DEGREE PROJECT IN TECHNOLOGY,
SECOND CYCLE, 15 CREDITS
STOCKHOLM, SWEDEN 2016

A Study of Weather's Impact on Consumption of Goods

For Certain Weather-Dependent Products at a Small Grocery Store

JOHAN KRYLSTEDT

ANDREAS WEIDLERTZ

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ENGINEERING SCIENCES

A Study of Weather's Impact on Consumption of Goods

For Certain Weather-Dependent Products at a Small Grocery Store

J O H A N K R Y L S T E D T
A N D R E A S W E I D L E R T Z

Degree Project in Applied Mathematics and Industrial Economics (15 credits)
Degree Progr. in Industrial Engineering and Management (300 credits)
Royal Institute of Technology year 2016
Supervisors at KTH: Johan Karlsson, Jonatan Freilich
Examiner: Henrik Hult

TRITA-MAT-K 2016:36
ISRN-KTH/MAT/K--16/36--SE

Royal Institute of Technology
SCI School of Engineering Sciences

KTH SCI
SE-100 44 Stockholm, Sweden

URL: www.kth.se/sci

Abstract

This bachelor thesis in applied mathematics and industrial engineering aims to determine if and how weather affects the consumption of goods at small grocery stores. To study this, we conducted a regression analysis based on sales data from an ICA Nära. We have collected one year's weather- and sales data and used mathematical statistics to determine how weather affects the sales for different product groups. Our belief is that weather does affect the consumption.

Several large actors in the industry have some sort of consumption of goods forecast. None of these takes weather into account when creating their sale forecast. Hopefully, this thesis will provide information aiding companies in deciding whether or not to use weather forecasts as a prediction parameter.

The results indicate a large effect on sales for some groups of products. The regression reveals how much the sale of a group increase along with an increase of one unit of the different measured weather factors. There is, most likely, not a perfect linear relation between our response variable and the explanatory variables. Therefore, one must interpret the results carefully. In addition, we discuss how a possible implementation affects the supply chain of a large grocery store company and the importance of flexibility in one's supply chain.

Keywords: Regression, Weather and consumption, forecasting consumption, grocery stores, supply chain flexibility

Sammanfattning

Detta kandidatexamensarbete i tillämpad matematik och industriell ekonomi syftar till att undersöka korrelationen mellan väderlek och försäljning av vissa sorters varor i en mindre matbutik. Genom att genomföra en regressionsanalys baserat på säljdata från en ICA Nära och väderdata från SMHI kan vi undersöka sambandet. Förhoppningsvis ger arbetet ett tydligare beslutsunderlag gällande huruvida företag bör implementera väderleksprognoser i sina varuåtgångsprognoser. Varuåtgångsprognoser är något flera stora aktörer på marknaden använder sig av, ingen av dessa tar väderprognoser i beaktning. Vi har antagit att väderlek påverkar försäljningen av varor.

Resultaten påvisar en stark relation mellan väder och försäljning för vissa varugrupper. Emellertid måste resultatet tolkas på rätt sätt, det visar en ökning i försäljning av en grupp varor då väderparametern ökar med en enhet. Detta skulle beskriva verkligheten perfekt förutsatt att det existerar ett linjärt samband mellan vädret och försäljningen, vilket det förmodligen inte gör. Följaktligen ska resultaten tolkas aktsamt.

Utöver genomförda regressionsanalys har ”supply chain” diskuterats. Matvaruföretags ”supply chains” påverkas om de förändrar sin prognostisering av varuåtgång. Den främsta effekten är att det sätter krav på hur nära inpå en butik kan beställa innan de får leveransen. Således blir ”supply chain flexibility” en viktig aspekt för att företaget ska ha möjlighet att implementera förändringar.

Nyckelord: Regressionsanalys, väderlek och varuåtgång, prognostisering, konsumption, matvarubutik, supply chain flexibility

1. Introduction

1.1. Background

The major companies in the grocery store industry in Sweden utilize different forecast engines (FEs) to create a forecast of future consumption of goods. This forecast is based on earlier sales and parameters such as seasons, events such as Valentine's Day, holidays, whether it is weekend or weekday, and campaigns. Campaigns are short periods of time where selected products are sold at a discounted price. However, the FEs do not take weather into consideration when forecasting. According to ICA's support service, they are frequently asked if their FE takes weather into consideration, indicating that there is a demand for the service.

1.1.1. Accuracy of Weather Forecasts

The accuracy of SMHI's one-day forecast is usually close to 85%, according to their own website. (Swedish Meteorological and Hydrological Institute, 2016)

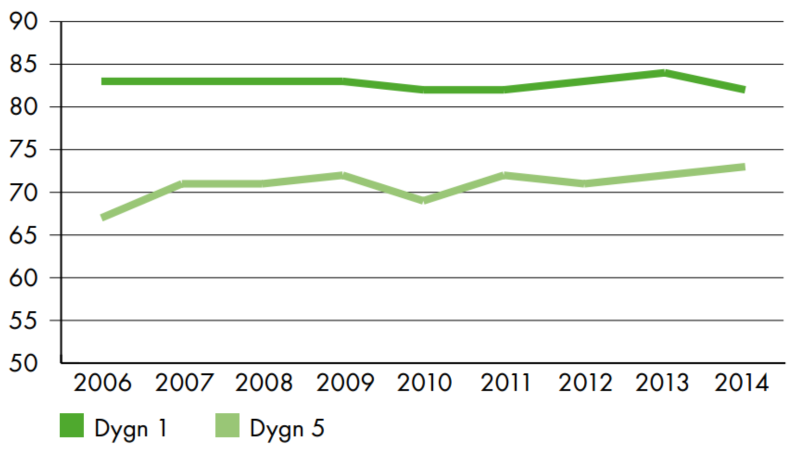

Figure 1 Forecast accuracy according to SMHI,
http://www.smhi.se/kunskapsbanken/meteorologi/hur-mats-prognosers-traffsakerhet-1.17383

Figure 1 shows the accuracy of SMHI's weather forecasts in percentage, this is measured by themselves. The upper one tracks the accuracy of their one-day forecast, and the lower one tracks the accuracy of their five-day forecast.

The Swedish newspaper *Dagens Nyheter* conducted their own study of SMHI's forecast-accuracy and the results indicated a successful forecast 45% of the time (DN, 2011). Therefore, it is difficult to determine how well a weather forecast performs. It would be more interesting to see an average error in temperature and sunshine hours.

1.1.2. Description of ICA

There are more than 1.300 ICA grocery stores in Sweden (ICA Gruppen, 2016). ICA Sweden manages a few central warehouses that service all ICA stores in Sweden.

1.1.3. Store Profiles

There are five different store profiles, ranging from largest to smallest; Maxi, Kvantum, Supermarket, Nära and To Go. ICA Nära are smaller grocery stores with the purpose of being the nearest alternative to as many customers as possible and offer a high service standard. Thanks to the small size, an ICA Nära can occupy space that larger ICA stores cannot, and therefore be more accessible to smaller communities.

1.1.4. Forecast Engine & AoB

All ICA stores use the same forecast engine. It is incorporated in a service called AoB (automatisk order i butik – automated order in store) which is an online service for ordering goods. This is accessed in each store, where they place their orders.

1.1.5. Store Order Process

The ordering procedure usually varies between different groups of products, and of course between stores. The most common division of product groups that are ordered together are colonial goods, i.e. goods that do not require any special handling, such as spices, canned foods or flour. Meat and various products requiring cool storage is ordered as one group and dairies as another. Fruits and vegetables are ordered together and the last major group is frozen products. Apart from these groups there are a few special products that are ordered separately, they are however negligible in this study.

1.1.6. Store Delivery Process

The delivery time is particularly important to this thesis since the closer in time you can place an order, the better are the weather forecasts. Time until delivery varies depending on the type of goods. However, most of the products are delivered the day after the order is placed, as long as it is placed before 12:00 AM. The delivery time may differ for other ICA stores, but the following table describes at what time an order needs to be placed before receiving the delivery at the studied ICA Nära, 2016.

Group	Order time
Colonial goods	1 day earlier before 13:00
Froze products	2 to 4 days earlier
Fruits and vegetables	1 day earlier before 12:00
Dairy	1 day earlier before 13:00
Meat and various refrigerated products	1 day earlier before 12:00

Table 1 Order Time for Different Groups of Articles

The difference in delivery time is mainly explained by the fact that the goods require different handling. They are loaded and sorted into trucks in different ways and the location of the central warehouses also affects delivery time.

1.2. Scope

The ambition of our work is to be able to describe the relationship between consumption of certain product types and weather. To do this we decided to sort certain products into groups that we believed to be weather-dependent, and one group of randomly selected goods for comparison. The sales data contains over 1400 products, it is unnecessary to include all in the regression to ensure a statistical significant result.

Note that we are merely observing the conditions for one store, not *ICA Sweden's* operations. The results from the regression analysis will describe one of the five store profiles (see *Description of ICA*) amply and to some extent the remaining profiles as well. Since we perform the regression analysis on only one store the results will naturally describe that store profile better because the store profiles vary in size and geographical location. However, the results will be relatively well transferable since the stores sell the same type of goods, with some variance in selection.

When studying the supply chain (SC), the interesting aspect is the larger picture. Therefore, we will focus on what is needed from a grocery store company's SC to easily conform to changes, like employing weather forecasts as a parameter in their FE. We will discuss generally what implementing the change implies and what is needed of the SC.

1.3. Aim and Research Questions

The aim of this thesis is to research if and how weather affects the consumption of goods in grocery stores. Our goal is to be able to estimate how much the consumption of certain product-groups fluctuate. We make the assumption that weather does affect consumption of some goods, in particular during the summer months. We state this assumption because it is our initial belief and it influences the thesis. The research question is:

"How does weather affect consumption of goods, for certain product groups, in grocery stores?”

Incorporating weather forecasts as a parameter in the consumption forecast might be difficult for different reasons. This implementation will affect the SC of any company deciding to realize the change, since it adds another supplier. Namely, a supplier providing weather forecasts. It will be discussed how this change affects the SC for an arbitrary business in the industry. Furthermore, we aim to describe what is needed of the SC when implementing this change, we consider this to be more interesting Hence, the second research question is:

“What is needed of a company’s supply chain when implementing weather forecasts as a parameter in their consumption forecast?”

2. Theoretical Background

2.1. Multiple Regression Analysis

Multiple linear regression is a method used to study the relationship between a dependent variable and independent variables. This can be represented as

$$y_i = \beta_0 + \sum_{j=1}^{k} x_{ij}\beta_j + e_i\,, \qquad i = 1, \dots, n$$

Where y_i is an element in the dependent variable $\boldsymbol{y}$, x_{ij} is an element in the independent variables, written in matrix form as $\boldsymbol{X}$, n is the number of observations and k is the number of covariates. This is in matrix form

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}, \qquad \boldsymbol{y} \in \mathbb{R}^{\mathrm{n}}, \boldsymbol{X} \in \mathbb{R}^{\mathrm{nxk}}, \boldsymbol{\beta} \in \mathbb{R}^{k}, \boldsymbol{e} \in \mathbb{R}^{\mathrm{n}}$$

i.e., $\boldsymbol{y}$ is dependent on a number of covariates that are columns in vector $\boldsymbol{X}$. The response variable

$$\boldsymbol{y} = [y_1, y_2, \dots, y_i]^T$$

and the covariates represented by each column in

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,k} \end{bmatrix}$$

are known. The residuals are stochastic variables that regulate the difference between observed data and the estimation of the model from the regression. The regression coefficients

$$\widehat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]^T$$

are estimated by the model, the betas are the partial derivatives of the dependent variable with respect to the independent variables. This beta vector describes the linear relationship between the covariates and the dependent variable in the observed data. The model obtained from the regression is built under the assumptions that:

- There exists a linear relationship between the independent and dependent variables.
- Little to no multicollinearity in the data, multicollinearity is developed if the variables are interdependent.
- Little to no autocorrelation in the data, which happens if the residuals show interdependency.
- The model is assumed to be homoscedastic meaning that the residuals have the same finite variance.
- All variables are to be multivariate normal distributed.
- There is weak exogeneity, meaning the predictors are seen as fixed values and not contaminated with measurement errors.

2.1.1. OLS

Ordinary least squares is a method that minimizes the Euclidian norm of residuals $\boldsymbol{e}$ by calculating a $\widehat{\boldsymbol{\beta}}$ often called OLS estimator, that best describes the relationship between the observed data and the regression model. It minimizes the difference (errors) between observed data and the estimated beta multiplied with the covariates, that is

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

Where the sum of squares (S) is simply

$$S = ||\hat{\boldsymbol{e}}||^{\mathbf{2}} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \hat{\boldsymbol{e}}_{\boldsymbol{i}}{}^{\mathbf{2}} \ .$$

An obvious problem arises since the square of the error term disregards the direction of the inaccuracy, whereas the error term alone does have a direction. The sum of squared errors is therefore instead calculated as

$$S = \sum_{i=1}^{n} \hat{e}_i{}^2 = \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)' \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) =$$

$$= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

The slope of this S-function is derived with respect to $\widehat{\boldsymbol{\beta}}$ and then set to zero to find minima.

$$\frac{\partial S}{\partial \widehat{\boldsymbol{\beta}}} = -\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\mathbf{y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = 0$$

$$\rightarrow \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

OLS estimates this beta that minimizes $\boldsymbol{e}$ (Kennedy P. , 2008)**;** (Hayashi, 2000)**.**

2.1.2. Dummy Variables

A dummy variable takes on the value one or zero to illustrate the existence or absence of a categorical effect that is expected to change the outcome (Gujarati, 2004). A common example of a dummy variable is for *gender*. Say the defined dummy variable represents females. If a female occurs in the sample, the dummy variable takes on the value one, otherwise it takes on the value zero (i.e. a male is observed). Thus, the intercept, or the *benchmark* is represented by males. If both dummies for *gender* are included in the regression model the OLS estimate will not yield a unique solution because the intercept and the two dummies are linearly dependent, this problem is known as *perfect multicollinearity* (Lang, 2015).

2.1.3. Multicollinearity

Multicollinearity is a problem that arises if two or more covariates are linearly dependent, i.e. $X'X$ is singular, thus, $(X'X)^{-1}$ and $\hat{\beta}$ are not defined. This phenomenon makes the OLS estimator non-unique and is called *strict multicollinearity* or *perfect multicollinearity*. However, this is rarely a problem since it is discovered quickly (Hansen, 2015).

A more important situation is *near multicollinearity*, often called *multicollinearity*. This problem occurs when $X'X$ is *near* singular and is caused when two or more covariates are *close* to linearly dependent. This definition is hard to understand since it is unclear what it means for a matrix to be “near singular”, thus, making multicollinearity hard to interpret (Hansen, 2015). However, multicollinearity is not a *specification error*, it is just a discomfort since it causes the standard errors of some regression coefficients to be very large. Hence making the point estimates for said coefficients to be imprecise. Since the standard errors decreases as the number of observations increases, multicollinearity can be aided by adding more observations (Lang, 2015).

Another solution to this problem can be to model the regression using dummy variables. Using one of the dummy variables as benchmark i.e. running the regression without the benchmark coefficient (Lang, 2015).

2.1.4. R-Squared

The measure of effect size given a regression of the independent variable $\boldsymbol{y}$ on some covariates can be calculated by computing the sum of residuals $||\hat{e}||^2$ and $\boldsymbol{y}$ on only one intercept. This in turn yields the residual sum of squares $||\hat{e}_*||^2$ i.e. sum of squares when $\boldsymbol{y}$ is regressed on the intercept only. The amount of variation explained by the covariates is then calculated as the difference between the two sums. Furthermore, the goodness of fit or the so called "coefficient of determination" is calculated as (Lang, 2015)

$$R^2 = \frac{||\hat{e}_*||^2 - ||\hat{e}||^2}{||\hat{e}_*||^2}.$$

It is clear from the definition of *R-squared* that a higher value implies a better model since the difference between the observed values and the values predicted by the model are smaller for a higher *R-squared*.

A flaw with *R-squared* is that the model assumes that every covariate explains the variation, thus, adding more covariates will raise the value of *R-squared*. This flaw is surpassed by adjusting for the number of covariates in the model by calculating the *adjusted R-squared* (Kennedy P. , 2008)

$$\bar{R}^2 = R^2 - (1 - R^2)\frac{k}{n-k-1},$$

where k is the number of covariates and n is the sample size. Hence, using the *adjusted R-squared* is an apt method to determine the goodness of fit.

2.1.5. Variance Inflation Factor (VIF)

The Variance Inflation Factor(VIF) is used to identify whether multicollinearity is present in an OLS regression and its severity. VIF measures how much the variance of an estimated regression coefficient increases because of multicollinearity. Running the OLS regression with a covariate as a function of the other covariates, one calculates the VIF as

$$VIF(\hat{\beta}_i) = \frac{1}{1 - R_i^2}.$$

Where R_i^2 is the goodness of fit for the model in which one of the covariates from the original model is regressed on all other covariates in the original model. The multicollinearity is considered high if $VIF(\hat{\beta}_i) > 10$, which is a problem (Lang, 2015).

2.1.6. Partial Eta-squared

Partial eta squared (η^2) is the measure of effect size from a covariate or a set of covariates (Lang, 2015) and is defined as the proportion of the total variation accounted by a covariate excluding other covariates from the total non-error variation (Choen, 1973); (Haase, 1983); (Kennedy J. , 1970).

$$\eta^2 = \frac{||\hat{e}_*||^2 - ||\hat{e}||^2}{||\hat{e}_*||^2} = \frac{R^2 - R_*{}^2}{1 - R_*{}^2},$$

where R^2 and $R_*{}^2$ are the coefficients of determination for the full and the “restricted” regression, $\hat{e}$ and $\hat{e}_*$ corresponding residuals (Lang, 2015).

2.1.7. AIC

The Akaike information criterion (AIC) is used to measure the quality of statistical models given a set of data. By itself the value of AIC has no meaning and first becomes interesting when it is compared between different models. The model with the lowest AIC will be the best model and is calculated by minimizing

$$AIC = nlog(||\hat{e}||^2) + 2k,$$

where n is the number of observations, k is the number of coefficients and $|\hat{e}|^2$ is the average residual sum of squares.

2.1.8. F-Test

The F-test is used under the null hypothesis that the means of a set of normally distributed populations are equal and the test statistic is F-distributed under the null hypothesis. The null hypothesis simply states that there is no relationship between the response variable and the covariates. The F-test is done by examining if all β-coefficients are equal to zero except the intercept (β_0). The F-statistic is calculated as

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{r}$$

where R^2 is the *coefficient of determination*, n is the total number of observations, k is the number of covariates and r the number of coefficients tested for zero. To determine the significance of the test one uses the p-value under a given significance level α to reject or keep the null hypothesis (Lang, 2015).

2.1.9. Confidence Interval

Confidence intervals are typically used to set the coverage probability with a certain validity, typically at 95%. The goal of a confidence interval is to contain the true value, e.g. $\theta \in C_n$ with a high probability, where θ is a real valued parameter and C_n are a collection of values in $\mathbb{R}^k$ (Hansen, 2015).

The confidence interval for β_j at level α is

$$\widehat{\beta_J} \pm \sqrt{F_\alpha(1, n-k-1)}\sigma(\widehat{\beta_J})$$

where $F_\alpha(1, n-k-1)$ is the α quantile of the F-distribution with one numerator degrees of freedom and n-k-1 denominator degrees of freedom and $\sigma(\widehat{\beta_J})$ is the estimate of the standard deviation of the estimator $\widehat{\beta_J}$.

Furthermore, a $F(1, n-k-1)$-statistic for the hypothesis $\beta_j = \beta_j{}^0$ is (Lang, 2015)

$$F = \left(\frac{\widehat{\beta_J} - \beta_j{}^0}{\sigma(\widehat{\beta_J})}\right)^2.$$

2.1.10.Endogeneity

Endogeneity is a problem that occurs when $E(e_i) \neq 0$. It arises when the expected value of e_i depends on the value of one or more of the covariates, in other words the residual is correlated with a covariate. This phenomenon cannot appear if β is defined by linear predictions and occurs when the regression is given a structural interpretation. Endogeneity causes the OLS estimation to produce inconsistent estimates, a negative correlation will underestimate the coefficient while a positive correlation will overestimate it (Lang, 2015)

2.1.11.Simultaneity

Simultaneity occurs if one or more covariates are influenced by the dependent variable, i.e. the cause and effect goes both ways (Lang, 2015).

2.1.12.Heteroscedasticity

To assume that all residuals e_i have the same standard error σ is a common assumption and is called *homoscedasticity*.

$$E(e_i) = 0 \; and \; E({e_i}^2) = \sigma^2$$

$e_i{:}\,s$ are independent between observations. This is however a rather unjustified assumption because the variances of the $e_i{:}\,s$ are generally not all the same, thus heteroscedasticity is present.

$$E \neq 0 \; and \; E({e_i}^2) \neq \sigma^2$$

If a model assumed to be homoscedastic displays heteroskedastic residuals, the standard deviations of the parameter estimates will be inconsistent, rendering the F-test invalid. Since OLS is inefficient for heteroskedastic models it is commendable to construct the regression model as close to homoscedastic as possible. There are more efficient estimators than OLS for regression models that are heteroscedastic, however they are not as robust and therefore OLS is the estimator of choice (Lang, 2015).

2.1.13.Missing Relevant Covariates

Missing relevant covariates occurs when the component of the residual that makes it correlate with a covariate can be identified. One might improve the regression by adding the missing covariate (Lang, 2015)

2.1.14.Logarithmic Transformation

Logarithmically transforming variables are frequently used when the relationship between the independent and dependent variables are non-linear. According to (Benoit, 2011) this makes the linear method applicable, while preserving the non-linear relationship. The log linear model is defined as

$$\log y_i = \alpha + \beta \mathrm{X_i} + e_i.$$

The interpretation of the log-linear model is that every increase in X by one-unit will result in an expected increase in $\log \boldsymbol{y}$ of $\hat{\beta}$ units. The expected value of $\boldsymbol{y}$ is multiplied by $e^{\hat{\beta}}$ for every one-unit increase in X. Furthermore, one can use that for small values of $\hat{\beta}$ ($e^{\hat{\beta}} \approx 1 + \hat{\beta}$) the approximation $100 \cdot \hat{\beta}$ is the expected percentage change in $\boldsymbol{y}$ for a unit increase in X (Benoit, 2011)

2.1.15. Root-mean-square error

The root-mean-square error (RMSE) is a measure of the difference between values predicted by an estimator or a model and the observed values. RMSE represents the sample standard deviation of the differences between predicted and observed values. These differences are called prediction errors when computed out-of-sample and are called residuals when the calculations are preformed over the data sample used for estimation. The RMSE provides a measure of accuracy to compare different forecasting models. However, it can only be used to compare forecasting errors for a particular variable and not between variables, as it is scale-dependent (Lehmann & Casella, 1998). The RMSE is calculated as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\hat{y}_i} - \boldsymbol{y_i})^2},$$

where $\boldsymbol{\hat{y}_i}$ is a vector of n predictions, and $\boldsymbol{y_i}$ is a vector of observed values.

3. Supply Chain

Supply chain (SC) is simply the flow of goods and services from extracting raw materials to delivery of the final product, usually described through one company's perspective. It includes facilities and distribution options that perform all necessary functions for a business (Harrison & Ganeshan, 1995), this involves all suppliers. A supplier is some entity that delivers any type of service, material or product to a business. Each company or organization in a SC have their own SC (Pettersson & Segerstedt, 2013).

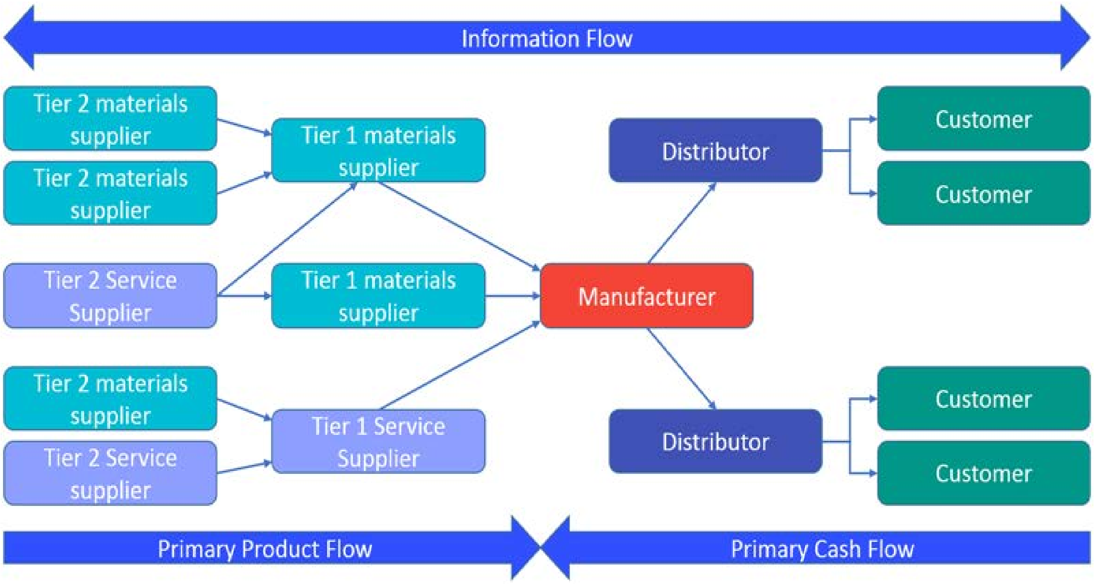

Figure 2 Simplified Supply Chain (Threadpunter, 2015)

3.1. Supply Chain Decisions

Supply chain decisions are often divided into two broad categories, either strategical or operational. Strategical decisions regard the long-term tactics and provide guidelines for which directions the company's supply chain management (SCM) should move towards. Operational decisions concern the day-to-day activities and govern the SC on activity level (Game Theory Lab, 1995).

3.2. Supply Chain Activities

There are several levels of activities in a SC. Firstly, there are logistic activities that include all sorts of movement and storage of materials as they move along the SC, e.g. procurement or purchasing, sales forecasting, receiving, stock control etc. (Waters, 2003). Secondly, it refers to all activities required to design a product until it is delivered and in certain cases the SC continues thereafter, since service is often offered depending on the type of product (Dolgui & Proth, 2010).

3.3. Supply Chain Logistics

SC logistics differ from SC in the sense that it is less extensive and deals mainly with activities in the organization, e.g. procurement, distribution and inventory management (What is logistics and supply chain management: Supply Chain Opz, 2016). SC logistics can yield competitive advantages since physical flows of material is costly, and how emergency logistic channels are handled directly affect the business's performance (Sànchez & Pérez Pérez, 2005).

3.4. Supply Chain Flexibility

Supply chain flexibility is a term describing a SC's ability to quickly reconfigure and align itself and its entities according to changing conditions, in a cost efficient way. There are many aspects to flexibility in a SC since the SC, generally, consists of several entities. The attributes of the SC can be divided into agility, adaptability, responsiveness and legality (Seebacher & Winkler, 2016). The components of the SC however, can, according to (Duclos, Vokurka, & Lummus, 2003), be represented by; operations system, market, logistics, supply, organisational and information systems. These are taken from research regarding manufacturing-, strategic- and supply chain flexibility.

More companies are working with SC flexibility, since it aids coping with changing markets and the demands placed on the quality of customer service. The SCs are becoming more dynamic and complex (Merschmann & Thonemann, 2010). However, having a flexible SC is costly and one should therefore determine a degree of flexibility that is sufficient (Pujawan, 2004).

3.5. Bullwhip Effect

The bullwhip effect is a phenomenon where forecasts yield supply chain inefficiencies and refers to shifts in customer demand, causing swings in inventory as you move up the supply chain. Customer demand is rarely perfectly stable and companies have to forecast the demand to accurately position inventory and other resources. Since said forecasts are based on statistics, they are rarely perfectly accurate, thus, companies tend to carry an inventory buffer called *safety stock*. Moving up the supply chain from end customer to supplier, one can observe a greater variation in demand, thus, a greater need for safety stock. As demand rises, down-stream participants increase orders. In periods of falling demand, down-stream participants reduce or stop orders. The effect is that variations in demand amplify as one move upstream in the supply chain (Herlyn, 2014) (Brauner, Runge, Groten, Shchuh, & Ziefle, 2013).

3.6. Stock Control

Stock control is used for formulating optimal strategy regarding inventory management. Some of the most important aspects to take into consideration are:

- Materials to store
- Overall investment
- Customer service
- Stock levels
- Order sizes
- Order timing

An optimal stock control is crucial for cutting costs without trailing behind in quality of customer service. The positive effects from improving stock control are many, e.g. less unnecessary inventory, better customer service, more efficient warehouse management etc. (Waters, 2003)

4. Method

4.1. Data

The data required to perform the regressions are sales figures from a grocery store and data regarding weather. Naturally, the data must correspond to the same time period.

4.1.1. Sales Data

We gathered one year of sales data from an ICA Nära via their servers, it contains extensive information regarding the sales figures. We were unable to collect a consecutive calendar year of sales data because ICA only stores the data one year.

The sales data stretches from the 12th of January 2015 to the 12th of January 2016. The data samples contain over 1400 different types of goods and the number of sold products each day during one year. However, we use 229 different product types since we deem it to be sufficient. The smaller sample size utilized contain goods selected partially at random and partially goods whose sales we intuitively believe to be correlated with weather. We only use name and number of sold goods, note that no profit margins or any monetary figures will be presented in this paper.

4.1.2. Weather Data

The Swedish Meteorological and Hydrological Institute (SMHI) stores weather history based on recorded data by their observation stations. More specifically, we collected data measured at observation stations in Stockholm's inner city and Bromma (Swedish Meteorological and Hydrological Institute, 2016).

From SMHI's website we have obtained weather data corresponding to the entire period of sales data. In addition to temperature and sunshine hours the data also includes wind speed and precipitation. The weather data is divided into the four categories as in ***Table 2***.

Weather	
Precipitation	**Total downfall in millimeters**
Temperature	**Average temperature**
Sunshine hours	**Total sunshine time in hours**
Wind speed	**Average wind speed**

Table 2 Description of Weather Data

4.2. Response Variable

The response variable $\boldsymbol{y} = [y_1, y_2, \dots, y_{3650}]^T$ is a vector with 3650 elements containing the sales data. The sales data for each group is represented in the vector, i.e. one year of sold beer is followed by the same year's sales of fruit, etc. Thus, y_i represents one day of sales of one of our selected groups of goods. E.g. y_1 to y_{365} represents number of sold products during one year, for one group and y_{366} to y_{731} represents the sales data during the same year for the next group, and so on.

4.3. Covariates

The covariates used for the construction of the initial model are: beer, fruit, grill, ice cream, odd drinks, pastry, meals ready to eat, large sodas, small sodas, events, winter, spring, summer, fall, weekend, precipitation, wind speed, sunshine hours, temperature and a control group of goods. The control group is the benchmark of choice for group of goods, containing goods picked at random. To avoid multicollinearity we also benchmark against winter. We denote the covariates as

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,18} \\ \vdots & \ddots & \vdots \\ x_{3650,1} & \cdots & x_{3650,18} \end{bmatrix}$$

where $X_{n,k}$ states which n: th day of sales and k: th covariate.

4.3.1. Goods

All groups of goods included in our regression analysis are summarized below, including their type and number of different products in each group.

Covariate	No. Of different goods
Beer	7
Fruits	16
Grill	20
Ice cream	36
Odd drinks	27
Pastry	23
Meals ready to eat	22
Large sodas	28
Small sodas	28
Control group	22

Table 3 Summary of the Product Groups, Covariates

4.3.2. Events

After contact with the AoB it was clear that a key element in their consumption forecast is to include events, such as valentine's day, Christmas, Midsummer etc. We decided to take into account all the big holidays in Sweden, namely: Christmas, New Year's Eve, Easter and Midsummer. Apart from these big events we also decided to include "Kanelbullensdag", "Fettisdagen", Valentine's day and the Swedish National Day since these four events have a considerable impact on sales for the studied *ICA Nära*. This covariate is defined as a dummy variable assuming the value one if a day falls under the category "event" and zero otherwise.

4.3.3. Season (Winter, Spring, Summer, Fall)

Season is a parameter that affects sales on ICA Nära, hence, four different season dummies were defined. The covariate winter stretches from the 1st of December to the 1st of March, spring stretches from the 1st of March to the 1st of June, summer continues from the 1st of June to the 1st of September and fall stretches from the 1st of September to the 1st of December. These dummy variables assume the value one if goods were sold on any of the days in their respective interval and zero otherwise.

4.3.4. Weekend

Another crucial parameter that can explain sales at ICA Nära is whether it is weekend. Knowing that the data is sorted from the 12th of January 2015 to the 12th of January 2016, we defined a dummy to assume the value one if goods were sold on a weekend and 0 otherwise.

4.3.5. Weather Related Covariates

The regression model is limited to four covariates describing weather. Since our aim is to understand how weather affects consumption of goods it is sufficient to be able to classify weather as "good" or "bad". The chosen weather covariates are precipitation, wind speed, sunshine hours and temperature. One might argue that weather depends on other parameters, such as humidity etc. But as explained, these four aspects are adequate to describe the weather for the purpose of the thesis.

- Precipitation: The data obtained was given in downfall per day, thus, making it impossible to arrange the precipitation data for the opening hours at ICA Nära. This covariate is defined as the precipitation in millimeters each day.
- Wind speed: This data was given by the hour and therefore arranged as a covariate defined as average wind speed for each day during the opening hours (8-22).
- Sunshine hours: This covariate was defined as the total amount of sunshine hours during the opening hours for each day.
- Temperature: Initially the data was given as the temperature each hour, however we decided to take the average temperature during the opening hours of the store each day.

4.4. Model Construction

All covariates described are used in an *initial model* that we evaluate by performing tests that indicate which covariates that impact the response variable. Through this, we will develop a *final model* that later is the basis for several regressions by including different *interaction terms*.

4.4.1. Initial Model

The initial model is constructed with the response variable and covariates described above. The log-linear model described in *Logarithmic Transformation* is applied to obtain the intercepts as a percentage and reduce the heteroscedasticity in our model. We use *control group* to benchmark, since we wish to examine if the other nine groups are weather dependent, in accordance with our initial beliefs. We also decided to benchmark against winter.

The covariates are not equally defined, therefore, they need to be interpreted differently. The first nine covariates are dummies describing each group of goods, defined as

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{1}} = \begin{cases} 1, if\ i \in beer \\ 0, otherwise \end{cases}$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{2}} = \begin{cases} 1, if\ i \in fruits \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{3}} = \begin{cases} 1, if\ i \in grill \\ 0, otherwise \end{cases}$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{4}} = \begin{cases} 1, if\ i \in ice\ cream \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{5}} = \begin{cases} 1, if\ i \in odd\ drinks \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{6}} = \begin{cases} 1, if\ i \in pastry \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{7}} = \begin{cases} 1, if\ i \in ready\ to\ eat \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{8}} = \begin{cases} 1, if\ i \in soda\ large \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{9}} = \begin{cases} 1, if\ i \in soda\ small \\ 0, otherwise \end{cases},$$

where $i = 1,2, \ldots ,3650$ and the tenth covariate in the regression is *events*

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{10}} = \begin{cases} 1, if\ i \in events \\ 0, otherwise \end{cases}.$$

The following three covariates are the ones representing *seasons*

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{11}} = \begin{cases} 1, if\ i \in spring \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{12}} = \begin{cases} 1, if\ i \in summer \\ 0, otherwise \end{cases},$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{13}} = \begin{cases} 1, if\ i \in fall \\ 0, otherwise \end{cases}.$$

The fourteenth covariate checks if a product is sold during the weekend

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{14}} = \begin{cases} 1, if\ i \in weekend \\ 0, otherwise \end{cases}.$$

The 15th, 16th, 17th and 18th covariate describes weather and are defined as

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{15}} = precipitation\ day\ i,$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{16}} = wind\ speed\ day\ i,$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{17}} = sunshine\ hours\ day\ i$$

$$\boldsymbol{X}_{\boldsymbol{i},\mathbf{18}} = temperature\ day\ i.$$

Thus, the initial model is defined as

$$\log(y_i) = \beta_0 + \sum_{j=1}^{18} \left(x_{i,j}\beta_j\right),\ i = 1, 2, \ldots, 3650.$$

Following the construction of this model we performed an OLS regression, yielding the following results.

	Estimate	Std.Error	Eta.sq	p.value	Lower	Upper
(Intercept)	3.3924	0.0261	0.8065	0.0000	3.3413	3.4436
beer	-0.9509	0.0247	0.2701	0.0000	-0.9993	-0.9026
fruits	1.9248	0.0192	0.6025	0.0000	1.8871	1.9625
grill	0.0594	0.0222	0.0014	0.0075	0.0159	0.1029
ice cream	0.1206	0.0305	0.0059	0.0001	0.0607	0.1804
odd drinks	0.1452	0.0182	0.0086	0.0000	0.1094	0.1809
pastry	1.0343	0.0210	0.3045	0.0000	0.9930	1.0756
ready to eat	-0.5158	0.0309	0.0982	0.0000	-0.5765	-0.4552
soda large	0.8733	0.0230	0.2379	0.0000	0.8283	0.9184
soda small	1.3176	0.0197	0.4153	0.0000	1.2791	1.3562
events	-0.2937	0.0431	0.0346	0.0000	-0.3783	-0.2091
spring	0.1224	0.0204	0.0102	0.0000	0.0824	0.1624
summer	0.0449	0.0311	0.0006	0.1488	-0.0161	0.1059
fall	0.1033	0.0216	0.0063	0.0000	0.0610	0.1457
weekend	0.1264	0.0132	0.0255	0.0000	0.1005	0.1523
precipitation	-0.0014	0.0013	0.0002	0.2816	-0.0039	0.0011
wind speed	-0.0086	0.0044	0.0011	0.0483	-0.0172	-0.0001
sunshine hours	0.0101	0.0019	0.0063	0.0000	0.0064	0.0137
temperature	0.0099	0.0016	0.0104	0.0000	0.0068	0.0130

Table 4 Values for Initial Model

4.4.2. Initial Model Evaluation

The R-squared is generally applied to determine the effect size. The initial model's R-squared value is 0.8535, thus, our covariates seem to explain 85.35% of the variance in our initial model. However, according to (Hansen, 2015) one cannot use R-squared to evaluate the model as it increases when the number of covariates in the model increases. Therefore, we also observe the adjusted R-squared for the initial model, that is 0.8528, indicating that 85.28% of the variance is explained by the covariates.

4.4.3. F-test for Initial Model

To further evaluate the initial model an F-test is performed for each of the coefficients, yielding the results described in ***Table 5***

	F	p	eta.sq	AIC	F.hom	p.hom
(Intercept)	16894.4200	0.0000	0.8065	5992.4190	15130.6400	0.0000
beer	1487.6040	0.0000	0.2701	1147.0370	1343.4400	0.0000
fruits	10009.9100	0.0000	0.6025	3365.6290	5504.3010	0.0000
grill	7.1618	0.0075	0.0014	3.2615	5.2379	0.0222
ice cream	15.6052	0.0001	0.0059	19.6422	21.5935	0.0000
odd drinks	63.4343	0.0000	0.0085	29.3392	31.3103	0.0000
pastry	2416.1800	0.0000	0.3045	1323.1820	1589.3880	0.0000
ready to eat	278.0859	0.0000	0.0982	375.1960	395.3065	0.0000
soda large	1445.8220	0.0000	0.2379	989.4056	1133.1830	0.0000
soda small	4483.0690	0.0000	0.4153	1957.0750	2579.4980	0.0000
events	46.3484	0.0000	0.0346	126.5398	130.1489	0.0000
spring	36.0225	0.0000	0.0102	35.4068	37.4034	0.0000
summer	2.0850	0.1488	0.0006	0.1789	2.1682	0.1410
fall	22.8886	0.0000	0.0062	20.8816	22.8340	0.0000
weekend	91.5033	0.0000	0.0255	92.1854	94.9145	0.0000
precipitation	1.1598	0.2816	0.0002	-1.1877	0.8081	0.3687
wind speed	3.9016	0.0483	0.0011	1.9034	3.8851	0.0488
sunshine hours	28.8880	0.0000	0.0063	20.8929	22.8453	0.0000
temperature	39.4232	0.0000	0.0104	36.1215	38.1218	0.0000

Table 5 Values for Each Covariate in Initial Model

The AIC in this case represents the difference in AIC as the covariate is removed from the regression. Thus, a negative AIC indicates that the overall model takes on a lower AIC if the covariate is excluded from the model. Following the result, one can observe that precipitation receives a negative AIC, that fact along with a high p-value and a low F-statistic is an indication that precipitation should be removed from the regression model. The results from *Table* 5 also show that the confidence interval for precipitation is both positive and negative. The covariate precipitation is to be removed, based by these results.

One might also notice that the covariate *summer* displays a high p-value together with a low F-statistic and a confidence interval that is both positive and negative, however, the AIC is positive. Once precipitation is removed, another F-test reveals that the covariate summer takes on a negative AIC and is therefore removed from the model.

4.4.4. VIF for Initial Model

The Variance Inflation Factor (VIF) for the initial model is displayed below. It is obvious from the chart that the VIF for every covariate is well below ten, the highest being 5.2. Thus, multicollinearity is probably not present.

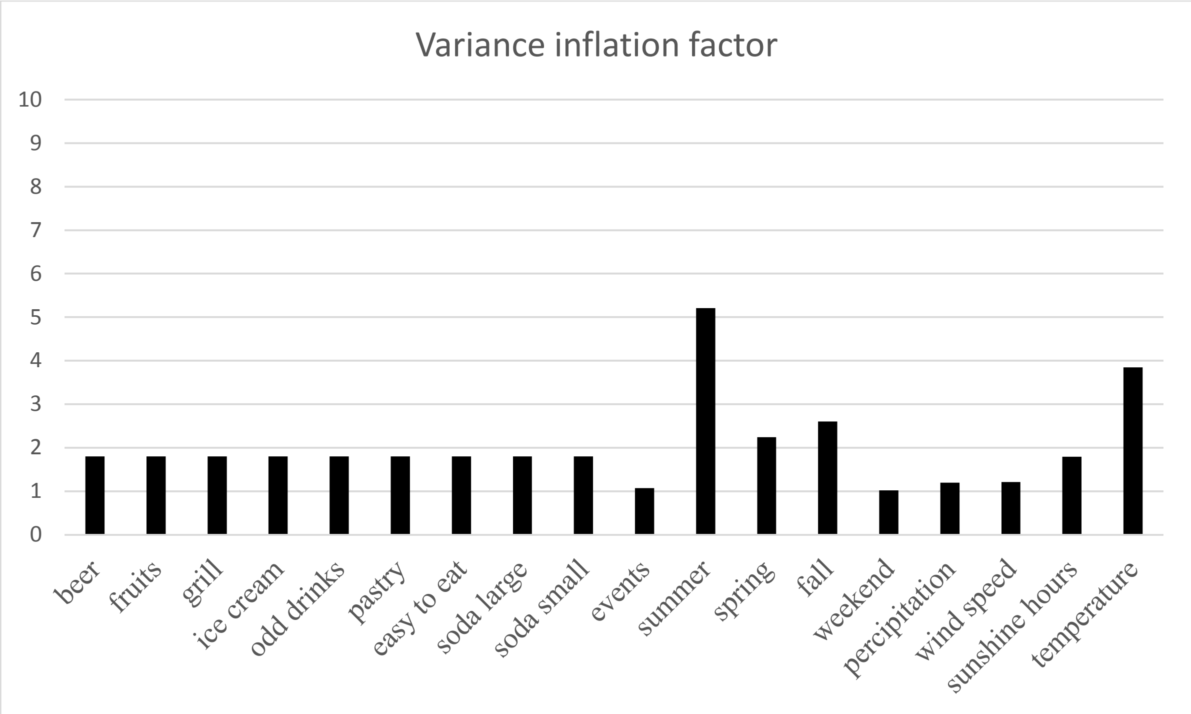

Figure 3 Variance Inflation Factor, Initial Model

4.4.5. Residuals and Q-Q Plot for Initial Model

Figure 4 displays the histogram of standardized residuals and by a visual inspection, one can see that the residuals are approximately normally distributed. If the two distributions compared in a Q-Q plot are similar, the points will follow the line. In the initial regression model we can see that this is the case for the majority of the points, however it exists an upper and lower tail. This might imply heteroscedasticity, but OLS is as mentioned a robust model.

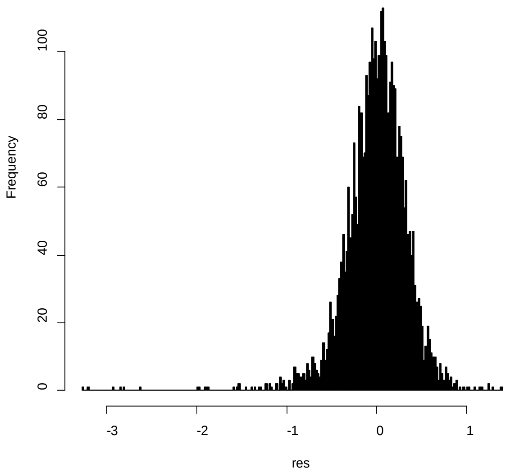

Figure 5 Histogram

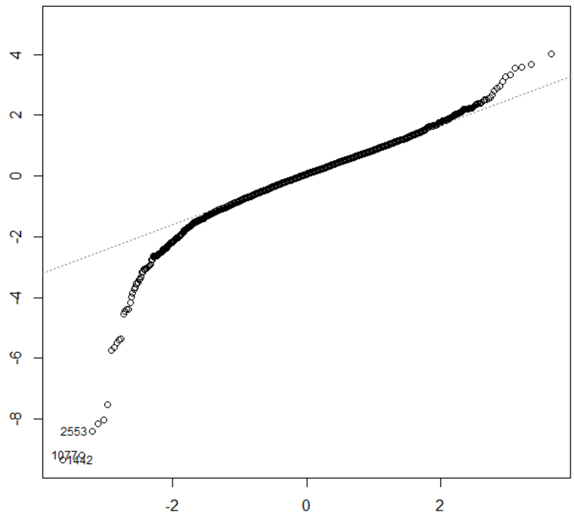

Figure 4 Q-Q Plot

4.4.6. Description of Final Model

After reducing the model, we arrive at a *final model* that includes the covariates beer, fruit, grill, ice cream, odd drinks, pastry, ready to eat, large sodas, small sodas, events, winter, spring, fall, weekend, wind speed, sunshine hours, temperature and a control group of goods. The response variable is kept as the logarithm of sold goods, thus, keeping it as a log-linear model to get the intercepts as a percentage and to reduce the heteroscedasticity in the model. Once again our benchmark is *control group* since the other nine groups are thought to be weather dependent. Furthermore, we still benchmark against winter. Resulting in the final model:

$$\log(y_i) = \beta_0 + \sum_{j=1}^{16} (x_{i,j}\beta_j),\ i = 1, 2, \ldots, 3650$$

4.4.7. Prediction using the final model

Now that the final model is established we want to investigate how the prediction accuracy compares between the models, i.e. the final model taking weather into account and the final model without the weather covariates. The final model without weather will be referred to as the *simple model*. To make the comparison we perform a regression analysis with three months data followed by one-week prediction of consumption of goods.

Regression	Prediction
12 January - 12 April	13 April - 19 April
20 April - 20 July	21 July - 27 July
1 August - 1 November	2 November - 8 November

Following the prediction, we compared the results with the actual consumption of goods the given week, thus, giving us an indication of the prediction accuracy for the different models. To further evaluate the models prediction accuracy we calculated the root-mean-square error.

4.4.8. Final Model Interactions

In addition to the final model, four more models are built to further investigate how goods consumption is affected by weather. These models are built by allowing each of the weather covariates interact with each group of goods. Since the same methods are used to reduce the interaction models, we choose to only present the final models of each interaction. The models with interaction terms including wind speed and precipitation have no significant effect. Therefore, we reduce the four models to two and we obtain one model containing the interaction between the type of goods and temperature, and another model with the interaction between type of goods and sunshine hours.

5. Results

5.1. Final Model

The final model is defined as

$$\log(y_i) = \beta_0 + \sum_{j=1}^{16} \left(x_{i,j}\beta_j\right),\ i = 1, 2, \dots, 3650$$

and performing an OLS regression yields the following result

	Estimate	Std.Error	Eta.sq	p.value	Lower	Upper
(Intercept)	3.4002	0.0249	0.8216	0.0000	3.3514	3.4490
beer	-0.9509	0.0246	0.2699	0.0000	-0.9992	-0.9026
fruits	1.9248	0.0192	0.6024	0.0000	1.8871	1.9625
grill	0.0594	0.0222	0.0014	0.0076	0.0158	0.1029
ice cream	0.1206	0.0306	0.0059	0.0001	0.0606	0.1805
odd drinks	0.1452	0.0182	0.0085	0.0000	0.1094	0.1809
pastry	1.0343	0.0210	0.3043	0.0000	0.9930	1.0756
ready to eat	-0.5158	0.0309	0.0981	0.0000	-0.5765	-0.4552
soda large	0.8733	0.0230	0.2377	0.0000	0.8283	0.9184
soda small	1.3176	0.0197	0.4152	0.0000	1.2790	1.3563
events	-0.2938	0.0435	0.0350	0.0000	-0.3790	-0.2086
spring	0.1013	0.0140	0.0129	0.0000	0.0738	0.1288
fall	0.0800	0.0133	0.0081	0.0000	0.0538	0.1062
weekend	0.1259	0.0132	0.0255	0.0000	0.1001	0.1517
wind speed	-0.0103	0.0043	0.0016	0.0165	-0.0186	-0.0019
sunshine hours	0.0109	0.0017	0.0086	0.0000	0.0075	0.0143
temperature	0.0115	0.0010	0.0349	0.0000	0.0096	0.0135

Table 6 Values for the Final Model

The final model's R-squared is lower than that of the initial model, however, this is reasonable since we removed two covariates. The adjusted R-squared is the same as before but the F-statistic is higher for the final model. Furthermore, the p-value for each of the covariates is very low with the majority being zero on the fourth decimal. In addition, there exist no two-sided confidence intervals.

It is clear from *Figure 6* that the VIF for some covariates have decreased in our final model.

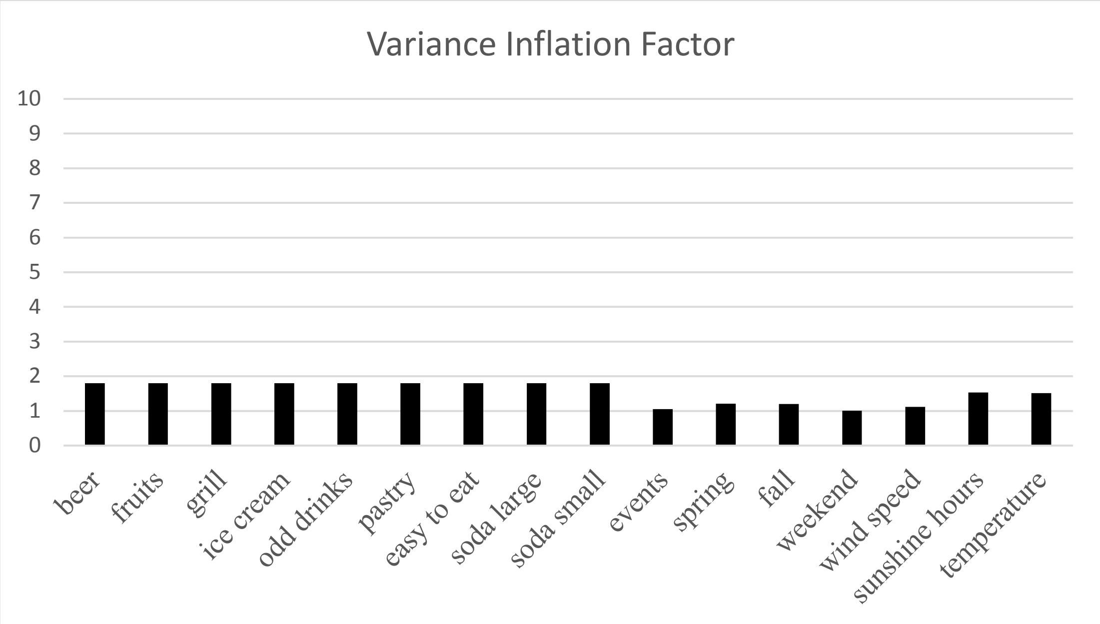

Figure 6 Variance Inflation Factor, Final model

From *Table 7* one can observe that there is no change in VIF for the covariates regarding goods. However, the other covariates all receive a reduction of VIF in the final model and it is likely that there exist no multicollinearity in the final model.

Covariate	VIF, initial model	VIF, reduced model	ΔVIF
beer	1.8000	1.8000	0.0000
fruits	1.8000	1.8000	0.0000
grill	1.8000	1.8000	0.0000
ice cream	1.8000	1.8000	0.0000
odd drinks	1.8000	1.8000	0.0000
pastry	1.8000	1.8000	0.0000
ready to eat	1.8000	1.8000	0.0000
soda large	1.8000	1.8000	0.0000
soda small	1.8000	1.8000	0.0000
events	1.0680	1.0550	-0.0130
spring	2.2450	1.2060	-1.0390
fall	2.6010	1.2000	-1.4010
weekend	1.0190	1.0080	-0.0110
wind speed	1.2100	1.1200	-0.0900
sunshine hours	1.7940	1.5270	-0.2670
temperature	3.8440	1.5140	-2.3300

Table 7 Difference in VIF from Initial to Final Model

5.2. Prediction Accuracy

When studying the results regarding prediction accuracy it is essential to note that our model is **not** constructed to predict the consumption. However, the prediction accuracy could be used as a basis when discussing whether or not to implement a new FE.

From the figure below one can observe that the final model is generally a better predictor than the simple model for all three of the predicted weeks. The final model "wins" over the simple model 60% of the time for week 16, 67% of the time for week 30 and 54% of the time for week 45. However, one cannot tell from the results in *Figure 7* how close the predicted value is to the real value. Therefore, we calculated the average error in prediction, which amounted to 28% for the final model and 30% for the simple model.

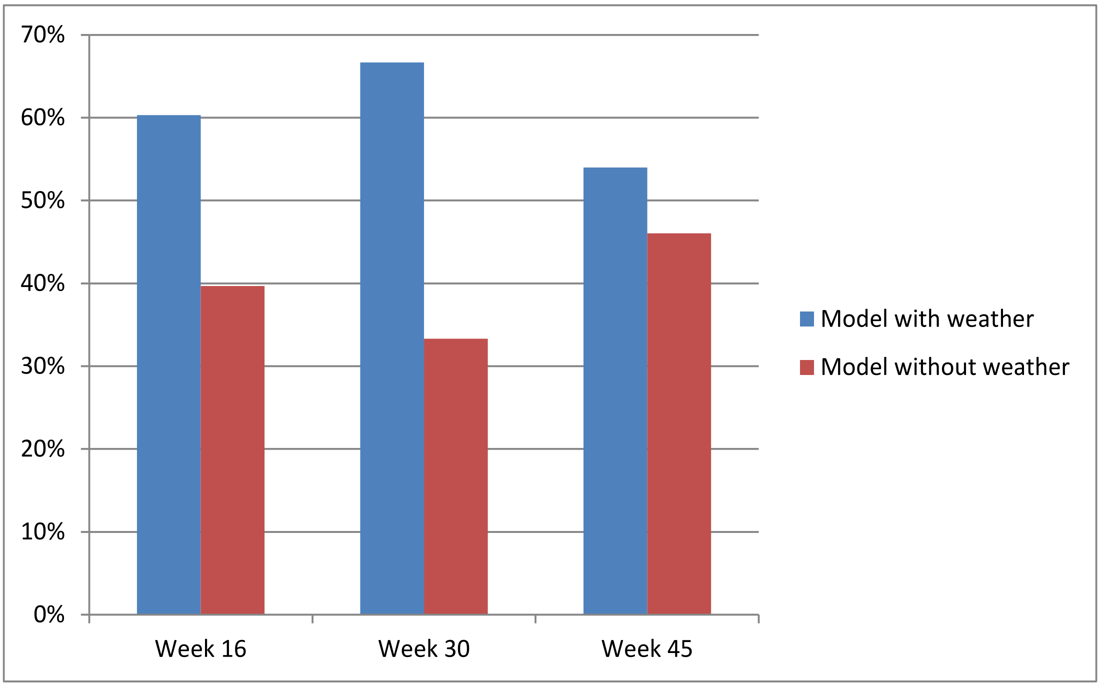

Figure 7 Comparison of Predictions

To further assess the quality of the models, the root mean square error is calculated. From *Table 8* one can observe the root-mean-square error for each model over the three different predicted weeks. One can also observe that the advanced model achieves a lower root mean square error for almost all of the covariates, all three predicted weeks.

	Week 16		Week 30		Week 45	
Covariate	**RMSE advanced**	**RMSE simple**	**RMSE advanced**	**RMSE simple**	**RMSE advanced**	**RMSE simple**
Beer	7.5061	7.6653	5.1791	5.7764	3.4823	3.3781
Grill	15.1306	15.6976	10.2248	9.9813	13.5118	13.0293
Ice cream	26.7556	27.5718	20.6092	23.0113	16.0062	18.6262
Odd drinks	4.8934	6.0233	7.0525	7.3902	12.4300	11.2105
Pastry	24.1188	24.9666	31.5078	31.8091	25.5249	25.6352
Ready to eat	6.6522	6.7848	10.7582	10.9926	6.9606	7.2416
Fruits	57.0292	57.7696	72.7725	73.7587	29.1726	33.8377
Soda large	28.9353	29.1103	20.2763	22.4496	29.3229	30.4499
Soda small	17.6294	18.3285	44.3216	48.0184	19.6086	25.6532

Table 8 Root Mean Squared Error for The Different Models

5.3. Interaction Models

The two interaction models we constructed are, one where temperature interacts with group of goods and one where sunshine hours interact with group of goods. The results from the regressions are presented below.

	Estimate	Std.Error	Eta.sq	p.value	Lower	Upper
(Intercept)	3.5292	0.0266	0.8020	0.0000	3.4771	3.5814
beer	-1.0170	0.0337	0.1788	0.0000	-1.0832	-0.9508
fruits	1.9248	0.0175	0.6344	0.0000	1.8906	1.9590
grill	0.0005	0.0311	0.0000	0.9876	-0.0604	0.0614
ice cream	-0.4216	0.0455	0.0361	0.0000	-0.5108	-0.3324
odd drinks	0.0569	0.0255	0.0007	0.0260	0.0068	0.1070
pastry	0.9906	0.0287	0.1712	0.0000	0.9343	1.0470
ready to eat	-0.7732	0.0606	0.1118	0.0000	-0.8920	-0.6544
soda large	0.7810	0.0364	0.1138	0.0000	0.7097	0.8523
soda small	1.0688	0.0337	0.1939	0.0000	1.0028	1.1348
events	-0.2916	0.0424	0.0392	0.0000	-0.3747	-0.2085
summer	0.0425	0.0297	0.0006	0.1526	-0.0157	0.1008
spring	0.1200	0.0192	0.0114	0.0000	0.0823	0.1576
fall	0.1022	0.0205	0.0070	0.0000	0.0620	0.1423
weekend	0.1252	0.0124	0.0288	0.0000	0.1009	0.1496
wind speed	-0.0086	0.0041	0.0012	0.0349	-0.0166	-0.0006
sunshine hours	0.0108	0.0016	0.0096	0.0000	0.0077	0.0139
temperature	-0.0056	0.0019	0.0019	0.0025	-0.0093	-0.0020
beer:temperature	0.0073	0.0031	0.0017	0.0202	0.0011	0.0135
grill:temperature	0.0065	0.0022	0.0013	0.0038	0.0021	0.0109
ice cream:temperature	0.0599	0.0035	0.1010	0.0000	0.0530	0.0667
odd drinks:temperature	0.0097	0.0019	0.0030	0.0000	0.0059	0.0136
pastry:temperature	0.0048	0.0022	0.0007	0.0292	0.0005	0.0092
ready to eat:temperature	0.0284	0.0045	0.0247	0.0000	0.0196	0.0373
soda large:temperature	0.0102	0.0027	0.0033	0.0001	0.0049	0.0155
soda small:temperature	0.0275	0.0026	0.0231	0.0000	0.0225	0.0325

Table 9 Interaction Between Goods and Temperature

	Estimate	Std.Error	Eta.sq	p.value	Lower	Upper
(Intercept)	3.4529	0.0250	0.8196	0.0000	3.4038	3.5020
beer	-0.9885	0.0320	0.1952	0.0000	-1.0513	-0.9257
fruits	1.9248	0.0185	0.6204	0.0000	1.8885	1.9611
grill	0.0594	0.0216	0.0016	0.0061	0.0170	0.1018
ice cream	-0.2386	0.0387	0.0139	0.0000	-0.3146	-0.1627
odd drinks	0.1452	0.0176	0.0092	0.0000	0.1107	0.1797
pastry	1.0343	0.0205	0.3206	0.0000	0.9942	1.0744
ready to eat	-0.6317	0.0512	0.0901	0.0000	-0.7321	-0.5313
soda large	0.8733	0.0225	0.2518	0.0000	0.8293	0.9174
soda small	1.1956	0.0283	0.2619	0.0000	1.1402	1.2510
events	-0.2916	0.0428	0.0370	0.0000	-0.3756	-0.2077
summer	0.0425	0.0301	0.0006	0.1577	-0.0165	0.1015
spring	0.1200	0.0197	0.0108	0.0000	0.0813	0.1586
fall	0.1022	0.0208	0.0066	0.0000	0.0614	0.1429
weekend	0.1252	0.0127	0.0272	0.0000	0.1004	0.1501
wind speed	-0.0086	0.0042	0.0011	0.0410	-0.0168	-0.0004
sunshine hours	-0.0041	0.0019	0.0009	0.0290	-0.0078	-0.0004
temperature	0.0098	0.0015	0.0110	0.0000	0.0068	0.0128
beer:sunshine hours	0.0088	0.0054	0.0008	0.1001	-0.0017	0.0194
ice cream:sunshine hours	0.0843	0.0059	0.0682	0.0000	0.0728	0.0959
ready to eat:sunshine hours	0.0272	0.0078	0.0076	0.0005	0.0120	0.0424
soda small:sunshine hours	0.0287	0.0038	0.0084	0.0000	0.0211	0.0362

Table 10 Interaction Between Goods and Sunshine Hours

The adjusted R-squared is 0.8712 and 0.8634, thus, the covariates can explain 87% and 86% of the variance respectively. In the first interaction model every interaction variable displays a p-value that is lower than 0.03 and some covariates have a significantly high partial eta squared. This together with one sided confidence intervals tends to suggest that it is a reasonable model. The same applies for the second model with the exception of the interaction between beer and sunshine hours. This interaction displays a high p-value of 0.1 and a two sided confidence interval, hence, no conclusions can be drawn from this interaction variable.

As illustrated below, the Variance Inflation Factor in both interaction models are below ten. Thus, multicollinearity is most likely not present in any of the interaction models.

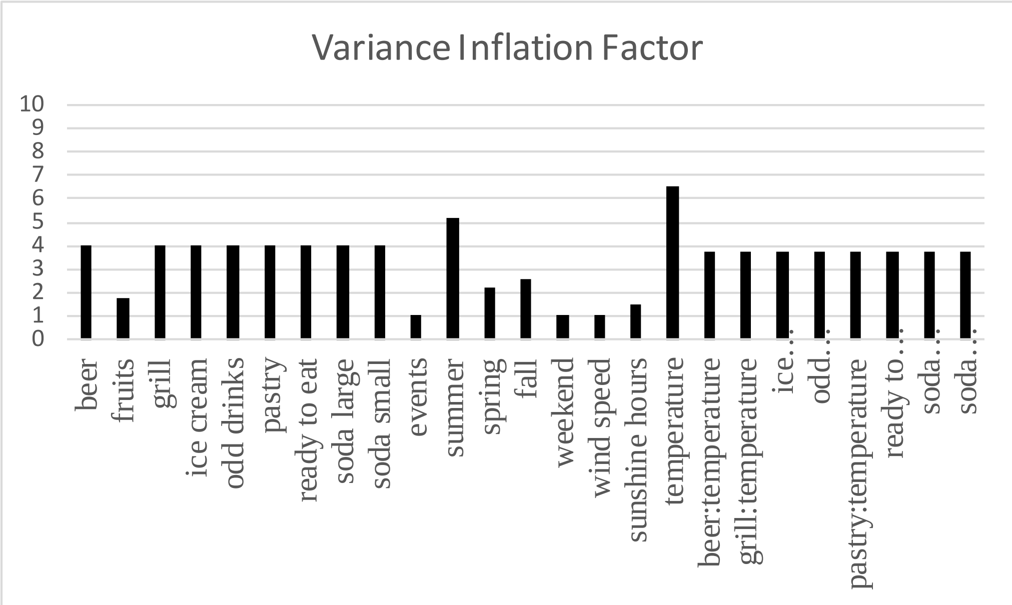

Table 11 Variance Inflation Factor, Interaction with Temperature

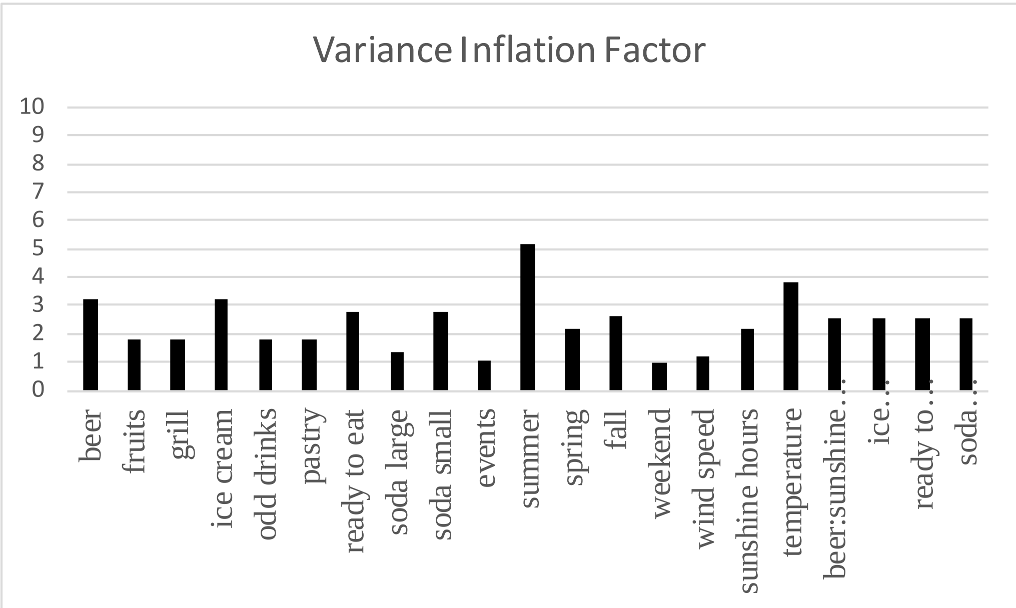

Table 12 Variance Inflation Factor, Interaction with Sunshine Hours

6. Discussion

6.1. Interpretation of Results

Since we made a logarithmic transformation in the model, the covariates can be transformed into percentages to easier interpret the results. The coefficients are transformed by

$$e^{\widehat{\beta}} - 1.$$

We are limited to answer as a percentage change per unit increase. However, such an answer is sufficient to draw the conclusion that ICA's sales suffer due to the forecast engine not taking weather into consideration.

6.1.1. Interpretation of Final Model

The model is constructed with *control group* as benchmark, thus, the intercept for each group of goods describes the difference in sales between the respective groups compared to the control group. Since each group consists of different amount of goods it is natural that the intercepts vary between positive and negative values. The estimates for those groups are not of interest, because they give no information regarding how consumption depends on weather. The results are presented in Table 13.

	Estimate	**Std.Error**	**Eta.sq**	**p.value**	**Lower**	**Upper**
(Intercept)	2896.9888%	0.0249	0.8216	0.0000	3.3514	3.4490
beer	-61.3612%	0.0246	0.2699	0.0000	-0.9992	-0.9026
easytoeat	585.3710%	0.0192	0.6024	0.0000	1.8871	1.9625
grill	6.1174%	0.0222	0.0014	0.0076	0.0158	0.1029
ice_cream	12.8126%	0.0306	0.0059	0.0001	0.0606	0.1805
odd_drinks	15.6236%	0.0182	0.0085	0.0000	0.1094	0.1809
pastry	181.3141%	0.0210	0.3043	0.0000	0.9930	1.0756
ready_to_eat	-40.2990%	0.0309	0.0981	0.0000	-0.5765	-0.4552
soda_large	139.4891%	0.0230	0.2377	0.0000	0.8283	0.9184
soda_small	273.4633%	0.0197	0.4152	0.0000	1.2790	1.3563
events	-25.4575%	0.0435	0.0350	0.0000	-0.3790	-0.2086
spring	10.6601%	0.0140	0.0129	0.0000	0.0738	0.1288
fall	8.3297%	0.0133	0.0081	0.0000	0.0538	0.1062
weekend	13.4137%	0.0132	0.0255	0.0000	0.1001	0.1517
windspeed	-1.0210%	0.0043	0.0016	0.0165	-0.0186	-0.0019
sunshine_hours	1.0972%	0.0017	0.0086	0.0000	0.0075	0.0143
temperature	1.1615%	0.0010	0.0349	0.0000	0.0096	0.0135

Table 13 Transformed Coefficients

Observing the variables describing weather, we can state that the results are in line with what we expected regarding weather's effect on consumption. The intercept for wind speed is negative and is interpreted as a decrease in consumption when wind speed increases. From *Table 13* we can observe that the intercept is approximately negative 1.02%. The opposite interpretation can be done for sunshine hours and temperature, as temperature and sunshine hours increase; the consumption of goods increases respectively.

The covariates not describing weather are irrelevant to our research question and aim of the thesis. However, they can be used to observe how well the model describes reality. Studying the intercept "weekend" we see that the consumption increase with more than 13%. This is consistent with reality since we know from experience and through communication with ICA, that consumption is higher during weekends. This is an indicator for how good our model is, but we can in no way state that it is definitely correct. There is a risk factor that is difficult to evaluate, due to the numerous error sources. However, it can be used in addition with the statistical tools described in the report to give an indication of how well the model describes reality. The fact that it is consistent with reality is of course a positive sign.

6.1.2. Discussion Regarding Prediction accuracy

By comparing the prediction accuracy between the final model and the simple model, we conclude that the final model is generally a better model to use when predicting the consumption forecast (see *Prediction Accuracy*). In average the final model predicts better than the simple model 60% of the time. The highest win percentage being 67% during week 30 and the lowest win percentage being 54% during week 45. The large difference in win percentage between the different weeks cannot be explained by the model. However, it is likely that weather affects the consumption of goods differently between seasons. Week 30 is in the middle of the summer while week 45 is during fall, therefore, it is likely that an increase in temperature or sunshine hours would have a bigger impact during week 30 than week 45. This is in line with our belief regarding how weather impact sales during the summer months. By the same assumption a decrease in wind speed would likely have a bigger impact during week 30 than week 45. Furthermore, one can observe that the root-mean-square error is lower for almost every covariate in the final model compared to the simple model, for each predicted week. Thus, one can conclude that the advanced model minimizes the prediction errors. This in conjunction with a 60% win ratio implies that the final model is a better predictor than the simple model.

The average error in prediction amounted to 28% for the final model and 30% for the simple model, which might sound high. However, it is essential to understand that our model is **not** constructed to be used as a forecast engine; the comparison between the models is merely a guideline. The simple model looks somewhat like ICA's current FE and the final model looks like their current FE but with parameters that take weather into account. Since the final model is a better predictor than the simple it adds additional credibility to arguments for developing a FE taking weather into account.

The error might depend on our model missing relevant covariates such as *campaigns*. However, this is out of our control due to lack of data. One might assume that the model would get better predictions if more data was accessible.

6.1.3. Interpretation of Interaction Models

By the results presented below, one can see that for every unit increase in temperature, the amount of sold goods of each group rises relative to the control group. While some groups, e.g. pastry only increases by 0.48% other groups are significantly more correlated with temperature, as expected. From *Table 14* we can observe that beer, grill, ice cream, odd drinks, pastry, ready to eat, soda large and soda small are significantly affected by an increase in temperature.

	Estimate	Std.Error	Eta.sq	p.value	Lower	Upper
(Intercept)	3309.7468%	0.0266	0.8020	0.0000	3.4771	3.5814
beer	-63.8317%	0.0337	0.1788	0.0000	-1.0832	-0.9508
fruits	585.3710%	0.0175	0.6344	0.0000	1.8906	1.9590
grill	0.0485%	0.0311	0.0000	0.9876	-0.0604	0.0614
ice cream	-34.4030%	0.0455	0.0361	0.0000	-0.5108	-0.3324
odd drinks	5.8544%	0.0255	0.0007	0.0260	0.0068	0.1070
pastry	169.2978%	0.0287	0.1712	0.0000	0.9343	1.0470
ready to eat	-53.8478%	0.0606	0.1118	0.0000	-0.8920	-0.6544
soda large	118.3638%	0.0364	0.1138	0.0000	0.7097	0.8523
soda small	191.1891%	0.0337	0.1939	0.0000	1.0028	1.1348
events	-25.2944%	0.0424	0.0392	0.0000	-0.3747	-0.2085
summer	4.3453%	0.0297	0.0006	0.1526	-0.0157	0.1008
spring	12.7462%	0.0192	0.0114	0.0000	0.0823	0.1576
fall	10.7553%	0.0205	0.0070	0.0000	0.0620	0.1423
weekend	13.3429%	0.0124	0.0288	0.0000	0.1009	0.1496
wind speed	-0.8557%	0.0041	0.0012	0.0349	-0.0166	-0.0006
sunshine hours	1.0854%	0.0016	0.0096	0.0000	0.0077	0.0139
temperature	-0.5614%	0.0019	0.0019	0.0025	-0.0093	-0.0020
beer:temperature	0.7324%	0.0031	0.0017	0.0202	0.0011	0.0135
grill:temperature	0.6525%	0.0022	0.0013	0.0038	0.0021	0.0109
ice cream:temperature	6.1708%	0.0035	0.1010	0.0000	0.0530	0.0667
odd drinks:temperature	0.9797%	0.0019	0.0030	0.0000	0.0059	0.0136
pastry:temperature	0.4833%	0.0022	0.0007	0.0292	0.0005	0.0092
ready to eat:temperature	2.8835%	0.0045	0.0247	0.0000	0.0196	0.0373
soda large:temperature	1.0251%	0.0027	0.0033	0.0001	0.0049	0.0155
soda small:temperature	2.7863%	0.0026	0.0231	0.0000	0.0225	0.0325

Table 14 Interaction with temperature

The same interpretation is made by the interaction model with sunshine, for every extra sunshine hour the amount of sold goods of each group rises, relative to the control group. Once again, some groups are significantly affected by weather, in particular ice cream, with a rise of 8.8% for every additional sunshine hour, relative to the control group. This is in line with our beliefs, since weather tends to be "better" as the amount of sunshine hours increases.

	Estimate	Std.Error	Eta.sq	p.value	Lower	Upper
(Intercept)	3059.2595%	0.0250	0.8196	0.0000	3.4038	3.5020
beer	-62.7873%	0.0320	0.1952	0.0000	-1.0513	-0.9257
fruits	585.3710%	0.0185	0.6204	0.0000	1.8885	1.9611
grill	6.1174%	0.0216	0.0016	0.0061	0.0170	0.1018
ice cream	-21.2284%	0.0387	0.0139	0.0000	-0.3146	-0.1627
odd drinks	15.6236%	0.0176	0.0092	0.0000	0.1107	0.1797
pastry	181.3141%	0.0205	0.3206	0.0000	0.9942	1.0744
ready to eat	-46.8311%	0.0512	0.0901	0.0000	-0.7321	-0.5313
soda large	139.4891%	0.0225	0.2518	0.0000	0.8293	0.9174
soda small	230.5610%	0.0283	0.2619	0.0000	1.1402	1.2510
events	-25.2944%	0.0428	0.0370	0.0000	-0.3756	-0.2077
summer	4.3453%	0.0301	0.0006	0.1577	-0.0165	0.1015
spring	12.7462%	0.0197	0.0108	0.0000	0.0813	0.1586
fall	10.7553%	0.0208	0.0066	0.0000	0.0614	0.1429
weekend	13.3429%	0.0127	0.0272	0.0000	0.1004	0.1501
wind speed	-0.8557%	0.0042	0.0011	0.0410	-0.0168	-0.0004
sunshine hours	-0.4098%	0.0019	0.0009	0.0290	-0.0078	-0.0004
temperature	0.9854%	0.0015	0.0110	0.0000	0.0068	0.0128
beer:sunshine hours	0.8869%	0.0054	0.0008	0.1001	-0.0017	0.0194
ice cream:sunshine hours	8.7990%	0.0059	0.0682	0.0000	0.0728	0.0959
ready to eat:sunshine hours	2.7580%	0.0078	0.0076	0.0005	0.0120	0.0424
soda small:sunshine hours	2.9066%	0.0038	0.0084	0.0000	0.0211	0.0362

Table 15 Interaction with Sunshine Hours

From the results it is clear that weather does affect the consumption of goods at ICA Nära. However, the results from our regression analysis do not describe what time during the year when temperature and sunshine hours affect consumption the most. This is due to the fact that it is most likely not a linear combination and we have performed an analysis with a linear model. It is not a linear combination since a change in temperature and sunshine hours probably impact consumption more at higher temperatures.

The conclusion that weather affects sales is drawn under the assumption that the general definition of “good” weather is low wind speed, high temperature and a high amount of sunshine hours. There are several error sources one needs to take into account when interpreting these results. However, it corresponds well with how weather and consumption of goods intuitively correlate, indicating that our model is fairly accurate. Making a consumption forecast without taking weather into consideration will therefore have a negative effect. E.g. making a consumption forecast based on seven days of cold and rainy weather will probably underestimate the actual demand of certain weather dependent goods, if the weather conditions change.

6.2. Data Error

There are several aspects to consider regarding our data. Let us part the weather, and sales data, respectively.

6.2.1. Error in Sales Data

ICA’s sales data is affected by many factors, campaigns and depleted stocks are most likely the ones that have largest impact on sales. Campaigns affect sale numbers in a positive way. Allowing the stock of a certain product be depleted on the other hand, affects sales negatively. This is a fortunate phenomenon, that one affects positively and one negatively, since they, in best case scenario, can cancel each other out if both events occur in the same group of articles. That is quite likely to occur, since it is hard to predict consumption during campaigns and the risk of stocks being depleted increase.

There are also three other factors that affects sales, thus, making forecasting consumption more complicated. Namely, products whose barcodes are poor, i.e. the cashier cannot register the product. A product may not be registered in the system i.e. the cashier must add the price manually, and lastly, returns. In the scenario where the cashier is unable to register the product, it is routine to add the cost to the receipt, this implies that not all sold articles are registered. This is something that is not corrected. Returns are naturally corrected since we have one year of sales data and returns during this time is accounted for simply by including both the day the item is bought and the day it is returned. We can assume a negligible small error from returns.

6.2.2. Error in Weather Data

SMHI performs a quality control of registered weather data, somewhat irregularly depending on what type of weather data. Therefore, some of the collected data have been controlled and received a quality marking. In our collected data there are 31.5%

data points that remains to be examined. However, they also state that these values have been roughly inspected. Implying that even those "not controlled" are quite reliable.

6.3. Weather Complications

Weather entails certain difficulties, firstly because it is hard to predict well and secondly because the effect on sales is almost certainly non-linear. With a non-linear relationship between sales and weather it becomes harder to predict the effect. Therefore, to examine this further, one should use a non-linear model, in contrast to what is used in this thesis.

6.3.1. Weathers Effect on ICA

There exist some problems regarding how weather affect ICAs sales since as mentioned above, it is most likely not a linear relationship between sales and weather. The results indicate that weather affects sales more during the summer months. This can be compensated for when forecasting by using sales history from the same time periods. The fact that weather forecast are uncertain is something a FE cannot correct perfectly. However, it can give the prediction weighted values which would improve the FE's performance.

6.3.2. Inaccuracy of Weather Forecast

From the information in *Accuracy of Weather Forecasts* we can deduct that it is difficult to say how well SMHI's forecast performs. One study indicates a 45% success rate while the other indicates roughly 84%. The gap in success rate is likely explained by the two studies having different criterions for classifying a forecast as correct. The fact that the weather forecasts accuracy is difficult to measure is a risk factor one needs to take into consideration when evaluating whether or not to implement these weather forecasts when creating a consumption FE.

7. Supply Chain Discussion

Changes in SCs is a common occurrence, how fast, and well, the SC adapts to changes is a measure of its adaptability and how flexible it is.

7.1. Supply Chain for Grocery Store Companies

A generic grocery store company's SC can be simplified as in ***Figure 8***. In reality there are many agents involved and different distribution channels exist in the SC. But the figure describes their SC sufficiently. The central warehouses are large storage facilities employed by the companies in the industry that service the grocery stores by delivering goods to them.

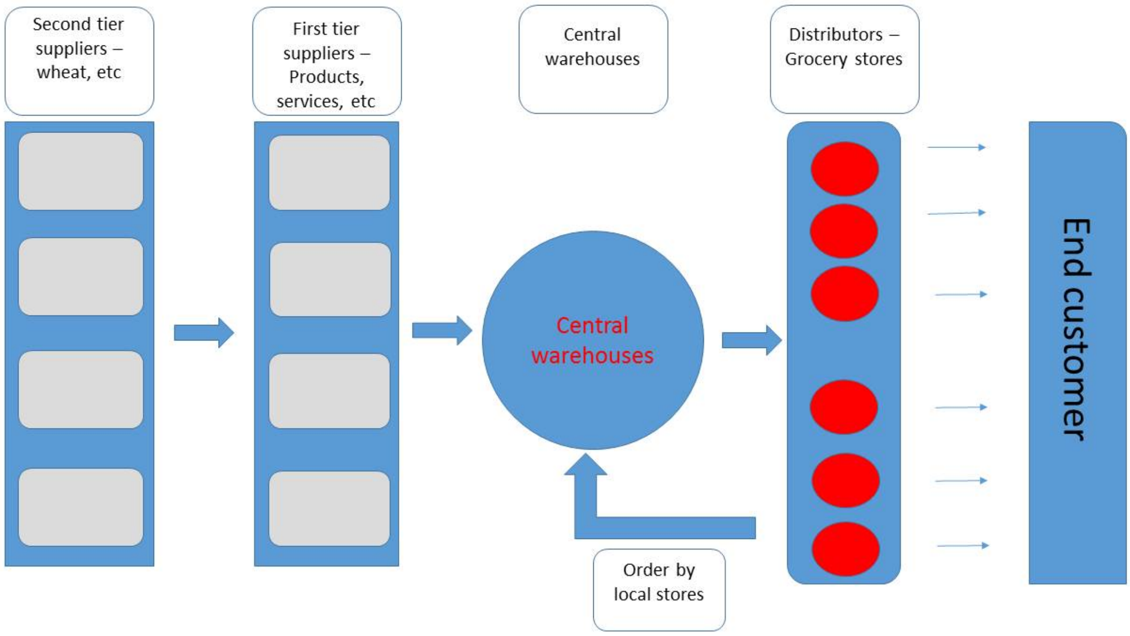

Figure 8, Simplification of General SCs for a Grocery Store Company

7.2. Effects on Supply Chain

How the SC is affected from this implementation is quite difficult to grasp. On one hand, it will experience an improved forecast and the positive effects that entails, and on the other hand it will be important to have short lead times regarding deliveries. This section mainly concerns the first part, the effects of improved consumption forecasts.

7.2.1. Effect on inventory

The results presented in this thesis indicate that the consumption forecast will improve by employing weather forecasts in the FE. Companies usually try to improve their forecast accuracy to reduce the need for inventory, however, they often fail to acknowledge the three rules of forecasting. The first and most important rule is that the future is never certain and therefore forecasts are always wrong, however, it can be more or less wrong. To manage this flaw in the forecast, a company deciding to use weather forecasts in their FE also needs to predict the confidence of the forecast. If the forecast is likely to over- or underestimate the demand with a high percentage the supply chain will suffer. This because the inventory will either be full due to an over estimation, or the inventory will be underused due to an under estimation, all of this leads to poor stock control. Therefore, it is essential to predict the confidence in the forecast and keep the confidence in a reasonable interval, this interval may vary depending on grocery stores size and location. If the company is able to keep the estimation within the given interval, the supply chain can absorb the error in prediction by simply adding safety stock, i.e. holding inventory in case of an under estimation in demand. In the opposite situation, if the estimation is higher than the actual demand one might add extra capacity to the inventory (Eades, o.a., 2010).

7.2.2. Purchasing

The new FE will also become a question of “more versus less”. Larger batches require a larger amount of inventory to be held, but the grocery store may take advantage of *bulk pricing,* i.e. buying larger quantities for lower unit prices than usual, this advantage might be lost when receiving deliveries more frequently. This may also increase the transportation costs. However, implementing the change in FEs will reduce the inventory because of the daily deliveries and smaller batch sizes. It will also reduce the need of safety stock since forecasts will predict shorter ahead in time, thus, the forecasts are likely to be more accurate. A more accurate forecast will likely reduce the *bullwhip effect* because of the reduced lead times, and it will lead to the supply chain not having to absorb as much uncertainty. Furthermore, receiving deliveries on a daily basis may also lead to less stock being thrown away due to expiring products.

7.2.3. Bullwhip Effect

Through contact with bigger grocery stores in Sweden it was revealed that the person in charge of placing an order had to take weather into account by manually altering orders, based on experience and past observations. This poses a problem because people appear to be overconfident in their own decisions and opinions as they make a forecast based on experience and past observations (Francesca & Gary, 2008). Thus, manual orders leads to errors that amplify the bullwhip effect. Why it is common to overestimate when ordering may be due to the fact that even when removing the profit argument, there is still the aspect concerning service to customers. Depleting the stock of a product a customer desires is considered poor service and affects the business image.

There will always be a demand variation apparent along the supply chain due to external factors, in this case weather. However, demand variation does not always have to originate from external factors (Taylor, 1999). According to (Taylor, 1999), demand variability may be initiated by an external factor but it is amplified by decision-makers tendency to over-react to the change in demand i.e. ordering too much. Following these arguments we assume that grocery stores in general will overestimate demand and therefore contribute to the bullwhip effect. Implementing a new FE taking weather into account would be beneficial for grocery stores, as decision-makers would no longer have to manually adjust for demand variation due to weather. The human error would be non-existent and over estimations would be less common. Participants within the SC would have to carry less safety stock to adjust for demand fluctuations, thus, implementing this new FE would reduce the *bullwhip effect* in the grocery stores supply chain.

7.3. Supply Chain Requirements

We will discuss what is required by a company's SC too be flexible and be able to implement weather forecasts as a parameter in a FE.

7.3.1. Importance of a Flexible Supply Chain

According to (Fisher, 1997) a flexible SC is mainly needed when dealing with innovative products, but we claim that it is required to be innovative in the grocery store industry. This is supported by (Kervenoael, Elms, & Hallsworth, 2014) who states that it is crucial also in this industry, to be innovative to attract spending. Therefore, we claim that a SC of a grocery store company should be flexible.

An inflexible SC will have ramifications when conditions change, due to its inability to adapt with changes like implementing weather forecasts as a parameter in a FE. According to the third rule of forecasting, we can predict near-term demand better than future demand (Eades, o.a., 2010). This rule also applies to the weather forecasts, the weather forecast accuracy decreases the further ahead in time they are forecasting. Therefore, it is crucial for a company implementing weather in their FE to predict only one or two days ahead in time to be able to use as reliable weather forecasts as possible. Predicting the consumption for only one or two day(s) ahead puts pressure on the company's supply chain. The company will have to design their supply chain with faster responses, and be able to deliver within one to two days after ordering. To change the SC in the proposed way cost effectively, the SC needs to be flexible.

7.3.2. Location

The location of the grocery stores may also have an impact on the implementation of this new FE. As mentioned, it may be necessary to demand shorter delivery times from ones suppliers, therefore, the SC may have to change. A grocery store located in, or close to bigger cities is likely to have easier access to suppliers whilst a store located far from cities is likely to be "locked" with a few suppliers. Therefore, the location of a store also impacts the supply chain flexibility, and in turn makes the implementation more or less possible. This is something the grocery store companies must take into consideration on numerous occasions. E.g. when choosing suppliers, distribution channels, etc.

A remedy for this problem is to use more central warehouses. Companies in the industry utilize larger storing facilities referred to as central warehouses, which service the grocery stores. Employing more of these will shorten the overall delivery time.

7.3.3. Decision making

For most companies, the decision to implement said change is an operational decision, since it concerns the management of an activity. The activity being forecasting consumption. This means that in most organizations, the decision can be made by middle- to lower-level management (Game Theory Lab, 1995). Because of this, the threshold to cross is not too great, and for companies in the industry it is not a major investment, implying that it is relatively easy to start the implementation process. However, it affects the entire SC, i.e. other parts of the company and its suppliers, this is important to take into consideration since it potentially could be costly and have a large impact on the company.

Decision making and implementation of changes to ones SC is highly dependent on the structure of the company. A flexible SC is considered to be complex and dynamic (Merschmann & Thonemann, 2010), however, it may also create friction between organizations in the SC when it comes to making these types of decisions. In this scenario, it is easier if one company controls the majority of the SC, parts like transportation, extracting raw materials, etc. However, it is easy to substitute entities in a complex and dynamic SC and the change itself is not as expensive since change comes naturally in a SC that is flexible, i.e. complex and dynamic. Therefore, it is preferable to have a flexible SC even though decisions may be exposed to different wills.

7.3.4. Dependency of Suppliers

Depending on the grocery store company, the change in SC might be difficult to implement, due to factors like product range and location. Such an implementation in the FE is likely to be easier to complete for a grocery store with a large range of products that are supplied by many different companies or organizations. The big product range and variety in suppliers make their supply chain more flexible in one aspect, because they can easily change their suppliers if some are unable to deliver products at required frequency. It follows that, grocery stores with a more niched range, that are dependent on their suppliers, may have a less flexible SC. If one or more of their suppliers are unable to deliver on a daily basis it is harder for them to switch suppliers because of the constellation of their SC. Thus, making it difficult for them to implement weather in their forecast engine.

8. Conclusions

Weather does affect the sales at *ICA Nära* and the results are well transferable to other grocery stores of different sizes, even though there of course will be smaller differences depending on location, size, etc. The statistical model with weather outperformed the simple model in average 60 % of the time during the predicted weeks. It performed even better when predicting during the summer, as expected. Remember that the results need to be interpreted with care.

We are unable to state precisely how much weather affects sales but we do estimate the effect, this was stated as our goal in the introduction. There is a clear pattern and it is possible to improve the forecast. To implement this change, one needs data on the sale pattern for more than one year.. The uncertainty contributed by weather forecasts however is difficult to improve. This is a factor one needs to take into consideration when evaluating how to weigh the effect from weather on sales.

Implementing weather forecasts as a parameter in a FE will affect grocery store companies SCs in several ways. Perhaps most significantly with a positive effect on inventory levels, since according to our research the accuracy of the consumption forecast will improve. Also, by adding significance to managing lead times. It puts pressure on lead times because being able to receive deliveries rapidly after ordering allows for utilizing more accurate weather forecasts, since the weather forecast performs better the shorter it predicts in time. Therefore, the pressure on lead time between order and delivery rises.

From our results we can anticipate that including weather forecasts as a parameter in FEs will contribute to a better stock control, this seems intuitively accurate. All the positive effects from better stock control mentioned in Stock can possibly be harvested by implementing the change. This implies that one's SC should be flexible for a smooth implementation of the new FE.

9. References

Works Cited

Benoit, K. (2011). *Linear Regression Models with Logarithmic Transformations.*

Brauner, P., Runge, S., Groten, M., Shchuh, G., & Ziefle, M. (2013). *Human Factors in Supply Chain Management.* German Research Foundation.

Bray, R. L., & Mendelson, H. (2012). Information transmission and the bullwhip effect: An empirical investigation. *Management Science*.

Choen, J. (1973). ETA-SQUARED AND PARTIAL ETA-SQUARED IN. *Educational and Psychological Measurement*.

Das, S. K., & Nagendra, P. (1997). Selection of routes in a flexible manufacturing facility. *International Journal of Production Economics*, 237-247.

DN. (2011). Retrieved from DN: http://www.dn.se/nyheter/sverige/smhi-bast-pa-ratt-vader/

Dolgui, A., & Proth, J.-M. (2010). *SUpply Chain Engineering.* Springer. Retrieved from http://catalogimages.wiley.com/images/db/pdf/R0471235172.01.pdf

Duclos, L. K., Vokurka, R. J., & Lummus, R. R. (2003). A conceptual model of supply chain flexibility. *Industrial Management & Data Systems, Vol 103 Iss: 6*, 446-456.

Eades, K. M., Laseter, T. M., Skurnik, I., Rodriguez, P. L., Isabella, L. A., & Simko, P. J. (2010). *The Portable MBA, 5th Edition.*

Fisher, M. (1997). What is the right supply chain for your product? *Harvard Business Review, 75* , 105-116.

Francesca, G., & Gary, P. (2008). Toward a Theory of Behavioral Operations. *Manufacturing and Service Operations Management*.

Game Theory Lab. (1995). Retrieved from http://lcm.csa.iisc.ernet.in/scm/supply_chain_intro.html

Gerwin, D. (1987). An Agenda For Research on the Flexibility of Manufacturing Processes. *International Journal of Operations & Production Management Vol. 7*.

Gujarati, D. N. (2004). *Basic Econometrics, 4th edition.*

Haase, R. F. (1983). Classical and Partial Eta Square in Multifactor ANOVA Designs. *Educational and Psychological Measurment*.

Hansen, B. E. (2015). *Econometrics.*

Harrison, T. P., & Ganeshan, R. (1995). An Introduction to Supply Chain Management.

Hayashi, F. (2000). *Econometrics.* Princeton University Press.

Herlyn, W. (2014). *The Bullwhip Effect in expanded Supply Chains and the Concept of Cmulative Quantities.*

ICA Gruppen. (2016). *About Ica Gruppen*. Retrieved 04 18, 2016, from Icagruppen: http://www.icagruppen.se/en/about-ica-gruppen/#!/

Kennedy, J. (1970). The eta coefficient in complex ANOVa designs. *Educational and Psychological Measurment*.

Kennedy, P. (2008). *A Guide to Econometrics.* Wiley-Blackwell.

Kervenoael, R. d., Elms, J., & Hallsworth, A. (2014). Influencing online grocery innovation: Anti-choice as a trigger for activity fragmentation and multi-tasking. *Futures, volume 62, Part B*, 155-163.

Lang, H. (2015). *Elements of Regression Analysis.*

Lehmann, E., & Casella, G. (1998). *Theory of Point Estimation, Second Edition.*

Merschmann, U., & Thonemann, U. W. (2010). Supply chain flexibility, uncertainty and firm performance: An empirical analysis of German manufacturing firms. *International Journal of Production Economics, Volume 130, Issue 1,* 43-53.

Pettersson, A. I., & Segerstedt, A. (2013). Measuring supply chain cost. *International Journal of Production Economics*, 357-363.

Pujawan, I. (2004). Assessing supply chain flexibility: a conceptual framework and case study. *International Journal of Integrated Supply Management, 1,* 79-97.

Sànchez, A. M., & Pérez Pérez, M. (2005). Supply chain flexibility and firm performance: A conceptual model and empirical study in the automotive industry. *International Journal of Operations & Production Management, Vol 25 Iss: 7,* 681-700.

Seebacher, G., & Winkler, H. (2016). A Citation Analysis of the Research on Manufacturing and Supply Chain Flexibility. *Interantional Journal of Production Research*.

Swedish Meteorological and Hydrological Institute. (2016, 04 20). *http://www.smhi.se/en*. Retrieved 04 20, 2016, from http://www.smhi.se/en: http://www.smhi.se/en

Taylor, D. (1999). Measurement and analysis of demand amplifications across the supply chain. *International Journal of Logistics Management*.

Threadpunter. (2015, October 16). Retrieved from http://www.threadpunter.com/supply-chain-management/basic-supply-chain/

Timothy R. Levine, C. R. (2002). Eta Squared, Partial Eta Squared and Misreporting of Effect Size in Communication Research. *Human Communication Research, Vol.28*.

Waters, D. (2003). *An Introduction to Supply Chain Management.* PALGRAVE MACMILLAN.

What is logistics and supply chain management: Supply Chain Opz. (2016, may). Retrieved from Supply Chain Opz: http://www.supplychainopz.com/2012/04/what-is-logistics-and-supply-chain-management.html

10. Appendix

10.1. Goods

Ciabatta Fitness	Re-Create Styl Wax	Fläder/lime
Trehör grov/rågr	Ris långkorn Boil	Blandsaft Jordgubb
Nötfärs finm 12%	Hasselnötsdryck	Fruit Müsli
Fläskfilé putsad	Strösocker	Tårtljus mix färg
Gouda 30%	Mellangrädde 27%	Spearmint påse
Tvål Cream Wash	Tonfisk	Multivitamin drink
Nappar Sour	Bar crunch	Leverpastej ugnsb
Majsväll dr.f 6m		

Figur 1 Control group

Apelsin	Päron Conference	Jordgubbar
Clementin	Melon Vatten kfr	Melon Galia
Citron	Avocado	Ananas
Äpple Granny Smith	Grapefrukt röd	ICA Banan Eko
Äpple Royal Gala	Melon Honung	I Love Eco Apelsin
Melon Cantaloupe		

Figur 2 Easy to eat fruit

Bague. Skag. m. räko	Fralla ost & skinka	Clubsandwich
Baguette Kyckling	Sandwich räk m ägg	Pastasall baco kyck
Bague ost & skink	Wraps kyck swe ch	Wrap Kyckling Bacon
Ciabat kyck salla	Pastasall fet & moz	Sandwich m. pesto
Grönsallad Tonfisk	Kyckling Curry Wrap	Fralla Ost
Pastasa. kyck mang	Frökusfralla ost sk	Grönsallad kyckling
Pastasallad ost/ski	Seafood Sallad	Baguette Texas Kyck
Pastasa kyck feta		

Figur 3 salads and sandwiches

drickkvarg	pucko	juice
Drickyogh jordgubb	Juice Apelsin	Ape.juice EKO KRAV
Mandarin Morning	Lemonad	yalla! Ban/Mang/Ap
Dr.kvarg apel/van	Smoothie Anan/Mang	yalla! Jord/Lime
Dr.kvarg jord/gr.ä	Juice Morot	Apelsinjuice
Pucko Original 1	Morot	Äppeljuice
Frukt smoot man/ap	Lemonad	Apelsin
Apelsin jucie	Blåbär/Svartvinbär	Apelsin/Jordgubb
Smoothie ap/mo/in	Passionad	Hallondryck

Figur 4 odd drinks

Sprite Zero ÅP	Spendrups Bright	Vatten Mango ÅP
Funky Orange ÅP	Crush Päron	Vatten Mango ÅP
Coca Cola Zero ÅP	Crush Hallon	Vatten Smultron ÅP
Cola Zero Läsk ÅP	Vatten Hallon/Lakr	Vatten Citrus ÅP
Cola Läsk ÅP	Vatten Persika	Trocadero
Cola Light ÅP	Vatten Rosa Skumsv	Lemon Fusion ÅP
Pepsi Max	Cider Päron ÅP	Russchian ÅP
Free Lemon/lime	Sockerdricka	Tonic Water ÅP
Cola regular	Päronsoda	Dr Pepper ÅP
Vatten Granatäp ÅP		

Figur 5 large bottles of water and sodas

Zero Calories	Orginal Min.vatten	Dryck Defence ÅP
Sprite Läsk	Vatten Hall/Björnb	Dryck Focus ÅP
Classic Orange	lime/Pin min kolsy	Dryck Reload ÅP
Cactus/Lime	Vatten Citron ÅP	Coca Cola ÅP
Mountain Dew	Vatten Päron-Äpple	Coca Cola Light ÅP
Sparkl Pear & Kiwi	Mineralvatten	Orange Läsk ÅP
Sparkl Melon&Lime	Stilla vatten	Sprite Läsk ÅP
Apelsinläsk ÅP	Mineralvatten	Cola Regular B
Min.vatten cit/cam	Stilla vatten	Coca Cola Zero B
Min.vatt anan/citr		

Figur 6 small bottles of water and sodas

Corona Öl 3,2%	Pripps Blå 3,5% B	Bright Öl 3,5% B
Carl Beer 3,5% B	Spendrups Bright	Norrl Guld 3,5%B
Öl Blå 2,8% B		

Figur 7 beer

Sandwich	Glass gammal krok	Magnum Vit
88:an pinne	Gräddgl Ap/Ban/Cho	Magnum Mandel
Nogger	Laktosfri Vanilj	Daimstrut
Favorit	Vaniljglass	GB Magnum yo fres
Päronsplitt	Gräddglass Polka	Magnum After Dinne
Piggelin	Sorbet Mango	Cornetto Jordg
Big Pack Tresmak	Sorbet Hallon Ekol	GB Klassiker
Glass gammal vanil	Carte d´Or Vanilla	Glass Pink
Glass gammal chokl	Twister	Glass Black
Glasstrut	Magnum Classic	Cinnamon Buns
Vanilj m kross cho	Gammaldags Vanilj	Cinnamon Buns
Glass Nogger	Vanilj m jordgubbs	Glass Chocolate

Figur 8 ice cream

TRITA -MAT-K 2016:36
ISRN -KTH/MAT/K--16/36--SE