which sales are realized. This can be used for deciding where to produce each of the items to be introduced in the store in a given time. For example, consider that a new item (e.g., within the T-shirt category) needs to be introduced in week 30 of a given year. The item can be designed and procured from East Asia in week 14 of the year (4 months before the introduction date), or in South Asia in week 18 (3 months before), or nearshore in week 26 (1 month before). Each choice determines a different level of speed of sales and uncertainty. It will also impact costs, which may be higher or lower depending on the opportunity cost of shelf space (which is occupied for a longer time with offshore production) and the procurement cost associated with the suppliers. This is a critical trade-off for the industry. As stated in H&M (2007), different items require different lead times: 'The time from an order being placed until the items are in the store may be anything from a few weeks up to six months. The best lead time will vary for a particular situation. For high-volume fashion basics and children's wear it is advantageous to place orders further in advance. In contrast, trendier garments in smaller volumes have to be in the stores much quicker.' Our study provides estimates for the drivers of these decisions.

Our study has several limitations that should be explored in further research. First, we work with a sample of one season in one company. The company is a European retailer that is representative of the best practices of fast fashion. It would be interesting to conduct similar studies in other fast fashion firms and a longer time span. Ideally, it could combine with controlled field experiments, as in Caro and Gallien (2010). Second, it would be valuable to validate and compare the results with other similar data sets, for example, for high-end fashion products, where brands try to create trends, rather than following them; or in other markets such as Asia or the USA. One should nevertheless note that this type of data is quite sensitive and firms tend to be reluctant to share it for research purposes. Third, more analytical work that explicitly model demand dynamics related to inventory availability seems necessary, as it seems to fit the data much better than models with exogenous stochastic demand.

Acknowledgements V. Martínez-de-Albéniz's research was supported in part by the European Research Council—ref. ERC-2011-StG 283300-REACTOPS and by the Spanish Ministry of Economics and Competitiveness (Ministerio de Economa y Competitividad)—refs. ECO2011-29536 and ECO2014-59998-P.

References

Amihud Y, Medenelson H (1989) Inventory behaviour and market power: an empirical investigation. Int J Ind Organ 7(2):269–280

Balakrishnan A, Pangburn M, Stavrulaki E (2004) “Stack them high, let’em fly”: lot-sizing policies when inventories stimulate demand. Manage Sci 50(5):630–644
Baron O, Berman O, Perry D (2011) Shelf space management when demand depends on the inventory level. Prod Oper Manage 20(5):714–726
Berfield S, Baigorri M (2013) Zara’s fast-fashion edge. Business week, Nov 14 (online)
Bernstein F, Martínez-de-Albéniz V (2014) Dynamic product rotation in the presence of strategic customers. Manage Sci Forthcoming
Butler S (2013) Inditex: Spain’s fashion powerhouse you’ve probably never heard of. The Guardian, Dec 15 (online)
Cachon G, Swinney R (2011) The value of fast fashion: quick response, enhanced design, and strategic consumer behavior. Manage Sci 57(4):778–795
Cachon GP, Terwiesch C (2009) Matching supply with demand: an introduction to operations management. McGraw-Hill/Irwin, New York, USA
Caldentey C, Caro F (2010) Dynamic assortment planning. Working paper, UCLA
Caro F (2011) ZARA: staying fast and fresh. Technical report, UCLA Anderson School of Management case 612-006-1
Caro F, Gallien J (2007) Dynamic assortment with demand learning for seasonal consumer goods. Manage Sci 53(2):276–292
Caro F, Gallien J (2010) Inventory management of a fast-fashion retail network. Oper Res 58 (2):257–273
Caro F, Gallien J, Daz M, Garca J, Corredoira J, Montes M, Ramos J, Correa J (2010) Zara uses operations research to reengineer its global distribution process. Interfaces 40(1):71–84
Caro F, Martínez-de-Albéniz V (2013) Operations management in apparel retailing: processes, frameworks and optimization. BEIO, Boletn de Estadstica e Investigación Operativa 29 (2):103–116
Caro F, Martínez-de-Albéniz V (2015) Fast fashion: business model overview and research opportunities. In: Agrawal N, Smith SA (eds) Retail supply chain management: quantitative models and empirical studies, 2nd edn, ed. Springer, New York
Caro F, Martínez-de Albéniz V, Rusmevichientong P (2014) The assortment packing problem: multiperiod assortment planning for short-lived products. Manage Sci 60(11):2701–2721
Chen H, Frank M, Wu O (2005) What actually happened to the inventories of American companies between 1981 and 2000? Manage Sci 51(7):1015–1031
Christopher M, Lowson R, Peck H (2004) Creating agile supply chains in the fashion industry. Int J Retail Distrib Manage 32(8):367–376
Çnar E, Martínez-de-Albéniz V (2013) A closed-loop approach to dynamic assortment planning. Working paper, IESE Business School
Corstjens M, Doyle P (1981) A model for optimizing retail space allocations. Manage Sci 27 (7):822–833
Curhan RC (1973) Shelf space allocation and profit maximization in mass retailing. J Mark 37 (3):54–60
El País (2008) Estrenar ropa dura segundos. El País, Jan 2 (online)
El País (2011) Inditex escapa a la crisis. El País, Jan 16 (online)
Eppen G, Iyer A (1997) Backup agreements in fashion buying—the value of upstream flexibility. Manage Sci 43(11):1469–1484
Feitzinger E, Lee H (1997) Mass customization at Hewlett-Packard: the power of postponement. Harvard Bus Rev 117(1):116–121
Ferdows K, Machuca JAD, Lewis M (2002) Zara. Technical report, ECCH case 603-002-1
Fisher M (2009) OR FORUM–rocket science retailing: the 2006 Philip McCord Morse Lecture. Oper Res 57(3):527–540
Fisher ML (1997) What is the right supply chain for your product? Harvard Bus Rev 75:105–117
Fisher ML, Rajaram K, Raman A (2001) Optimizing inventory replenishment of retail fashion products. Manuf Serv Oper Manage 3(3):230–241
Fisher ML, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. Oper Res 44(1):87–99

Foroohar R (2006) A new fashion frontier. Newsweek International, Mar 19 (online)
Gaur V, Fisher M, Raman A (2005) An econometric analysis of inventory turnover performance in retail services. Manage Sci 51(2):181–194
Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. Manage Sci 52(10):1528–1543
Ghemawat P, Nueno JL (2003) ZARA: fast fashion. Technical report, Harvard Business School Multimedia Case 9-703-416
Hammond JH, Kelly MG (1990) Quick response in the apparel industry. Technical report, Harvard Business School Note 9-690-038
Heath DC, Jackson PL (1994) Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. IIE Trans 26(3):17–30
H&M (2007) Annual report
Iyer AV, Bergen ME (1997) Quick response in manufacturer-retailer channels. Manage Sci 43 (4):559–570
Keeley G, Clark A (2008) Zara overtakes gap to become world's largest clothing retailer. The Guardian, Aug 11 (online)
Kesavan S, Gaur V, Raman A (2010) Do inventory and gross margin data improve sales forecasts for U.S. public retailers? Manage Sci 56(9):1519–1533
Lee H, Tang C (1997) Modelling the costs and benefits of delayed product differentiation. Manage Sci 43(1):40–53
Martínez-de-Albéniz V, Roels G (2011) Competing for shelf space. Prod Oper Manage 20(1):32–46
McAfee A, Dessain V, Sjöman A (2004) ZARA: IT for fast fashion. Technical report, Harvard Business School case 9-604-081
Musalem A, Olivares M, Bradlow E, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. Manage Sci 56(7):1180–1197
Olivares M, Cachon G (2009) Competing Retailers and Inventory: an empirical investigation of general motors' dealerships in isolated U.S. markets. Manage Sci 55(9):1586–1604
Rumyantsev S, Netessine S (2007) What can be learned from classical inventory models? A cross-industry exploratory investigation. Manuf Serv Oper Manage 9(4):409–429
Smith S, Achabal D (1998) Clearance pricing and inventory policies for retail chains. Manage Sci 44(3):285–300
Song J-S, Zipkin PH (2012) Newsvendor problems with sequentially revealed demand information. Naval Res Logistics 59(8):601–612
Ton Z, Raman A (2010) The effect of product variety and inventory levels on retail store sales: a longitudinal study. Prod Oper Manage 19(5):546–560
Wang Y, Gerchak Y (2001) Supply chain coordination when demand is shelf-space dependent. Manuf Serv Oper Manage 3(1):82–87