The Top-Dog Index: a new measurement for the demand consistency of the size distribution in pre-pack orders for a fashion discounter with many small branches

Sascha Kurz · Jörg Rambau · Jörg Schlüchtermann · Rainer Wolf

Published online: 31 October 2014
© Springer Science+Business Media New York 2014

Abstract We propose the new Top-Dog-Index, a measure for the branch-dependent historic deviation of the supply data of apparel sizes from the sales data of a fashion discounter. Our approach individually measures for each branch the scarcest and the ampest sizes, aggregated over all products. This measurement can iteratively be used to adapt the size distributions in the pre-pack orders for the future. A real-world blind study shows the potential of this distribution free heuristic optimization approach: The gross yield measured in percent of gross value was almost one percentage point higher in the test-group branches than in the control-group branches.

Keywords Revenue management · Size optimization · Demand forecasting · Top-Dog-Index · Field study · Parallel blind testing

Mathematics Subject Classification Primary: 90B05 · Secondary: 90B90

1 Introduction

The financial performance of a fashion discounter depends on its ability to predict the customers’ demand for individual products. More specifically: trade exactly what you can sell
to your customers. This task has two aspects: offer what your customers would like to wear because the product as such is attractive to them and offer what your customers can wear because it has the right size.

In this paper, we deal with the second aspect only: meet the demand for sizes as accurately as possible. The first aspect, demand for products, is a very delicate issue: Products in a fashion discounter are never replenished because of lead times of around three months. Therefore, there will never be historic sales data of an item at the time when the order has to be submitted (except for the very few “never-out-of-stock”-items, NOS items, for short).

When one considers the knowledge and experience of the professional buyers employed at a fashion discounter—acquired by visiting expositions, reading trade journals, and the like—it seems hard to imagine that a forecast for the demand for a product could be implemented in an automated decision support system at all. We seriously doubt that the success of fashion product can be assessed by looking at historic sales data only. In contrast to this, the demand for sizes may stay reasonably stable over time to extract useful information from historic sales data.

In the historic sales data the influences of demand for products and demand for sizes obviously interfere. Moreover, it was observed at our partners’ branches that the demand for sizes seems not to be constant over all around 1,200 branches.

The main question of this work is: how can we obtain information about the demand for sizes (individually for each branch or for a class of branches) that can be exploited to improve the supply of branches with apparel in different sizes?

Finding the right stock level of a product in order to maximize profit by sales minus cost for inventory is a classical problem of Operations Research. How to utilize historical data is a core aspect of it, and in our problem class this is a particular delicate issue.

1.1 Related work

Interestingly enough, we have not found much work that exactly deals with our task. It seems that, at first glance, the problem of determining the size distribution in delivery pre-packs can be considered as a simple regression once you have historic sales data: Just estimate the historic size profile and fit your delivery to that.

In our problem, however, the historic sales data is not necessarily equal to the historic demand data, and it is interesting how to find the demand data in the sales data in the presence of unsatisfied demand and very small delivery volumes per branch and per product.

There is a study dealing with the size-dependent successive replenishment problem in Caro and Gallien (2010). In that study, the authors try to improve on the size distribution of the supply for a brand of large Spanish fashion group. The biggest difference in the problem setting is that the authors develop a method for a successive replenishment. In the replenishment problem, the forecasting aspect is much less pronounced than in the single-delivery problem that we face. This is mainly because the size-dependent forecasts can rely partly on sales data of the product under consideration whereas we have to forecast size-dependent demands from historic sales of possibly related but different products with different overall success. The methodology they use for their field study, however, is very similar in the sense that they use test and control groups to assess the impact of introducing a new method. The main difference is that they distinguish test and control articles. We chose not to classify articles but branches because we observed extreme volatility in the success of an article. Consequently, very large test and control groups would have been needed to obtain significant results. In contrast to Caro and Gallien, we explicitly do not want to estimate
success and demand for sizes at the same time because we have no success data for the article under consideration. They want to estimate success, too, because they need to decide on an absolute replenishment quantity. We remark that their field study and our field study have been designed and performed completely independently of each other. In Sect. 7.6 we will compare some of their principal findings to ours.

There is a recent PhD thesis describing the theoretical background of that planned case study on demand forecasting and distribution optimization (Garro 2011). For a more detailed literature review, especially for the involved demand forecasting in the fashion retailing industry, we refer to this thesis and the contained list of references, see also Caro and Gallien (2010).

The problem of estimating demands from historic sales data in the fast changing fashion industry is also studied in some newer publications. Whenever pre-ordering information can be collected (which cannot be done in our case), then one can improve the demand estimation; this was investigated in Mostard et al. (2011). In Ni and Fan (2011) forecasting demands by ART (auto-regressive time series) depending on a high-dimensional attribute space is separated into long-term and short-term forecasting stages, the latter being the disturbance data to the former. In Vulcano et al. (2012) a more general proposal is made how to capture substitution effects (including stock-out substitution) among a number of products in demand forecasting. In Yu et al. (2011) it is shown how forecasting demands by an ANN (artificial neural network) can be accelerated to cope with the variety of different stock-keeping units in fashion industry. In Aksoy et al. (2012) demand data for clothing manufacturers (a much higher aggregation level as a branch in our case) is estimated by an ANFIS (adaptive network based fuzzy inference system) system, a combination of fuzzy logic and neural networks. All these efforts are essentially complementary to the proposal in this paper. We have too little data per attribute to adopt the suggested methods directly. Moreover, we are giving away information because we are only interested in branch-dependent demands for sizes. Therefore, we will specify a new entity that can particularly well be used for the distribution of supply over sizes in each branch because it can be estimated independently of the many other attributes of a product.

The type of research closest to ours is assortment optimization. In a sense, we want to decide on the start inventory level of sizes in a pre-pack for an individual product in an individual branch, see e.g. van Ryzin and Mahajan (1999). (Let us ignore for a moment that these unaggregated inventory levels are very small compared to other inventory levels, e.g., for grocery items.)

The most successful approaches to demand estimation deal rather with NOS items than with perishable and not replenishable fashion goods. For example, assortment optimization in the grocery sector (Kök and Fisher, 2007, Sect. 4) — one of the very few papers documenting a field study — can usually neglect the effects of stockout substitution in sales data, which make demand estimation from sales data much more reliable. There is work on the specific influence of substitution on the optimization of expected profit (see, e.g., Mahajan and van Ryzin 2001), but the problem of how to estimate demand parameters from low-volume sales data in the presence of stock-out substitution remains.

As an example of an overview article on revenue management models and software for the increasing business areas where it is applied, we refer the reader to Quante et al. (2008). Specific peculiarities of the fashion industry can e.g. be found in Zhou and Xu (2008).

Much more work has been published in the field of dynamic pricing, where in one line of research pricing and inventory decisions are linked. See Elmaghraby and Keskinocak (2003), Chan et al. (2006) for surveys.
A common aspect of all cited papers (and papers cited there) that separates their research from ours is the following: those papers, in some sense, postulate the possibility to estimate a product’s demand in an individual branch directly from sales data, in particular from sales rates. In our real-world application we have no replenishment, small delivery volumes per branch, lost sales with unknown or even no substitution, and sales rates depending much more on the success of the individual product at the time it was offered than on the size. Therefore, estimating future absolute demand data from historic sales data directly seems to need extra ideas, except maybe for the data aggregated over many branches.

1.2 Our contribution

All revenue management methods sketched above require at least stochastic knowledge about the demand in absolute terms. This means in our case, we need to know information about the distribution for the demand for a T-shirt in a certain size in a certain branch. This may be possible to obtain, however, so far we have not found any method that is able to estimate these small integers (demand for T-shirts in a branch in a size) on the basis of right-censored observations (sales for similar products in the same branch and the same size).

We go into a different direction: we propose to use an adaptive optimization technique that requires no absolute demand values but only information about which sizes have been the scarcest and which sizes have been the amplest ones in the past. This information can be obtained by measuring the new Top-Dog-Index (TDI).

The TDI can be utilized in a dynamic heuristic optimization procedure, that adjusts the size distributions in the branches’ corresponding pre-packs accordingly until the difference between the scarcest and the amplest sizes can not be improved anymore. The main benefit of the TDI: it measures the consistency of historic supply with sizes with the historic demand for sizes in a way that is not influenced by the attractivity of the product itself. This way, we can aggregate data over all products of a product group, thereby curing the problem with small delivery volumes per branch and product and size.

The potential of our TDI-approach is shown in a blind-study with 20 branches and one product group (womens’ outer garments). Ten branches randomly chosen from the 20 branches (test-branches) received size pre-packs according to our heuristics’ recommendations, ten received unchanged supply (control-branches). The result: One percentage point increase of gross yield per merchandise value for the test-branches against the control-branches. A conservative extrapolation of this result for our partner would already mean a significant increase of revenue.

Meanwhile, our technique for field studies of this type and the TDI as a means to evaluate the quality of a distribution was successfully applied to test a more sophisticated optimization procedure based on two-stage stochastic programming taking into account the dynamic pricing during the sales process. For this, alternative relative estimates of demands had to be computed (Kießling et al. 2012). During that project, the TDI has been used as a measurement for the consistency of demand for sizes and supply. This was meaningful because the results of the field study in this paper showed a correlation between well-balanced TDIs and economic success.

Beyond Caro and Gallien (2010) discussed above, we have not seen any field study of this type documented in the literature so far. [In Bitran et al. (1998) the success of the method is tested on historical data only.]
We consider the proposal of parallel blind testing based on test and control branches a contribution in its own right. In our problem setting, a comparison of gross yield data of a fashion discounter across seasons is problematic, because the differences between the yields in different seasons caused by general economic influence factors can be much larger than the differences caused by a better size distribution. The use of test and control articles like in Caro and Gallien (2010) was tried in a different, unpublished study but turned out to be less significant because there were not enough comparable articles to find significant results. This were the reasons for us to use the parallel blind test based on branches instead.

1.3 Outline of the paper

In Sect. 2, we briefly restate the real-world problem we are concerned with. In Sect. 3, we introduce the new Top Dog Index, which is utilized in Sect. 6 in a heuristic optimization procedure. Plausibility checks of the Top Dog Index are outlined in Sects. 4 and 5. Section 7 is the documentation of a field study containing a blind testing procedure among two groups of branches: one supplied with and one supplied without the suggestions from the first step of the optimization heuristics. We summarize the findings in Sect. 8, including some ideas for further research.

2 The real-world problem

In this section, we state the problem we are concerned with. Before that we briefly provide the context in which our problem is embedded.

2.1 The supply chain of a fashion discounter

As in most other industries the overall philosophy of supply chain management in fashion retailing is to coordinate the material flow according to the market demand. The customer has to become the “conductor” of the “orchestra” of supply chain members. Forecasting the future demand is, therefore, crucial for all logistics activities. Special problems occur in cases like ours, when the majority of inventory items is not replenished, because the relationship between lead times and fashion cycles makes replenishments simply impossible. The resulting “textile pipeline” has strong interdependencies between marketing, procurement, and logistics.

The business model of our real world problem bases on a strict cost leadership strategy with sourcing in low cost countries, either East Asia or South East Europe. The transportation time is between one and six weeks, economies of scale are achieved via large orders.

Our industrial partner uses an order process based on pre-packed assortments. A pre-pack of an article contains a certain number of pieces in each size. A combination of such numbers for all sizes is called a lot-type; an actual physical pre-pack containing pieces according to a lot-type is called a lot. An order is placed at a supplier in a low-wage country in terms of numbers of lots of certain lot-types. The ordered lots are then delivered to the central warehouse in Germany (high-wage country). There, only complete lots are distributed to the branches. The advantage of such a lot-based supply-chain is that the number of picks necessary in the central warehouse in the high-wage country is reduced considerably, while the additional effort of prepacking the supply in a low-wage country is usually accepted by the suppliers. The disadvantage is that this process poses combinatorial restrictions on the size distributions that can be placed in the various branches. At most five different lot-types
are allowed for each order because of supplier restrictions and spatial capacity restrictions in the central warehouse. According to our industrial partner, the advantages outweigh the disadvantages by far. While the lot-based supply process is very important for the detailed implementation of our supply optimization in the field study, the value of the Top-Dog Index as a measurement tool for relative demands for sizes is not affected.

2.2 Internal stock turnover of pre-packs

The material flow in our problem is determined by a central procurement for around 1,200 branches. All items are delivered from the suppliers to a central distribution center, where a so called “slow cross docking” is used to distribute the items to the branches. Some key figures may give an impression of the situation: 32,000 square meters, 80 workers, 30,000 tons of garment in 10 million lots per year. Each branch is delivered once a week with the help of a fixed routing system. This leads to a sound compromise between inventory costs and costs of stock turnover.

There are two extreme alternatives for the process of picking the items. The retailer can either work with one basic lot and deliver this lot or integer multiples of it for every article to the branches, or he develops individual lots for the shops.

2.3 The problem under consideration

Recall that the stock turn-over is accelerated by ordering pre-packs of every product, i.e., a package containing a specified number of items of each size. We call the corresponding vector with a non-negative integer for every size a lot-type.

In this environment, we focus on the following problem: Given historic delivery data (in terms of pre-packed lots of some lot-type for each branch) and sales data for a group of products for each branch, determine for each branch a new lot-type with the same number of items that meets the relative demand for sizes more accurately.

In particular, find out from historic sales data some information about the relative demands for sizes. We stress the fact that stockout substitution in the data can not be neglected since replenishment does not take place (unsatisfied demand is lost and does not produce any sales data). We also stress the fact that we are not trying to improve the number of items delivered to each branch but the distribution of sizes for each branch only.

3 The Top-Dog-Index

Our new idea throws overboard the desire to estimate an absolute size profile of the demand in every branch. Instead, we try to define a measure for the scarcity of sizes during the sales process that can be estimated from historic sales data in a stable way.

The following thought experiment is the motivation for our distribution free measure: Consider a product, for which in a branch all sizes are sold out at the very same day. This can be regarded as the result of an ideal balance between sizes in the supply. Our measure tries to quantify the deviation of this ideal situation in historic sales data. How can this be done? In the following, we extract data of a new type from the sales process. From a more general angle our approach shows some similarities with the concept of stopping times from probability theory, see e.g. Fisher (2013).

Fix a delivery period $\Delta := [0, T]$ from some day in the past normalized to day 0 up to day $T$. Let $B$ be the set of all branches that are operating in time interval $\Delta$, and let $P$ be
the set of all products in a group delivered in time interval $\Delta$ in sizes from a size set $S$. We assume that in each branch the product group can be expected to have homogeneous demand for sizes throughout the time period. Fix $b \in B$. For each $p \in P$ and for each $s \in S$ let $\theta_b(p, s)$ be the stockout-day of size $s$ of product $p$, i.e., the day when the last item of $p$ in size $s$ was sold out in branch $b$.

Fix a size $s \in S$. Our idea is now to compare for how many products $p$ size $s$ has the earliest stockout-day $\theta_b(p, s)$ and for how many products $p$ size $s$ has the latest stockout-day $\theta_b(p, s)$. These numbers have the following interpretation: If for many more products the stockout-day of the given size was first among all sizes, then the size was scarce. If for many more products the stockout-day of the given size was last among all sizes, then the size was ample.

In order to quantify this, we use the following approach. (In fact, it is not too important how we exactly quantify our idea, since we will never use the absolute quantities for decision making; we will only use the quantities relative to each other.)

**Definition 1 (Top-Dog-Index)** Let $s \in S$ be a size and $b \in B$ be a branch. The Top-Dog-Count $W_b(s)$ for size $s$ in branch $b$ is defined as

$$W_b(s) := \left| \left\{ p \in P \mid \theta_b(p, s) = \min_{s' \in S} \theta_b(p, s') \right\} \right| \quad (1)$$

and the Flop-Dog-Count $L_b(s)$ for size $s$ in branch $b$ is defined as

$$L_b(s) := \left| \left\{ p \in P \mid \theta_b(p, s) = \max_{s' \in S} \theta_b(p, s') \right\} \right|. \quad (2)$$

Moreover, for a fixed dampening parameter $C > 0$ let

$$TDI_b(s) := \frac{W_b(s) + C}{L_b(s) + C} \quad (3)$$

be the Top-Dog-Index (TDI) of size $s$ in branch $b$.

In the data of this work, we used $C = 15$. Of course a suitable choice of $C$ depends on the typical magnitude of the values $W_b(s)$ and $L_b(s)$. Fortunately it turned out that the results are not too sensitive in $C$.

**4 Plausibility of the Top-Dog-Index**

Before we collect experimental evidence for the usefulness of the TDI, we consider a simple special case in order to show that the TDI is a sound concept.

All considerations in this paper assume that products can be treated individually, i.e., substitution effect across products are neglected.

Consider the special case in which in a branch there are only two sizes $s_1$ and $s_2$, and both sizes are supplied with one item of a product. Assume, moreover, that the sales process is a Bernoulli process with time step one day and success probabilities $p_1$ and $p_2$, resp., for a sale on a day.

Good estimates of $p_1$ and $p_2$ compared to each other would give us information whether one of the sizes is scarce or ample. However, for different products we may get completely different estimates for $p_1$ and $p_2$, since these absolute probabilities depend more on other aspects than size. If we assume that all other success factors are the same for $s_1$ and $s_2$, we could try to estimate the ratio $\frac{p_1}{p_2}$ or the so-called ratio of odds $\frac{p_1 p_2}{p_2 q_1}$ (with the usual notation
\( q_1 = 1 - p_1 \) and \( q_2 = 1 - p_2 \) instead. In the following we show, how the TDI is connected to these entities.

We estimate the probability \( R(s_1, s_2) \) that \( s_1 \) is sold strictly before \( s_2 \):

\[
R(s_1, s_2) = \frac{p_1 q_2}{1 - q_1 q_2} = \sum_{k=0}^{\infty} (q_1 q_2)^k \quad \text{success for} \ s_1 \text{only}
\]

\[
\text{no success for both}
\]

Thus, the ratio of \( R(s_1, s_2) \) and \( R(s_2, s_1) \) is

\[
\frac{R(s_1, s_2)}{R(s_2, s_1)} = \frac{p_1 q_2}{p_2 q_1},
\]

which is the ratio of odds. In this simple example, \( \frac{W_b(s_1)}{N} \), where \( N \) is the number of products observed, is an estimator for \( R(s_1, s_2) \), and \( \frac{L_b(s_1)}{N} \) is an estimator for \( R(s_2, s_1) \). Note that the dependence on \( s_2 \) is hidden in \( W_b(s_1) \) and \( L_b(s_1) \), respectively. Therefore, we can (consistently) estimate the ratio of odds by

\[
\frac{W_b(s_1) + C}{L_b(s_1) + C} = \frac{W_b(s_1) + C}{L_b(s_1) + C} = \text{TDI}_b(s)
\]

for every constant \( C > 0 \).

The choice for the constant \( C > 0 \) leads to a biased estimator. However, really important for the assessment of scarceness is only the relative order of these values, and the constant \( C > 0 \) does not change this order.

In this simple special case the role of the constant \( C \) can also be interpreted as a means to estimate \( \frac{p_1}{p_2} \) instead of \( \frac{p_1 q_2}{p_2 q_1} \). We want to find a constant \( D > 0 \) such that

\[
\frac{R(s_1, s_2) + D}{R(s_2, s_1) + D} = \frac{p_1}{p_2}.
\]

Plugging in the formula for \( R(s_1, s_2) \) yields:

\[
\frac{R(s_1, s_2) + D}{R(s_2, s_1) + D} = \frac{\frac{p_1 q_2}{1 - q_1 q_2} + D}{\frac{p_2 q_1}{1 - q_1 q_2} + D} = \frac{\frac{p_1 q_2 + (1 - q_1 q_2) D}{p_2 q_1 + (1 - q_1 q_2) D}}{\frac{\frac{p_1 q_2 + (1 - q_1 q_2) D}{p_2 q_1 + (1 - q_1 q_2) D}} {\frac{p_1 - p_1 p_2 + (1 - q_1 q_2) D}{p_2 - p_1 p_2 + (1 - q_1 q_2) D}}}
\]

\[
= \frac{p_1}{p_2} \quad \text{for} \quad D = \frac{p_1 p_2}{1 - q_1 q_2}.
\]

From (7) we get that

\[
\frac{\frac{W_b(s_1)}{N} + D}{\frac{L_b(s_1)}{N} + D} = \frac{W_b(s_1) + ND}{L_b(s_1) + ND} = \text{TDI}_b(s)
\]
if the TDI is computed with $C = ND$. The term $ND = N \frac{p_1 p_2}{1 - q_1 q_2}$ can, moreover, be interpreted as the expected number of products for which both sizes were sold on day 1 among the products for which at least one size was sold on day 1.

In this special case we see: The TDI is a direct observation of the success of sizes with no distortion by other sales factors. Thus, it estimates interesting parameters of the sales process. It is clear to us that the real sales process is more complicated. Since an analytical description of the real sales process seems out of reach, we just tried the TDI in the real world in the remaining sections.

5 Statistical significance of the Top-Dog-Index

We want to analyze the significance of the proposed Top-Dog-Index. Since this method is supposed to be applied to a real business case, we analyze the statistical significance in more detail. We utilize seven subsets of the data $D_i$. These sets are formed as follows: We assign a random number in $\{1, 2, 3, 4\}$ to each different product. The sets are composed of the data of products where the corresponding random number lies in a specific subset of $\{1, 2, 3, 4\}$, see Table 1 for the assignment. For the interpretation we remark that the pairs $(D_1, D_2)$, $(D_3, D_4)$, and $(D_5, D_6)$ are complementary. The whole data set is denoted by $D_7$.

Since the Top-Dog-Index is designed to provide mainly ordinal information, we have to use another statistical test to make sure that it yields some significant information. Let $TDI_b(s, D_i)$ denote the Top-Dog-Index in branch $b$ of size $s$ computed from the data in Data Set $D_i$. We find it convincing to regard the ordinal information generated by the Top-Dog-Index as robust whenever we have

$$TDI_b(s, D_i) \gg TDI_b(s', D_i) \iff TDI_b(s, D_j) \gg TDI_b(s', D_j)$$

(14)

for each pair of sizes $s, s'$, each pair of data sets $D_i, D_j$, i.e., $1 \leq i, j \leq 7$ and $i \neq j$, and all branches $b \in B$. In words: the order of Top-Dog-Indices of various sizes does not change significantly when computed from a different sample. The following is a sufficient condition for this to happen:

$$\frac{\sum_{j=1}^7 TDI_b(s, D_j)}{\sum_{j=1}^7 TDI_b(s, D_j)} \approx const$$

(15)

for all $(b, s) \in B \times S$ and all $1 \leq i \leq 7$.

Our first aim is to provide evidence that the $TDI_b(s)$ values are robust measurements in this sense. There is a nice way to look at Eq. (15) graphically. Let us plot a column of the relative values $\frac{TDI_b(s, D_j)}{\sum_{j=1}^7 TDI_b(s, D_j)}$ for all branch-size combinations $(b, s) \in B \times S$ and for all $1 \leq i \leq 7$. The columns corresponding to the same branch-size combination but different

| Table 1 Assignment of test sets |
|---------------------------------|
| $D_1$                           | {1, 2}   |
| $D_2$                           | {3, 4}   |
| $D_3$                           | {1, 3}   |
| $D_4$                           | {2, 4}   |
| $D_5$                           | {3}      |
| $D_6$                           | {1, 2, 4}|
| $D_7$                           | {1, 2, 3, 4} |
samples are stacked on top of each other. This way, each column stack has height one in total, and the size relations of the measurements based on the different samples can be assessed right away. To show the robustness of the mean and the median, we have added an additional column for each of them: The median of the seven estimates corresponds to the height of the top-most column, the mean to the height of the second-to-top-most column. The goal now is not to read off the exact values, but to get an intuitive impression how heavily an estimate depends on the data subset it was computed from.

The Top-Dog-Indices are plotted this way in Fig. 1.

In order to provide some intuitively clear reference data to compare to the plot of Fig. 1, we present the corresponding plots for the two extreme cases of deterministic numbers (i.e., $\text{TDI}_b(s, D_i) = \text{TDI}_b(s, D_j)$ for all $1 \leq i, j \leq 7$ and all branch-size combinations $(b, s) \in B \times S$) and totally random numbers in Figs. 2 and 3. In the complete deterministic case the areas of same color form perfect rectangles. In the random case the areas of same color corresponding to the data subsets form zig-zag lines; the median has fewer zig-zag than the mean, but both are quite stable because the random numbers are all from the same distribution.

It is immediately obvious that the plot of Fig. 1 looks more like the plot in the deterministic case as the plot in the random case. Although Eq. 15 might seem rather restrictive at first sight, it is indeed satisfied by our real-world data. It is interesting to note that the dampening parameter in the computation of the TDI does indeed influence the amount of noise in the plot but had almost no influence on the order of TDI values, which is what we intend to use.

We have analyzed a direct estimation of size distributions at a fixed day for each branch (look at the average size distribution of sales after a fixed number of sales days). In order to be able to directly compare the results of that attempt to Fig. 1, we plot the resulting average size distributions for the same seven data subsets. This is shown in Fig. 4 for an estimate from the sales up to the end of day 0 (the first day in the sales period) and in Fig. 5 for an estimate from the sales up to the end of day 12.

It can be seen that day 0-estimates are extremely dependent on the data subset even if only the relative values are taken into account, i.e., the estimates are not robust even in the
Median\(\bar{j}=1\) Const\(_b\)(s,\(D_j\))

Mean\(\bar{j}=1\) Const\(_b\)(s,\(D_j\))

Const\(_b\)(s,\(D_7\))

Const\(_b\)(s,\(D_6\))

Const\(_b\)(s,\(D_5\))

Const\(_b\)(s,\(D_4\))

Const\(_b\)(s,\(D_3\))

Const\(_b\)(s,\(D_2\))

Const\(_b\)(s,\(D_1\))

\((b,s)\in B\times S\)

Fig. 2 Relative distribution of a constant number, the same for each of the seven data subsets, for all branch-size combinations; the two top-most columns are median and mean, resp. This represents the case of deterministic information

Median\(\bar{j}=1\) Rand\(_b\)(s,\(D_j\))

Mean\(\bar{j}=1\) Rand\(_b\)(s,\(D_j\))

Rand\(_b\)(s,\(D_7\))

Rand\(_b\)(s,\(D_6\))

Rand\(_b\)(s,\(D_5\))

Rand\(_b\)(s,\(D_4\))

Rand\(_b\)(s,\(D_3\))

Rand\(_b\)(s,\(D_2\))

Rand\(_b\)(s,\(D_1\))

\((b,s)\in B\times S\)

Fig. 3 Relative distribution of uniformly distributed random numbers in [0, 1], drawn separately for each of the seven data subsets and for all branch-size combinations; the two top-most columns are median and mean, resp. This represents the case of totally uncertain information

weak sense measured in the plot. The day-12 estimates are more robust, but not even close to what the TDI achieves in Fig. 1. Moreover, as we said, the day-12 estimates are already quite close to the supply because of unsatisfied demand, and will therefore fail to measure the size distribution of the demand.

So far we have argued that the Top-Dog-Index produces size-related information in a robust way. In the next section we describe a procedure to harmonize demand and supply with respect to the size distribution. In Sect. 7 we provide evidence via a real-world blind study that this procedure helps to raise the gross yield in reality; thus, a well-balanced Top-
Dog-Index is correlated to the economic success caused by a demand-consistent supply with sizes. Thus, by computing the TDI we can assess the historic quality of the supply with sizes and need not continuously perform expensive field studies to determine gross yields.

6 The heuristic size optimization procedure based on the TDI

Interesting for us is not the absolute TDI of a size but the TDIs of all sizes in a branch compared to each other, i.e., the ordinal information implied by the TDIs. The size with the maximal TDI among all sizes can be interpreted as the scarcest (the one that was sold the fastest) size in that branch; the size with the minimal TDI among all sizes can be interpreted as the amplest size in that branch. Of course, we have the problem of deciding whether or not a maximal TDI is significantly larger than the others. Since the absolute values of the TDIs have no real meaning we did not even try to assess this issue in a statistically profound way.
Our point of view is that absolute forecasting is too much to be asked for. Therefore, we resort to a dynamic heuristic optimization procedure: sizes with “significantly” large TDIs (Top-Dog-Sizes) should receive larger volumes in future deliveries until their TDIs do not improve anymore, while sizes with “significantly” small TDIs (Flop-Dog-Sizes) need smaller supplies in the future. Whenever this leads to oversteering, the next TDI analysis will show this, and we go back one step. This is based on the assumption that the demand for sizes does not change too quickly over time. If it does then optimization methods based on historic sales data are useless anyway.

Let us describe our size distribution optimization approach in more detail.

We divide time into delivery periods (e.g., one quarter of a year). We assume that the sales period of any product in a delivery period ends at the end of the next delivery period (e.g., half a year after the beginning of the delivery period).\footnote{The assumption of a maximal sales period of six months is a management decision ensuring a frequent assortment change, i.e., whenever a product is not sold out within this time frame, prices are cut down in advance.} Recall that a size distribution in the supply of a branch is given by a pre-pack configuration: a package that contains for each size a certain number of pieces of a product (see Sect. 2.2).

We want to base our delivery decisions for an up-coming period on

– the pre-pack configuration of the previous period and
– the TDI information of the previous period giving us the deviation from the ideal balance

According to given restrictions from the distribution system, we assume that only one pre-pack configuration per branch is allowed. We may use distinct pre-pack configurations for different branches, though.

Since we are only dealing with the size distribution of the total supply but not with the total supply for a branch itself, the total number of pieces in a pre-pack has to stay constant. Since the TDI information only yields aggregated information over all products in the product group, all products of this group will receive identical pre-pack configurations in the next period, as desired.

In order to adjust the supply to the demand without changing the total number of pieces in a pre-pack, we will remove one piece of a Flop-Dog-Size from the pre-pack and add one piece of a Top-Dog-Size instead. At the end of the sales period (i.e., at the end of the next delivery period), we can do the TDI-analysis again and adjust accordingly.

Given the usual lead times of three months this leads to a heuristics that reacts to changes in the demand for sizes with a time lag of nine month to one year. Not exactly prompt, but we assume the demand for sizes to be more or less constant over longer periods of time.

The most interesting question for us was how much, in practice, could be gained by performing only one step of the heuristics explained above.

7 A real-world blind study

In this section we describe the set-up and the results of the blind study carried out by our industrial partner, who prefers not to be mentioned by name in publications. In order to put the activities of our partner into context, we briefly describe the business model and the principle organizational structure at the time of the field study in more detail.

Our partner runs over a thousand branches in a couple of European countries. Supply is produced in the far-east and delivered in pre-packed lots by ship with lead times of around three months. The lots are unpacked only after they have arrived in the branches. The stock
Table 2  Summary of parameters of the blind study

| Parameter                      | Description                                                                 |
|--------------------------------|------------------------------------------------------------------------------|
| Test period                    | April through June 2006 (3 Months)                                           |
| Data collection period         | April through September 2006 (6 Months)                                      |
| Branches                       | 20 branches with unbalanced TDIs randomly classified into 10 test and 10 control branches |
| Pieces of merchandise          | Approx. 4,000 pieces for all test and control branches                        |
| Merchandise value              | Approx. 30,000 EUR                                                           |

turn-over happens in a central warehouse, where only one pick per lot is required for the distribution instead of one pick per piece. Each branch receives an integral number of lots, all of which are of the same lot-type. At most four to five lot-types can be used per article.

The company serves the mass market segment by offering a broad range of apparel under own labels. There are some never-out-of-stock (NOS) articles, but the products that are of interest in this study are seasonal and offered only once. During a season, slowly selling products are marked-down. At the end of a season, the prices of left-over articles are cut-down even more, so that all of them are eventually sold, but in some cases only at a salvage value. It was observed that branches differ not only in their total demands for an article but also in their demands for sizes relative to each other.

Since the total demand for a product depends on the success of the article, we wanted to find out whether the TDI can help to improve the size distribution of the supply for each branch. To this end, we decided that we would not change the total numbers of pieces that were delivered to the branches but only the distribution of sizes. Thus, we planned to assess the impact of one iteration of the optimization procedure of the previous section for the TDI and for some economic key performance indicators (KPI).

Since the influence of the general success of an article on KPIs is usually much more dominant than its size distribution, any comparison “before–after” would have shown predominantly the market success of the different articles in the different sales periods. We chose instead to keep constant most influence factors apart from the size distribution. We were only possible if the comparison was performed simultaneously on identical sets of articles. Our idea was to select a subset of branches and to randomly classify them into test and control branches. Test branches would receive their supply according to the result of one iteration of the optimization procedure based on the TDI analysis of the previous supply. Control branches would receive their supply as usual. Neither the test nor the control branches knew in which class they were. Since the classification into test and control branches was random, systematic differences in their performances like the skills of their staff should average out. This is reminiscent of a clinical blind study. Table 2 shows the basic parameters of the experiment.

7.1 Selection of branches

A reasonable selection of branches for a test and a control group had to meet essentially three requirements: first, only those branches should be chosen whose TDI indicated that, in the past, the supply by sizes did not meet the demand for sizes; secondly, no branch should be chosen, where other tests were running during the test period; thirdly, the assignment of branches to test- and control group should be completely random. The reason for the third aspect was that this way all other influences on the gross yield than the selection of sizes would appear similarly in both the test group and the control group and, thus, would average out evenly.
We suggested a set of 50 branches with interesting Top-Dog-Indices to our partner. Out of these 50 branches, our partner chose 20 branches where a potential re-packing of pre-packs would be possible. This set of 20 branches was fixed as the set of branches included in our blind study. After that, a random number between 0 and 1 was assigned to each of the 20 branches. The 10 branches with the smallest random numbers were chosen to be the test group, the rest was taken as the control group.

Fortunately, any mark-downs are centrally organized for all branches so that test and control branches were always selling at identical prices.

In order to provide evidence for the assertion that the test and control branches were not systematically different beyond their supply strategy, it would have been desirable to swap their roles and repeat the field study. This, however, would have doubled the cost and the duration of the (expensive and long lasting) field study. A parallel study with reverted roles would have been desirable, too, but a re-packing process in both directions would have required a random classification of all articles on top. Implementing an article-dependent re-packing process would have become more complicated and error-prone. It was decided to stick to the blind study as described. That we relied on the randomness of the classification into test and control branches is certainly a compromise between effort and outcome.

7.2 Handling of pre-packs

We had to specify the modifications to the size distribution in pre-packs on the basis of the TDI information. It turned out that additional side constraints had to be satisfied: Whenever a product would appear in an advertisement flyer, the pre-pack had to contain at least one piece in each of the four main sizes S, M, L, and XL. That means, although sometimes the TDI suggested that S was the amplest size, we could not remove the only piece in S from the pre-pack. We removed a piece of the second amplest size (M or L, of which there were two in the unmodified pre-packs) instead. To all branch deliveries, one additional piece of XL was packed, because this was the scarcest size in every branch in the test group. This way, the total number of pieces was unchanged in every pre-pack, as suggested in Sect. 6.

Since all orders had been placed well before the decision to conduct a blind study, our partner re-packed all pre-packs for the test branches according to Table 3.

Note that the re-packing of lots was only necessary for the purpose of this field study. It was the only way to change the size distributions of supply that was already delivered to the central warehouse. Of course, this meant that some of the branches that were not chosen for the test received possibly worse “left-over” lots than before. If the optimization procedure were put into practice, then for each branch the supply would be ordered directly in terms of the lot-types derived by the TDI analysis. We chose to re-pack instead because we did not want the test to interfere with the service of the suppliers. Moreover, re-arranging some of the already delivered lots instead of ordering based on new lot-types accelerated the results of the field study by approximately three months. It is important to note that it was clearly always possible to collect the necessary additional pieces for the ten test branches from the lots of the over a thousand branches not selected for the test.

7.3 Time frame

The test included two relevant time periods: the first period from which the TDI data was extracted and the test period in which the recommendations based on the TDI data were implemented for the test group.
The TDI data was drawn from a delivery period of nine month (January through September 2005) and a sales period of twelve month (January through December 2005).

The test data was drawn from a delivery period of three months (April through June 2006) and a sales period of six months (April through September 2006).

### 7.4 Data collection

In order to eliminate contaminated data easily, our partner agreed to take stock to check inventory data for correctness every month. To receive a good estimate for the financial benefit of the supply modification proposed by the procedure described in the previous section we had defined some criteria how to detect contaminated data automatically via a computer program.

### 7.5 Data analysis

In Fig. 6 we have depicted the initial Top-Dog-Indices for the test branches and in Fig. 7 the initial Top-Dog-Indices for the control branches.

The analysis of our field study was intended to answer the following two main questions: are the Top-Dog-Indices better distributed in the test branches than in the control branches, and, if yes, does this have a significant monetary impact?

To investigate the latter, we had to analyze monetary variables. The most important monetary indices for our partner are the gross yield and the last price. Since the values of different products vary widely, we only consider relative values for the gross yield. Hence, the gross yield is defined as the sum of achieved sales prices over all articles divided by the sum of start sales prices over all articles. The last price is defined as the price at which a final piece of an article is sold. This is usually the minimal price in the whole sales process.

The gross yield directly shows how much turnover was lost using a price cutting strategy to sell out all items. The last price tells us how far one was forced to mark-down items provoked by an inadequate size distribution of the supply.

Since we had to deal with a large amount of lost or inconsistent data, we have applied two ways of evaluating gross yield and last price. Imagine that your data says that you have sold 10 items of a given article in a given branch but that the supply was only 8 items. Or vice versa, that the initial supply was 10 items, during the time 8 items were sold, but all items are gone. Both of the described situations occurred significantly often in our data set.

### Table 3  How the pre-packs were modified in the test group

| Branch | Special advertising | No advertising |
|--------|---------------------|----------------|
|        | Remove   | Add   | Remove | Add   |
| 1      | L        | XL    | L      | XL    |
| 2      | M        | XL    | M      | XL    |
| 3      | L        | XL    | S      | XL    |
| 4      | L        | XL    | S      | XL    |
| 5      | M        | XL    | M      | XL    |
| 6      | M        | XL    | M      | XL    |
| 7      | M        | XL    | M      | XL    |
| 8      | L        | XL    | S      | XL    |
| 9      | M        | XL    | M      | XL    |
| 10     | M        | XL    | M      | XL    |
Our first strategy to evaluate the given data was to “eliminate” inconsistent data. In the first case, 8 sale transactions are consistent. For the remaining two items the corresponding supply transaction is missing. So, we eliminated these transactions. In the other case we would eliminate the supply of two items.

The alternative to the elimination of inconsistent data is to “estimate” it from the rest of the data set. As an example, we would simply assume that there was a supply of 10 items instead of 8 items at the same price level in the first case. In the second case we would assume that the remaining 2 items were also sold. Maybe they were shoplifted, some sort of selling for a very cheap prize. So, we need an estimation for the sales prize of the two missing items.
Here, we have used the last sales price, i.e., the minimal sales price, over all sizes for this product in this branch as an estimate.

Neither evaluation method reflects reality exactly. Our hope was that both estimations encompass the true values. At least our partner accepts both values as a good approximation of reality. The truth may be somewhere in between both values. We remark that the amount of inconsistent data in our data set was about 5%.

7.6 Results

We have depicted the new Top-Dog-Indices after applying our proposed repacking in Fig. 8 for the test branches and in Fig. 9 for the control branches.

We can see that it is rather hard to compare the Top-Dog-Indices of the same branches before and after the blind study. The situation on the real market almost never stays constant over time. There are so many influences not considered in our study that it would have been a bad idea to measure a possible raise of earnings directly. For this reason, the simultaneous observation of a test group and a control group makes all outer effects appear in both.

Comparing Figs. 8 and 9 based on the same time period, it appears that the Top-Dog-Indices of the test group have improved more.

More specifically, looking at the individual branches, we can see that in some of the test branches size XL is no longer too scarce, while it remains too scarce in other branches. Moreover, on average over all test branches, the Top-Dog-Indices of sizes S, M, and L are better balanced, whereas in the control branches the corresponding Top-Dog-Indices differ more. We remark that the demand for sizes is rather branch dependent.

While the former achievement might have been equally possible on the basis of a statistics aggregated over all branches, it seems that the latter result was made possible only by the branch dependent information from the Top-Dog-Index, since different sizes were removed from the pre-packs in favor of XL. Moreover, some of the test branches still need more pieces in size XL, some test branches already had enough of XL. That is, in the next optimization step, the branch dependent information becomes vital also for the consistent supply with size XL.
Top-Dog Index (TDI) for Various Samples
(Mean of Control Branches after Field Study)

Fig. 9 TDI in the control branches in the test period

Fig. 10 Average last prices/gross yields per merchandise value over all articles in all test/control branches

But is an improved Top-Dog-Index really an improvement for the business? To answer this question, we have quantified the gross yields and the last prices in the test group and the control group, resp.

Whereas in Caro and Gallien (2010) economic KPIs like revenue increase are estimated in a second step after the statistical analysis of other metrics like sales numbers, we assessed the monetary impact directly in the statistical experiment.

In Fig. 10 we have compared the average values of the gross yield and the last price for the control and the test branches for both evaluation methods. The gross yield of the test branches is 98.0% using the elimination-method and 97.2% using the estimate method. For the control branches we have gross yields of 97.2% (elimination) and 95.7% (estimate). This corresponds to improvements of 0.85 and 1.5 percentage points, resp.

If we compare this with the slightly large improvements of 3-4% on sales documented in Caro and Gallien (2010) for the size optimization of their replenishment process, this is quite encouraging. First of all, our assessment is about a single iteration of an improvement algorithm; thus, more improvement may be possible in further iterations. Second, our optimization must rely on historic sales data of different articles (which caused the need for the
The improvement for the last price is even larger. The test branches show a last price level of 94.2% (elimination) and 94.1% (estimate); the control branches exhibit a last price level of 92.5% (elimination) and 92.7% (estimate). This corresponds to improvements of 1.7 and 1.4 percentage points, resp.

The drastically improved results for the respective “loser branches” (see Fig. 11) and the reduced standard deviation (see Fig. 12) in the data of the test branches provides evidence for the fact that our procedure was able to reduce the risk of a very low last price or a very low gross yield in an individual branch. This effect is desirable beyond the better earnings, as very low last prices undermine the image of the retailer.

7.7 Statistical evidence of the improvement of the gross yield

As in the previous sections we analyze our results concerning the gross yield from the statistical point of view. Faced with the fact of a widely varying gross yield over the branches with no appropriate theoretical sales model, we have to restrict ourselves to distribution-free statistics.

Therefore, we adapt the Wilcoxon rank sum test to our situation. This is a test method to find out whether or not two data sets are drawn from the same distribution. We sort in increasing order the gross yields of the 20 branches participating in our blind study. We associate the largest value with Rank 1, the second largest with Rank 2, and so on. Then we form the rank sums of the test branches and the control branches, resp. The more the rank sums differ, the less likely is the event that our method did not influence the gross yields/last prices at all.
Table 4  Ranks sums and certainties of improvements of the gross yield

|         | \( i-0.00 \) | \( e-0.00 \) | \( i-0.25 \) | \( e-0.25 \) | \( i-0.50 \) | \( e-0.50 \) | \( i-0.75 \) | \( e-0.75 \) | \( e-1.00 \) | \( e-1.50 \) |
|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Control group | 121          | 131          | 118          | 128          | 105          | 124          | 94           | 120          | 112          | 107          |
| Test group   | 89           | 79           | 92           | 82           | 105          | 86           | 116          | 90           | 98           | 103          |
| Certainty (%) | 87.6         | 97.4         | 82.4         | 95.5         | 48.5         | 91.7         | 19.7         | 86.0         | 68.5         | 54.4         |

For the Wilcoxon rank sum test, it is vital that we have partitioned the 20 branches for the blind study independently at random into test and control branches.

It is intuitively clear that a smaller rank sum for the gross yield/last price is more likely if the corresponding expected values are better. A lower rank sum can indeed occur by pure coincidence, but the probability decreases with the rank sum. As an example, the rank sum for the test branches regarding the gross yield measured by the (elimination) method is 89. If we had not changed anything, the chance to receive a rank sum of 89 or lower would have been 12.4%. So, we have a certainty of 87.6% that our proposed re-packing improved the situation. (More formally: the probability that the gross yields of the test branches and the gross yields of the control branches stem from the same distribution, i.e., nothing has changed systematically, is at most 12.4%.)

Now we consider different scenarios. Let \( y_i(b) \) be the gross yield measured with the method (elimination) of branch \( b \) and \( y_e(b) \) the gross yield measured with the method estimate. By \( i_c \) we denote the scenario where we consider the values \( y_i(b) \) for control branches and the values \( y_i(b) + \frac{c}{100} \) for test branches. Similarly, we define the scenarios for \( e_c \) utilizing \( y_e(b) \) instead of \( y_i(b) \). In Table 4 we have given the rank sums and the certainties of some scenarios.

How can we interpret these numbers? The first two columns of Table 4 show that with a certainty of 87.6% (elimination) and 97.4% (estimate) that our proposed modification increased the expected gross yield. In Scenario \( i-0.25 \) we artificially decrease the gross yield (elimination) values by 0.25 percentage points. The monetary value associated with this specific decrease can be interpreted, e.g., as the implementation and consultancy costs of the modification. So, by a look at Table 4 we can say that with a certainty of 82.4% our proposed modification yields an improvement of the gross yield (elimination) by at least 0.25 percentage points.

8 Conclusion and outlook

The distribution of fashion goods to the branches of a fashion discounter must meet the demand for sizes as accurately as possible. However, in our business case, an estimation of the relative demand for apparel sizes from historic sales data was not possible in a straightforward way.

Our proposal is to use the TDI, a measure that yields basically ordinal information about what were the scarcest and the amplest sizes in a product group in a historic sales period. This information was utilized to change the size distributions for future deliveries by replacing one piece of the amplest size by a piece of the scarcest size in every pre-pack (this can be seen as a sub-gradient improvement step in an iterative size distribution heuristics based on the TDI analysis).

Empirical evidence from a blind study with twenty branches (ten of them, randomly chosen, were supplied according to TDI-based recommendations; ten of them were supplied...
as before) showed a significant increase in gross yield: On average, the increase in the gross yield in our blind study was around one percentage point. The probability that gross yield improvements of at least 0.25 percentage points occurred is at least 87.6 (even 95.5 % if inconsistent data is repaired in a plausible way). And: this was the result of a single iteration of the optimization procedure, which did not result in perfectly balanced Top-Dog-Indices.

Given the large economies of scale of a fashion discounter, we consider the TDI a valuable contribution to revenue management tools in this business sector. Moreover, to the best of our knowledge, our blind study is the first published study that evaluates a revenue management method in the apparel retailer industry by comparing simultaneously obtained business results of test-branches (optimized) and control-branches (no action).

The draw-back of the TDI is its lack of information about the cardinal expected revenue for a given size distribution of the supply. This is partly due to the fact that the loss of a bad size distribution is closely related to the markdown policy of the discounter. This markdown policy, however, is itself subject of revenue management methods. Therefore, we regard the integration of size and price optimization as interesting.

Meanwhile, a method based on this work and extensions is applied in practice routinely by our industry partner.

Acknowledgments This research was supported by the Bayerische Forschungsstiftung. Project DISPO—A Decision Support for the Integrated Size and Price Optimization. The authors thank Lutz Schade and Tobias Metze as well as an anonymous referee for valuable comments on an earlier version of the manuscript.

References

Aksoy, A., Ozturk, N., & Sucky, E. (2012). A decision support system for demand forecasting in the clothing industry. International Journal of Clothing Science and Technology, 24(4), 221–236.

Bitran, G. R., Caldentey, R., & Mondschein, S. V. (1998). Coordinating clearance markdown sales of seasonal products in retail chains. Operations Research, 46(5), 609–624.

Caro, F., & Gallien, J. (2010). Inventory management of a fast-fashion retail network. Operations Research, 58(2), 257–273.

Chan, L. M. A., Max Shen, Z. J., Simchi-Levi, D., & Swann, J. L. (2006). Coordination of pricing and inventory decisions: A survey and classification. In D. Simchi-Levi, S. D. Wu, Z. J. Max Shen (Eds.), Handbook of quantitative supply chain analysis: Modeling in the e-business era, Chap. 14 (pp. 335–392). Dordrecht: Kluwer.

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

Fisher, T. (2013). On simple representations of stopping times and stopping time sigma-algebras. Statistics & Probability Letters, 83(1), 345–349.

Garro, A. (2011). A new product forecasting and distribution optimization: A case study at Zara PhD thesis, MIT, Massachusetts, USA, 194 pp.

Kießling, M., Kurz, S., & Rambau, J. (2012). The integrated size and price optimization problem. Numerical Algebra, Control and Optimization, 2(4), 669–693.

Kök, A. G., & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. Operations Research, 55(6), 1001–1021.

Mahajan, S., & van Ryzin, G. (2001). Stocking retail assortments under dynamic consumer substitution. Operations Research, 49(3), 334–351.

Mostard, J., Teunter, R., & de Koster, R. (2011). Forecasting demand for single-period products: A case study in the apparel industry. European Journal of Operational Research, 211(1), 139–147.

Ni, Y., & Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. Expert Systems with Applications, 38(3), 1529–1536.

Quante, R., Meyr, H., & Fleischmann, M. (2008). Revenue management and demand fulfillment: Matching applications, models, and software. International Journal of Services Sciences, 1(2), 127–147.

van Ryzin, G., & Mahajan, S. (1999). On the relationship between inventory costs and variety benefits in retail assortments. Management Science, 45(11), 1496–1509.
Vulcano, G., van Ryzin, G., & Ratcliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Operations Research, 38*(6), 7373–7379.

Yu, Y., Choi, T.-M., & Hui, C.-L. (2011). An intelligent fast sales forecasting model for fashion products. *Expert Systems with Applications, 60*(2), 313–334.

Zhou, J., & Xu, Q. (2008). Analysis of system dynamics model in garment supply chain. *Journal of Textile Research, 29*(12), 122–125.