

Measuring the impact of stock-keeping unit attributes on retail stock-out performance

Abstract: Stock-outs are one of retail chains' biggest problems because they lead directly to lost sales, reduced profits, and the potential loss of customers. This research applied probit regression to determine the relationship between various stock-keeping unit (SKU) attributes and retail stock-out performance. The data sample came from a large grocery retailer in Serbia and included two high-risk product categories consisting of a total of 115 SKUs and 98 stores. For the identification of stock-outs, a perpetual inventory aggregation method was used. Regardless of the category observed, the variables that were identified as having a detrimental impact on stock-out performance include stock-out at the distribution center, promotion, and high sales speed. On the other hand, a beneficial effect in terms of a reduced number of stock-outs was observed when the ordering process was performed using an automated ordering system.

Keywords: retail, stock-outs, stock-keeping unit, probit regression

1 Introduction

Long-term satisfaction is an essential prerequisite for building customer loyalty to a specific brand and retail chain. Similarly, loyalty and customer satisfaction are two of the most important factors in increasing the sales and profitability of each product category (Colacchio et al. 2003). While maintaining sufficient inventory levels and ensuring the availability of each stock-keeping unit (SKU) are among the best ways of increasing satisfaction and loyalty (Gruen and Corsten 2007), the out-of-stock phenomenon still occurs too frequently in the sales of consumer products (Berger 2002).

Academics and practitioners consider retail stock-outs to be one of the key challenges associated with the retail business, irrespective of the category of goods being sold (Gruen et al. 2002). Studies conducted since the 1960s have shown that the average stock-out rate (percentage of the SKUs not in stock at the time of the audits) is generally constant at a level of around 7 to 8% (Aastrup and Kotzab 2010). This can result in an average financial loss of up to 4% of the total sales (Corsten and Gruen 2003). Regardless of the research perspective, most of the studies indicated that managers are constantly trying to reduce the number of unavailable SKUs (Berger 2003).

Corsten and Gruen (2003) stated that three key forces create pressure and urgency in addressing the shortage of supplies. First, with the increasing amount of information and a large number of retail stores and shopping channels, consumers are increasingly switching to competitor stores. Second, the options for solving this problem are growing – potential solutions are becoming increasingly valuable, since most of them can be applied globally. Third, technology provides new ways of tackling the problem of stock-outs beyond the traditional solutions, which involve higher labor costs and safety stocks.

Problems related to retail stock-outs are usually examined from two key perspectives: the measurement of stock-out rates and the examination of customers' reactions to the stock-outs (Zinn and Liu 2001). This research aims to shed some light on the less explored aspect of this problem – the influence of SKU attributes that can lead to an inventory shortfall. In this regard, the research will be limited to the examination of certain product characteristics that can contribute to the increase in stock-outs.

Although some insights into the significance of retail stock-out drivers can be found in the literature, they are mostly based on a limited type and number of examined SKUs, stores, or time

periods. The objective of this study is to fill this gap by using a different stock-out identification method and a larger data set to measure the impact of potential stock-out drivers on product categories that experience different stock-out effects, further enabling the comparison of the results and improved managerial implications.

The reminder of the paper is organized as follows: the second part provides an overview of the literature and the results of the previous research in the field of retail stock-outs and their causes; the third part of the paper describes the methodology, presenting a brief overview of the statistical methods used in the study as well as the sources of data and the method for measuring retail stock-outs; the fourth section provides the results of the probit regression and a brief interpretation; and the last section concludes the paper and provides recommendations for practitioners with the aim of increasing their attention to potentially high-risk SKU attributes and improving the cooperation between manufacturers/suppliers and retailers.

2 Literature review

Several researchers have tried to reveal the causes of retail stock-outs and the factors that may contribute to the increased stock-out rates (Berger 2003; Corsten and Gruen 2003; Angerer, 2005; Usman 2008; Ehrenthal and Stölzle 2013). Most of the studies on the causes of stock-outs have attributed this issue largely to the practice of ordering and forecasting at the store level. Managers of retail outlets must simultaneously manage thousands of SKUs, hundreds (often thousands) of which may be on promotion and may experience fluctuating demand. Furthermore, retailers have recently faced additional challenges, such as reduced shelf space due to increasing competition and product proliferation.

Corsten and Gruen (2003) concluded that about three-quarters of the causes of stock-outs at the global level originate from practices at the retail store level, while one-quarter to one-third originates at the distribution center (DC) or the headquarters level. From the perspective of retail processes, almost half of the causes of insufficient stock can be associated with ordering problems (i.e., late retail orders or insufficient quantity ordered), often due to imprecise and unreliable predictions.

Ehrental and Stölzle (2013) conducted research with the goal of increasing the understanding of the causes of stock-outs at the retail level. A two-year period of field research at the store level found that the causes of stock-outs are retailer-, store-, category- and SKU-specific. However, few studies exist on SKU characteristics as drivers of retail stock-out performance. The attributes usually examined are promotion activities (Ettouzani et al. 2012), product price (DeHoratius and Raman 2008), availability at the DC (Usman 2008), sales speed (Gruen et al. 2002; Angerer 2005), variation in demand (Mattsson 2010), case pack size, and shelf life (Angerer 2005).

In the study conducted by Andersen Consulting (1996), SKUs with a higher turnover rate had a higher stock-out rate. It is clear that faster SKUs must be checked and ordered more often because the risk of stock-out is much higher. On the other hand, a similar situation occurs with very slow SKUs. Very slow products are often overlooked when ordering is performed manually. Stölzle and Placzek (2004) found that the highest rate of stock-outs corresponded to the products with the lowest inventory turnover.

When the demand for a particular SKU varies significantly over time, it is very difficult to determine the right quantity to order and keep in stock. In this case, the ordering system must be flexible enough to order different amounts in subsequent periods, taking into account the

limitations in terms of the case pack size. When demand variations occur as a result of strong seasonal effects, ordering becomes very complex and the importance of safety stock increases (Mattsson 2010).

Several studies aimed to reveal the impact of price on stock-out performance, achieving conflicting results. On the assumption that more expensive goods have higher margins, managers tend to maintain a higher inventory level to avoid stock-out and customer dissatisfaction (Van Donselaar et al. 2010). DeHoratius and Raman (2008) concluded that inaccurate records are more likely to be kept for inexpensive products than expensive ones, as expensive SKUs receive more attention. Accordingly, expensive goods should have a lower stock-out rate than cheap products.

If a store generates a demand for an SKU and it is not available at the DC to be shipped to the store on time, a stock-out situation at the distribution center is created. Usman (2008) concluded that a stock-out at the DC may have a detrimental downstream impact at the retail store level. If there is a minimal downstream impact at the store level regarding stock-outs and eventually sales, then the inventory levels at the DC can be kept to a minimum, achieving cost savings.

One of the factors that can have a critical impact on the availability of a product is promotional activity. This factor is usually associated with significant variation in the demand for the SKU, which may further lead to a shortage of supplies (Ettouzani et al. 2012). Regardless of the type of ordering system, the ability to prevent a stock-out is directly related to the ability to forecast the demand during promotions.

Another product attribute that can cause an increase in the stock-out rate is the product size. There is no clear consensus in the literature on this issue (Andersen Consulting 1996). On

the one hand, there is an argument that larger products require more space on the shelf and thus fewer are available to the customer. This implies that smaller items always have greater availability and a lower stock-out rate. An analysis conducted on four categories of consumer products showed that products with larger packaging have a higher stock-out rate than those with smaller packaging (Angerer 2005).

Angerer (2005) used manual audit methods for stock-out identification, as well as linear regression and ANOVA, to examine the impact of ordering system automation on stock-out performance. He found that variation in demand, along with sales speed, price, and case pack size, have a significant impact on manual but not automated ordering systems. He concluded that even simple automatic replenishment systems are able to reduce the average shelf stock-out rate independently of the product characteristics.

In most of the studies mentioned, stock-out identification was performed with the use of manual audit methods or point-of-sale estimation (POS) methods (the use of models that estimate stock-outs based on SKU sales trends). Since the manual audit method is time and resource intensive and the POS method is applicable to fast-sale items only, the results and conclusions were mostly based on either a small number of SKUs and retail stores or a short time period.

In the present study, we applied the perpetual inventory aggregation method (the use of store-level compiled inventory records), which allowed the analysis of a larger number of SKUs and stores in an extended time frame. The data set structure and the identification method employed also enabled the measurement of the impact of several literature-based stock-out drivers on two product categories with different stock-out effects (consumer reactions). This further enabled the comparison of the results obtained from one category with the results

obtained from the other one, which provided recommendations for both manufacturers and retailers.

3 Research methodology

Probit regression was selected for the main statistical analysis. In addition to a description of the selected statistical method and its suitability for this research, this section provides a brief profile of the retail chain examined, the sources and structure of the acquired data, and the method used for measuring the stock-out rate.

3.1 Data set description

Retailer_A is a grocery retailer that operates 130 retail stores in Serbia and one distribution center (DC) in Belgrade. The assortment consists of over 10,000 active SKUs, which are ordered in two ways: automatic and manual ordering. Automatic ordering is performed by a computer, which uses algorithms to define the order quantities and ordering time for each store. Manual ordering is performed by store managers, who audit the store inventory and order quantities as needed. The store orders are collected, summarized, and ordered by suppliers. The total quantity is first delivered to the DC and then forwarded to the retail stores.

One of the key prerequisites for drawing conclusions from a given set of data is the relevance and consistency of the data. This involves proper pre-processing of the data for selected methods and techniques. The main data source was the central database of the retail company, which records information on the sales and inventory of all the products and retail stores. Different reports, generated from the database, were pre-processed and compiled into a

data set of high granulation. Preparation and analysis were performed using the statistical software Stata 12.

The first part of the pre-processing included the selection of SKUs. From the assortment of over 10,000 active SKUs, two high-risk product categories were selected for this research. First, the baby diapers product category (comprising 69 SKUs) was selected for the research since it is considered to be one of two categories that possess the largest potential for a direct loss for retailers. In their global study, Corsten and Gruen (2003) showed that, in the case of a stock-out, around 40% of baby diaper customers worldwide change stores to acquire the desired item. This is due to the great loyalty that exists between the customer and the product and customers are generally not willing to compromise in the form of a substitute of any kind.

For the second data set, the paper towels and toilet tissue category (comprising 46 SKUs) was selected due to the opposite effect that it produces when it comes to stock-outs. Corsten and Gruen (2003) showed that this type of product carries the lowest risk for the retailer, but the highest risk for the manufacturer. They found that toilet tissue and paper towels, along with the salted snacks category, involve the highest level of brand substitution. In the case of a baby diapers stock-out, 39% of consumers change stores (retailer loss) and 20% of consumers change brands (manufacturer loss); in the case of paper towels, 37% of consumers change brands (manufacturer loss) and 13% of customers change stores (retailer loss).

These two categories (retailer risk and manufacturer risk) represent a potential field of conflict regarding the manner in which the manufacturers and retailers approach stock-out issues. Therefore, examining the potential drivers of stock-outs within product categories with different stock-out effects can be beneficial for both retailers and manufacturers, especially if some of the drivers turn out to be significant in both cases. This could lead to cooperative tracking of

products with certain characteristics with the aim of reducing the negative effects of stock-outs. If this appears not to be the case, recommendations could be made separately for retailers and manufacturers.

The next part of the pre-processing included the selection of retail stores. To ensure data consistency, only those retail stores that had been operating for a whole year were included. In addition, the stores that recorded incomplete data in certain periods due to technical problems with the system were excluded from the sample. In this regard, of more than 130 stores, the data set included 98 retail stores that had consistent sales during the entire calendar year.

The final step of the pre-processing stage included the integration of some electronic data from different sources into a single SKU/store/day-level spreadsheet for the selected SKU and store groups. The following electronic data were integrated: inventory reports at the DC and store level, daily sales and delivery reports, assortment tables at the store level, promotion periods, brand, price, case pack size information, and tables of automatically ordered SKUs. The sales and inventory data were used to derive new information regarding sales speed and sales variation. The final data set for both categories included 115 SKUs in 98 stores over the period from January 1, 2012 to December 31, 2012. The complete data set was structured with high granulation and consisted of a total of 2,953,027 observations at the SKU/store/day level.

3.2 Measurement of stock-outs

One of the biggest challenges in the data collection process is the measurement of stock-outs. According to Gruen and Corsten (2007), three different kinds of stock-out identification could be used: (1) the manual audit method, (2) point-of-sale estimation, and (3) perpetual inventory aggregation.

Due to the challenges involved in obtaining comprehensive and meaningful stock-out data, perpetual inventory aggregation metrics were used. This system implies that the inventory in the information system corresponds to a real physical state, which, for various reasons, is often not the case. As a consequence of non-compliance, there can be a surplus situation in the information system (phantom stock) or a surplus in a physical state (hidden stock), and the results of the study by DeHoratius and Raman (2008) indicate that the disagreement may be as much as 65%.

To cope with this problem, the examined retailer tends to carry out physical inventory checks every two weeks. The inventory data were double-checked to reduce the potential error and raise the accuracy of the data used for the analysis. This included the cross-checking of sales and delivery data with the data on assortment and inventory at the SKU/store/day level. The biggest advantage of the proposed identification method is that it can provide useful insights regarding stock-out rates at the store, SKU, or entire system level. It also enables the analysis of a larger number of SKUs and stores during a long time period, which is not the case with the other two methods.

The proposed model registers inventory at the level of SKU/store/day, which we used to analyze the stock-out performance. The stock-out rate can be calculated as a percentage of the maximum possible SKU availability (full stock), as illustrated in Table 1. For example, if we observed three stores, two SKUs, and a period of two days, and for one SKU in one store for one day the inventory level at the end of the day was zero, the stock-out rate in this case would be 8.3% (1/12).

[Insert Table 1 about here]

The mathematical expression of the stock-out rate obtained in this way would be:

$$\text{Stockout rate (\%)} = \frac{\sum_i^n \sum_j^m \sum_k^l \text{stockout}_{ijk}}{\sum_i^n \sum_j^m \sum_k^l \text{maximum SKU availability}_{ijk}} \quad (1)$$

where i represents the SKU for which the stock-out rate is measured; n is the total number of SKUs; j is the store for which the stock-out rate is measured; m is the total number of observed stores; and k and l are the first and the last day, respectively, when the SKU appeared in a certain store. Bearing in mind the frequent changes in assortment, the time range is defined from k to l to exclude the possibility that the period when the SKU was not in the store assortment is considered as a period in which the SKU was out of stock.

In addition, it is extremely important to note that this method takes into account not the availability of the product on the shelf, but its availability in the whole store, which implies that the actual shelf stock-out rate may be higher than calculated. The analysis of stock-out performance at the SKU/store level (by using the stock-out rate as a dependent variable) can indicate differences in performance among SKUs or stores, as well as oscillations over time. Even though the stock-out rate provides a good measure of performance, we decided to analyze the data at a more detailed level, since this allows the inclusion and analysis of different variables, such as the promotional effect and the availability at the DC.

3.3 Probit regression analysis

Probit regression analysis can be used for modelling the occurrence of a retail stock-out. The dependent variable will take the value of 1 when there is a stock-out of a certain SKU on a certain day in a certain store, and the value 0 is taken when there is at least 1 unit in stock. Therefore, it is possible to develop a probit regression model that predicts the occurrence of a stock-out for 115 SKUs and 98 stores for a period of 1 calendar year. The dependent variable of a retail stock-out for any SKU–store–time period combination to be modelled is given as:

$$stockout_{ijk} = f(SKU \text{ variables}) \quad (2)$$

where i is 1 out of 115 selected SKUs, j is 1 out 98 selected stores, and k is the day in the year for which the analysis will be carried out. The SKU variables are: the availability of the SKU at the DC, promotional activity, average price, sales speed, sales variation, ordering type, and SKU size.

In addition to probit regression, several similar statistical approaches were considered in the search for the appropriate method of data analysis. Some provided similar results and conclusions, while others were rejected due to certain restrictions. For example, an ordinary least squares regression can be used with a binary dependent variable for the description of conditional probability, but it can lead to errors when testing hypotheses (Long 1997). On the other side, logistic regression provides similar results to probit regression, so we decided to measure the relative quality of the two models and use the one with a better fit.

[Insert Table 2 about here]

We used two popular measures for comparing maximum likelihood models – the Bayesian information criterion (BIC) and Akaike information criterion (AIC). The AIC and BIC can be viewed as measures that combine fit and complexity. Fit is measured negatively by $-2 \ln(\text{likelihood})$; the larger the value, the worse the fit. Complexity is measured positively, either by $2k$ (AIC) or $\ln(N)k$ (BIC). The results in Table 2 show a better fit (lower value) for the probit model in both the observed categories; therefore, probit regression was selected for this research.

4 Results and interpretation

Probit regression was performed at the SKU/store/day level using the statistical software Stata 12. The dependent variable that indicates a stock-out was depicted in binary form, so 1 indicates a stock-out and 0 the existence of SKU inventory in a certain store on a certain day. In Tables 3

and 4, which illustrate the descriptive statistics of the variables for both data sets, it can be seen that the average stock-out rate for the retailer risk category is 1.85% and that for the manufacturer risk category is 3.41%. The stock-out rate represents the ratio of the number of observations for which stock-out = 1, and the total number of observations (maximum possible SKU availability).

To exclude the independent variables with significant dependence from the model, a correlation analysis was conducted. The inclusion of significantly dependent variables may reduce the accuracy of the model and lead to misconceptions regarding the influence of the independent variables. Seven different variables were measured for both data sets and five of them were included in the final models: sales speed, product price, stock-out in a DC, promotional effect, and ordering type.

[Insert Table 3 about here]

[Insert Table 4 about here]

The variable product price had a high correlation with the SKU size (0.80 for the retailer risk category and 0.57 for the manufacturer risk category); thus, the variable SKU size was not included in the model. The variable sales variation had a high correlation with the sales speed (0.89 for the retailer risk category and 0.87 for the manufacturer risk category); thus, the variable sales variation was not included in the model. The variables of sales speed and average price were analyzed as continuous variables, while the variables of stock-out in a DC, promotion, and ordering type were analyzed as Boolean (binary) ones.

The sales speed variable represents an average daily quantity of each SKU sold at each of the observed stores. Due to the difference in demand between the stores, the calculation of the sales speed at the SKU/store level increases the accuracy of the predictions in relation to the

SKU sales speed at the retail chain level. The selected categories are not seasonal and the demand is generally constant, except for weekday or promotion oscillations. Oscillations from average weekday sales ranged from -14% to +13% for the retailer risk category, and from -10% to +6% for the manufacturer risk category. Promotion oscillations included average daily sales increase of 62% for the retailer risk category, and 24% for the manufacturer risk category. The product price variable was calculated as the average selling price of a product at the retail chain level, expressed in euros. The stock-out in a DC variable is a binary one, for which 1 indicates a stock-out and 0 the existence of SKU inventory on a certain day. The inventory levels for each SKU at the DC and store levels were recorded at the end of each day.

The promotion variable is binary, with 1 indicating that the SKU was on promotion and 0 that it was regularly sold in a certain store on a certain day. Promotions usually include discounted sales. From Tables 3 and 4 we can see that about 11% of all the observations account for promotions for the retailer risk category and around 16% for the manufacturer risk category. For the retailer risk category, 57 of 69 SKUs were on promotion during 2 periods (from January 20 until March 14 and from October 15 until November 14), and all the SKUs in the manufacturer risk category were on promotion during 2 periods lasting for 4 weeks (from February 15 until March 14 and from October 15 until November 14). These were the only times when the products were on promotion.

The ordering type variable is a binary one, where 1 indicates that the SKU was ordered automatically and 0 indicates that it was ordered manually at the retail chain level. Automated ordering means that a computer suggests the time and quantity of SKUs to be ordered (up to a defined level), which has to be approved by employees in the logistics and commercial sector. In the retailer risk category, 43 out of 69 SKUs were automatically ordered (accounting for 68% of

all the observations), and in the manufacturer risk category, 17 out of 46 SKUs (accounting for 36% of all the observations).

Tables 5 and 6 show the probit model for both product categories. *Prob > chi2* is the probability of obtaining the chi-square statistic if there is no effect of the combination of independent variables on the dependent variable. In other words, this is the p-value, which is compared with a critical value (0.01) to determine whether the overall model is statistically significant. In both categories, the models are statistically significant as are all the model coefficients.

[Insert Table 5 about here]

[Insert Table 6 about here]

4.1 Retailer risk category

To interpret the results obtained, we introduced the concept of marginal effects, which provides a measure of the association of the independent variables with the outcome variable. The margin, i.e., average predicted probability of a stock-out, is a statistic based on a fitted model calculated over a data set in which some or all of the covariates are fixed at different values from their real ones. Figure 1 shows the average predicted probabilities of a stock-out with the 95% confidence interval for the continuous variables sales speed and product price. Another way of conducting the evaluation would be to analyze the predicted probabilities evaluated at the means of the covariates, which provided very similar results.

[Insert Figure 1 about here]

Based on a probit regression of stock-outs, the average probability of a stock-out for average sales of 1 was 0.034 and increased to 0.31 for average sales of 5 (Figure 1). Since the

probability of a stock-out almost doubles with an increase of 1 in the average sales, we can conclude that the sales speed has a significant impact on the probability of a stock-out for the retailer risk category.

On the other side, this does not apply to the product price. The average probability of a stock-out for average price of 3 euros was 0.016 and increased to 0.020 for average price of 12 euros. In relation to the retailer risk product category, we can conclude that the product price increases the probability of a stock-out but does not have a significant impact on the stock-out performance. Next, we examined the impact of the binary independent variables on stock-outs for the retailer risk category (Table 7).

[Insert Table 7 about here]

The average probability of a stock-out in the case of SKUs' full availability at the DC would be 0.016, compared with the opposite situation (a total lack of SKUs at the DC), and almost 5 times higher predicted probability of 0.079. Another variable that has a significant impact on stock-outs at the store level is promotion. The predicted probability of a stock-out in the case of the retailer risk category is about 50% higher when the SKU is sold during a promotion. The independent variable that decreases the probability of stock-outs is the automation of the ordering system. The predicted probability of a stock-out at the retail level is more than 3 times lower in the case of automatically ordered SKUs.

To examine the marginal effects of the interaction of binary variables, we generated Table 8. The first line in the table reports the marginal probability for a stock-out with $DC = 0$, $promotion = 0$, and $ordering\ type = 0$. Namely, it reports the estimated probability if every observation in the data was treated as if the SKU was available in the DC, regularly sold, and manually ordered. Obviously, the best-case scenario is 0-0-1, with the predicted stock-out

probability of 0.008 (item available at the DC, regularly sold, and automatically ordered), while the worst-case scenario is 1-1-0, with a stock-out probability of 0.17 (item out of stock at the DC, sold on promotion, and manually ordered). It is also interesting that the negative effect of the stock-out at the DC alone ($101 = 0.051$) is almost equal to the combined negative effect of promotion and manual ordering ($010 = 0.047$).

[Insert Table 8 about here]

4.2 Manufacturer risk category

To compare the results obtained, the same type of analysis was applied to the second data set – the manufacturer risk category. Figure 2 shows the average predicted probabilities with the 95% confidence interval for the continuous variables sales speed and product price. The average probability of a stock-out for average sales of 1 was 0.036 and increased to the four times higher probability of 0.151 for average sales of 15. We can conclude that the sales speed has a rather significantly positive impact on the probability of stock-out for the manufacturer risk category, which is still lower than that in the case of the retailer risk category.

[Insert Figure 2 about here]

Concerning the product price, the average probability of a stock-out for average price of 1 euro was 0.03 and increased to 0.06 for average price of 4 euros (Figure 2). Even the probability of a stock-out for high priced products is two times greater than for low priced products, it still remains below 0.1. Regarding the manufacturer risk product category, we can conclude that the product price has a greater impact on the stock-out performance than in the case of the retailer risk category, but it is not as significant as the sales speed. Furthermore, we examined the impact of the binary independent variables on stock-outs for the manufacturer risk category (Table 9).

[Insert Table 9 about here]

The average probability of a stock-out in the case of SKUs' full availability at the DC would be 0.023, compared with the opposite situation (a total lack of SKUs at the DC), and almost four times higher predicted probability of 0.085, which is similar to the result for the retailer risk category. Slightly different results were obtained with regard to promotion. The predicted probability of a stock-out would be about 20% higher when the SKU is sold during a promotion, which is the half of the increase obtained in the retailer risk category. In relation to the retailer risk category, the automation of the ordering system within the manufacturer risk category also decreases the probability of a stock-out by 40%, in comparison with the three times lower probability in the case of the retailer risk category.

The marginal effects of the interaction between the binary variables for the second category are shown in Table 10. The best-case scenario is again 0-0-1, with the predicted stock-out probability of 0.015 (item available in the DC, regularly sold, and automatically ordered), while the worst-case scenario is again 1-1-0, with the probability of 0.109 (item out-of-stock in the DC, sold on promotion, and manually ordered). In this case, the negative effect of promotion and manual ordering ($010 = 0.032$) is almost twice as low as the negative effect of the stock-out at the DC alone ($101 = 0.062$).

[Insert Table 10 about here]

The overall finding is that promotional activity and the automation of the ordering system have a lower impact on the stock-out performance in the manufacturer risk category than in the retailer risk category, while the impact of a stock-out at the DC is very similar. The same conclusion can be drawn from Figure 3, which shows the increase in the stock-out probability in the case of a sales speed increase, a decrease in the stock-out probability in the case of the

presence of the favorable variable (automation of the ordering system), and an increase in the stock-out probability in the case of the unfavorable variables (stock-out at the DC and promotion) for both product categories. In this case, the average price variables were fixed at their average values (8.86 and 1.49).

[Insert Figure 3 about here]

5 Conclusion

Stock-outs in the retail business can be used as a performance measure of the supply system and as such they can be determined and monitored. The analysis confirmed that some of the stock-out drivers are product/category dependent, while some of them are general. Stock-outs at the distribution center level have the most prominent impact on stock-outs at the store level, regardless of the product category. Stock-outs in the distribution center increase the probability of stock-outs in retail stores by four to five times. To cope with this problem and prevent lost sales, retail chains should work with manufacturers/suppliers to ensure higher product availability at the DC, taking into account higher inventory holding costs.

The probit regression showed that promotion activity also has a significant influence on stock-outs, which is twice as high in the case of the retailer risk category as in the manufacturer risk category. Since the effect of promotional activity also depends on the price elasticity of demand, it is expected that higher-priced categories will be more susceptible to negative effects of promotions. Therefore, retail managers and manufacturers have to focus more on cooperative planning to be able to predict the fluctuations in demand that occur as a result of promotional activities, especially in the case of fast-selling and expensive SKUs.

Although the analysis in both categories showed that an increase in the product price leads to an increase in the probability of a stock-out at the retail store level, we can conclude that there is not enough evidence to treat this increase as significant. It is important to note that both categories consisted of low- to medium-priced SKUs, so the price range may have not been wide enough to express the full-scale negative effect, as was the case with some examples found in the literature. Furthermore, a stronger effect of price on product availability was recorded in the case of the generally cheaper product category.

Although fast-selling SKUs usually attract more managerial attention, the analysis showed that a higher probability of a stock-out can be expected for high-speed SKUs. This is especially true for the retailer risk category, in which the predicted stock-out probability is several times higher for very fast SKUs, compared with half the size effect in the manufacturer risk category. Due to the large impact on profits, the prevention of fast-item stock-outs can produce significant improvements in financial results for retailers and manufacturers.

A particularly negative effect can be expected in the case of the combination of several negative attributes, such as a high-sales-speed SKU, a stock-out at the DC, and promotion. The combination of a stock-out at the DC, as the most dangerous factor, and a high-demand SKU, such as fast-selling items and items on promotions, leads to an almost certain stock-out at the store level. Retailers should improve the replenishment and auditing procedures for those items that fall into one of these categories, which would improve the forecasting and ordering processes, resulting in a smaller number of stock-outs at the DC and store levels.

One factor that factor that can mitigate the risk of a stock-out at the retail level is the automation of the ordering process. Automatically ordered SKUs recorded a significantly lower number of stock-outs in both categories, even though this effect was double-sized in the retailer

risk category. Although technology cannot usually respond to all the challenges of the dynamic retail business, it can provide assistance to retailers and contribute to greater efficiency in their part of the supply chain. In order to obtain these benefits, key prerequisites such as data integrity and development of correct forecasting and order algorithms must be met.

References

- Aastrup J, Kotzab, H (2010) Forty years of out-of-stock research – And shelves are still empty. *International Review of Retail, Distribution and Consumer Research* 20:147-164.
- Andersen Consulting (1996) Where to look for incremental sales gain. The retail problem of out-of-stock. The Coca-Cola Research Council, Atlanta.
- Angerer A (2005) The impact of automatic store replenishment systems on retail. PhD thesis. University of St. Gallen.
- Berger R (2002) Full-shelf satisfaction: Reducing out-of-stocks in the grocery channel. Grocery Manufacturers of America (GMA), Washington, DC.
- Berger R (2003) Optimal shelf availability – Increasing shopper satisfaction at the moment of truth. Kontich. ECR Europe and Roland Berger, Belgium.
- Colacchio F, Tikhonova O, Kisis J (2003) Consumer response to out-of-stock: Decision-making process and influencing factors. In ECR Europe Berlin Conference.
- Corsten D, Gruen, T (2003) Desperately seeking shelf availability: An examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management* 31:605-617.
- DeHoratius N, Raman A (2008) Inventory record inaccuracy: An empirical analysis. *Management Science* 54:627-641.

- Ehrenthal JC, Stölzle W (2013) An examination of the causes for retail stockouts. *International Journal of Physical Distribution & Logistics Management* 43:54-69.
- Ettouzani Y, Yates N, Mena C (2012) Examining retail on shelf availability: Promotional impact and a call for research. *International Journal of Physical Distribution & Logistics Management* 42:213-243.
- Gruen TW, Corsten DS (2007) A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry. Grocery Manufacturers of America, Washington, DC.
- Gruen TW, Corsten DS, Bharadwaj S (2002) Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses. Grocery Manufacturers of America, Washington, DC.
- Long JS (1997) Regression models for categorical and limited dependent variables. Sage Publications, Thousand Oaks, CA.
- Mattsson SA (2010) Inventory control in environments with seasonal demand. *Operations Management Research* 3:138-145.
- Stölzle W, Placzek T (2004) Umsetzung von Optimal Shelf Availability – Messkonzepte und Standardisierungspotenziale. Presentation at the BVL Congress.
- Usman K (2008) Determination of drivers of stock-out performance of retail stores using data mining techniques. PhD thesis. Massachusetts Institute of Technology.
- Van Donselaar KH, Gaur V, Van Woensel T, Broekmeulen RA, Fransoo JC (2010). Ordering behavior in retail stores and implications for automated replenishment. *Management Science* 56:766-784.
- Zinn W, Liu PC (2001) Consumer response to retail stockouts. *Journal of Business Logistics* 22:49-71.

Table 1 Proposed model for stock-out identification

Inventory level	Store 1	Store 2	Store 3	Store n
SKU 1, day 1	10	45	14	---
SKU 1, day 2	2	7	0	---
SKU 2, day 1	1	5	5	---
SKU 2, day 2	6	1	32	---
SKU n, day n	---	---	---	---

Table 2 Model selection statistics

Data set	Number of obs.	Model	AIC	BIC
Retailer risk category	1,389,172	Logit	242,406.2	242,479.1
		Probit	241,880.0	241,952.8
Manufacturer risk category	1,563,855	Logit	433,931.0	434,004.5
		Probit	433,888.1	433,961.7

Table 3 Descriptive statistics of the variables for the retailer risk category

Variable	Obs.	Mean	Std dev.	Min.	Max.
avg_sales	1389172	0.2330158	0.3637913	0	5.044025
avg_price	1389172	8.8611810	2.8438430	3.2	12.24
st_out_dc	1389172	0.0410489	0.1984034	0	1
promotion	1389172	0.1137591	0.3175186	0	1
auto_or dr	1389172	0.6797517	0.4665721	0	1
stock_out	1389172	0.0185477	0.1349212	0	1

Table 4 Descriptive statistics of the variables for the manufacturer risk category

Variable	Obs.	Mean	Std dev.	Min.	Max.
avg_sales	1563855	0.4971689	0.9021178	0	14.91257
avg_price	1563855	1.4928780	0.8573341	0.486092	3.970331
st_out_dc	1563855	0.1577934	0.3645473	0	1
promotion	1563855	0.1633783	0.3697106	0	1
auto_or dr	1563855	0.3574065	0.4792361	0	1
stock_out	1563855	0.0341950	0.1817298	0	1

Table 5 Probit model for the retailer risk category

Probit regression	Number of obs. =		1,389,172	Prob > chi2 =	0.0000
Log likelihood =	-120,933.99	LR chi2(5) =	14,662.21	Pseudo R2 =	0.0572
Stock-out	Coef.	Std err.	z	P > z	[95% conf. interval]
avg_sales	0.3543335	0.0057038	62.12	0.000	0.3431542 0.3655128
avg_price	0.0117355	0.0010084	11.64	0.000	0.0097591 0.0137118
1.st_out_dc	0.7499024	0.008953	83.76	0.000	0.7323549 0.7674499
1.promotion	0.1776517	0.0073312	24.23	0.000	0.1632829 0.1920205
1.auto_or dr	-0.5321096	0.0058817	-90.47	0.000	-0.543637 -0.520582
_cons	-2.051561	0.0090279	-227.25	0.000	-2.069255 -2.033867

Table 6 Probit model for the manufacturer risk category

Probit regression		Number of obs. =	1,563,855	Prob > chi2 =	0.0000
Log likelihood =	-216,938.05	LR chi2(5) =	32,261.40	Pseudo R2 =	0.0692
Stock-out	Coef.	Std err.	z	P > z	[95% conf. interval]
avg_sales	0.0582953	0.0019406	30.04	0.000	0.0544918 0.0620988
avg_price	0.1056493	0.0024302	43.47	0.000	0.1008862 0.1104124
1.st_out_dc	0.6287118	0.004369	143.9	0.000	0.6201487 0.6372749
1.promotion	0.0929786	0.0051285	18.13	0.000	0.0829268 0.1030303
1.auto_ordr	-0.2186341	0.0051905	-42.12	0.000	-0.228807 -0.208461
_cons	-2.14338	0.0057246	-374.41	0.000	-2.154601 -2.13216

Table 7 Marginal effects for the binary variables – retailer risk category

		Margin	Std err.	z	P > z	[95% conf. interval]
st_out_dc	0	0.0163415	0.0001088	150.15	0.000	0.0161282 0.016555
	1	0.0790114	0.0012022	65.72	0.000	0.0766551 0.081368
promotion	0	0.0175012	0.0001179	148.39	0.000	0.0172700 0.017732
	1	0.0262521	0.0003911	67.12	0.000	0.0254855 0.027019
auto_ordr	0	0.0387674	0.0003377	114.79	0.000	0.0381055 0.039429
	1	0.0115041	0.0001042	110.36	0.000	0.0112998 0.011708

Table 8 Interaction of binary variables for the retailer risk category

st_out_dc#promotion #auto_ordr	Margin	Std err.	z	P > z	[95% conf. interval]
0 0 0	0.0325739	0.0003077	105.87	0.000	0.0319709 0.0331769
0 0 1	0.0089916	0.0000971	92.60	0.000	0.0088013 0.0091819
0 1 0	0.0475099	0.000727	65.35	0.000	0.0460850 0.0489349
0 1 1	0.014186	0.0002562	55.37	0.000	0.0136838 0.0146882
1 0 0	0.1345417	0.0020395	65.97	0.000	0.1305443 0.1385391
1 0 1	0.0515535	0.0008889	58.00	0.000	0.0498114 0.0532956
1 1 0	0.1762801	0.0028848	61.11	0.000	0.1706260 0.1819342
1 1 1	0.0728417	0.0014165	51.42	0.000	0.0700654 0.0756180

Table 9 Marginal effects for the binary variables – manufacturer risk category

		Margin	Std err.	z	P > z	[95% conf. interval]
st_out_dc	0	0.0232560	0.0001333	174.41	0.000	0.0229947 0.023517
	1	0.0851929	0.000553	154.06	0.000	0.0841091 0.086277
promotion	0	0.0330061	0.0001548	213.19	0.000	0.0327026 0.033310
	1	0.039979	0.0003741	106.86	0.000	0.0392457 0.040712
auto_ordr	0	0.0381381	0.0001831	208.31	0.000	0.0377793 0.038497
	1	0.0238148	0.0002423	98.29	0.000	0.0233399 0.024290

Table 10 Interaction of the binary variables for the manufacturer risk category

st_out_dc#promotion #auto_ordr	Margin	Std err.	z	P > z	[95% conf. interval]	
0 0 0	0.0257309	0.0001777	144.83	0.000	0.0253827	0.0260791
0 0 1	0.0151894	0.0001743	87.14	0.000	0.0148478	0.0155310
0 1 0	0.0317908	0.0003626	87.68	0.000	0.0310802	0.0325014
0 1 1	0.0191038	0.0002829	67.53	0.000	0.0185494	0.0196583
1 0 0	0.0930810	0.0006261	148.68	0.000	0.0918539	0.0943080
1 0 1	0.0618307	0.0007043	87.80	0.000	0.0604504	0.0632110
1 1 0	0.1094444	0.0010176	107.55	0.000	0.1074500	0.1114389
1 1 1	0.0739396	0.0009691	76.30	0.000	0.0720403	0.0758390

Fig 1 Marginal effects for sales speed and price – retailer risk category

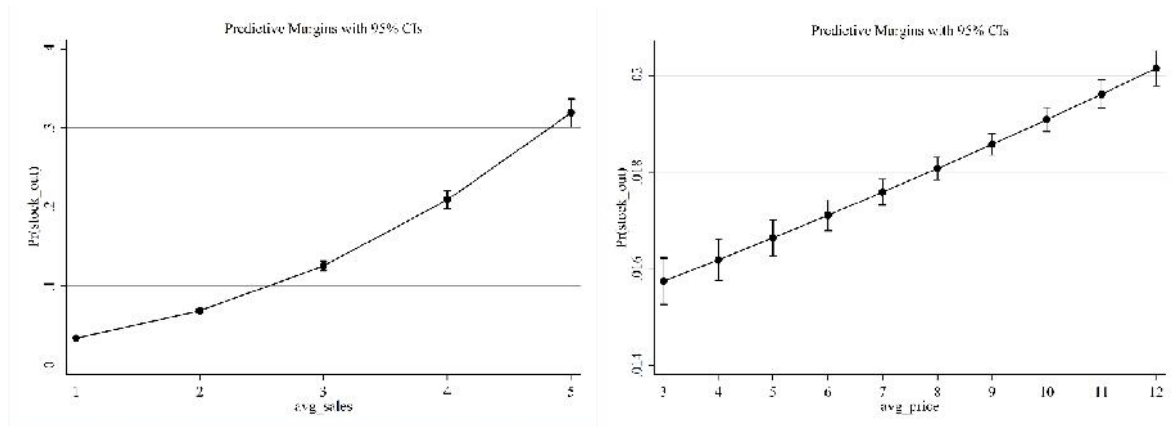


Fig 2 Marginal effects for sales speed and price – manufacturer risk category

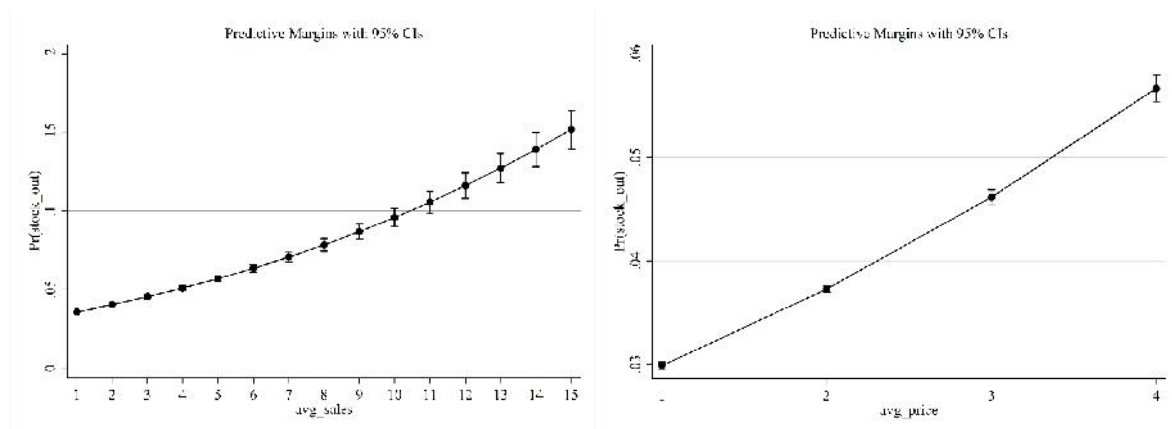


Fig 3 Predicted probabilities of a stock-out in the case of a fixed product price

