Modeling handling operations in retail stores : an empirical analysis

Modeling Handling Operations in Retail Stores: an Empirical Analysis

Alina Curşeu, Tom van Woensel, Jan Fransoo, Karel van Donselaar and Rob Broekmeulen
Technische Universiteit Eindhoven, Department of Technology Management, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
April 3, 2006

Abstract
Shelf-stacking represents the daily store process of manually refilling the shelves with products from new deliveries. For most retailers, this handling operation is usually labor-intensive and often very costly. This paper presents an empirical study of the shelf-stacking process in retail stores. We examine the complete process at the level of individual sub-activities and study the main factors that affect the execution time of this common store operation. Based on insights from different sub-activites, a simple prediction model is developed that allows estimating the total stacking time per order line, only on the basis of the number of case packs and consumer units. The model is tested and validated using real-life data from two European retailers and may serve as a useful tool for evaluating the amount of workload the usual shelf-stacking operation requires. Furthermore, we illustrate the benefit of the model in quantifying analytically the potential time savings in the stacking process, which offers interesting opportunities for extending existing inventory control rules with a handling component.

Keywords: retail operations, stacking process, store processes

1. Introduction
In today's highly competitive market environment, many retailers are concentrating on controlling costs, as a means of achieving operational excellence and their business success as a whole. In a recent logistics survey (Butner, 2005), an overwhelming 83% of participants ranked logistics cost reduction as their primary objective, competing with the permanent strive to provide a high customer service. Proper control of store operating expenses typically requires balancing transportation, inventory, shelf space and handling costs. Currently, models that assess the overall operational costs in retail stores on multiple dimensions are not available. Existing research in retail operations mainly concentrates on inventory, marketing, or planograming decisions (Corstjens and Doyle, 1981; Drèze *et al.,* 1994; Urban, 1998; Cachon, 2001; Hoare and Beasley, 2001). Typically, in these models the handling time and its related costs are not considered explicitly (see e.g. Themido *et al*. 2000, where handling costs are treated in an aggregate way). This research focuses exclusively on the handling cost component of retail operations, an area, we believe, still largely overlooked.

For most retailers, the store handling operations are not only labor-intensive, but also very costly. Empirical studies (see e.g. Saghir and Jonson, 2001, Broekmeulen *et al.,* 2004) suggest that the handling costs in the retail chain represent the largest share of operational costs with high shares in the retail stores. There is however, a general lack of understanding of what drives handling costs in retail stores, and little evidence exists in the academic literature on this topic. An early study that considers both inventory and handling costs comes from 1960's (Chain Store Age, 1963). SLIM (Store Labor and Inventory Management), a system widely promoted in the mid–1960's, focused on minimizing store handling expense, by reducing backroom inventories and the double handling of goods (Chain Store Age, 1965). Two other studies carried out by the Swedish group DULOG in 1976 and 1997, measured package handling time in the store, in order to gather information about the impact of the type of package on handling efficiency in the grocery retail supply chain (DULOG, 1997).

More recently, Van Zelst *et al.* (2005) showed that significant efficiency in terms of shelf stacking time could be gained once the impact of most important drivers is well understood.

This paper builds upon the analysis of Van Zelst *et al*. (2005) by extending and deepening their analysis: Van Zelst *et al.* (2005) looked at the total stacking time per consumer unit, while we extend this approach towards the individual sub-activities constituting the total stacking time. On top of this, analytical expressions for the potential gains are derived, while the paper of Van Zelst *et al.* (2005) was strictly empirical. In this study, we specifically focus on the shelf–stacking process in retail stores and study the key factors that drive the execution time of this store operation. Shelf-stacking represents the daily process of manually refilling the shelves in the store with products from new deliveries. As most manual activities, such process is often time consuming and costly. Furthermore, unless clear and reliable work standards are implemented, such activities may well suffer from a lot of variation, which possibly will negatively affect the overall store performance.

We conduct an empirical analysis in which, by means of a traditional motion and time study (Barnes, 1968), we examine the entire shelf-stacking process at the level of individual subtasks and propose a regression-based methodology for predicting order-stacking times in retail stores. Similar time-study approaches are sometimes reported in the warehouse operations research for estimating order-picking times. Gray (1992) uses basic multiple regression to derive estimations of the necessary time to pick all items from a pick list for a customer order, and applies it for establishing labor productivity standards. Gray *et al.* (1992) consider the general problem of warehouse design and operation, and propose a model in which order-picking time includes three components: walking, stopping and grabbing. Varila *et al.* (2004) uses order-picking in a warehouse as a case activity to illustrate, using regression analysis, that a time-based accounting system is often suitable in tracing the cost behavior of an activity, especially when this is directly proportional to time.

The main contributions of this paper are threefold:
First, we gain a deeper insight into the characteristics of the shelf-stacking process itself by analyzing the impact of different logistical drivers (number of case packs and consumer units) on the shelf-stacking time. While warehouse handling operations received considerable attention in the literature (Rouwenhorst *et al.*, 2000, Tompkins *et al.*, 2003), there is still much opportunity for research in the field of store handling operations. This study looks into the specifics of the shelf-stacking process and analyzes, using empirical data, the effects of key variables on the shelf-stacking times. Compared to previous work, the sub-activities for the complete process are also analyzed in detail.

Secondly, we investigate whether it is possible to derive a reliable estimation of the shelf-stacking time beforehand, using only a set of key time-drivers. Using multiple regressions, a simple prediction model is developed, which allows estimating the shelf-stacking time to a large extent only on the basis of the number of case packs per order line and the number of consumer units. Real-life data was used to test the model and assure it has face validity. Overall, this study suggests a simple, inexpensive and adaptable tool for quantifying the amount of workload the regular shelf-stacking operation requires. Moreover, for each sub-activity available in the data, the drivers are quantitatively analyzed.

Thirdly, closed-form analytical expressions for the expected gains are developed and analyzed in detail. The expressions are important as it gives a general idea of the potential gains that can be achieved in the stacking process. Moreover, these analytical expressions can be used to augment currently available inventory models with a handling component, which is an interesting path from a future research point of view.

The remainder of the paper is organized as follows: In Section 2, we describe the process of restacking the retail shelves and derive a conceptual model for estimating the time required to fulfill this common store activity based on a set of potential variables. Section 3 introduces the methodology we used to test the proposed model and describes the dataset supporting our analyses. Section 4 presents the results of our study, while the last sections of the paper are devoted to discussions and conclusions.

2. Conceptual model and hypotheses development

This study focuses on improving our understanding about the handling activities in the retail stores, more specifically the process of refilling of shelves. Generally, each store undertakes the following process of replenishing the products on the shelves: upon the arrival of a new shipment, the truck is unloaded; next, the store clerks move the deliveries into the store and then restock the shelves with the newly arrived products. This process usually consumes a significant amount of store resources (such as paid workforce), which makes this activity rather costly. Therefore, it is beneficial that store management identifies the key drivers of handling workload in order to better control the related costs.

In the present study, we concentrate on one specific aspect of the store handling process: the shelf stacking, and study the main factors that affect its efficiency. The shelf stacking process starts after the incoming products are moved into the store and are taken to the shelves' areas (usually by rolling containers). Next, for each article (or Stock Keeping Unit (SKU)), the store clerks unpack the case packs and stock the consumer units on the shelves at the assigned shelf location (as indicated in the planogram, which is a diagram of fixtures and products that illustrates where and how every SKU should be displayed on a store shelf, in order to increase customer purchases (Levy and Weitz, 2001)). The shelf maintenance is an important sub-activity in this process: in preparing the shelves, for some products, the store clerks need to check the 'best before' date of the products on the shelf and remove old inventory, if necessary, before one can stack new items on the shelves. Also, to promote First-In-First-Out (FIFO) retrievals from the shelves, and to improve the display, consumer units are sometimes rearranged on the shelves placing the oldest inventory in front. For each SKU, the shelf stacking process ends with disposing the empty case packs.

Although apparently simple, the shelf stacking process at retail stores is manually executed and thus may suffer from a lot of variation. If time drives costs, then it becomes valuable to understand what drives time. In this study, we are aiming at developing a model for the estimation of the total time required for a worker to refill the shelves with the deliveries from a shipment. We are particularly interested in estimating the *Total Stacking Time per order line* (*TST*) (i.e. for each individual SKU), based on a reduced set of underlying factors. A similar approach used in motion and time studies (Barnes, 1968) is adopted here to better examine the causes and effects of time variation, by examining the total stacking process at the levels of individual subtasks. Breaking down the entire operation into small components allows, on the one hand, assessing the contribution of each individual sub-activity to the *TST*, and on the other hand, gives a better indication of the potential variables affecting the *TST*.

Therefore, we have divided the shelf-stacking activity, as previously described, into seven subtasks: *grabbing/opening* a case pack (G), *searching* for the assigned location (S), *walking* to the assigned location (W), *preparing* the shelf for stacking the new items (P), *filling new inventory* on the shelves (Fn), *filling* the *old inventory* back on the shelves (Fo) and *disposing the waste* package (D). The difference between filling old versus new inventory is important as depending upon the inventory level just before filling, the activity filling old inventory will become important for higher levels. The total time for stacking an order line on the shelves (*TST*) has thus been divided into seven time components, and for each component, the key variables that could logically influence the execution time of each subtask were identified. It is expected that the time needed to stack new inventory on the shelves depends on the number of units being handled, while grabbing and unpacking a case pack, traveling within the shelf aisle to and from the right location, or disposing the wasted case packs depend on the number of case packs being handled per order line. Lastly, searching for the right shelf location, preparing the shelf or restacking old inventory if necessary, are normally executed only once, for each SKU, independent of the number of case packs or consumer units. The set of potential drivers of time variation identified for each individual sub-activity are summarized in Table 2.1.

Table 2.1. Potential drivers of time variation, for each sub-activity

	Sub-activity	Order line information		Product information
		Number of *CP*	**Number of *CU***	**Product category**
1	Grabbing/Opening (G)	x	x	x
2	Searching (S)	-	-	-
3	Walking (W)	x	-	x
4	Preparing (P)	-	-	x
5	Filling New Inventory (Fn)	x	x	x
6	Filling Old Inventory (Fo)	-	-	-
7	Disposing waste (D)	x	-	-

In reality, there could be many other potential factors affecting the duration of shelf-stacking time (such as SKU volume, weight or type of packaging, the distance traveled within the aisle, the old inventory position just before new replenishment, the labor, the environment, etc.). In our subsequent analysis, we concentrate only on order line-related and product-related characteristics, as the key drivers of the time variation of the shelf staking process: the *number of case packs* (*CP*) and the *number of consumer units* (*CU*) per order line, while also taking into account possible variation between different *product categories*. While *CP* and *CU* are rather straightforward factors, the product subgroup variable captures any time variation that could be attributed to differences in product-related characteristics not measured specifically in this study (such as total weight of volume of products being handled, or the type of packaging). In general, the order line information refers to the number of items (case packs or consumer units) being handled, while the product information approximates the difficulty in handling products from different categories. These variables are selected as potential predictors in our subsequent analyses.

The dependent variables are the individual times per sub-activities (T^s, with $s \in \{G, S, W, P, Fn, Fo, D\} = A$) and the Total Stacking Time (*TST*), all expressed in seconds. The explanatory variables are hypothesized to have the following influence on the execution time of each sub-activity:

Hypothesis 1: The number of case packs (*CP*) has a positive effect on the individual times T^G, T^W, T^{Fn}, T^D and *TST*.

Hypothesis 2: The number of consumer units to be stacked (*CU*) has a positive effect on sub-activities' execution times T^G, T^{Fn} and *TST;*

We expect that *CP* and *CU* have no significant effect on T^S, T^P, T^{Fo}. Under these hypotheses, the *Search*, *Prepare* and *Fill Old* sub-activities could be regarded as fixed activities, while only the remaining activities are variable, depending on the set of hypothesized factors.

3. Study design and data description

Data collection

The store operation under study is the replenishment of shelves in the retail stores with new items. Two grocery retail chains (denoted here by A and B) agreed to participate in this study. Empirical data on the stacking process in the two retail companies were collected using a motion and time study approach (Barnes, 1968). Data from chain A are used to test the hypotheses, and data from chain B are used to validate the results. In four stores, (two for each supermarket chain) employees who were familiar with the operation, were videotaped during the shelf stacking process. The product subgroups were selected such that items:

- contain both fast- and slow movers;
- contain different case pack sizes;

- contain SKU's for which sufficient shelf space is available to accommodate more than one case pack in a delivery (see also Broekmeulen *et al.*, 2004).
- All selected product groups should contain items that are comparable in terms of the handling process and productivity. For this reason, we did not consider product groups such as soft drinks, beers as well as dairy products.

Finally, we note that the data collection period did not include days with peak or dropping demand, and the stores were consistent in their operations.

The stacking of items on the shelves is observed for each delivery (i.e. an order line) in the store. Each order line consists of taking a case pack from a rolling container, unpacking the case pack and placing the consumer units on the shelf at the assigned location. The entire stacking process was divided into smaller sub-activities, such that the elements are as short as possible and can be accurately timed and that constant elements can be separated from the variable ones (Barnes, 1968). The shelf stacking process was broken down into the following subtasks:

- grab and unpack the case pack;
- search for the assigned location on the shelf;
- walk to the shelf;
- prepare the location on shelf for stacking and check the shelf life of the inventory on the shelf;
- fill the new inventory on the shelf;
- fill the old inventory back on the shelf;
- dispose the waste;

Appendix 1 gives a complete description of each sub-activity. After the recording process, the execution time of each individual sub-activity and the Total Stacking Time per order (*TST*) line was registered using a computerized time registration tool, and results were recorded into a database. Additional information necessary to identify the stacking process for each order line was added as well, such as the SKU type, the number of case packs and case pack size per order line or the product category each SKU belong to.

Data description

The final dataset contains 1048 observations, for chain A, across nine product categories, and 563 observations, for chain B, across five different product categories (see Tables 2.1 and 2.3 from Appendix 2). The first set of data is selected to develop a predictive model of the total time required by a store clerk to stack an order line onto the shelves. The second dataset is used to test and validate the results.

Table 2.1 (Appendix 2) contains descriptive statistics of the variables used in this study, for the first dataset. The average total time to stack an order line into the shelves is 57.31 seconds, ranging from a low 10 seconds per order line (personal care category) to a high 334 seconds (coffee), with a standard deviation of 36.6 seconds. This reveals the degree of variation that exists in the *TST* between different order lines and this study aims at gaining a better understanding of the factors underpinning this variation. We further note that some degree of variation exists also between the *TST* corresponding to different product categories. The average *TST* line varies between 35.47 seconds (products of personal care) and 80.86 seconds (coffee milk). With reference to the explanatory variables of this study, we note that the average number of case packs per order line varies between 1 *CP* (all categories) to 9 *CP* (coffee), with an average of 1.3 *CP* and a standard deviation of 0.7 *CP*. The average number of consumer units per order line exhibits quite some variation, ranging from 3 *CU* (personal care) to 135 *CU* (coffee), with an average of 16.78 *CU* per order line.

Based on this empirical data, we also derive the distribution of the Total Stacking Time and the relative contribution of each individual sub-activity to the *TST*, as illustrated in Figure 3.1. We note that the most time consuming sub-activity in the shelf-stacking process is the Stacking of new inventory (Fn) (about 48% of the *TST*), followed by the Grabbing and unpacking the case packs (G) (about 20% of *TST*) and Disposing the waste (D) (about 13% of *TST*), respectively. Together, they account for almost 81% of the *TST*. Table 2.2 (Appendix 2) provides descriptive statistics of the dependent variables used in this study. The corresponding average times for execution of the three most time consuming sub-activities are 27.32 seconds (Stacking new inventory), 11.65 seconds (Grabbing / opening a case pack) and 7.28 seconds (Disposing waste) respectively.

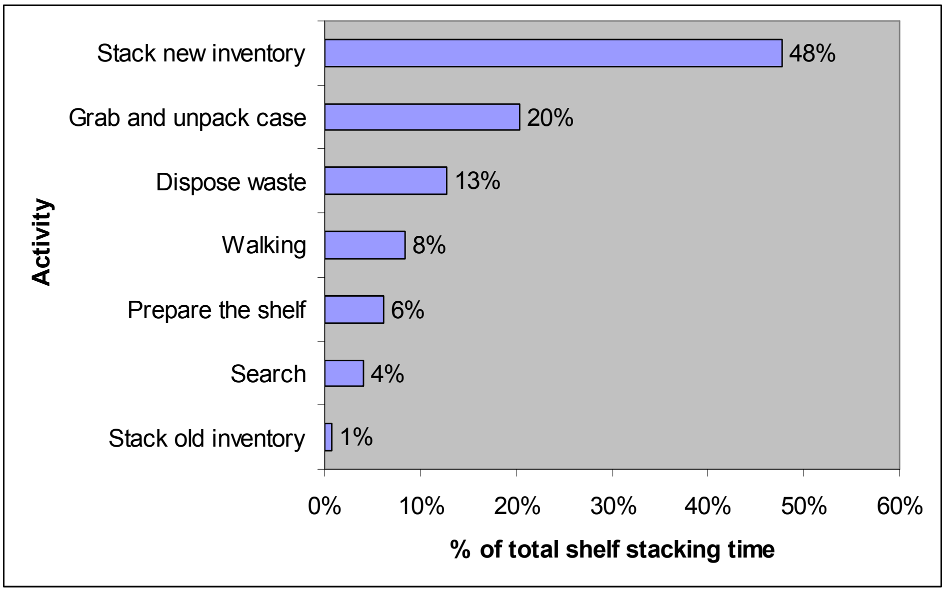

Figure 3.1 Distribution of the Total Stacking Time (Chain A)

4. Analysis and Results
4.1 Model Testing

In order to test our hypotheses and analyze the duration of each individual sub-activity and of the entire shelf stacking process, we performed several regression analyses with T^s ($s \in \{G, S, W, P, Fn, Fo, D\}$) and *TST* as dependent variables. In this study, we adopt two different strategies for estimating the Total Stacking Time per order line (*TST*), which we shall refer to as *sequential regression* and *overall regression*, respectively. Both approaches allow one to predict the *TST* as a function of the identified drivers of time variation using multiple linear regressions. The general models for each case are introduced next.

Sequential regression

$$TST = \sum_{s \in A} T^s , \tag{1}$$

where the duration of each individual sub-activities T^s per order line is estimated using the following general linear regression model:

$$T^s = b_0^s + b_1^s CP + b_2^s CU + \sum_{pc=1}^{PC-1} \alpha_{pc}^s \overline{D}_{pc} + \varepsilon , \tag{2}$$

for every sub-activity $s \in A$ and where PC represents the number of different product categories considered in the analysis. One set of dummy variables is used to account for differences between product categories $\{\overline{D}_{pc}\}$.To avoid perfect multicolliniarity, which will make the OLS parameter estimation impractical, one category from the group of product categories must stand as a reference for others (Gujarati, 1995).

Overall regression

$$TST = c_0 + c_1 CP + c_2 CU + \sum_{pc=1}^{PC-1} \gamma_{pc} \overline{D}_{pc} + \varepsilon , \tag{3}$$

Both models, (1)-(2) and (3), give an estimation of the expected *TST* (in seconds) required by a store clerk to handle the workload of restacking an order line onto the shelves. *Sequential regression* requires the *TST* be estimated in two steps: first, an estimation of individual subactivites' times per order line is necessary, which then add up naturally into the Total Stacking Time according to (2). The *overall regression* on the other hand, allows one to predict the *TST* directly on the basis of the key drivers identified. If (2) gives an accurate estimation of the time for each individual subtask, then (3) follows as a consequence of models (1) and (2), with the following relationships between coefficients of models (1)-(3):

$$c_0 = \sum_{s \in A} b_0^s \, , c_1 = \sum_{s \in A} b_1^s \, , c_2 = \sum_{s \in A} b_2^s \, , \, \gamma_{pc} = \sum_{s \in A} \alpha_{pc}^s \text{ (for } pc = 1, 2, ..., PC). \tag{4}$$

Therefore, under the assumption of accurate estimation of individual times, we expect both models will perform equally well.

Starting from the *sequential regression* model formulation introduced by equation (2), we derived three predictive models to test the effect of each explanatory variables used in this study. We first estimate each model for the first dataset (chain A) and then validate the results on the second dataset (chain B). The tested models for each individual sub-activity are specified next. Similar models are used for the analysis of the total stacking time, too.

Model 1: $T_i^s = b_0^s + \sum_{pc=1}^{PC-1} \alpha_{pc}^s \overline{D}_{pci} + \varepsilon_i$,

Model 2: $T_i^s = b_0^s + b_1^s CP_i + b_2^s CU_i + \sum_{pc=1}^{PC-1} \alpha_{pc}^s \overline{D}_{pci} + \varepsilon_i$,

Model 3: $T_i^s = b_0^s + b_1^s CP_i + b_2^s CU_i + \varepsilon_i$,

where ε_i is the error term for each order line $i = 1 : N$.

Model 1 is an ANOVA model with only the product category identifier as an explanatory variable, which is modeled here by the group of dummy variables $\overline{\{D_{pc}\}}_{pc=1:PC-1}$. To avoid perfect multicollinearity, one product category is used as a reference, as detailed in our subsequent analysis. Therefore, this model estimates differences in execution time across products categories and is used as a reference in our analysis. Model 2 includes the main effects of the number of case packs (*CP*) and the number of consumer units (*CU*) per orderline, respectively. Thus, this model tests the effect of the explanatory variables from our Hypotheses, while controlling for differences across product categories. Model 3 is a simple regression model with only *CP* and *CU* as explanatory variables. It is derived from Models 2 by removing the product subgroup-effect. Thus, Models 2 and 3 by comparison show if the product grouping has a significant effect on the execution times.

4.1.1 Sequential regression

For the derivation of the *TST*, we carried out a two-phase sequential analysis. First, for each individual sub-activity, we tested the regression models 1 to 3 and derived estimates of the execution times T^s for each sub-activity. These estimates are then used to approximate the *TST*, as indicated by equation (1). Separate analyses for each individual sub-activity correspond to our motivation of identifying which sub-activities are mostly affected by the selected order line- and product-related factors. The final derivation of the *TST* is in line with our purpose of deriving a predictive model for estimating the total time necessary to stack the products from an order line into the shelves.

Estimating the time for individual sub-activities

Models 1 and 2 are analyzed using hierarchical regression. The group of dummy variables representing the merchandising category was considered as a control variable, and it was introduced in the first step of hierarchical regressions. The reference category was chosen to be the one with the largest number of samples in the dataset. The first empirical dataset contains nine product subgroups and the largest category in this dataset is *Personal care* (see Appendix 2, Table 2.1). In the second step of hierarchical regression, we added together the main effects *CP* and *CU*.

The results of ordinary least squares estimation for the first data set are presented in Table 4.1. Relevant collinearity diagnosis (such as coefficient of correlation, variance inflation factors) indicated no significant problems with respect to multicollineariry. For each of the three models, Table 4.1. gives the standardized coefficient estimates, for each individual sub-activity together with some measures of goodness of fit. Overall, results for Model 1 indicate that the product category variable explain only a small proportion of the total variance in the execution times of corresponding sub-activity. The three largest adjusted R^2, obtained for *Fill New*, *Prepare* and *Dispose* in this sequence, varies from 10% to almost 17%. We also note that although some product categories dummies are not significant predictors, the group of dummies is overall significant (as confirmed by the overall F-statistics), and this holds true for every individual sub-activity.

Results from the 2nd regression step indicate that Model 2 explains a significantly higher proportion of the variance in sub-activities' times. The adjusted R^2 ranges from .008 (for *Search* sub-activity) to as high as .679 (for *Fill New* sub-activity). The three largest proportion of variance in the dependent variable accounted for by the explanatory variables of Model 2 belong to *Fill New* (R^2adj equals 67.9%), *Grab and unpack* (R^2adj of 41.6%) and *Dispose* (R^2adj of 31.6%) sub-activity, respectively. Recall from Figure 3.1 that these are also the three most influential sub-activities with respect to their relative contribution to the Total Stacking Time. The overall F-statistics indicate a significant joint contribution of the variables in predicting the execution times for all sub-activities (at $p \leq .05$). However, we notice that the explanatory variables *CP* and *CU* do not contribute significantly in explaining the time for *searching*, and have only a marginal contribution in explaining the time for *preparing* the shelves, *filling old inventory* and *walking*, respectively (R^2change of 0.011 and 0.057).

Further, we notice that when the subgroup effect is removed from the analysis (Model 3), the adjusted R^2 for the Fn, G and D drops marginally from the previous model to 63.3%, 39.8% and 25.4%, respectively. The F-statistics show that the joint contribution of *CP* and *CU* is statistically significant for Fn, G and D and their standardized coefficients are both positive, thus showing support for our hypotheses for these sub-activities. Note that these results are also rather consistent between models 2 and 3. Comparing models 2 and 3 we also find no support for S being affected by *CP* or *CU*. Although the results show a statistically significant effect of *CP* or *CU* for W, P and Fo, by inspecting the adjusted R^2 we conclude that the impact of these variables on the execution times of the aforementioned sub-activities is weak. This result is consistent with our prediction that *CP* and *CU* do not affect the *Search*, *Prepare* and *Fill old* sub-activities.

In summary, we conclude that the results provide evidence that the execution time for *Fill new*, *Grab/unpack* and *Dispose* are mostly explained by *CP* and *CU*, while we found little evidence that these variables affect substantially the other sub-activities.

Table 4.1. Regression results for each individual sub-activity (standardized coefficients)

Step	Variables	Model 1						
		G	S	W	P	Fn	Fo	D
1	Baby food	.033	-.039	-.052	**.075***	**.065***	.028	**.065***
	Chocolate	**.205*****	**-.085***	**.111****	**.207*****	**.238*****	**.083***	**.284*****
	Coffee	**.271*****	-.065	-.027	**.320*****	**.368*****	**.103****	.051
	Coffee milk	**.106****	-.043	**.062***	**.165*****	**.283*****	**.083***	**.181*****
	Candy	**.076***	.007	**.201*****	.042	**.247*****	.004	**.165*****
	Sugar	.045	-.040	**-.060***	**.079****	**.165*****	-.002	.040
	Canned meat	**.080***	**-.091****	**.138*****	**.090****	**.231*****	-.003	**.271*****
	Canned fruits	**.071***	.011	-.032	.040	**.161*****	-.003	**.080****
	R^2	.072	.017	.071	.109	.172	.018	.122
	R^2adj	.065	.010	.064	.103	.166	.010	.115
	Mean SS Err.	111.167	13.520	13.855	49.434	405.135	12.581	39.847
	Overall F	10.117***	2.284*	9.933***	15.949***	27.048***	2.329*	17.998***
	df	8, 1039	8, 1039	8, 1039	8, 1039	8, 1039	8, 1039	8, 1039

Statistical significance at *p≤.05, also **p≤.01, *** p≤.001; Reference category: Personal care (N = 285)

Table 4.1. (continued) Regression results for each individual subactivity (standardized coefficients)

Step	Variables	Model 2						
		G	S	W	P	Fn	Fo	D
2	Baby food	.031	-.039	-.052	**.077***	**.057****	.031	**.064***
	Chocolate	.037	**-.084***	.051	**.262*****	**-.086*****	**.146*****	**.195*****
	Coffee	**.124*****	-.062	**-.084***	**.338*****	**.147*****	**.128*****	-.044
	Coffee milk	.001	-.041	.023	**.188*****	**.104*****	**.111****	**.119*****
	Candy	-.005	.006	**.175*****	**.092***	.043	.057	**.135*****
	Sugar	-.032	-.037	**-.092****	**.072***	**.083*****	-.005	-.019
	Canned meat	**-.055***	**-.087*****	**.085****	**.077****	**.052****	.008	**.177*****
	Canned fruits	-.030	.014	**-.072***	.262	.030	.004	.008
	CP	**.422*****	-.023	**.187*****	**.338*****	**.184*****	**.144****	**.401*****
	CU	**.253*****	.003	.082	**.188*****	**.647*****	**-.171****	**.090***
	R^2	.422	.018	.128	.121	.682	.028	.322
	R^2adj	.416	.008	.120	.112	.679	.019	.316
	R^2 change	.350	.000	.057	.011	.510	.011	.200
	F change	313.815***	.214	33.932***	6.710**	832.809***	5.678**	153.239***
	Mean SS Err.	69.387	13.540	13.029	48.896	155.751	12.468	30.816
	Overall F	75.730***	1.867*	15.237***	14.241**	222.847***	3.016***	49.266***
	df	10, 1037	10, 1037	10, 1037	10, 1037	10, 1037	10, 1037	10, 1037

Statistical significance at *p≤.05, also **p≤.01, *** p≤.001; Reference category: Personal care (N = 285)

Table 4.1. (continued) Regression results for each individual subactivity (standardized coefficients)

Variables	Model 3						
	G	S	W	P	Fn	Fo	D
CP	**.409*****	-.025	**.085***	**.120****	**.241*****	**.090***	**.313*****
CU	**.271*****	-.030	**.183*****	-.007	**.606*****	-.072	**.232*****
R^2	.399	.003	.063	.013	.634	.004	.256
R^2adj	.398	.001	.061	.011	.633	.002	.254
Mean SS Err.	71.616	13.643	13.894	54.455	178.035	12.681	33.568
Overall F	346.724***	1.361	35.157***	7.007***	905.874***	2.118	179.629***
df	2, 1045	2, 1045	2, 1045	2, 1045	2, 1045	2, 1045	2, 1045

Statistical significance at *p≤.05, also **p≤.01, *** p≤.001

Estimating TST

We derive the *TST* simply as the sum of the estimated execution times for each individual subactivity, derived under models 2 and 3. Thus the estimated *TST* is derived as follows:

$$TST = \hat{T}^G + \hat{T}^S + \hat{T}^W + \hat{T}^P + \hat{T}^{Fn} + \hat{T}^{Fo} + \hat{T}^D ,$$

where $\hat{T}^s$, $s \in A$ stands for the estimated execution time of the corresponding sub-activity, as given by Models 2 or 3. To estimate the accuracy of this prediction we compare the estimated *TST* with the actual *TST* (obtained from empirical data) and the results are included in Table 3.1 from Appendix 3. Both variables have the same mean (57.31 seconds) as confirmed by a paired-samples t-test. The correlation coefficient between the predicted and the measured *TST* is .819 (Model 2) and .798 (Model 3) and thus 67% (respectively 63.7%) of the variance in the measured *TST* per order line is explained by the sum of time estimates for individual sub-activities. Thus, results show a slightly better performance of Model 2 as compared with Model 3 but the increase in adjusted R^2 is marginal. Therefore, given the simplicity of Model 3, we recommend choosing it for forecasting purposes.

4.1.2 Overall regression

Similar regression models as described by Models 1-3 are analyzed for the dependent variable *TST*. The prediction models 1 and 2 are also analyzed using hierarchical regression in two steps and then compared with the prediction model 3. The results of ordinary least squares estimation for the first dataset (1048 observations) are presented in Table 4.3. Collinearity tests (correlation coefficients, variance inflation factors) indicated no significant problems with respect to multicollinearity for the estimated models. In addition, upon preliminary inspection of the results, no significant outliers or influential points were detected, and thus the results included in Table 4.3 reflect the entire dataset. An alternative model formulation where interactions between the explanatory variable and the product category were included did not improve the model specification and were not significant. (Aiken and West, 1991).

Table 4.3. Regression analyses results for *TST* (Chain A)

		Model 1			**Model 2**			**Model 3**		
Step	Variables	Unstd. Coeff.	Std. Err.	Std. Coeff.	Unstd. Coeff	Std. Err.	Std. Coeff.	Unstd. Coeff	Std. Err.	Std. Coeff
1	(Constant)	35.474***	1.973		3.902 *	1.785		10.240***	1.447	
	Baby food	14.978 *	6.298	.069	13.898***	3.995	.064			
	Chocolate	30.872***	3.239	.310	5.896 *	2.375	.059			
	Coffee	38.063***	3.270	.377	18.471***	2.166	.183			
	Coffee milk	45.383***	4.868	.279	21.527***	3.227	.132			
	Candy	19.968***	2.892	.232	7.932***	2.007	.092			
	Sugar	35.360***	8.093	.126	10.517 *	5.176	.037			
	Canned meat	41.697***	5.243	.236	12.016***	3.418	.068			
	Canned fruits	29.401***	6.209	.138	2.922	3.998	.014			
2	CP				19.614***	1.471	.375	19.052***	1.396	.364
	CU				1.180***	.081	.442	1.327***	.071	.496
	R^2	.178			.670			.637		
	R^2adj	.172			.667			.636		
	R^2 change	.178			.492			.637		
	F change	28.130***			773.434***			916.178***		
	Mean SS Err.	1108.989			445.936			487.185		
	Overall F	28.130***			210.651***			916.178***		
	df	8, 1039			10, 1037			2, 1045		

Statistical significance at *p≤.05, **p≤.01, *** p≤.001; Reference category: Personal care (N=285)

Results for Model 1 indicate that the product category variable alone explains in proportion of 17.2% the variance in *TST*, while Model 2 yields a significantly larger adjusted R^2 (66.7%), with a significant 49.2% of the total variability in *TST* accounted for by the group of variables *CP* and *CU*. Model 3 shows that the explanatory variables *CP* and *CU* together have a significantly high joint contribution in predicting the *TST*, accounting for 63.7% of the variability in the *TST*. The *F*-statistics for all three models are statistically significant (p≤.001). Thus, results provide evidence that the *TST* is systematically explained in large measure by our model. Furthermore, comparing models 2 and 3, we note that when the subgroup-effect is excluded from the model, the adjusted R^2 decreases only marginally and in both cases the hypotheses of our study are supported.

The coefficient estimates presented in Table 4.3. show that *TST* is positively correlated with *CP* and *CU*, thus providing support for our hypotheses at p≤.001. They also are fairly consistent between models 2 and 3. The standardized values of the coefficients indicate that most of the explanatory power comes from the variables *CP* and *CU*, with a higher influence of *CU*. While all significant in Model 1, the coefficients of dummy variables for product category remain significant (at p≤.05), with one exception (Canned fruits), in Model 2. Compared with *CP* and *CU* however, they indicate a relatively small explanatory power. Also, recall that in our modeling we used *Personal care* as a reference category for the group of dummy variables (with the largest number of observations), and therefore the positive coefficients for the dummy variables confirm our expectations from previous descriptive statistics (see Table 2.1, Appendix 2): according to this dataset, the personal care category is the fastest to handle on order line basis.

Based on these results, we conclude that models 2 and 3 explain already a large portion of the Total Stacking Time and the variables *CP* and *CU* are important predictors of *TST*. Due to the simplicity of the model and its good accuracy, we recommend using Model 3 for forecasting *TST*.

4.2 Validation of results

To validate the results from the previous section, we use the empirical data from chain B and replicate the analysis conducted for chain A. We do this in order to verify the reliability of the previously obtained results and the accuracy of the predictive models (Wang, 1994).

Summary statistics for the variables in this study using the second dataset are included in Tables 2.3, 2.4 from Appendix 2. The average *TST* across all five product categories is 49.29 seconds with a standard deviation of 27.06. The smallest average *TST* is recorded for the products from the wine subgroup (39.69 seconds), while the most time consuming products in this set for handling are those from category cookies (mean *TST* equals 60.62 seconds). The *TST* shows significant variation between order lines, ranging from a minimum of 6 seconds (wine) to a maximum of 212 seconds per order line (canned vegetables). The variable *CP* ranges from 1 *CP* (all categories) to 8 *CP* (cookies) with an average of 1.22 *CP* across all categories, and a standard deviation of 0.6 *CP*. The variable *CU* has an overall mean of 15.5 consumer units (standard deviation = 8.86), ranging from 6 to 80 *CU* per order line.

To assure the general applicability of the approach proposed in this study, we are interested in how consistently the previous results replicate for the second data set. Regarding individual sub-activities, regression results (see Table 4.1 from Appendix 4) for models 1 to 3 confirm to a high extent previous findings: the *CP*, *CU* have a positive effect on execution times of sub-activities Fn, G and D and are the most important predictors of time variation for these sub-activities (adjusted R2 is 52.6%, 48.3% and 20.4%, respectively for Model 2, and 50.9%, 47.6% and 19.6%, respectively for Model 3). When sequential regression is then used to estimate the *TST* (see Table 4.2 from Appendix 4), we found that the correlation coefficient between the predicted and the measured *TST* is 0.724 (R2 = 52.4%) using Model 2 and 0.684 (R2 = 46.7%) for Model 3. The high values of these correlation coefficients indicate that the *TST* for chain B is also explained to a large degree by the chosen models. Furthermore, comparing results for models 2 and 3, we are again in favor of the simplest model 3 to be used for deriving good estimations of the *TST*.

For forecasting purposes, the overall regression for estimating *TST* provides a simple and less time-consuming procedure. Therefore, we tested the reliability of the results on the second data set as well. We found consistent support for our hypotheses regarding *TST* (see Table 4.3 from Appendix 4). While the group of dummy variables related to product category have a significant, but small contribution in predicting *TST* (adjusted R2 about 10%), the most explanatory power comes again from the group of variables *CP* and *CU*, which affect significantly and positively the *TST*. Compared to Model 2 (R2adj = 51.9%), *CP* and *CU* alone explain 46.5% of the variance in *TST*, thus indicating only a marginal decrease in adjusted R2. Their coefficients are both statistically significant at $p \leq .001$ and consistent between the two models. Moreover, note that they have also comparable sizes with coefficients' estimates for *CP* and *CU* derived for the first dataset (see Table 4.3). We can, therefore, be confident that the effects of both *CP*, and *CU* are needed to model the *TST* to a large extent and that Model 3 represents a simple and reliable alternative for predicting *TST*.

We performed a final verification of the results, in which we used the data collected for chain B and the coefficients estimates for Model 3 derived for chain A (see again Table 4.3) to compute predicted values of *TST* for each order line. We restate here for reference the model used for prediction:

$$TST = a_0 + a_1 CP + a_2 CU \,, \quad \text{where } a_0 = 10.240,\ a_1 = 19.052 \text{ and } a_2 = 1.327.$$

Table 4.4 presents the results obtained by comparing the predicted and the measured values of the *TST*. The high values of the correlation coefficient between the measured and predicted *TST* (R = 0.683, R^2 = 46.6%) provide additional evidence that the *TST* for chain B is also explained to a large degree by our model.

Table 4.4. Validation results for *TST* using coefficient estimates of Model 3 from the first dataset and empirical values for *CP* and *CU* for chain B

	Regression results	
	Intercept (Std. Err)	Slope (Std. Err.)
	1.104 (2.331)	.891***(.040)
Correlation coefficient	.683	
R²	.466	
R²adj	.465	
Mean SS Err.	391.423	
Overall F	490.139***	
df	1, 561	

Overall, the results confirm that Model 3 represents a simple and reliable method for predicting the *TST*. The stacking-times for each order line can be estimated inexpensively in this way, and ultimately be used for management decisions. For example, one can assess the amount of work necessary during a day to execute the restocking of the shelves, or can assess the individual labor performance of store employees.

5. Analytical expressions for efficiency gains

We showed that the shelf-stacking time per order line (i.e. for each SKU) can be estimated based on the number of case packs (*CP*) and consumer units (*CU*) per order line, according to the following model:

$$TST = a_0 + a_1 CP + a_2 CU \,, \tag{5}$$

where the parameters a_0 , a_1 and a_2 have been estimated and are shown to be positive (a_0 = 10.240, a_1= 19.052 and a_2 = 1.327). Considering that $CU = CP \cdot Q$, with Q the case pack size, and rewriting (5) we obtain:

$$TST(CP, Q) = a_0 + a_1 CP + a_2 CP \cdot Q \tag{6}$$

The *TST per SKU* is thus linearly dependent on *CP* and *CU*, but not directly proportional with *CP* and *CU* (due to the constant parameter a_0). For each order line, a fixed 'setup time' (a_0) is incurred, additionally to the positive time related to the number of units being handled. This structure of the *TST* can be exploited in order replenishment decisions, for a better time and cost management of the overall store operations. For example, the model allows quantifying the time savings obtained when it is possible to reduce the frequency of the replenishments, by ordering more products at once, rather than the same amount multiple times. Let *n* (*n* = 1, 2, 3, ...) be the number of order lines for the same SKU in a replenishment order in the subsequent analysis. Two situations can then be considered:

Order more case packs per order line

The effect of reducing the replenishment size (i.e. the number of order lines) by ordering more *CP* per order line, while keeping the same case pack size (*Q*) can be evaluated. We compare the time savings obtained if, instead of ordering *n* order lines with *CP* of size *Q* per order line, it is possible to order the entire amount at once (i.e. in one order line), by ordering *nCP*, each of size *Q*.

The total time needed for stacking *n* order lines with *CP* of size *Q* per order line can be written as:

$$TT(CP, Q) = nTST(CP, Q) = na_0 + na_1 CP + a_2 nCP \cdot Q \,. \tag{7}$$

The total time needed for stacking the same amount at once is expressed as:

$$TST^1(CP, Q) = TST(nCP, Q) = a_0 + a_1 nCP + a_2 nCP \cdot Q \,. \tag{8}$$

Then the time savings can be derived as:

$$TT - TST^1 = (n-1)a_0 > 0, \text{ for } n > 1, \tag{9}$$

which implies that we may save stacking time if we order, in each replenishment, more case packs at once, instead of ordering one case pack at a time, and this saving is due to the constant 'setup time' a_0. The efficiency gain, compared with the case of multiple replenishments is then:

$$S^1(CP,Q,n) \equiv \frac{TT - TST^1}{TT} = \frac{(n-1)a_0}{na_0 + na_1CP + a_2nCP \cdot Q} \cdot 100\% \text{ , for } n > 1. \tag{10}$$

Increase the case pack size Q

We evaluate the time savings obtained if, instead of ordering *n* order lines with *CP* of size *Q* per order line, it is possible to order the entire amount at once (i.e. in one order line), by ordering *CP* case packs, each of size *nQ*. In this case, the shelf-stacking time is derived as follows:

$$TST^2(CP,Q) = TST(CP,nQ) = a_0 + a_1CP + a_2CP \cdot nQ. \tag{11}$$

Then the time gains are now:

$$TT - TST^2 = (n-1)a_0 + (n-1)a_1CP > 0, \text{ for } n > 1, \tag{12}$$

and the percentage of time saving is then:

$$S^2(CP,Q,n) \equiv \frac{TT - TST^2}{TT} = \frac{(n-1)a_0 + (n-1)a_1CP}{na_0 + na_1CP + a_2nCP \cdot Q} \cdot 100\%, \text{ for } n > 1. \tag{13}$$

Again, in this case, reducing the frequency of the replenishments may result in time savings and efficiency gains as given by (12) and (13). Furthermore, by comparing equations (10) and (13), we notice that the time saving is always higher in the second case, when the strategy is to increase the case pack size (*Q*) instead of the number of case packs (*CP*) per order line.

The efficiency gains derived from (10) and (13) are illustrated in Figure 3.2, for two particular choices of *CP* and *Q*. Typically, case pack sizes take values of 6, 12 or 24 consumer units. In Figure 3.2, the effect of *n* on S^1 and S^2 is illustrated for one and respectively four case packs per order line, each of size six. We note that the higher the reduction in the number of order lines, the higher the savings. Reducing one order line for the same SKU (*n = 2* in Figure 3.2) as a consequence of ordering two times more case packs results in efficiency gains of 13% (if *CP* = 1) and 4% (if *CP* = 4), respectively. Alternatively, we observe higher potential gains, up to 40%, when it is possible to place orders for higher case pack sizes.

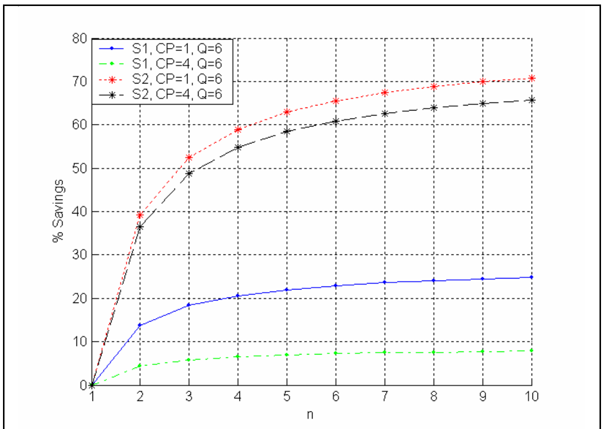

Figure 3.2 Effect of *n* on S^1 and S^2 for two particular choices of *CP* and *Q*

Note again from (10) and (13) that *n*, *CP* and *Q* have a combined effect on S^1 and S^2. Particular joint effects are illustrated in figures 3.3 and 3.4. Generally as *n* increases, S^1 and S^2 reach steady values, with a maximum around 30% (for S^1) and 80% (for S^2), respectively. However, as *CP* and *Q* increases, the efficiency gains are decreasing. Notable is the behavior of S^1 with respect to n and CP, when the savings may drop significantly as CP increases, indicating that in formula (10), the estimated time due to CP and Q, far outweighs the fixed 'setup-time' (due to a_0).

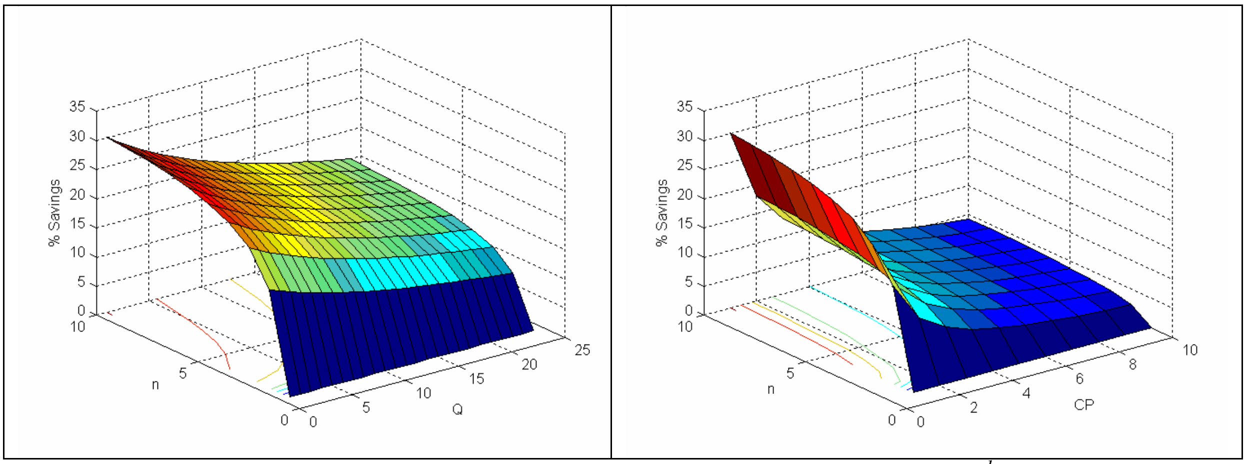

Figure 3.3 Effect of (n, Q) (left; CP=1) and (n, CP) (right; Q = 1) on S^1

Although in practice the case pack sizes are usually set by the manufacturers, it is still valuable to recognize the impact of reduced sizes on handling efficiency, perhaps especially for retailers that also carry their own private labels.

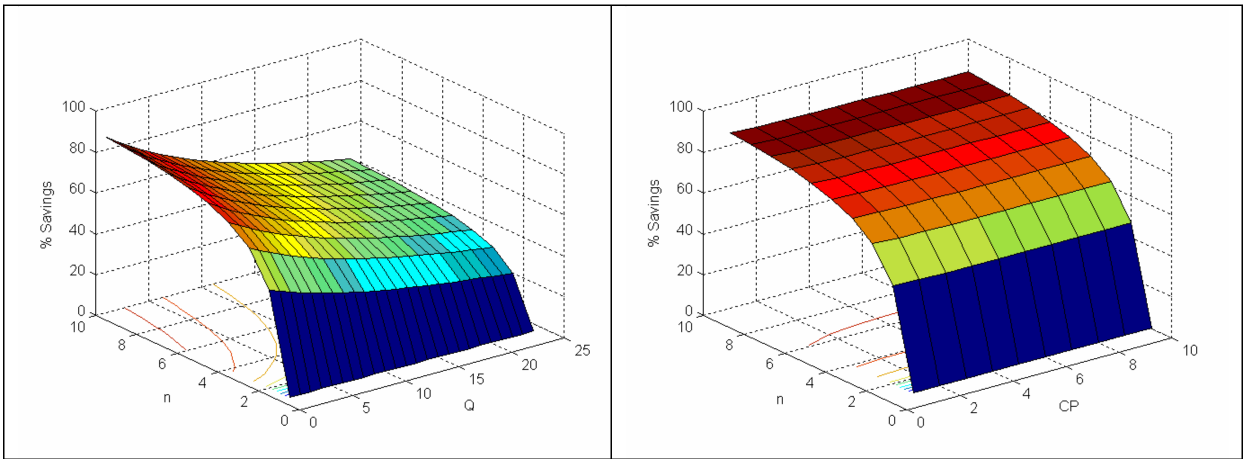

Figure 3.4 Effect of (n, Q) (left; CP = 1) and (n, CP) (right; Q = 1) on S^2

These preliminary insights into the potential efficiency gains derived from the proposed model, offer interesting opportunities for developing adapted inventory control rules that take into account the handling component. Building new inventory replenishment policies that recognize the handling efficiency, should of course consider the possible tradeoffs (such as the shelf space availability and physical constraints of the shelves, the demand pattern, or restrictions with respect to possible case pack sizes). We recognize this as an interesting area for further research.

6. Conclusions and Discussions

In this study, we focus on the shelf stacking process in retail stores and study the factors that drive the shelf stacking time. Three major contributions are recognized: first, a conceptual model based on the actual process is described; secondly, this conceptual model is tested and verified based on the empirical data collected at two retailers; thirdly, analytical expressions are derived to quantify the time gains. All analysis presented in this paper is based on both the aggregate total stacking time as well as on the individual sub-activity times available.

We proposed a methodology based on multiple regressions for deriving a prediction model for the Total Stacking Time. To gain a better understanding of the shelf stacking process and of the underlying factors affecting this process, we adopted two strategies for estimating the *TST* per order line (sequential vs. overall regression). On one hand, the two approaches may serve two different practical purposes. The sequential approach, allows one for a better insight into the details of the process of shelf-stacking, identifying those sub-activities that are likely to be mostly affected by the number of items being handled, and those for which the variation in workload is potentially affected by other factors. At the same time, detailed insight into the shelf staking process indicates which sub-activities contribute mostly to the total variation in the stacking time of a new order line. The three most relevant are: stacking new inventory, grabbing and opening of a case pack, and waste disposal, in this order.

On the other hand, the overall regression strategy offers a less time consuming procedure for predicting the *TST* per order line, since it allows one to predict the *TST* on the basis of a set of key factors using one model only. In this study, we found enough support to conclude that a simple prediction model, depending only on the number of case packs and the number of consumer units, offers already a reliable estimate of the *TST*. Results from testing and validation show that the model is stable and it explains the *TST* to a large extent. Therefore, this model offers a useful tool for forecasting the *TST* per order line.

Limitations of this study and future research

In this study, we were primarily interested in estimating the time necessary to execute the restacking of shelves with new deliveries in the stores, based on a reduced set of item-related characteristics. Therefore, while we illustrate the impact of some key drivers on time variation, we recognize there are more potential factors that may affect the execution time of a certain activity, which worth further investigation. It seems reasonable for example to assume that the type of package, the distance traveled within the aisle, the SKUs volume or weight, and the inventory level of the products on the shelves just before restocking might have an impact on the Total Stacking Time. Such analysis requires however further investigation.

References

Aitken L S and West S G (1991). *Multiple regression: testing and interpreting interactions.* Sage Publications.

Barnes R M (1968). *Motion and time study design and measurement of work*. John Wiley & Son: New York.

Broekmeulen R, van Donselaar K, Fransoo J and van Woensel T (2004). Excess shelf space in retail stores: An analytical model and empirical assessment. *BETA working paper 109*, Technische Universiteit Eindhoven, Eindhoven.

Butner K (2005). *The GMA 2005 Logistics Survey: Supply chain performance in food, grocery and consumer product*, IBM Institute for Business Value study, January 2005.

Cachon G (2001). Managing a Retailer's Shelf Space, Inventory, and Transportation. *Manufacturing & Service Operations Management* **3**: 211–229.

Corstjens M and Doyle P (1981). A model for optimizing retail shelf space allocations. *Management science* **27**: 822-833.

Chain Store Age (1963). Cifrino's Space Yield Formula: A Breakthrough for Measuring Product Profit, **39**.

Chain Store Age (1965). Shelf allocation breakthrough **41**:77-88.

DULOG (1997). Handling Costs in the Swedish Grocery Retail Industry, DULOG: Dagligvaruhandlarnas Utvecklings och Logistikgrupp, Sweden.

Drèze X, Hoch S J and Purk M E (1994). Shelf Management and Space Elasticity. *Journal of retailing* **70**: 301-326.

Gray Ann E, Karmarkar U S and Seidmann A (1992). Design and operation of an order-consolidation warehouse: Models and application**.** *European Journal of Operational Research* **58**: 14-36.

Clifford G F (1992). An integrated methodology for dynamic labor productivity standards, performance control, and system audit in warehouse operations. *Production and Inventory Management Journal* **33**: 63-67.

Gujarati D (1995). *Basic Econometrics*. McGraw-Hill Health Professions Division, pp 849.

Hoare NP and Beasley JE (2001). Placing boxes on shelves: a case study, *Journal of the Operational Research Society* **52**: 605-614.

Levy M and Weitz B A (2001). *Retailing management*, Irqwin: Chicago.

Rouwenhorst B, Reuter B, Stockrahm V, van Houtum G J, Mantel R J and Zijm W H M (2000). Warehouse design and control: Framework and literature review. *European Journal of Operational Research* **122**: 515-533.

Saghir M and Jonson G (2001). Packaging handling Evaluation Methods in the Grocery Retail Industry. *Packaging Technology and Science* **14**: 21-29.

Tompkins J A, White J A, Bozer Y A, Frazelle E H, Tanchoco J M A and Trevino J (2003). *Facilities Planning* (second edition). John Wiley & Sons: New York.

Urban T L (1998). An Inventory-Theoretic Approach to Product Assortment and Shelf-Space Allocation**.** *Journal of Retailing* **74**(1): 15-35.

Themido I, Arantes, A, Fernandes C and Guedes AP (2000). Logistics costs case study – an ABC approach. *Journal of the Operational Research Society*, **51**: 1148-1157.

Van Zelst S, van Donselaar K, van Woensel T, Broekmeulen R and Fransoo J (2005). A Model for Store Handling: Potential for Efficiency Improvement. *BETA working paper 137,* Technische Universiteit Eindhoven, Eindhoven.

Varila M, Seppänen M and Suomala P (2004). *Automatic data collection in logistics accounting: Analyzing and causes and effects of variation.* 4th Conference on New directions in Management Accounting: Innovation in Practice and Research, Brussels, Belgium.

Wang G C S (1994). What should you know about regression based forecasting. *The Journal of Business Forecasting Methods and Systems* **12**(4): 15-21.

Appendix 1

Sub-activity	**Starting/ Ending point of sub-activity**	
Grab/ open case pack (**G**)	*Start*	The filler stands in front of the rolling container and reaches for a case pack.
	End	The filler prepares to walk away from the rolling container (case pack is or is not opened).
or	*Start*	The filler has arrived at the shelf location and starts opening the case pack.
	End	The filler is ready with opening the case pack and an other sub-activity starts.
Search (**S**)	*Start*	The filler starts with checking the product and he/ she lookes for the right shelf location.
	End	The filler sees the right shelf location and prepares to approach it (walk).
Walk (**W**)	*Start*	The filler prepares to walk away from the rolling container or walks after searching the right shelf location.
	End	The filler stands still in front of the shelves.
and	*Start*	The filler prepares to walk away from the shelf location or waste disposal place, to the rolling container.
	End	The filler stands in front of the rolling container and reaches for a case pack.
Prepare the shelves/ check 'best before' date (**P**)	*Start*	The filler reaches for the old inventory on the shelves and start to check the 'best before' date (if needed).
	End	The filler is ready with preparing the shelves. This means that old inventory is straightened or is removed from the shelves.
Fill new inventory (**Fn**)	*Start*	The filler reaches for the new inventory in the case pack.
	End	The filler reaches for the old inventory or grabs the empty box or plastic.
Fill old inventory (**Fo**)	*Start*	In case old inventory was removed from the shelves, the filler starts with putting old inventory back on the shelves.
	End	The filler is ready with putting old inventory back on the shelves en grabs the empty box or plastic.
Waste disposal (**D**)	*Start*	The filler holds an empty box (or plastic) and starts to flatten it (sometimes the box is preserved for customers).
	End	The moment the filler prepares to leave the waste disposal place (a trolley or a place near the rolling container).
Extra (**E**)	Any activity not part of the first sub-activities, e.g. help a customer, customer is in the way, get or put away crate, process inventory remainder, organise labels, general cleaning, discuss with a colleague, take away waste, bring empy boxes for customers to check out area, get a new rolling container, take away misplaced products, repare a broken product, remove cord from rolling container, take a product to the kiosk, straighten seperation plate.	

Nota bene:

* Grabbing and opening the case pack are taken together, because the individual activities were difficult to seperate.

** Walking does not include walking with the rolling container from the storage area to the right aisle or walking with the rolling container between the aisles. But it does include (in exceptional cases) walking with the rolling container when the rolling container is moved to bring certain case packs to the right shelf location (e.g. heavy products).

*** It is possible that a filler performs multiple sub-activities at once, e.g. walking while opening the case pack , searching or disposing waste. When this took place, the following reasoning was used: if the walking time was significantly influenced by the attention focused on opening the case pack (or searching or waste disposal), the time for e.g. opening the case pack was measured as sub-activitity "G", and the remaining time as sub-activity "W". If the walking time was not significantly influenced by one of these sub-activities, then the total time was measured as walking time (W).

Appendix 2. Descriptive statistics of the empirical datasets

Table 2.1 Descriptive statistics for the first sample (N=1048): Chain A

Category	Number of Order Lines	Number of CP per Order Line				Number of CU per Order Line				TST per Order Line [sec]			
		Avg.	SD	Min	Max	Avg.	SD.	Min	Max	Avg.	SD	Min	Max
Baby food	31	1.13	0.34	1	2	8.90	4.03	4	16	50.45	17.47	20	88
Chocolate	168	1.36	0.72	1	4	25.26	16.58	6	80	66.35	41.22	15	294
Personal care	285	1.14	0.37	1	3	7.80	3.86	3	36	35.47	14.93	10	94
Coffee	163	1.47	1.08	1	9	18.88	18.74	6	135	73.54	48.62	20	334
Coffee milk	56	1.45	0.81	1	5	22.93	12.27	10	60	80.86	35.34	34	211
Candy	248	1.15	0.39	1	3	17.92	9.17	8	72	55.44	26.69	16	74
Sugar	18	1.83	0.99	1	4	17.33	9.43	8	40	70.83	32.25	22	151
Canned meat	47	1.77	0.96	1	5	22.55	15.20	6	72	77.17	51.23	12	245
Canned fruit	32	1.72	0.81	1	4	20.63	13.10	6	48	64.88	31.09	11	125
Aggregate statistics	**1048**	1.30	0.70	1	9	16.78	13.69	3	135	57.31	36.59	10	334

Table 2.2 Descriptive Statistics of the response variables: Chain A

	N	Mean	Std. Dev.	Std. Error Mean
TST	1048	57.31	36.59	1.13
Grab/Open	1048	11.65	10.91	0.34
Search	1048	2.31	3.70	0.11
Prepare	1048	3.54	7.42	0.23
Fill New	1048	27.32	22.04	0.68
Dispose	1048	7.28	6.71	0.21
Walking	1048	4.77	3.85	0.12
Fill Old	1048	0.44	3.57	0.11

Table 2.3 Descriptive statistics for the second sample (N = 563): Chain B

Category	Number of Order Lines	Number of CP per Order Line				Number of CU per Order Line				TST per Order Line [sec]			
		Avg.	SD	Min	Max	Avg.	SD.	Min	Max	Avg.	SD	Min	Max
Sandwich spread	56	1.25	0.47	1	3	17.11	9.71	8	48	48.73	25.02	18	148
Canned vegetables	57	1.28	0.67	1	5	14.74	7.83	8	60	52.67	31.88	17	212
Cookies	179	1.22	0.68	1	8	18.41	9.60	8	80	60.62	27.34	19	194
Candy/Chocolate	142	1.13	0.35	1	3	18.05	7.22	6	50	42.59	20.04	13	132
Wine	129	1.29	0.69	1	5	8.29	4.36	6	30	39.69	26.29	6	168
Aggregate statistics	**563**	1.22	0.60	1	8	15.50	8.86	6	80	49.29	27.06	6	212

Table 2.4 Descriptive Statistics of the response variables: Chain B

	N	Mean	Std. Dev.	Std. Error Mean
TST	563	49.29	27.06	1.14
Grab/open	563	7.49	6.92	0.29
Search	563	0.67	3.09	0.13
Prepare	563	5.87	7.51	0.32
FillNew	563	21.94	12.97	0.55
Dispose	563	4.56	4.57	0.19
Walking	563	7.26	5.78	0.24
FillOld	563	1.50	4.93	0.21

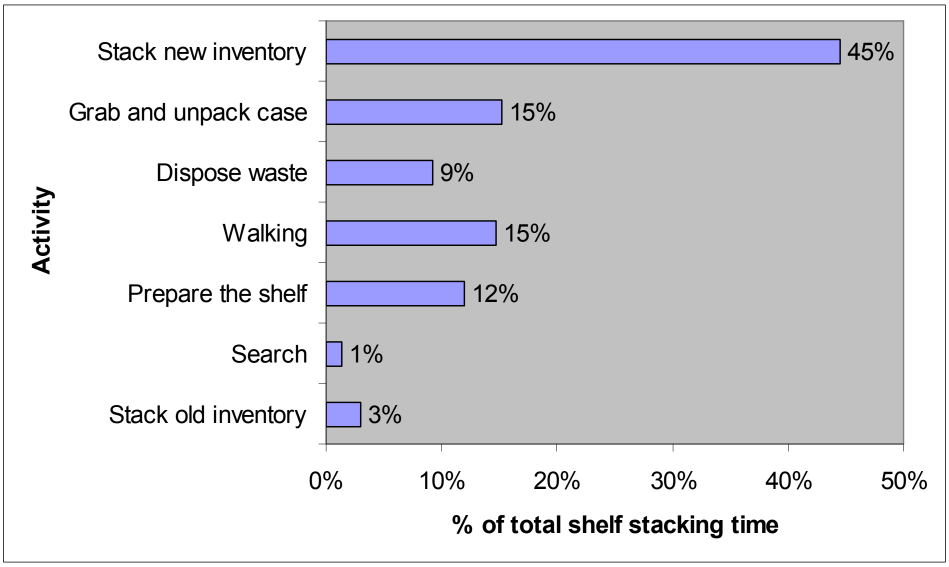

Figure 2.1. Distribution of Total Stacking Time (Chain B)

Appendix 3: Regression results

Table 3.1. Estimating *TST* by sequential regression using models 2 and 3(Chain A)

	Model 2			Model 3		
	Unstd. Coeff.	Std. Err.	Std. Coeff.	Unstd. Coeff	Std. Err.	Std. Coeff.
(Constant)	5.008E-08	1.403		1,163E-08	1,502	
SumPRE_Subactivities	**1.000***	**.022**	**.819**	**1,000***	**,023**	**,798**
R	.819			,798		
R²	.670			,637		
R²adj	.670			,636		
Mean SS Err.	442.099			486,720		
Overall F	2124.795			1834,109***		
df	1, 1046			1, 1046		

Appendix 4: Validation of results for the second dataset

Table 4.1. Regression results for each individual subactivity (standardized coefficients)

step	Variables	Model 1 (Only Product categories)						
		G	S	W	P	Fn	Fo	D
1	Sandwich spread	-.073	-.002	**-.188***	-.056	-.033	**-.218***	-.009
	Canned vegetables	-.002	.058	**-.152***	-.016	-.050	**-.257***	.076
	Candy/Chocolate	-.053	.002	**-.106***	**-.376***	**-.168***	**-.351***	-.028
	Wine	.030	**.241***	**-.130***	**-.333***	**-.339***	**-.343***	-.091
	R²	.009	.056	.041	.157	.094	.142	.016
	R²adj	.002	.049	.034	.151	.088	.136	.009
	Mean SS Err.	47.715	9.069	32.253	47.865	153.305	20.975	20.722
	Overall F	1.298	8.279***	6.015***	25.955***	14.546***	23.039***	2.340
	df	4, 558	4, 558	4, 558	4, 558	4, 558	4, 558	4, 558

Statistical significance *p≤.05, also **p≤.01, *** p≤.001; Reference category = Cookies (N =179)

Table 4.1 (continued). Regression results for each individual subactivity (standardized coefficients)

step	Variables	Model 2 (Product categories, CP, CU)						
		G	S	W	P	Fn	Fo	D
2	Sandwich spread	**-.075***	-.006	**-.194*****	-.053	-.014	**-.223*****	-.015
	Canned vegetables	-.001	.049	**-.165*****	-.005	.008	**-.271*****	.064
	Candy/Chocolate	-.011	.007	-.088	**-.372*****	**-.143*****	**-.345*****	.001
	Wine	.076	**.212*****	**-.159****	**-.288*****	**-.099***	**-.389***	**-.106***
	CP	**.591*****	.084	**.282*****	.031	**.245*****	**.117**	**.436*****
	CU	**.155****	-.051	-.032	.097	**.523*****	-.083	.014
	R²	.489	.060	.109	.169	.531	.148	.213
	R²adj	.483	.050	.100	.160	.526	.139	.204
	R² change	.480	.004	.068	.012	.437	.007	.196
	F change	260.917***	1.075	21.264***	4.036*	258.759***	2.200	69.314***
	Mean SS Err.	24.702	9.067	30.069	47.350	79.686	20.885	16.646
	Overall F	88.644***	5.879***	11.389***	18.837***	104.909***	16.159***	25.046***
	df	4, 558	4, 558	4, 558	4, 558	4, 558	4, 558	4, 558

Statistical significance *p≤.05, also **p≤.01, *** p≤.001; Reference category = Cookies (N =179)

Table 4.1 (continued). Regression results for each individual subactivity (standardized coefficients)

Variables	Model 3 (Only CP, CU)						
	G	S	W	P	Fn	Fo	D
CP	**.634*****	**.193*****	**.220*****	-.004	**.238*****	.008	**.388*****
CU	**.089***	**-.211*****	.054	**.180*****	**.546*****	.091	.087
R²	.478	.033	.066	.032	.510	.009	.199
R²adj	.476	.030	.062	.028	.509	.006	.196
Mean SS Err.	25.053	9.256	31.326	54.780	82.599	24.128	16.825
Overall F	256.308***	9.594***	19.646***	9.137***	291.821***	2.588	69.384***
df	2, 560	2, 560	2, 560	2, 560	2, 560	2, 560	2, 560

Statistical significance *p≤.05, also **p≤.01, *** p≤.001;

Table 4.2. Sequential regression for estimating *TST* models 2 and 3(Chain B)

	Model 2			**Model 3**		
	Unstd. Coeff.	Std. Err.	Std. Coeff.	Unstd. Coeff	Std. Err.	Std. Coeff.
(Constant)	-8.443E-14	2.133		2,337E-14	2,373	
SumPRE_Subactivities	**1.000*****	**.040**		**1,000*****	**,045**	**,684**
R	.724			,684		
R²	.524			,467		
R²adj	.523			,466		
Mean SS Err.	348.965			390,733		
Overall F	618.031***			491,994***		
df	1, 561			1,561		

Table 4.3. Overall regression results for chain B: comparison of Models 1-3.

Step	Variables	Model 1			Model 2			Model 3		
		Unst Coeff.	Std. Err.	Std. Coeff.	Unstd. Coeff	Std. Err.	Std. Coeff.	Unstd. Coeff	Std. Err.	Std. Coeff
1	(Constant)	**60.620*****	1.923		**19.981*****	2.361		**9.960*****	1.971	
	Sandwich spread	**-11.888****	3.940	**-.132**	**-11.367*****	2.883	**-.126**			
	Canned vegetables	**-7.953***	3.914	**-.089**	**-5.929***	2.920	**-.066**			
	Candy/Chocolate	**-18.029*****	2.892	**-.290**	**-15.897*****	2.113	**-.255**			
	Wine	**-20.930*****	2.972	**-.325**	**-13.282*****	2.661	**-.206**			
2	CP				**19.890*****	1.887	**.441**	**18.526*****	1.732	**.411**
	CU				**.892*****	.143	**.292**	**1.079*****	.117	**.353**
	R²	.102			.524			.467		
	R²adj	.095			.519			.465		
	R² change	.102			.422			.467		
	F change	15.820***			246.753***			245.558***		
	Mean SS Err.	662.247			352.103			391.431		
	Overall F	15.820***			102.087***			245.558***		
	df	4, 558			6, 556			2, 560		

Statistical significance *p≤.05, also **p≤.01, *** p≤.001; Reference category = Cookies (N =179)