Aalto University
School of Science

MS-E2177

Seminar on Case Studies in Operations Research

Modelling the procurement costs of subcontracted metal parts in machine manufacturing

Final report

Team:
Benoist Deroo
Eero Kere (Project manager)
Matias Linnankoski
Niklas Hantila

Client:
Normet

May 25, 2022

Contents

1 Introduction

1.1 Background

When companies define their strategy, they often conduct a SWOT analysis. The analysis basically consists of listing the Strengths and Weaknesses the company possesses on one side; and the Opportunities and Threats posed by its external environment. For a company that focuses on industrial production or assembly, threats mostly come from their dependencies. Notably, companies doing an assembly or production activity depend on their suppliers to feed their processes. As seen in the year of 2021, crises such as the metal shortages block companies from operating as usual. It thus leads to an explosion of the prices that is hard to deal with for customers. At some point, steel prices were 300% above their pre-pandemic levels [1]. But such crises can also turn into an opportunity because it can highlight other problems companies have been facing for years without knowing it, or at least without being alert enough about it.

In supply chains generally, the Supply Chain Management (SCM) has grown in importance as the cost management in supply chains has become a strategic business issue [2]. Knowing the components of costs (i.e., transportation, inventory cost, production cost etc.) can prove to be a competitive advantage for firms when dealing with their suppliers.

The metal shortage crisis has led some companies to rethink their contracts with their steel parts suppliers. Several questions have been asked: On what basis are the prices of the parts changing or if they actually have changed over time? What are the factors that make the prices fluctuate? And more importantly, how can we forecast the evolution of the prices to get more flexibility? These are questions that have been asked to us by the specific company in this study: Normet Group (called Normet from here onwards).

Normet works in the underground mining and tunnelling business. Over 1600 professionals operate over 50 locations in 33 countries worldwide to improve Normet customers' processes in terms of safety, productivity and sustainability. They have designed and produced over 14 000 underground machines that are still serviced and supported. Their service centers assembly and remanufacture machines after they have received the subcontracted metal parts from suppliers (mainly Finnish), with the main manufacturing facility and R&D center being located in Finland. [3] In 2020, Normet Group made around 305 million euros of revenue [4].

The subcontracted metal parts are priced in the market. It is now organized quarterly for most suppliers due to the rapid changing raw material costs; in normal market conditions, the prices would be typically agreed on once or twice a year. Also, unit prices for metal parts could be fixed, dependent dynamically on ordering quantity or ruled by price breaks. In the latter case, it means the prices are not proportional to ordering quantity and that they are generally fixed for different "intervals" of ordering quantity.

1.2 Motivation

Metal parts represent a big part of Normet's overall cost, which they sometimes struggle to understand because of a lack of transparency with some suppliers. These costs generally depend on different elements such as raw material, labour, electricity, order quantity etc. Furthermore, smaller suppliers buy from distributors that themselves sell big factory lots. This causes delays in how the raw material prices affect the prices of metal parts to Normet. Overall, there are two types of suppliers:

- The suppliers who belong to the "white box". Normet and they agree over the prices depending on the raw material prices. These are the most transparent type of suppliers for Normet.

- The "grey box" ones. There is not the same type of transparency but the prices follow the same dynamics as above. There is a clear reaction to raw material prices but it is not fully clear for Normet.

Normet thus face the challenge of forecasting the development of their material purchasing costs when raw material costs change. They have provided us data (notably purchase order data and steel index data) that we analysed to understand the phenomenon. When Normet understand how the costs are constituted, they can forecast their cost development better and they can also spot possible pricing errors.

1.3 Objectives and scope

The objectives of our project were determined by the project brochure and clarified with Normet in meetings. The objectives are the following:

1. Develop a model that explains the procurement cost for selected metal parts as a function of major cost elements to forecast future cost development.

2. Determine methods to identify parts which exhibit abnormal cost development, in which costs are higher or varying more than expected.

In the first objective, the model can be tailored to different types of items, as there are differences in the cost drivers across different types of items. We expect that the most important cost drivers are steel price and weight of the item. For some items, the ordering quantity also affects the unit price, either continuously or through price breaks. Besides, we expect that electricity price and labor index have an effect on the unit prices of some items, but the magnitude of these factors is expected to be smaller than that of steel price and weight. The result of the first objective is to produce a mathematical model which Normet can use for forecasting the price development.

Regarding the second objective, the methods to identify parts with abnormal cost development are determined as a side product from our data analysis and modelling process. The original objective was to simply use our cost model to detect parts with abnormal cost development. During exploratory data analysis, it turned out that there are also other potential methods, which can detect abnormal cost development in the past. Thus, the objective allows now also the development of other methods to identify those interesting parts.

At the beginning, the scope included the selected subcontracted metal parts from which Normet provided data to us, including 18 suppliers and 12 521 unique items. The scope was narrowed based on our exploratory data analysis to the suppliers and items whose unit price variation can be explained with the set of predictors we have data from. Besides, Normet's opinions were considered so that we are focusing on the suppliers which are the most interesting to them. In the scoping, we also filtered out items which have been ordered during less than ten months in our observation period of 46 months to ensure enough data points for each of the items included in our modelling. The final scope includes four suppliers and 1783 unique metal parts.

2 Literature review

In this section, we provide a literature review on the key concepts of this project. Subsection 2.1 discusses the characteristics of supply chains in the steel industry. Subsection 2.2 describes the principles of cost modelling in procurement. Subsection 2.3 introduces the dynamics of price development in the metal industry.

2.1 Steel industry and its supply chains

Since the European Industrial Revolution, during which European industries needed more and more steel, the steel industry has grown. Furthermore, the expansion of global markets and the geopolitical circumstances since the 1980s have given Chinese, Russian, Brazilian and Indian steelmakers the opportunity to grow their market shares due to increasing urbanization and industrialization demands. [5]

This created a more competitive market as new markets decided to maximize output towards price competitiveness. On the other hand, historical North American and Western European markets chose to focus on quality and portfolio specialization. The disequilibrium between these two markets took a new dimension when wages rose in developed countries. Wages, mixed with over-regulation, increasing tax burdens and prior currency exchange issues hindered North American and Western European steel to regain their former levels of market share. As a result, their competitiveness in the global market dropped and developing nations became the main steelmakers. All those events, coupled with significant price drops led North American and Western European markets to slow down investments and deploy financial protection strategies. [5]

Having realized that cutting costs on manpower alone could stir the European political scene against the steel industry, European steelmakers developed their Supply Chain Integration to avoid compromising shareholder value. Supply Chain Integration (SCI) became an increasingly important strategy when it comes to improving collaboration within a supply chain, among its stakeholders and to improving the management of intra- and inter-organization processes. [5]

In order to help addressing costs, performance and risks, SCI requires intense exchange and cooperation between the involved stakeholders and subcontractors. It can be approached in three different ways [5]:

1. Horizontal integration: information, strategies, decisions and flows are shared but ownership and management of each company in the supply chain remain independent or decentralized.

2. Vertical integration: capital, ownership and management are also shared or centralized by means of mergers, acquisitions and equity efforts.

3. Hedging: this can occur either vertically or horizontally, but focuses mostly on ensuring profitability across markets, by having different branches of a supply chain's operation be more or less active than others so as to adjust to market variations.

European steelmakers are often key manufacturers in their respective supply chains and tend to focus on supplier side integration, mostly vertically, due to the commoditization of steel in the global market.

These strategies have helped build Supply Networks. In most Supply Networks, a customer forecasts its demand in terms of requested quantity and time of delivery. It then has some time to change this forecasted demand after which it is committed to its demand. To satisfy their customers' needs, manufacturers usually have to order materials in advance to produce their products without knowing the exact changes in customer demand. The materials can be ordered either from a standard supplier or from an emergency supplier, if there is not enough material in stock or if there is not enough time for a delivery from the standard supplier. [6]

A Supply Network is considered robust if it has consistent performance in an uncertain environment with very little variation in its output [6]. In that regard, the steel Supply Network - European or not - cannot be considered robust enough as it is suffering from crises such as the steel shortage crisis observed in the year of 2020 that has seen steel prices skyrocket. That is why it is strategic for companies, like Normet, to understand their procurement costs for metal parts. Finding out what importance do the steel extraction and transformation and the supply chains have in the pricing can give Normet a possibility to forecast its costs and thus provide them a big competitive advantage and flexibility.

Having stated that, it should be highlighted that our study focused more on understanding the costs from a product point of view and not from a supply chain one. In that regard, lead times and the fact that some suppliers could be emergency or "new" suppliers were not considered in our study. Also, the importance of warehousing costs could not be quantified, mostly because of a lack of data. Finally, we assumed that all suppliers are reliable in quantity, quality and delivery delay. [6]

2.2 Procurement cost modelling

In every industry, optimizing supply chain is a challenge for both suppliers and manufacturers. It is always about consolidating quality, costs and delivery time to optimize customer satisfaction while also reducing costs. In recent years, SCM (Supply Chain Management) has focused on managing the costs in supply chains. Indeed, good management of costs can bring more profits to firms, and as a result, give a non-negligible competitive advantage. [2]

In that regard, cost modelling is an important capability in purchasing organizations because it allows firms to understand the total cost system of their stakeholders. Understanding purchasing costs changes the supply-demand relationship between each stakeholders [2], as it allows organizations to negotiate better prices with suppliers, improve sourcing strategies, and optimize product designs [7]. As the purchase prices of metal parts account for a major share of Normet's equipment manufacturing costs, understanding the procurement costs is particularly crucial for Normet.

Besides cost elements, cost models should use cost drivers as explanatory variables. Cost elements - directly captured by accounting systems - are easier to incorporate as data is easily available. However, considering also the drivers behind these cost elements is important, because they affect the cost elements (and single driver can affect multiple elements). Thus, understanding cost drivers enables to understand the true dynamics behind the cost development better, helping to anticipate and prepare for changing costs. Using of indirect sources is also encouraged, such as industry experts, suppliers' facility tours and public statistics. These sources might provide great additional insights to the model. Sophisticated cost models account for total cost of ownership, such as shipping, acquisition, quality, and inventory carrying costs, beyond the costs constituting the actual purchase price paid to the supplier. [7]

However, cost models should be as simple as possible and complexity should be added only if needed - as the complexity increases, implementing the model in practice with good quality information becomes more difficult [7, 8]. Overfitting is a typical problem in cost modelling, as complex models seems attractive due to high R^2 values and low standard errors; these models fit sample data set well but are not accurate with unseen data. In the other end, there is underfitting where model omits relevant information and does not explain variation in the dependent variable well enough even in the sample data set. One way to break this trade-off is to conduct cross-validation by splitting the data set to multiple partitions (of training and validation data sets) and investigate the average results of fitness. [8]

Cost models should be commodity-specific to capture the difference in the cost drivers between different types of items. Purchases should be aggregated to logical groupings conducive to cost modelling, best practice being to aggregate them by supply industry or supplier process technology - and not by end-product application - as supply dynamics drive the costs. Here, cooperation with suppliers plays a key role: understanding the relative size of the cost components at each supplier helps to understand how the changes in their cost drivers affect the total price. Typically, suppliers are willing to provide details from cost structure in terms of how labor, materials, manufacturing overhead as well as general and administration are distributed. Cost models should be first built at commodity level, after which at supplier-facility level. This is typically enough for developing sourcing strategies. However, for cost estimation and target setting at individual item level, the model should incorporate information from individual items instead of just supplier-facility averages. [7]

Cost drivers can be divided to four categories to help identify all relevant factors affecting purchase prices [7]:

- **Design-related drivers:** Product design might differ across suppliers even for a product with the same end use purpose, potentially resulting in differing raw material costs and working hours needed to manufacture the product. Thus, considering product design is important to ensure that cost structures are similar for items with same modelling logic.

- **Facility-related drivers:** The most important facility-related drivers are typically the size of the facility, equipment and process technology. Larger facilities typically exhibit scale effects, and thus costs should be expected to be lower, while automated production technology implies higher dependency on energy prices and lower on labor costs. Besides, degree of vertical integration can affect the raw material costs and the lag with which raw material cost changes affect the purchase price for the buyer.

- **Geography-related drivers:** Important drivers are typically the distance to the factory (affecting transportation costs) and proximity to major transportation infrastructure. Here, the weight and size of the item play also important roles, as heavier and larger parts are more costly to transport, depending also on the mode of transportation. Labor costs also often differ depending on location.

- **Operations-related drivers:** Derived from the operational functionality at the supplier, such as how productive the factory is, in how many shifts it is operated (evening and night shifts are more expensive), and how high are the yield and scrap rates. This kind of drivers can be best captured by engaging in strong cooperation with supplier.

Thus, the keys to successful procurement cost modelling are to identify the relevant cost drivers with a significant effect on purchase price of each item type and to fit a model as simple as possible, explaining the variation in purchase price well enough while not fitting noise. Here, finding the suitable level of aggregation plays a key role as well - if there are significant item-level differences in pricing dynamics, those should be considered but otherwise staying at higher levels of aggregation is better due to the principle of simplicity.

2.3 Costs in the metal industry

Some industries have markets that follow a random walk, i.e., in which the commodities are being correctly priced at equilibrium. In those cases, the market is considered efficient. In contrast, inefficient markets are subject to twists in the capital and risk pricing which can completely change the financial resource allocation and economic development of those. When analysing the metal industry (for all metals), traditional unit root tests reveal that prices of all the metal are not stationary at all. Indeed, they are usually disturbed by structural changes, non-linear adjustments and price bubbles. [9]

- **Structural changes** occur when uncertain events such as economic crises, financial crises, political instability, changes in global production and consumption, health crises happen. These events create a shift in the fundamental functioning of the market that completely redefines the way prices are calculated.

- **Non-linear adjustments** are situations in which prices are not directly proportional to measure of quantity demand. These non-linear adjustments are a result of market frictions, transaction costs (bid-ask spread, short selling and borrowing constraints), heterogeneous agents' interactions and beliefs.

- **Price bubbles** are said to occur when there is a rise and then a tragical slump in the original asset price in a specific period. They are usually a result of speculation, currency movements, variations in supply and demand conditions, interest rate, financial crisis and climate change. In that sense, they can be a direct result of a structural change. This is why some methods and tests use only structural changes and non-linear adjustments to analyse and explain the fluctuations in metal prices.

For this study, Normet indicated non-linear adjustments for each metal part they bought that was subject to it as "price breaks". Otherwise, the literature leads us to reflect on our results by considering recent crises (Covid-19 and Ukraine). Indeed, it is likely that there will be structural changes or bubbles in the prices of metals in the following months. The slowdown of global production and consumption of commodities and the interruption of information and transportation flows as well as government interventions have made the market even more unstable than he was before the Covid crisis [9].

3 Data and methods

This section provides an extensive overview how our data analysis and model formulation progressed. Section 3.1 summarizes data sets based on which we are conducting our modelling. Section 3.2 describes how we pre-processed the data before conducting analysis. In section 3.3, our exploratory data analyses are explained, including how we determined the items whose cost we could model based on this data. Section 3.4 describes our process of finding the satisfactory models. Finally, section 3.5 provides the validation of our models.

3.1 Summary of data

3.1.1 Data from Normet

We got the following data sets from Normet, which are further described below one by one:

- Purchase order data
- Steel index data
- Item weights
- Price breaks

Purchase order data. We got purchase data from subcontracted metal parts from Normet from time period 4/2018 - 1/2022. The data set contained 107 686 purchase order rows. Each row contained an order for a certain item from a certain supplier. The most important variables in the data set are supplier, purchase order type, unit price and total amount of the order, ordered quantity, purchase order date, status of the row, item number and description, as well as technical item group. Figure 1 shows the overall cost development during the observation period of the data.

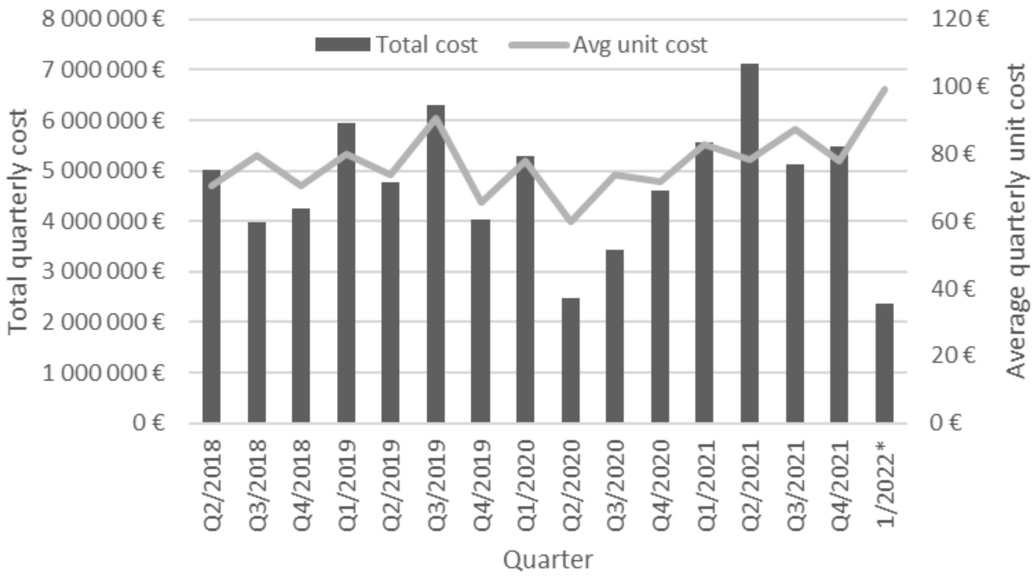


Figure 1: Quarterly total cost and average unit cost of subcontracted metal parts, time period 4/2018 - 1/2022.

There is significant variation in the total cost, probably caused by changing ordering quantities, as the average unit costs exhibit less variation. Average unit cost has stayed relatively stable but there is an increasing trend detectable starting from Q2/2020, and during the first month of 2022, the average unit cost has risen sharply. However, the sample size to calculate the average unit cost is smaller than for the other quarters.

With a closer look at the data, a general observation is that the bigger the order quantity, the lower the item price. Nevertheless, the metal industry can behave a bit differently. Indeed, the metal industry is subject to structural changes (due to crises), price bubbles (due mostly to speculation) and nonlinear adjustments (called price breaks) in which prices are not directly proportional to measure of quantity demand. [9] To counter that effect, Normet has also provided us the price breaks for metal parts that have seen them applied.

Steel index data. Normet provided us steel index data from the Management Engineering Production Services, showing the prices of raw steel since April 2018 for Europe, China and India. These prices also depend on the type of metal part (i.e., hot rolled coil, hot rolled plate etc.). Figure 2 shows the price development for European Hot Rolled

Coil (Europe HRC) and stainless steel (Europe 304 HRC) from April 2018 to December 2021. These were the indexes that were used in our modelling. Prices for both types of metal have been largely stable until December 2020 where they started skyrocketing (especially for Stainless Steel), which can be explained by the beginning of the steel crisis.

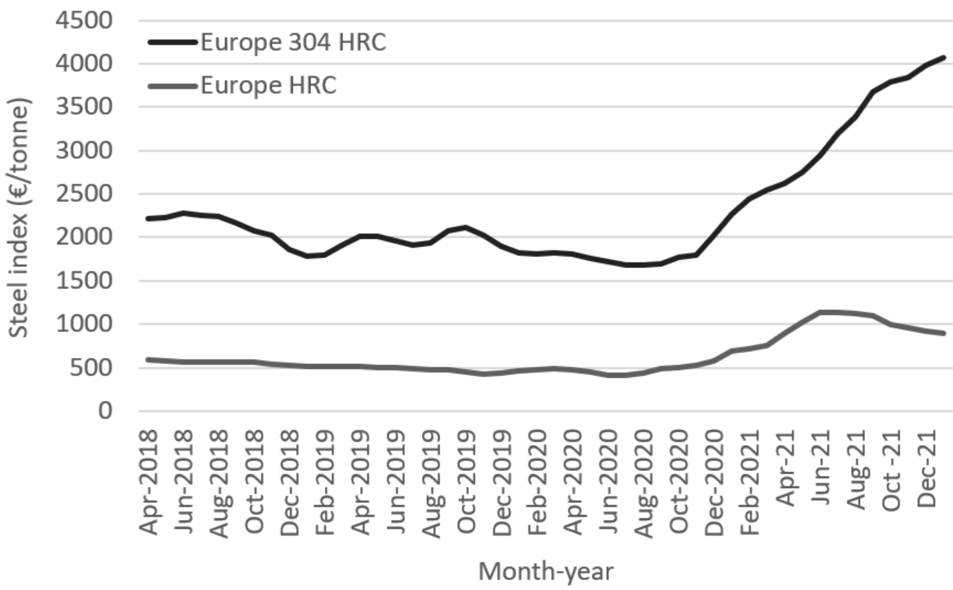


Figure 2: Steel index development for European Hot Rolled Coil (Europe HRC) and stainless steel (Europe 304 HRC), time period 4/2018 - 1/2022.

Item weights. All metal parts ordered by Normet have different weights. Usually, the heavier, the more expensive the metal parts are. The weights of the items ranges from nearly 0 kg to approximately 2 800 kg.

Price breaks. Normet provided us information on the pricing logic in the metal part industry. Three types of pricing exist.

Unit prices can depend dynamically on order quantity. In other words, unit prices are proportionally dependent on the order quantity. This is the case for suppliers G, H, I.

Nevertheless, as explained in the literature review, non-linear adjustments called "price breaks" also exist. They imply that some unit prices are not proportional to the order quantity. In the data provided, eight suppliers over 18 are concerned by price breaks. Over the four suppliers we built a model for (C, D, I and M), two (D and M) have price breaks.

Finally, some item prices are fixed. They are neither ruled by order quantity nor by price breaks, but agreed on typically once or twice a year, while with current market conditions mostly on a quarterly basis.

3.1.2 External data

We gathered the following external data about electricity prices and labor indexes, which are described more in detail below.

Electricity prices. Monthly electricity prices since 2017 were gathered from Statistics Finland's website [10]. We used the prices for enterprise and corporate clients according to the estimated energy consumption of the supplier, where Normet's insights were exploited. Most of the suppliers fell into the price category 20 - 499 MWh/year, while couple of very large suppliers were estimated to belong to the 2 000 - 19 999 MWh/year category. Figure 3 shows the electricity price development in each category. Electricity prices have been relatively stable except the large increase in the last months of 2021. This trend is expected to contribute to rising unit prices of purchased metal parts.

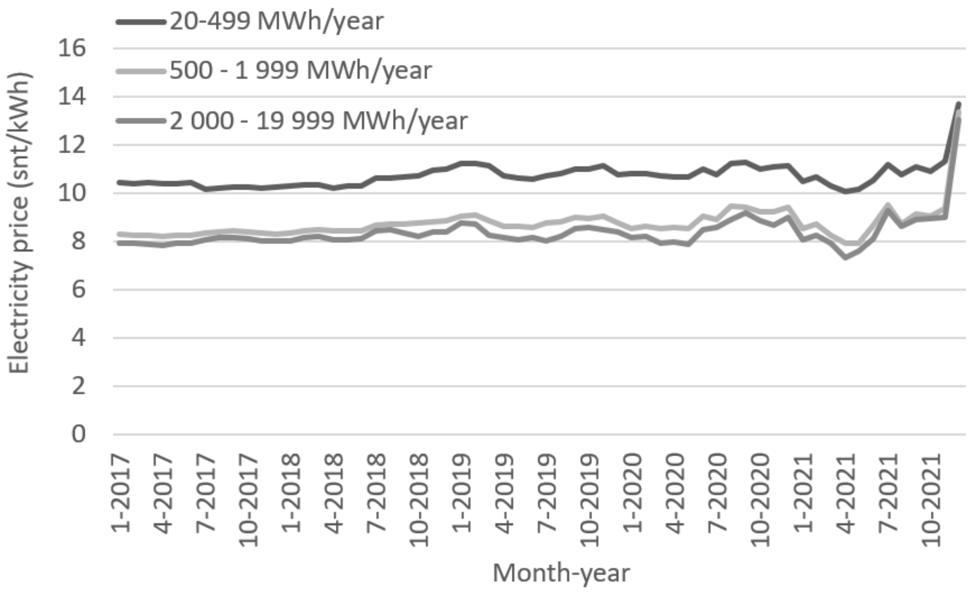


Figure 3: Electricity price development for enterprise and corporate clients in different categories of yearly energy consumption, time period 1/2017 - 12/2021.

Labor indexes. Quarterly labor index data since the first quarter of 2007 were gathered from Statistics Finland's website [11]. We used the seasonally adjusted labor cost index excluding one-off items in the metal industry, as we assumed that the metal industry does not exhibit seasonal variation in the wages. Figure 4 shows the labor index development. Labor index has stayed relatively stable until the end of 2018, after which there was increase and drop, before in 2021 labor index has increased significantly. This is expected to increase the unit prices of purchased metal parts.

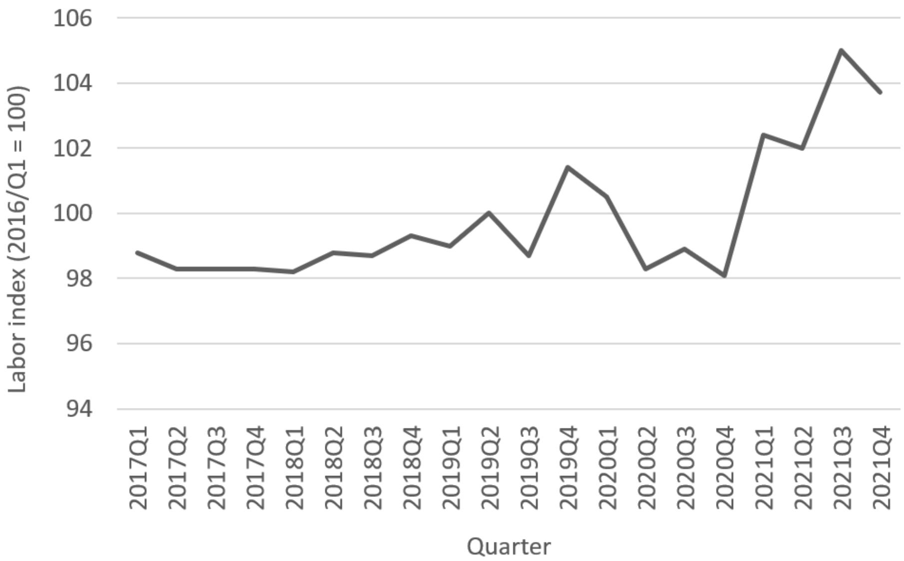


Figure 4: Seasonally adjusted quarterly labor cost index excluding one-off items in the metal industry, time period Q1/2017 - Q4/2021.

3.2 Data pre-processing

In this section, we explain how we filtered and integrated the data for our data analysis and modelling. First, we removed the purchase order rows, which were biased according to their sales origin, such as rushed orders, return orders, or rebuilds. Next, we removed the cancelled orders. We also removed the orders with non-positive unit price. To have enough data points for each item in our modelling, we filtered out items which were ordered during less than ten different months.

Then, we integrated the steel index data to the purchase order data. We got information from Normet about the steel indexes that each supplier follows and their historical monthly values (up to 4/2018). For 3 from 18 suppliers we did not get this information so they were excluded from the analysis. As changes in the steel index affect purchase prices of metal parts often with a lag, we wanted to include also lagged value of the steel index. This was done by converting the purchase order date to month, and adding columns with steel index value for each purchase order row with lags from 0 to 12 months. We carried out a similar integration for the labor index and electricity price data with lags up to 12 months. We integrated also the item weights for each PO row.

3.3 Exploratory data analysis

In exploratory data analysis, our goal was to improve our understanding of the characteristics of and relationships between the unit price and explanatory variables. Our analysis was divided to two streams according to our objectives:

1. **Finding outliers:** Aimed towards developing methods to detect items with abnormal cost development. This part involved analysis of the deviation in unit price and steel index as well as difference in minimum and maximum unit price of items. It helped us also improve our understanding of the pricing dynamics for the modelling part.

2. **Building the cost model:** Aimed towards building the cost model. This part involved correlation analysis of unit price and explanatory variables. This was first done on a more aggregated level for supplier selection, continued by more detailed analysis of the selected suppliers to determine initial model parameters.

3.3.1 Deviation in unit price and steel index

In this section, we analyze the variations of unit price and steel index values at the time of the purchase. We used the pre-processed data, which is described above.

We were given price break points, which are fixed, for seven different suppliers. In price break points there are fixed prices for unit prices in agreed intervals. The unit price may change according to the steel index or manufacturing costs. Due to some suppliers having price break points, we have divided the deviation analysis between unit price and steel index into two different parts: deviation with price break points and deviation without price break points.

Deviation with price break points The seven suppliers with price break points are: K, F, M, B, L, D, P and J. The price break points were delivered to us in an Excel file, in which the total amount of intervals per company are three to six, except one company where there was only one interval. Before executing the code, we simplified the Excel sheets by copying the orders from suppliers from pre-processed data to an empty Excel. Then data was sorted from lowest to highest in quantities. In this step, we could check if there were any zero quantity orders and if there were any, those were deleted. After this, quantities according to given price break points were copied to another sheet in the Excel file. This was done with each price break point intervals. Lastly, each sheet was sorted from lowest to highest according to item numbers. This grouped the items together in the Excel, which was necessary for the code.

The code was done in R-language in RStudio, where a data frame is created from a sheet for a specific supplier. The code worked in a for-loop for each different item numbers, which was read from the Excel. It knows the starting point in the array for the item number and searches the last point for the item number. This gives us the information of position of first and last points for this item, which was efficient. Then we calculated the deviation for the unit price of the item and divided it with the mean of the unit price of the item. This was done in order to normalize the deviation, which enables us to compare items equally. After having the normalized deviation of the unit price, we moved on to the deviation of steel index. The normalized deviation of unit price is

$$SD_{NUPi} = \frac{SD_{UPi}}{Mean_{UPi}}. \tag{1}$$

In the equation (1) NUP is referring to the normalized unit price, UP is the unit prices and i is referring to the item number.

Even though most of the items used the same steel index, we found an interesting way to find the deviations in the index. For each purchase we know the index value at the time and this created a new column with steel indexes. So if there was item which was bought six times since 2018, we take those six purchase dates and find the right index, from which the deviation for steel index were created. The code used the same search engine as before and in the deviation there was the steel index values. All of these were divided with the mean of the used index values. We used this to create normalized deviation values which can be compared to another normalized index deviations. The equation for normalized steel index is

$$SD_{NSIi} = \frac{SD_{SIi}}{Mean_{SIi}}. \tag{2}$$

In the equation (2) SI refers to the steel index, SI is the Steel indexes and i is the item number.

Combining the two deviation metrics gave us a 2-dimensional plot, in which we could see how the price of the unit price behaved according to index deviation. A minor problem with this was that it did not determine if the unit price rose or dropped. The plots for most important suppliers D and M are shown in the Figures in the Appendix starting from page 41.

Deviation without price break points In this section, we wanted to see how the unit price behaves according to steel index with the suppliers which did not have price break points. The problem is that the order quantities can have a major effect on the unit prices. The ordering quantity can be taken into account in the analysis by calculating the deviation of the order quantities, which then will be added to the deviation of unit price and steel index. This creates a 3-dimensional plot, in which we can detect items whose unit prices behaves abnormally. The preparation needed for running this code is to create new Excels, for each of the suppliers and sorted the items according to item number. This could be also done in the main data source by filtering the unwanted suppliers out and the sorting the data according to item numbers.

The code worked the same way as in the section of 3.3.1 deviation with price break points. The third metric, normalized deviation of quantity orders were calculated by deviation of the order quantity divide by the mean of the ordered quantity, which is equation (3). With the third metric, we could plot these in a 3-dimensional plane in RStudio. In the program we could move freely in the plot and have a closer look to different points, but this is not possible outside of the program. The plots for most important suppliers I and C can be found in Appendix. Taking a closer look to the data of supplier C, we found that the prices do not vary according to the order quantity. So the variation of the unit prices are explained by other factors, for example steel index, which applies to some items in the deviation Figure of supplier C.

$$SD_{NOQi} = \frac{SD_{OQi}}{Mean_{OQi}}. \tag{3}$$

In the equation (3) NOQ refers to the ordering quantity, OQ is the ordering quantities and i is the item number.

3.3.2 The difference in maximum and minimum value of an item

We next look into the percentage difference in minimum and maximum values for an item. The deviation metric will not show the potential spikes in the data, which have occurred in the year 2022. With the minimum and maximum metric, we will be aware of the potential spikes in the data and it will be a fine addition to detecting outlying items. The data used in this analysis was done with the filtered data that includes the four selected suppliers: C, D, M and I. The data was arranged according to the item numbers.

The code worked by reading the data from Excel and saved it as a data frame. The code memorized the first position of new item number and then it searched the last one, which was then memorized. With this, the computer knows the positions of the item. After this, the minimum and maximum was found from the unit price of the item and it calculated the percentage difference. The equation used for this is:

$$Dif_i = \frac{Max_i - Min_i}{Max_i}. \tag{4}$$

In equation (4), Max_i is the maximum unit price value of the item i and Min_i is the minimum unit price value of the item i.

The results of minimum and maximum metric, with the four suppliers, are in the Appendix in the page 51 under the section minimum and maximum Figures. The Figures show that for items with price breaks, the minimum and maximum values will increase in some items, but in the other items they do not change at all. For example in the Figure in the page 56 and in the price break 3, the prices of the items do evolve only for few items and the evolvement is minor. For suppliers which do not have price break points, the minimum and maximum values vary a lot. This is due to the ordering quantity having a large effect on the unit price and there is no separation of orders in to different categories. For example in the Figure on the page 51, the price difference is not ranging between 0 % and 100 %, which can be seen in the Figure on the page 57.

Combining the minimum and maximum metric with the deviation metric, we are able to detect the items, that have varying price behavior. This can be used for detecting outlying items in the past and these detected items may be interesting for Normet.

3.3.3 Correlation analysis

In this section, correlation analysis between item purchase price and the potential exploratory variables was conducted for all suppliers for which we had steel index data (15/18 suppliers). The purpose was first to determine those suppliers, whose unit prices correlate strongly with steel index, indicating that steel index is a good predictor for the unit price. This helped us to scope our modelling efforts to the suppliers and items, whose costs we can model with steel index, which is known to be the most significant cost element for the purchase price of the metal parts. Besides, we evaluated the correlation between unit price and labor index as well as between unit price and electricity price to compare the magnitude of these correlations. The second part of the correlation analysis was to define the correlation-maximizing lag for each selected supplier for each explanatory variable.

We expected the dependency between unit price and the predictors to be linear, so we used the Pearson correlation coefficient for calculating the correlations between the unit price and predictors:

$$r_{UPis,Xisl} = \frac{\sum_{o=1}^{n}(UP_{iso} - \overline{UP_{is}})(X_{islo} - \overline{X_{isl}})}{\sqrt{\sum_{o=1}^{n}(UP_{iso} - \overline{UP_{is}})^2 \sum_{o=1}^{n}(X_{islo} - \overline{X_{isl}})^2}}, \tag{5}$$

where UP refers to unit price, X to predictor (steel index, labor index, or electricity price). Indices are as follows: i refers to unique items, s to suppliers, l to lag from the ordering month and o to purchase order, n is the number of purchase orders of item i. So, we calculated the correlation for each item-supplier combination exhibited in the data with the predictor values at different lags (from 0 to 8 months).

We analyzed the obtained correlation coefficients both at individual item level as well as aggregated at the supplier level by the average values. The average correlations by supplier are shown in the Figure below.

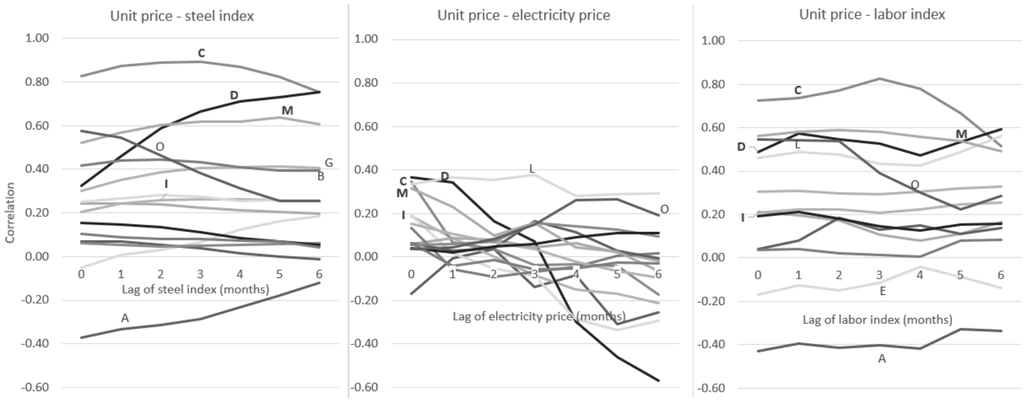


Figure 5: Average correlations between unit price and each predictor by supplier. Lags calculated from the ordering month, time period 4/2018 - 1/2022.

Overall, the graph shows that unit price correlates stronger with steel index than with other predictors; for most suppliers, the unit price - steel index -correlation is clearly above zero and for a couple of suppliers even close to one. Electricity price is the least correlated with the unit price, correlation being close to zero for most of the suppliers. Unit price - labor index -correlation is positive for most of the suppliers.

Unit prices of supplier C exhibits the strongest correlation with the steel index. The parts manufactured by this supplier are heavy parts consisting mostly of steel, so this pattern is reasonable. The highest correlation is at lag of three months, which also makes sense as suppliers use their inventories before they procure raw material with the changed prices - reflected at then later in the purchase price for Normet. Thus, supplier C was selected as a supplier for which we build the cost model.

Unit prices of supplier D exhibit the second strongest correlation with steel index. The long lag is explained by the fact that supplier D is a smaller supplier that buys with more intermediate players in the supply chain compared to larger suppliers. However, the volumes from supplier D are large enough to consider it as a relevant supplier for our modelling (346 purchase order rows for 20 unique items). Based on this information, supplier D was selected for further modelling.

Supplier M has the third highest unit price - steel index -correlation, peaking with the lag of five months. The relatively long lag could be explainable by the relatively small size of the supplier, however with large enough volumes for our modelling (47 unique items, 1394 PO rows). Thus, supplier M was selected for further modelling.

Suppliers O, G, and B exhibit the next highest correlations between unit price and steel index. However, we discussed with Normet that one of these suppliers is known not to follow the steel index, so we decided to not include that in our modelling. The other two suppliers were less interesting for Normet and have relatively small volumes, and considering that the next supplier with slightly smaller correlation (I) was interesting and has huge volumes (approximately half of the POs and total quantity of ordered parts were from this supplier). Thus, supplier I was chosen for further modelling.

We agreed to focus on these four suppliers which exhibited high correlation with steel index (it was known to be the key driver), had large enough volumes, and were interesting to Normet. This way, we could keep the modelling workload reasonable.

The unit prices of all four selected suppliers exhibited positive correlations with all predictors significantly above zero at least with some lag. Thus, we decided to first try to model the unit prices of items supplied by these suppliers including also electricity price and labor index as explanatory variables with the correlation-maximizing lags. To define the correlation-maximizing lags for each predictor, we looked at the correlation distribution across all items for each supplier. See the Appendix page 71 for boxplots of these distributions. Table 1 summarizes the correlation-maximizing lags for the selected suppliers.

Supplier	Steel index	Elec. price	Labor index
C	3	0	3
D	8	1	7
M	5	0	1
I	3	0	6

Table 1: Correlation-maximizing lags (months from ordering date) and pricing logic for the selected suppliers.

3.4 Modelling

Modelling was conducted by following the principles discussed in the literature review and knowledge obtained through data analysis to forecast the future cost development of the subcontracted metal parts for the four selected suppliers (C, D, M, I).

3.4.1 Modelling logic

Cost models should be first built on an aggregate level. We examined the supplier-level monthly average unit price and total cost development, but we concluded that there was too much price deviation caused by the changing mix of ordered items (with different absolute prices) - even when investigated at a quarterly basis (see Figure 6 below). Even though for supplier I and C the deviation was relatively well smoothed out at quarterly level, for suppliers M and D the deviation was still too high. Next, purchases should be aggregated to logical groupings based on the supply dynamics, and the best criteria for grouping that we had was the technical item group division. When we investigated the average unit prices and total cost per supplier for each relevant technical item group, it turned out that the mix of (differently priced) ordered items still causes too much deviation for the aggregation level to be reasonable in our modelling. Thus, we decided to model individual items for each four suppliers, where the changing mix of ordered items is not disturbing our model. Modelling at the item-level enabled also to use the model to detect items with abnormal cost development, as will be explained in Chapter 4.2.

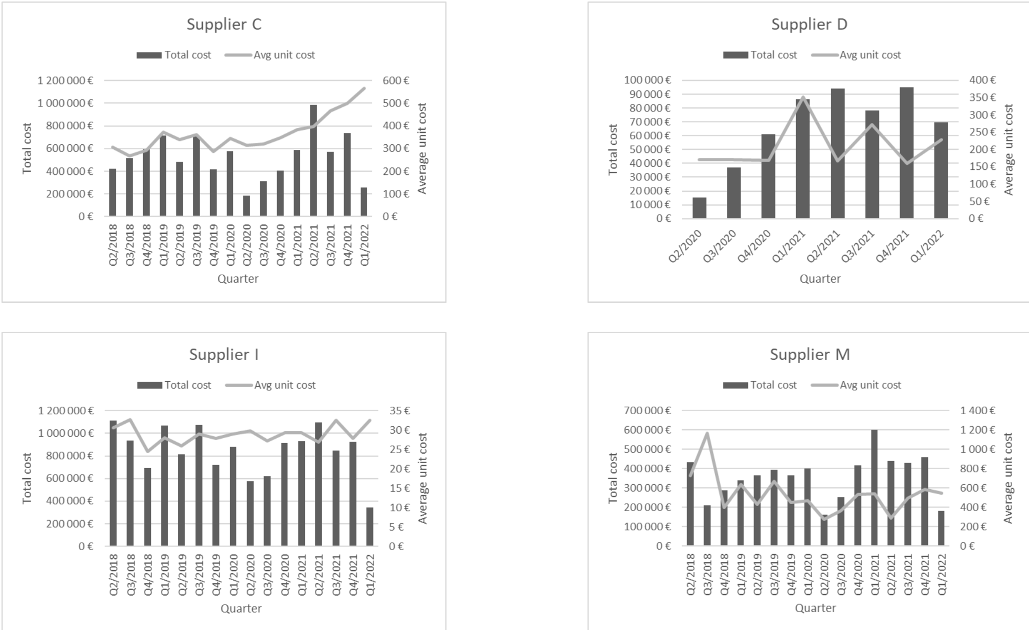


Figure 6: Total cost development and average unit costs by quarter for each supplier.

Cost modelling should be kept as simple as possible and complexity should be added only if needed. We also knew that the relationship between the unit prices of items and explanatory variables is expected to be linear: unit prices increase linearly as functions of increasing steel index, electricity price or labor index with some lags. Thus, we chose to use a multiple linear regression model, which is a robust method minimizing the risk of overfitting, while the risk of underfitting with the data available could not have been reduced significantly with a more complex model.

Despite modelling at individual item level per supplier, our model has parameters which are common for all items ordered by the same supplier. The lags with which steel index, electricity price, and labor index affect the unit price were selected at the supplier level, despite for some items, the correlation might be higher for some other lag. This simplification makes the model more implementable.

3.4.2 Quantity discount function

We knew that for suppliers C, D, and M the ordering quantity affects the unit price, so the ordering quantity should be accounted for when modelling the unit price. We decided to do this by normalizing the unit prices with a discount function. To keep the model simple enough, we used the continuous discount function also for suppliers with price breaks (D and M).

The quantity discount will be modelled with a exponential function

$$QD(q) = ae^{-c(q-1)} + (1-a) \; ,$$

where a is the terminal discount and c is the slope and q the quantity. The function is fitted to each item-supplier combination using unit prices normalized by the maximum value as the y-values. After fitting the unit prices, we set criteria to validate if the fits were satisfactory. The chosen criteria were that the R-squared of the fit has to be greater than 0.25 and that the parameters have to be within the 95%-CI compared to all other items parameters. If the criteria is not met the product will be assumed to not have any kind of quantity discount.

Figure 7 shows how the quantity discount function is fitted to the data. The items were chosen such that the good fit has a R-squared near 1, mediocre just above the threshold and bad is below threshold. In this case the first two fits are valid and the last one is treated with a flat discount function. Figure 8 shows the ratio of good and bad fits between suppliers.

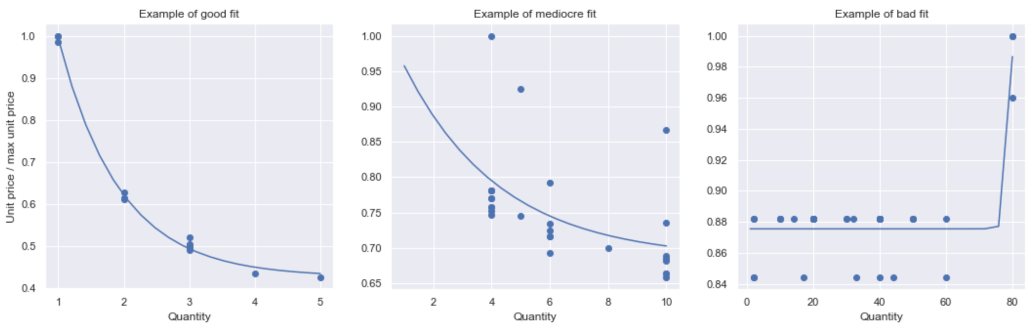


Figure 7: Examples of good, mediocre and bad fits for the quantity discount function.

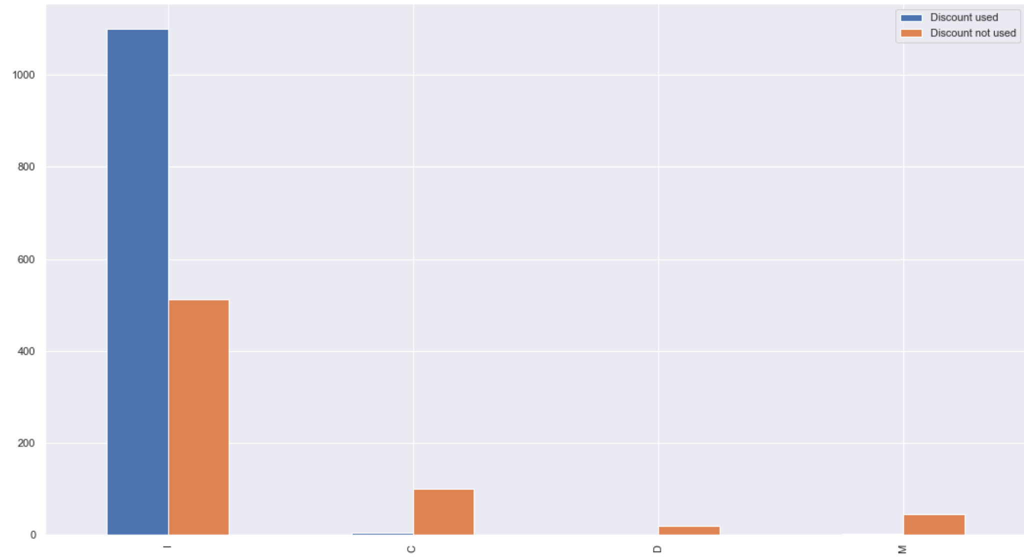


Figure 8: Distribution of good and bad fits between suppliers.

3.4.3 Regression model

While conducting the correlation analysis, we recognized that even for some suppliers with high average correlation between unit prices and steel index, there were items which did not correlate strongly with steel index. These items needed to be removed from our analysis, as the individual item level modelling would not have resulted in unit price variation explained well. We used the correlation threshold of 0.7 for including an item in our modelling. Besides, for some items there were not enough data points to fit the model and were thus not included in our modelling (we used the earlier mentioned limit that item needed to be ordered at least during ten different months to be included in the modelling). Thus, we obtained the final cost model as follows:

$$NUP_{ist} = \beta_{0,is} + \beta_{1,is} W_i SI_{t-l_{SI,s}} + \beta_{2,is} E_{t-l_{E,s}} + \beta_{3,is} L_{t-l_{L,s}} + \epsilon_{ist}, \qquad (6)$$

where the variables and indices are as follows:

- NUP_{ist} normalized unit price (€) of item i ordered from supplier s in month t.

- Coefficients $\beta_{0,is}$ (intercept, also the sales margin), $\beta_{1,is}$, $\beta_{2,is}$, $\beta_{3,is}$ are constants for item i ordered from supplier s.
- W_i weight of item i in kilograms.
- $SI_{t-l_{SI,s}}$ steel index (EHRC) in month $t-l_{SI,s}$, where t is current month and $l_{SI,s}$ supplier-specific lag for steel index (months).
- $E_{s,t-l_{E,s}}$ electricity price (snt/kWh) for supplier s in month $t-l_{E,s}$, where t is current month and $l_{E,s}$ supplier-specific lag for electricity price (months).
- $L_{t-l_{L,s}}$ labor index in month $t-l_{L,s}$, where t is current month and $l_{L,s}$ supplier-specific lag for labor index (months).
- ϵ_{ist} (random) error terms.

The normalized unit price, that is obtained from this equation, does not yet account for the quantity discount based on the ordering quantity. Therefore, the discount function needs to be applied:

$$UP_{ist} = NUP_{ist} QD_{is}(q), \tag{7}$$

where QD_{is} is the supplier-item specific discount function with order quantity q.

3.5 Validation

We implemented the model (6) for the selected items of the selected four suppliers in Python and estimated the regression parameters with ordinary least squares method. We evaluated the model fit with coefficient of determination R^2, which measures the proportion of variance in the unit price that is predictable from the explanatory variables.

The distribution of R^2 for the fitted models was as follows for each supplier in Figure 9.

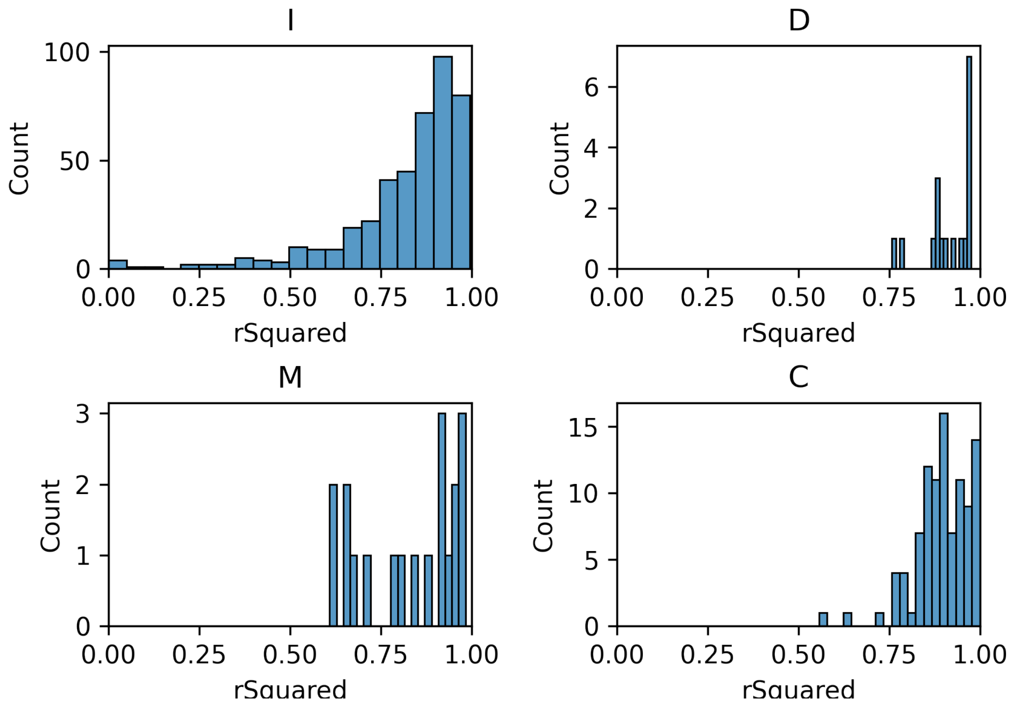


Figure 9: R-squared distribution of regression fits.

We see that the R-squared distribute nicely and are skewed to the right, which implies that the model has worked in-sample. In addition, the residuals could be further examined but are left out of this study since they should be inspected for each item separately. Besides, the correlation between the explanatory variables could be investigated to ensure that they are linearly independent. For example, we found a correlation between the labor index and steel index, but decided to leave them as is due to our straightforward approach and time limitations set by the course.

4 Results

In what follows, Subsection 4.1 presents the forecasted future costs and Subsection 4.2 the results of detecting abnormal cost development.

4.1 Forecasting cost development

Figure 10 shows the total predicted cost of the modelled items versus realized costs using in-sample data. The fits are exceptionally good, meaning that the model has found good fits and the noise has canceled itself out.

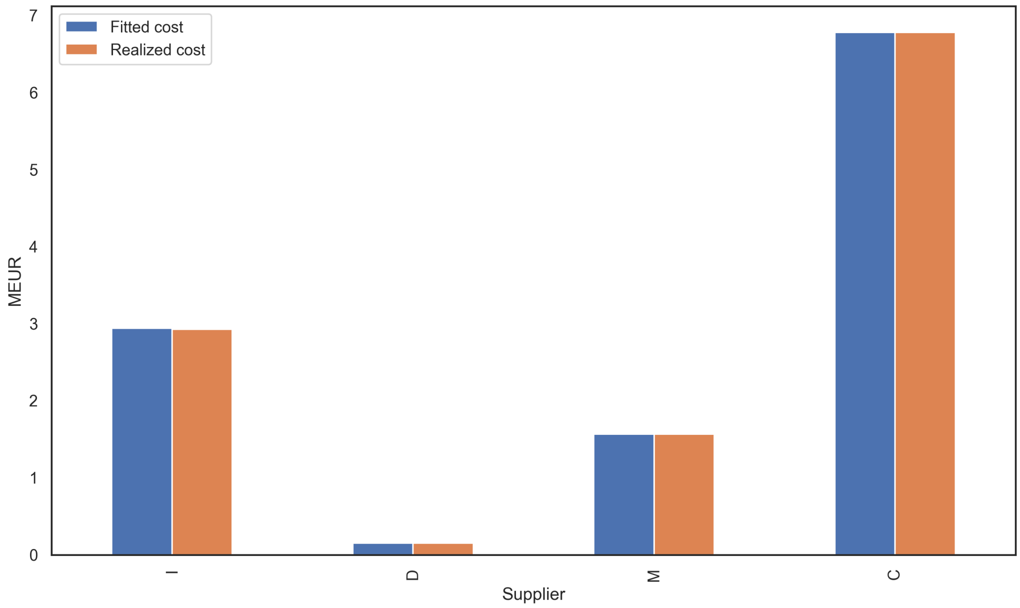


Figure 10: Total costs calculated by in-sample predictions.

However, the quality of the predictions weakens when making out of sample predictions, shown in Figure 11. The model systematically predicts lower costs with all suppliers. The reason is most probably caused by the rally in the steel price during 2021, which caused item prices to go up. The data used for model fitting, therefore might have had a different covariance between item and raw material prices, and lags may have been adjusted more responsively. The largest relative difference was with supplier C, for which we estimated 20% smaller costs than what was realized. However, we consider this as a relatively good result taken the circumstances, data and approach.

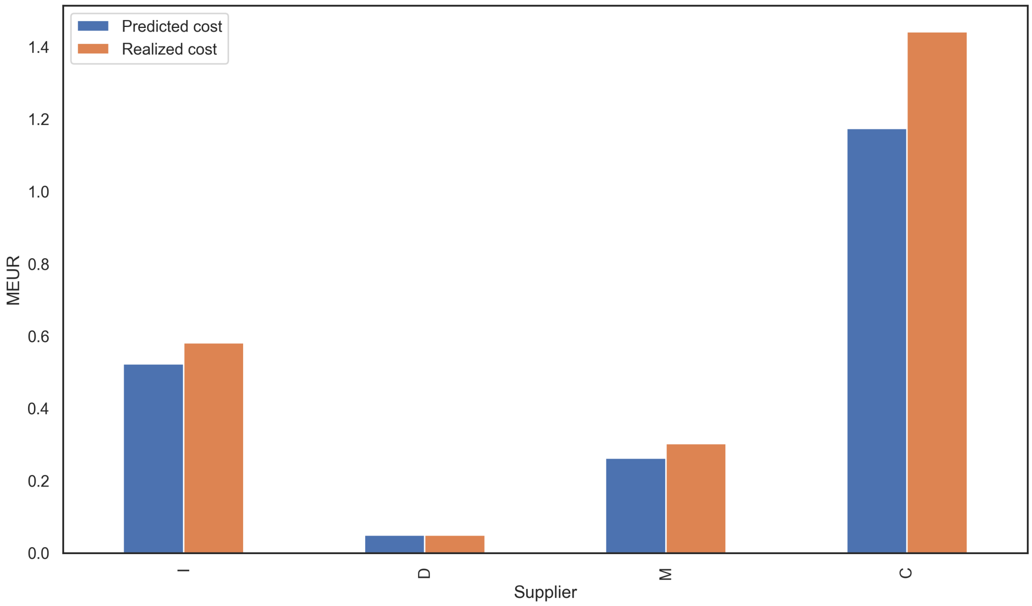


Figure 11: Out of sample predictions for 6/2021 onwards.

Tables 2 through 5 show statistics of the model parameters for different suppliers. The data was normalized so the parameter coefficients shows the significance. The columns 1-4 show the values of the coefficients, columns 5-8 their significance (the lower, the better), and column 9 the R-squared value. One can conclude that the least explaining variable was the electricity price, which was followed by labor index. The most significant driver of the price was the steel, which is natural both in raw material perspective and also though the filtering of the data. There was also correlation between the steel and labor indexes, but we will not take a view on its reason.

Table 2: Regression summary for supplier I. Sample size 429.

	const	Steel	Elec	Labor	pconst	pSteel	pElec	pLabor	rSquared
mean	64.5	4.3	0.5	0.1	0.0	0.0	0.3	0.2	0.8
std	115.6	9.9	3.2	5.1	0.0	0.1	0.3	0.3	0.2
min	0.2	-13.9	-13.0	-82.8	0.0	0.0	0.0	0.0	0.0
max	1558.2	127.1	46.2	23.2	0.0	1.0	1.0	1.0	1.0

Table 3: Regression summary for supplier D. Sample size 18.

	const	Steel	Elec	Labor	pconst	pSteel	pElec	pLabor	rSquared
mean	260.3	5.1	-1.2	2.6	0.0	0.0	0.3	0.0	0.9
std	320.7	9.9	2.7	4.3	0.0	0.0	0.3	0.1	0.1
min	45.3	0.0	-8.5	0.0	0.0	0.0	0.0	0.0	0.8
max	1119.5	32.2	0.2	14.2	0.0	0.0	0.9	0.2	1.0

Table 4: Regression summary for supplier M. Sample size 19.

	const	Steel	Elec	Labor	pconst	pSteel	pElec	pLabor	rSquared
mean	2153.4	98.1	31.9	77.3	0.0	0.1	0.3	0.2	0.8
std	2062.3	128.1	78.9	140.2	0.0	0.2	0.3	0.3	0.1
min	113.5	2.0	-5.8	-0.1	0.0	0.0	0.0	0.0	0.6
max	8630.3	457.0	349.5	628.2	0.0	0.6	0.9	0.9	1.0

Table 5: Regression summary for supplier C. Sample size 99.

	const	Steel	Elec	Labor	pconst	pSteel	pElec	pLabor	rSquared
mean	845.4	102.6	10.9	45.7	0.0	0.0	0.4	0.1	0.9
std	1467.0	215.3	101.1	86.7	0.0	0.1	0.3	0.2	0.1
min	4.4	-65.0	-113.0	0.1	0.0	0.0	0.0	0.0	0.6
max	7412.9	1392.2	955.9	417.6	0.0	0.7	1.0	0.9	1.0

4.2 Detecting parts with abnormal cost development

Abnormal cost development can be detected in various ways. Each of these has robust features and weaknesses. Our methods focused on a fairly simple math, which gives transparent information. The real challenge, which comes with our method, is what we are looking for in the metrics. Each supplier behaves differently and the same thresholds do not work for all suppliers. Another interesting way to detect abnormal cost development is to vary the thresholds and see how do the results change. This would strengthen the assumption that the specific item has strong assumption on being an outlying item.

Thresholds for each of the four suppliers were defined by looking at the original Figures and then refitting suitable thresholds. Finding suitable thresholds for each supplier is challenging due to each supplier behaving differently. The process starts with looking at the original Figure and then defining a rough threshold, for example filtering out the items that have the quantity variation lower than 0.5. The objective is to find items that have high variation in unit price, high value in the minimum and maximum value, low variation in the steel index and low variation in the quantity. The challenging part is to fit the thresholds to the quantity and unit price because they have a correlation in the pricing. The threshold values are in Table 6.

Supplier	Min-Max	Unit price var.	Index var.	Quantity var.
C	>0.3	>0.1	<0.8	<0.6
M	>0.2	>0.1	<0.8	-
I	>0.8	>0.5	<0.8	>0.2

Table 6: Thresholds used in the analysis.

After defining the thresholds, we took some items and searched those in the data. We looked into these points and tried to see visually some abnormal cost developments. If there were no abnormal cost developments, we tried new threshold values. This process was a feedback loop until the results were acceptable. In the supplier I, it was hard to detect these due to unit prices being defined according to quantity, which makes detecting outlying items this way harder. These Figures for the three suppliers can be found in the appendix on the page 61. The supplier D was uninteresting in this section due to the unit price varying only a little and this is due to the unit price increasing a little by the end of the year 2021. For example taking an item number 1547 in supplier I, which stays above the thresholds. Having a closer look at the data, we can see that the item does not behave naturally with the variables, but the unit price changes radically.

The thresholds were varied as an example for supplier I, see the Appendix on page 67. We varied four different thresholds and plotted the remaining items for each of varied threshold. Some of the same items stay there and those items are most likely to be outlying items. The variables that were varied are the min-max, deviation, steel index and deviation in quantity. These variables help us to detect the abnormal cost development in the past.

One effective way to detect abnormal cost in real time is to use the model. It is currently one the best ways to see if the pricing has been priced correctly. The model can give a warning for the change in a price or it can raise a suspicious for supplier's pricing. It is good to know that the model does not tell the absolute truth and if there have been some changes in the supplier's inside operations, it may give a false alarm for Normet.

5 Discussion

Even though our cost model excludes a significant share of items for some suppliers, it is still usable for forecasting the future cost development for the selected suppliers, as we can calculate the share of total costs per supplier that the items included in our modelling account for (see Figure 12 below). For supplier C, the items in our modelling account for vast majority of total costs, while for other suppliers, it is vice versa. However, the unit prices of items not included in our modelling are not correlating strongly with steel index, and thus those costs are not expected to react so strongly to steel index changes. Here is also important to notice that part of the items not included in our modelling were excluded due to the lack of observations (i.e., they were ordered during less than ten different months), not just due to the low correlation with steel index.

Figure 12: Total costs (before any filtering) by quarter for each supplier, showing also costs that items in our modelling account for.

This way, Normet can understand the magnitude of changes in the unit prices of subcontracted metal parts - and when knowing the ordering quantities, also the changes in total costs - as a result changes in steel index (and partly in labor and electricity price). Understanding this dynamic is crucial, as for many metal parts, material costs and particularly that of steel is one of the most important cost drivers. Currently, understanding this magnitude is even more crucial as steel crisis has lead steel prices to increase extremely high levels and estimating the cost increases can help to prepare contingent actions on-time.

The literature suggests to model cost drivers besides cost elements. Our model addresses this well, as our explanatory variables are drivers of the cost elements which are captured in the cost breakdown of items: steel index and weight drive raw material costs, labor index drives labor cost and electricity price drives the cost of machine hours. Knowing these drivers enables to react to upcoming cost changes faster, as change is anticipated when the value of a key driver is changing, instead of detecting the change only from the changing cost element (e.g., increasing raw material costs).

From the four categories of cost drivers framework suggested by literature, our model captures well some important design-related drivers. By modelling at individual item level, we account for the differences in raw material costs between items by using item weight and steel index as explanatory variables. By comparing the magnitude of coefficient β_1 of one item to other items, the relative importance of steel costs to the total costs of that item can be seen. Similarly, magnitude of the coefficient β_3 for an item (theoretically) indicates the magnitude of working hours needed to manufacture that item - however, this dependency was weak for many items as labor index was not so accurate proxy for the labor cost of working hours. From facility-related drivers, the coefficient β_2 captured dependency on electricity price.

However, our model could have been improved by including more specific data from suppliers, to better categorize the items depending on process technology used in manufacturing them. Technical item group categorization (which was tried as a level of aggregation for modelling) was not optimal, as it is more end-product focused. Our model did not include any geography- or operations-related drivers, which could be significant. Besides, our model did not consider the total cost of ownership - accounting also for costs beyond the purchase price - such as logistics and inventory costs - which could reveal interesting drivers.

6 Conclusions

Modelling the procurement cost turned out to be more challenging than expected. Steel index as the main driver was not able to explain the variation in unit prices for many suppliers well enough to be able build a cost model on that. To overcome this problem, correlation analysis was used to focus the modelling on the suppliers which best follow the steel index. Besides, there were so many different items with different price levels and behavior - likely with different underlying cost structure - that building a model on aggregated level with our knowledge was not possible. This challenge was overcome by conducting the modelling at individual item level.

Accounting for the quantity discounts was also challenging for supplier I, as the quantity discount varied by item and the original idea to fit a supplier-specific discount function did not work: for some items, there was significant discount and the pattern varied item by item, while for some items there was no discount at all. Thus, the discount functions needed to be determined at item-level. For suppliers D and M (which were supposed to have price breaks) we could not fit a satisfactory discount function (like was the case for supplier C, as supposed).

The item level models were created for the four suppliers for the items, whose unit prices exhibited the highest correlation with steel index (with supplier-specific lag). The models are relatively accurate for the selected items, and by knowing the cost share of these items from total costs, the model allows to forecast the magnitude of the total cost changes as a result of steel index changes with a reasonable accuracy.

The metrics used for detecting abnormal costs development are useful for spotting outlying items in the past. The problem with the metric is that it includes quantity driven items in there, which makes it harder to spot outlying items. Adding more complexity is not a great idea, due to it making the interpretation of outlying items more difficult. For spotting outlying items in real time, comparing the prices of suppliers and our model is the best option. If there is a sudden increase in price and there is no explanation for that, then it can interpreted as an abnormal cost development.

Finding abnormal cost development for the supplier I is challenging. There were many false alarms for items and most of them showed a trend of having one or multiple orders which included quantity of one, low weight. These items are the items that has a major effect on the metrics in the supplier I. Filtering out the items that had quantity of one and weight lower than 2 kilograms, got rid most of these unclear orders. After applying the same threshold as before, the results showed only the top of the top outlying items in the supplier I. These items were send to Normet as an example of the end product of the metrics analysis.

For further research, our recommendation is to look at the quantity discounts more thoroughly since they play a very large role in the product pricing. Ideally we would want to inspect the price behaviour during a time period where the raw material prices do not change much. However, we did not find a method to do this since the sample size would have decreased too much if looking at a narrow time period. Besides, the discount function could be estimated directly from the price list of items for those suppliers who have the list.

In retrospect, the products could be clustered and not analyzed in a item-supplier specific level. This could reveal reasonable levels of aggregation to build models which are easier to implement. We chose not to do this due to our limited knowledge of the products and also of clustering methods.

We would also encourage Normet to ask suppliers for price breakdowns of their items, in order to understand the cost structure better. This could be one way to group the items and create different models for different cost structures. For example, the items with low correlations between unit price and steel index, have most probably different cost structure than those with high correlations. Thus, different types of models could be created for those.

Another interesting topic of further research could be to investigate the operations-related cost drivers. Investigating the internal operations at suppliers' facilities could reveal how automated processes they use, i.e., how dependent they are on electricity and labor costs. This could help to set weights for model parameters with some real information, instead of just estimating them directly from data. Besides, the inventory levels and practices at suppliers' facilities can affect the pricing and lag with which steel index affects their prices. Understanding those practices could help to define the suitable lag for steel index parameter in the model. We also suggest investigating geography-related drivers with a total cost of ownership approach incorporating also logistics costs, such as distance to suppliers' facilities and mode of transport. This can change the profitability of each procurement option significantly, particularly now when the fuel prices are also very high.

References

[1] A. Miller, "How a U.S. steel shortage created a market bubble," *CNBC.com*, vol. 1, no. 1, pp. 1–1, 2021.

[2] C. Li, "An analytical method for cost analysis in multi-stage supply chains: A stochastic network model approach," *Applied Mathematical Modelling*, vol. 38, no. 1, pp. 2819–2836, 2014.

[3] NORMET GROUP, "Normet.com : About us." `https://www.normet.com/about/#4e1d6f74`. Accessed: 06.05.2022.

[4] N. G. Management, "Normet Group Financial Statements 2020," *Normet Group Financial Statements*, vol. 1, no. 1, p. 5, 2020.

[5] J. T. Pinto and A. Diemer, "Supply chain integration strategies and circularity in the European steel industry," *Resources Conservation Recycling*, vol. 153, no. 1, pp. 1–16, 2020.

[6] D. Petrovic, M. Kalata, and J. Luo, "A fuzzy scenario-based optimisation of supply network cost, robustness and shortages," *Computers Industrial Engineering*, vol. 160, no. 1, pp. 1–16, 2021.

[7] J. A. Ask and T. M. Laseter, "Cost Modeling: A Foundation Purchasing Skill," *Strategy + Business*, vol. 4, no. 1, pp. 1–14, 1998.

[8] Hamaker, Joseph and Smart, Christian B. and Sanchez, Sam and Sick, Eric, "The art and science of cost modeling." `https://galorath.com/wp-content/uploads/2019/01/The-Art-and-Science-of-Cost-Modeling-NASA-CS-2018.pdf`. Accessed: 26.04.2022.

[9] B. A. Wahab and A. O. Adewuyi, "Analysis of major properties of metal prices using new methods: Structural breaks, non-linearity, stationarity and bubbles," *Resources Policy*, vol. 74, no. 1, pp. 1–16, 2021.

[10] Statistics Finland, "Price of electricity by type of consumer." `https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__ene__ehi/statfin_ehi_pxt_12gx.px/`. Accessed: 11.03.2022.

[11] Statistics Finland, "Labour cost index by industry." `https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__pal__tvki__nj/statfin_tvki_pxt_123m.px/`. Accessed: 11.03.2022.

Appendix

Deviation Figures

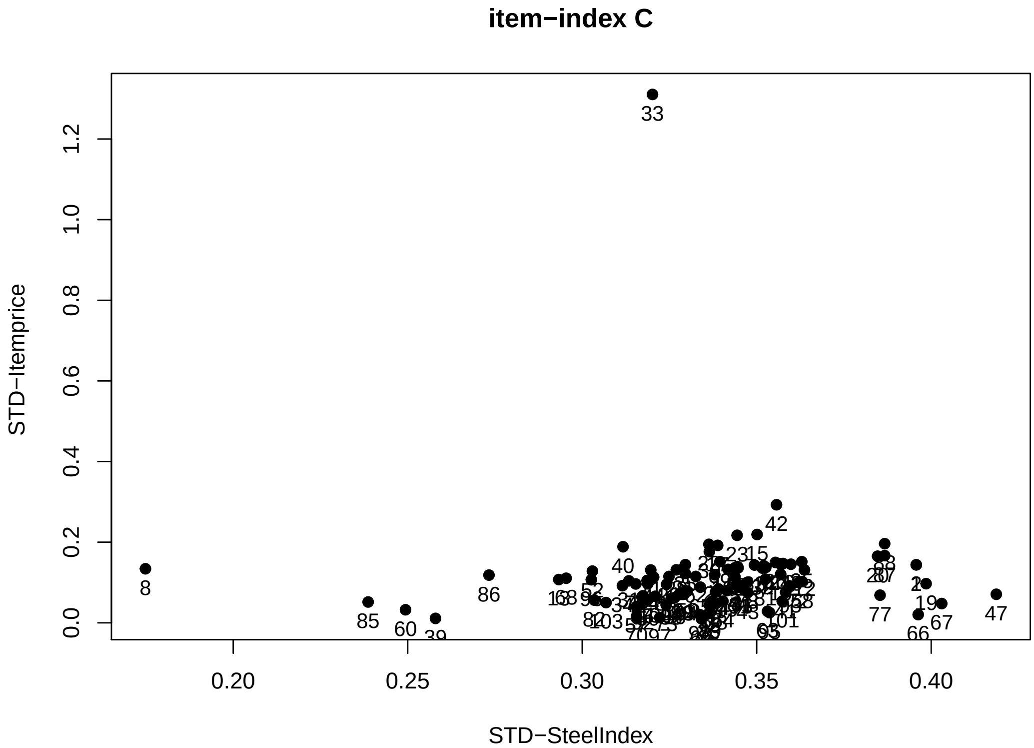


Figure 13: Two dimensional deviation photo of the items for supplier C.

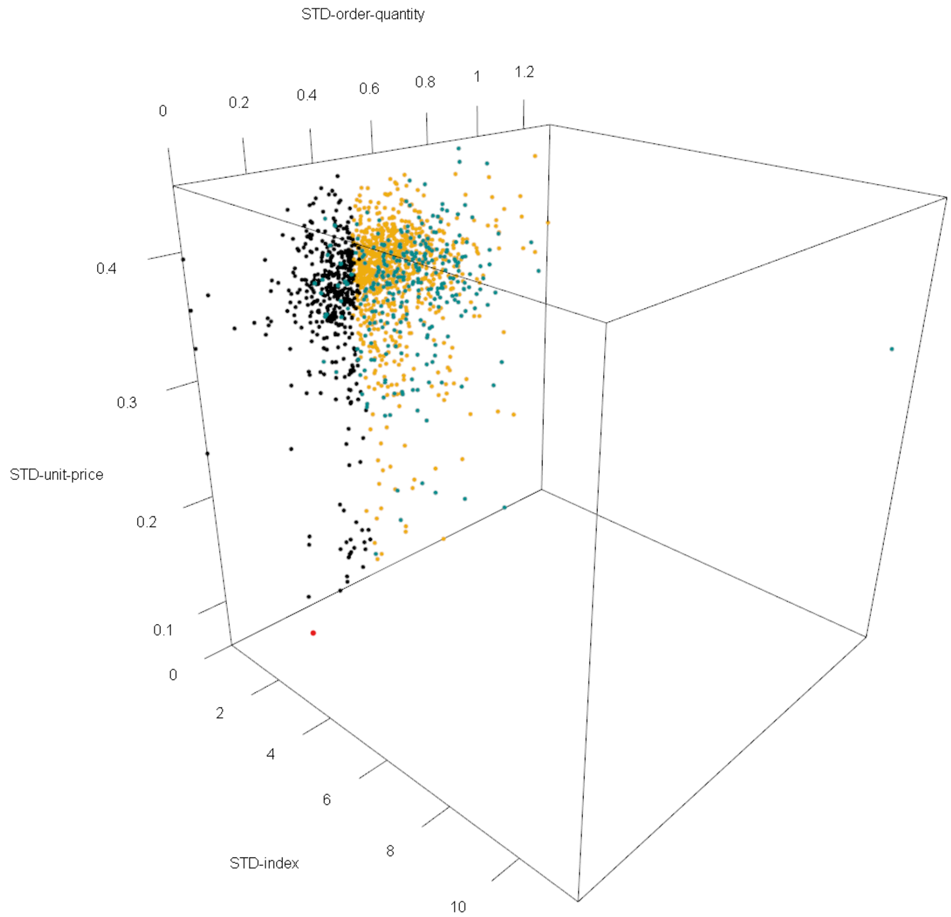
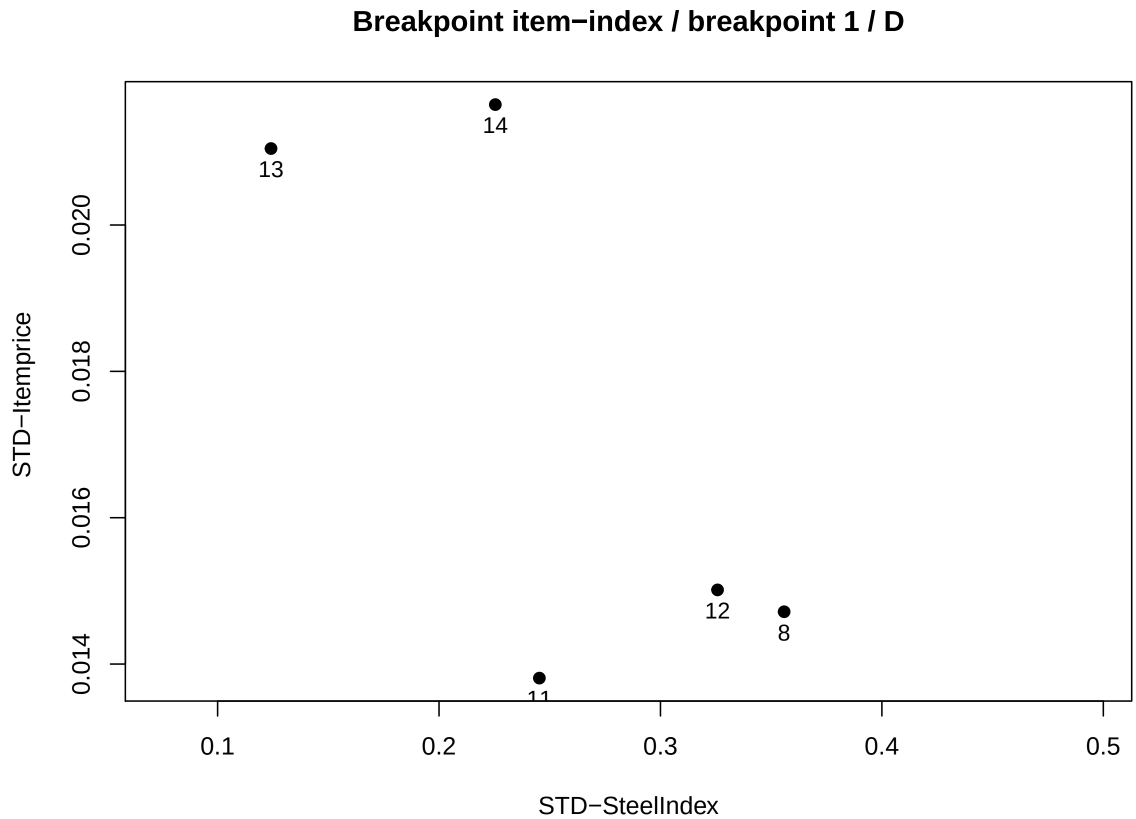
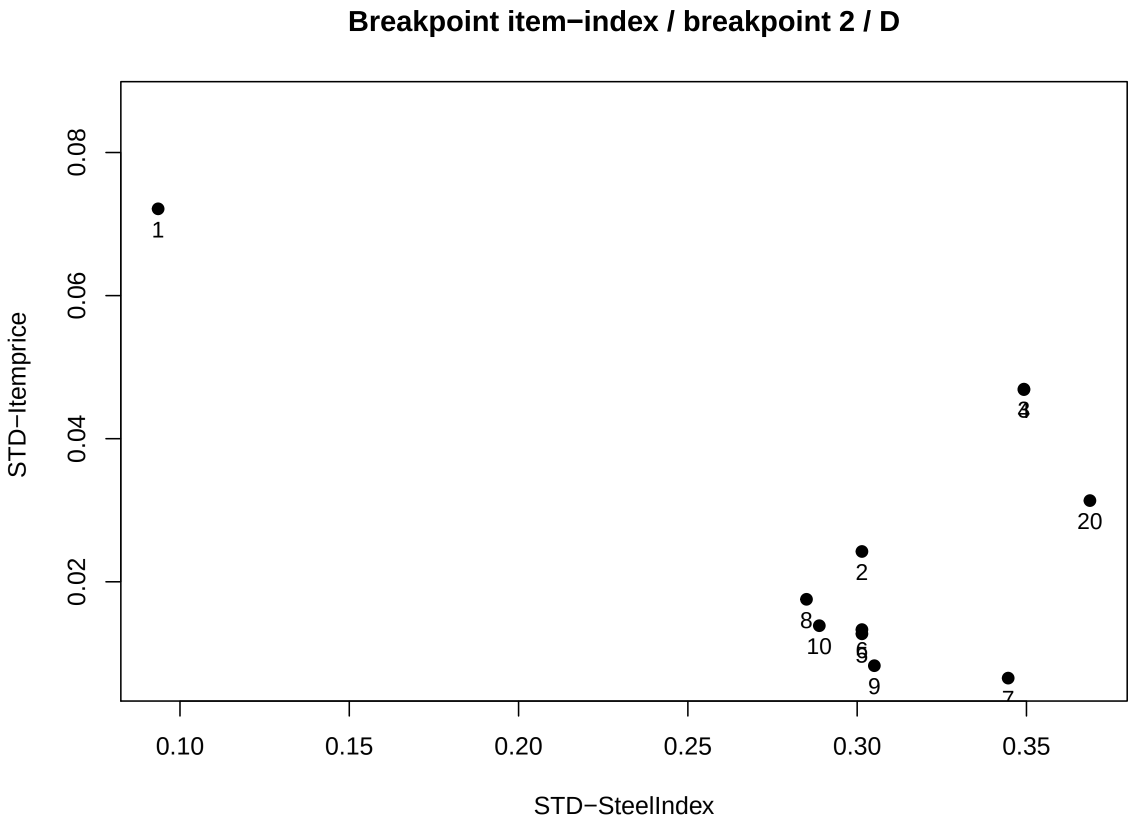
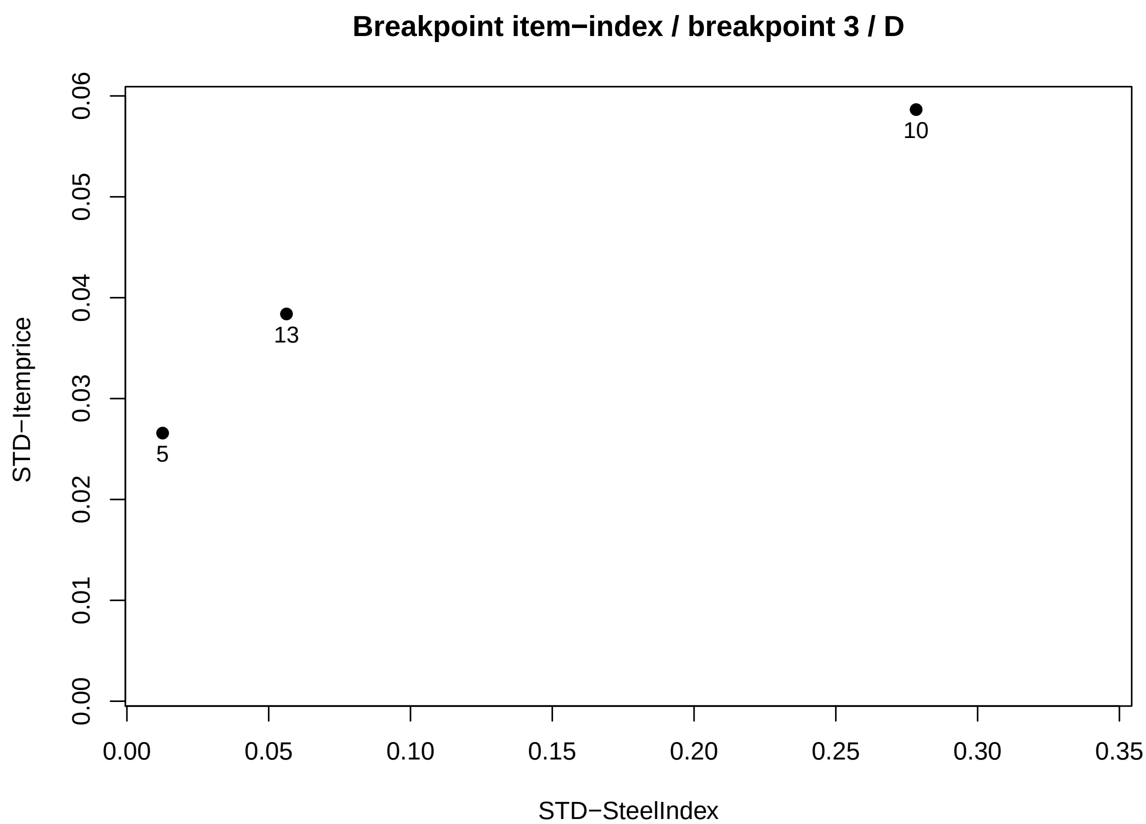


Figure 14: Three dimensional deviation photo of the items for supplier I. Figure does not include fake item numbers due to them making the Figure unreadable.

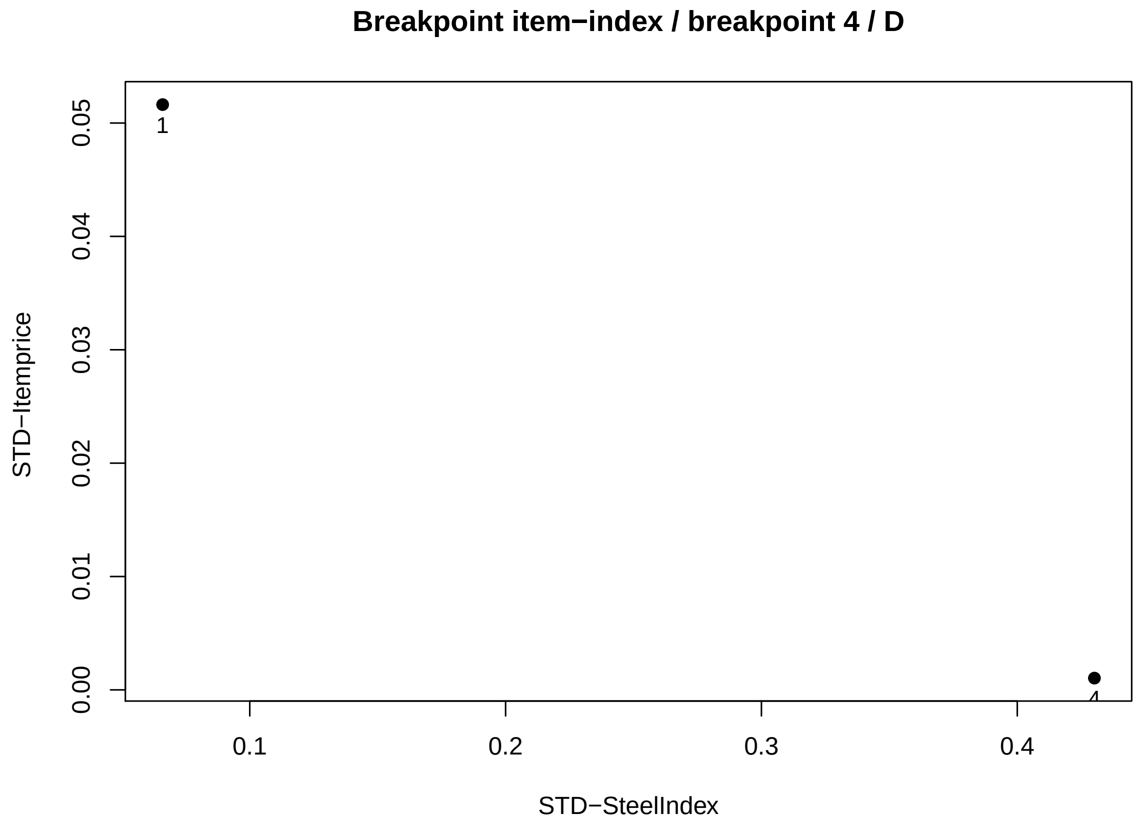


Breakpoint item−index / breakpoint 1 / D
STD−Itemprice
STD−SteelIndex
0.014
0.016
0.018
0.020
0.1
0.2
0.3
0.4
0.5
13
14
12
8
11

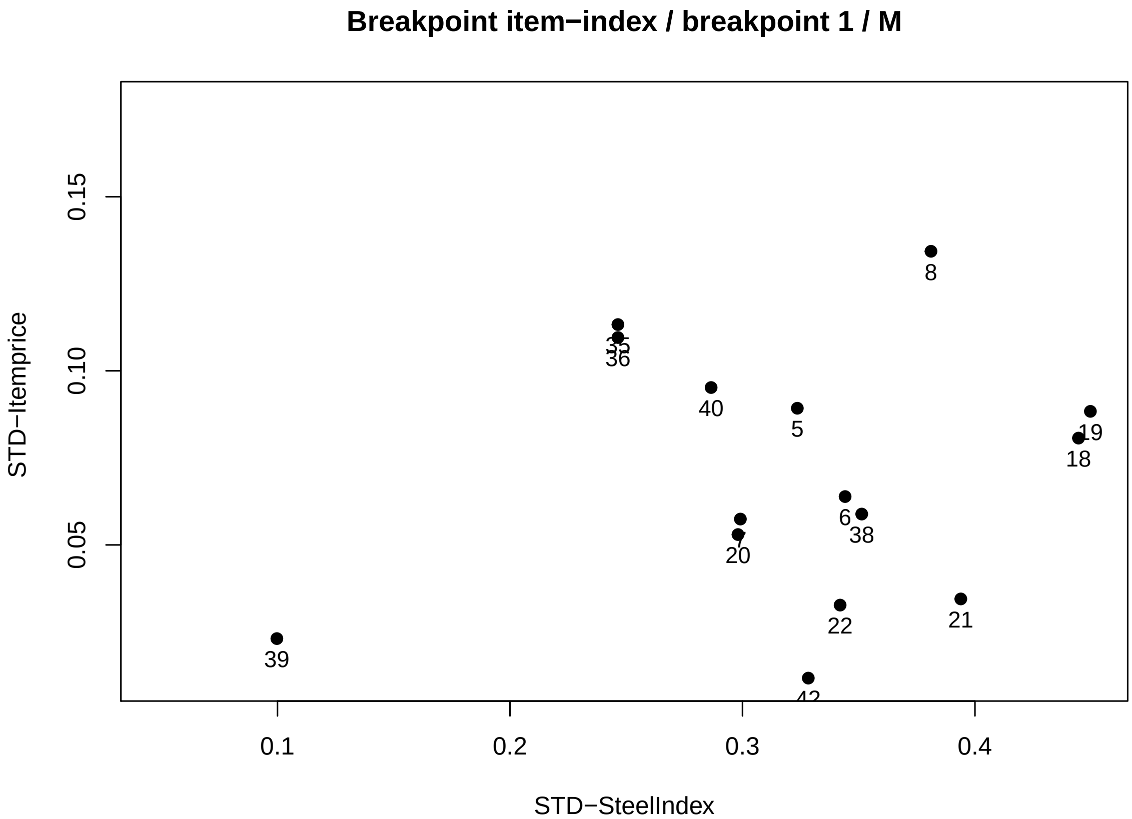

Breakpoint item−index / breakpoint 2 / D
STD−Itemprice
0.08
0.06
0.04
0.02
0.10
0.15
0.20
0.25
0.30
0.35
STD−SteelIndex
1
2
8
10
9
7
20



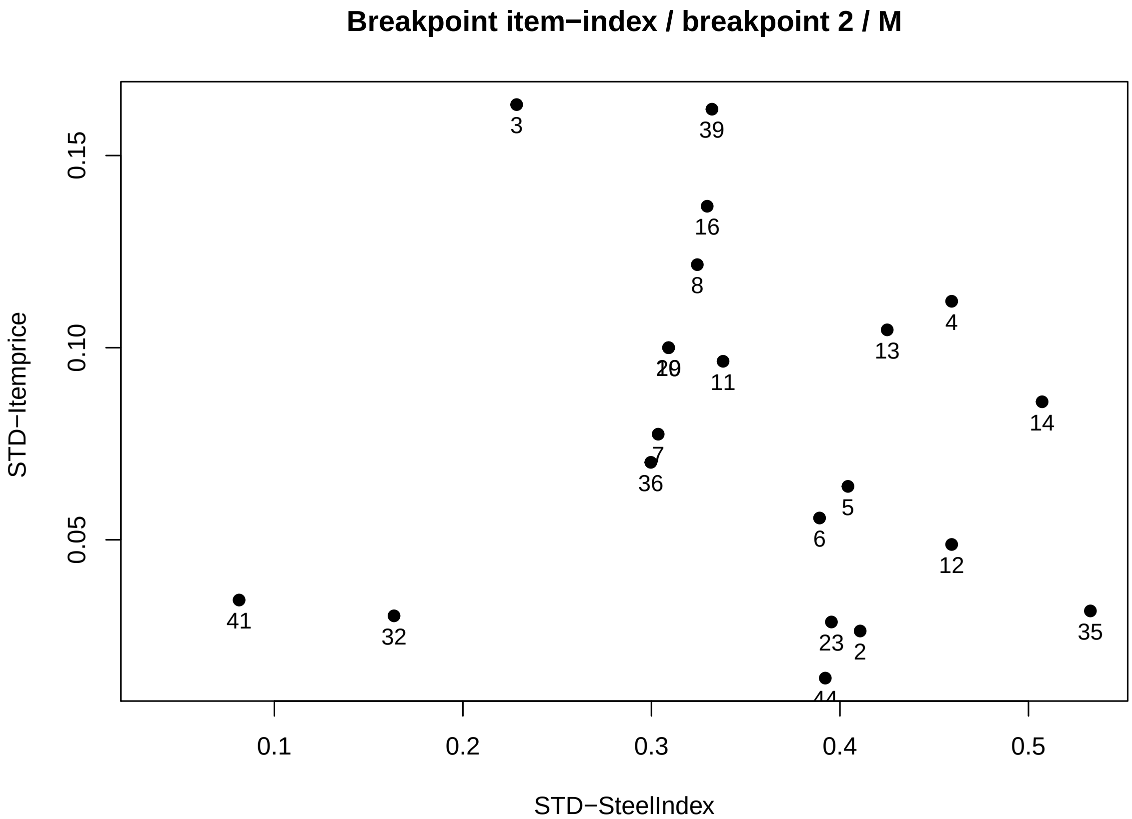
Breakpoint item−index / breakpoint 3 / D
STD−Itemprice
STD−SteelIndex
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
5
13
10

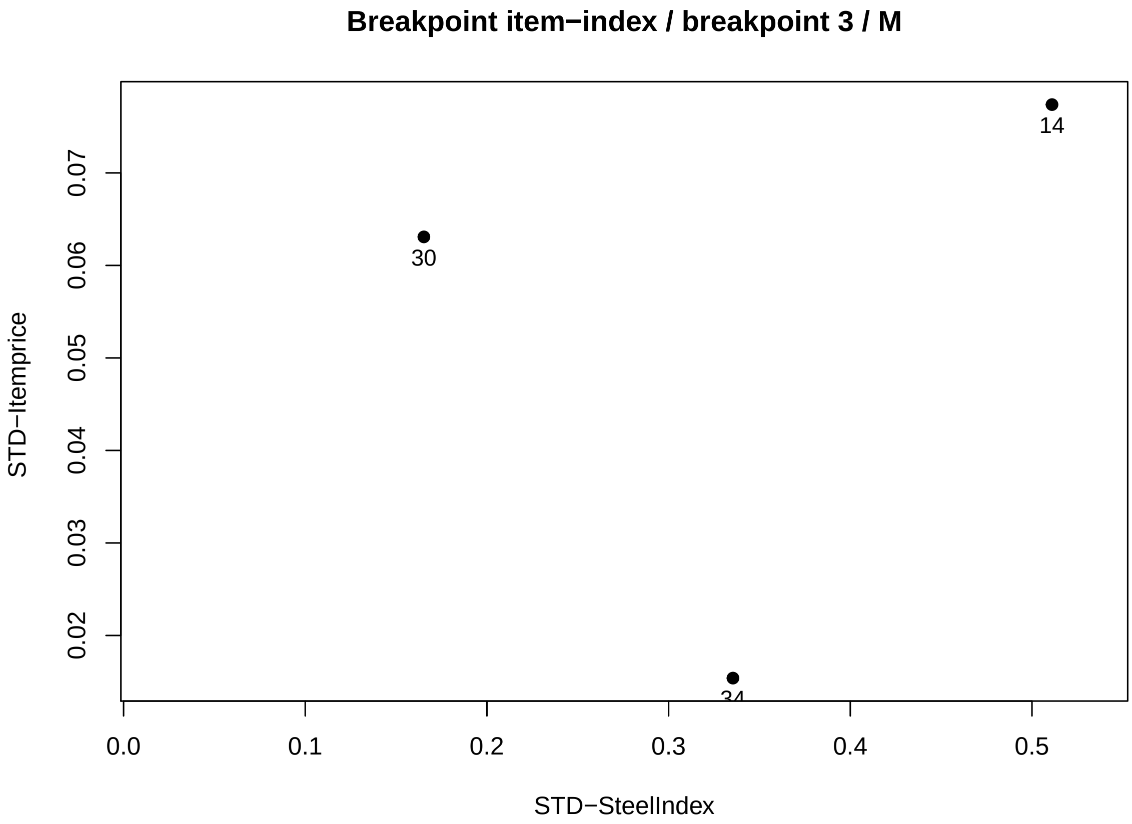


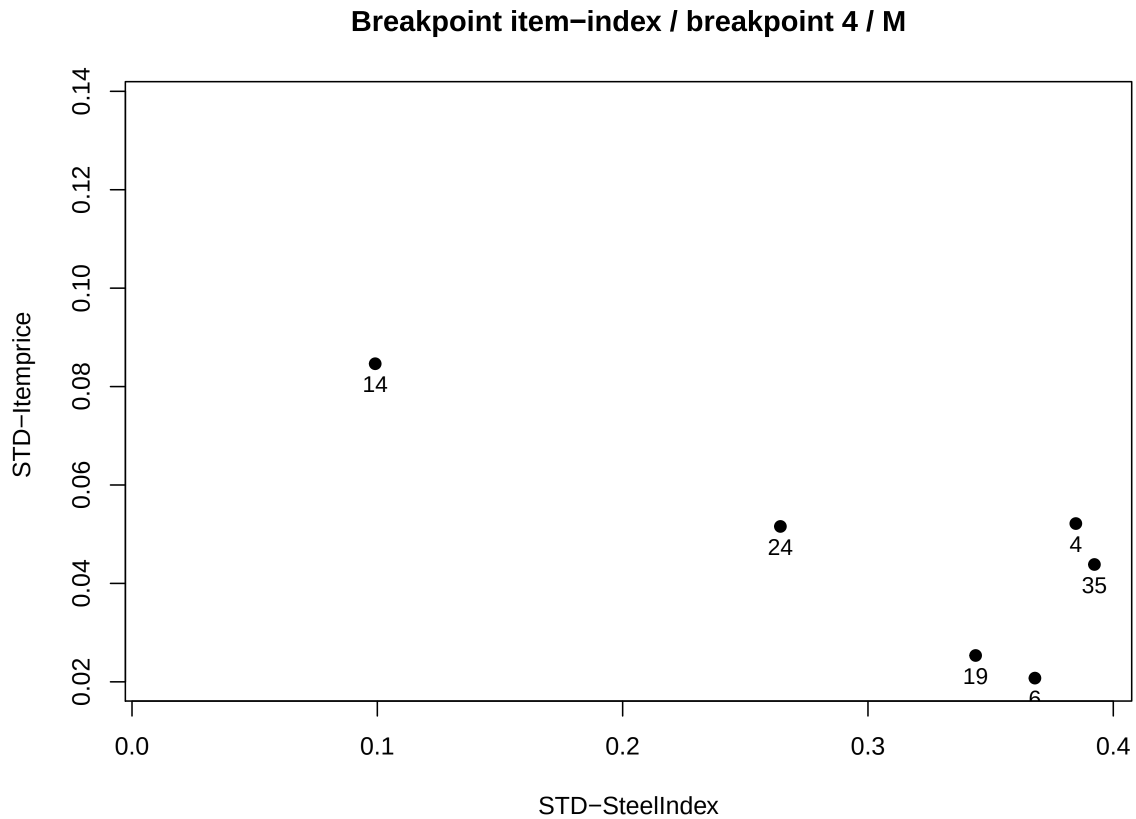
Breakpoint item−index / breakpoint 4 / D
STD−Itemprice
0.00
0.01
0.02
0.03
0.04
0.05
1
4
0.1
0.2
0.3
0.4
STD−SteelIndex



Breakpoint item−index / breakpoint 1 / M
STD−Itemprice
STD−SteelIndex
0.05
0.10
0.15
0.1
0.2
0.3
0.4
8
35
36
40
5
19
18
6
38
7
20
22
21
39
42

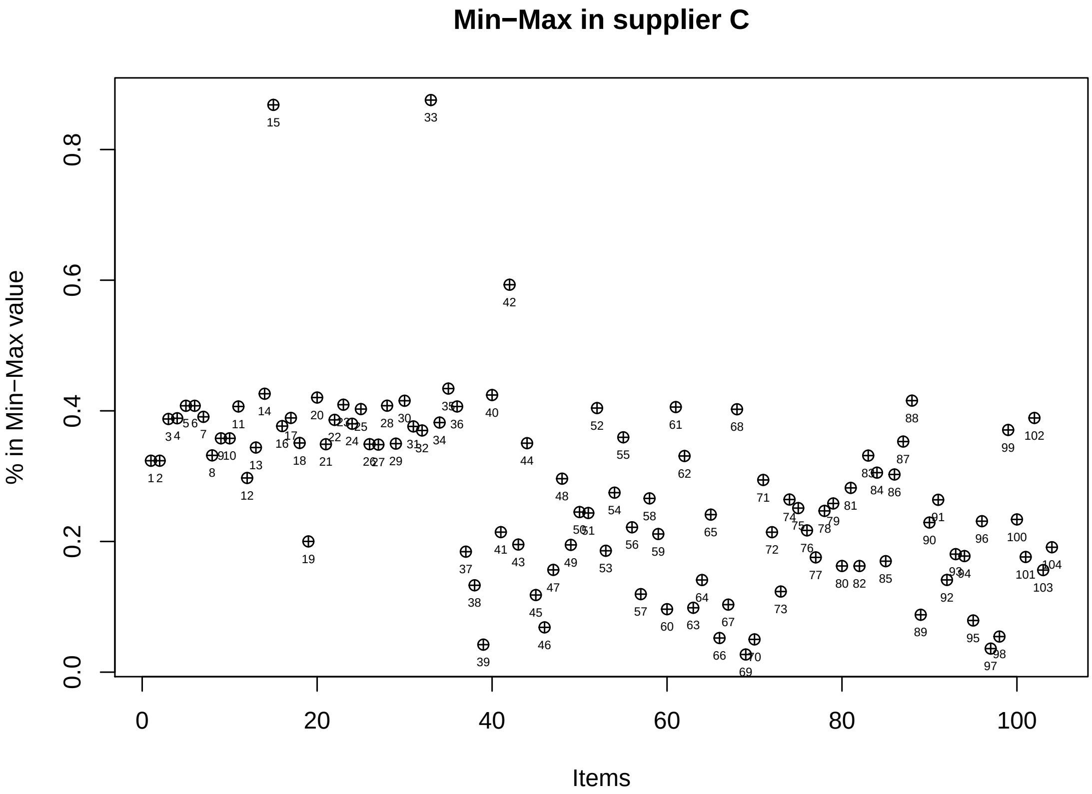

Breakpoint item−index / breakpoint 2 / M
STD−Itemprice
0.05
0.10
0.15
STD−SteelIndex
0.1
0.2
0.3
0.4
0.5
3
39
16
8
4
13
10
20
11
14
7
36
5
6
12
41
32
23
2
35
44

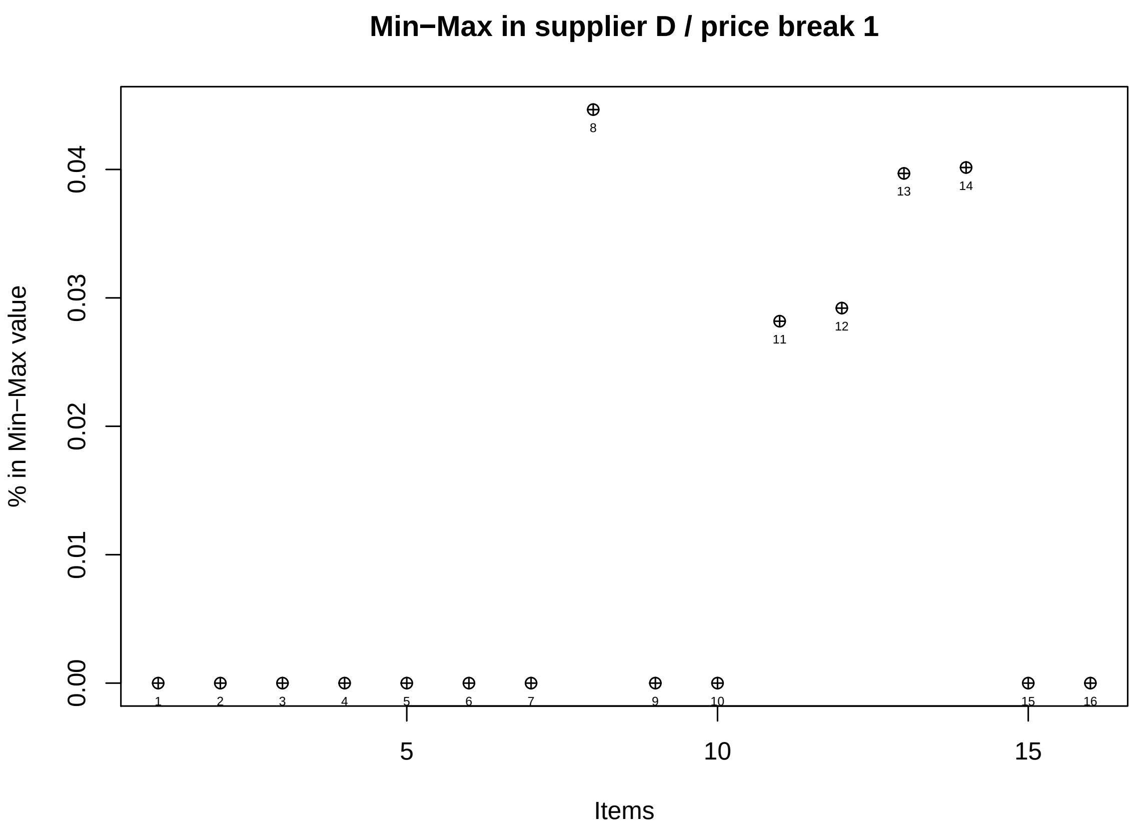

Breakpoint item−index / breakpoint 3 / M
STD−Itemprice
0.02
0.03
0.04
0.05
0.06
0.07
30
14
34
0.0
0.1
0.2
0.3
0.4
0.5
STD−SteelIndex



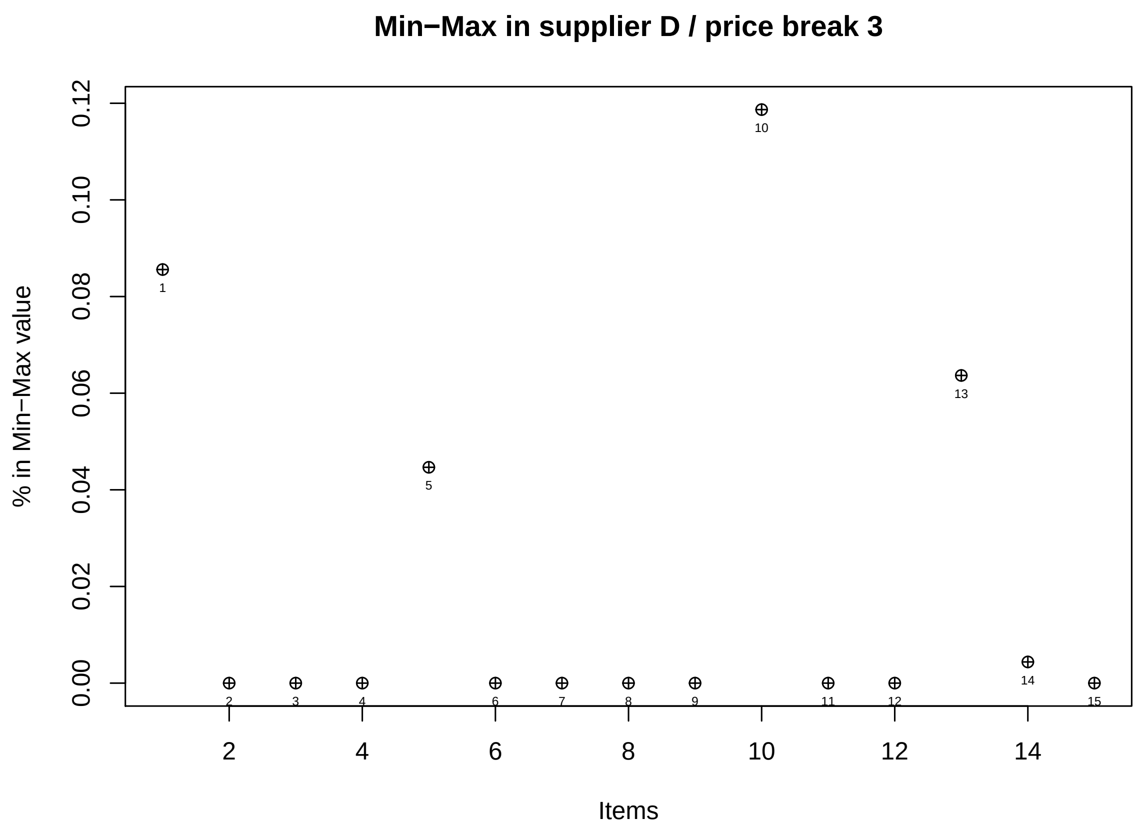
Breakpoint item−index / breakpoint 4 / M
STD−Itemprice
0.02
0.04
0.06
0.08
0.10
0.12
0.14
STD−SteelIndex
0.0
0.1
0.2
0.3
0.4
14
24
19
6
4
35

Minimum and maximum Figures

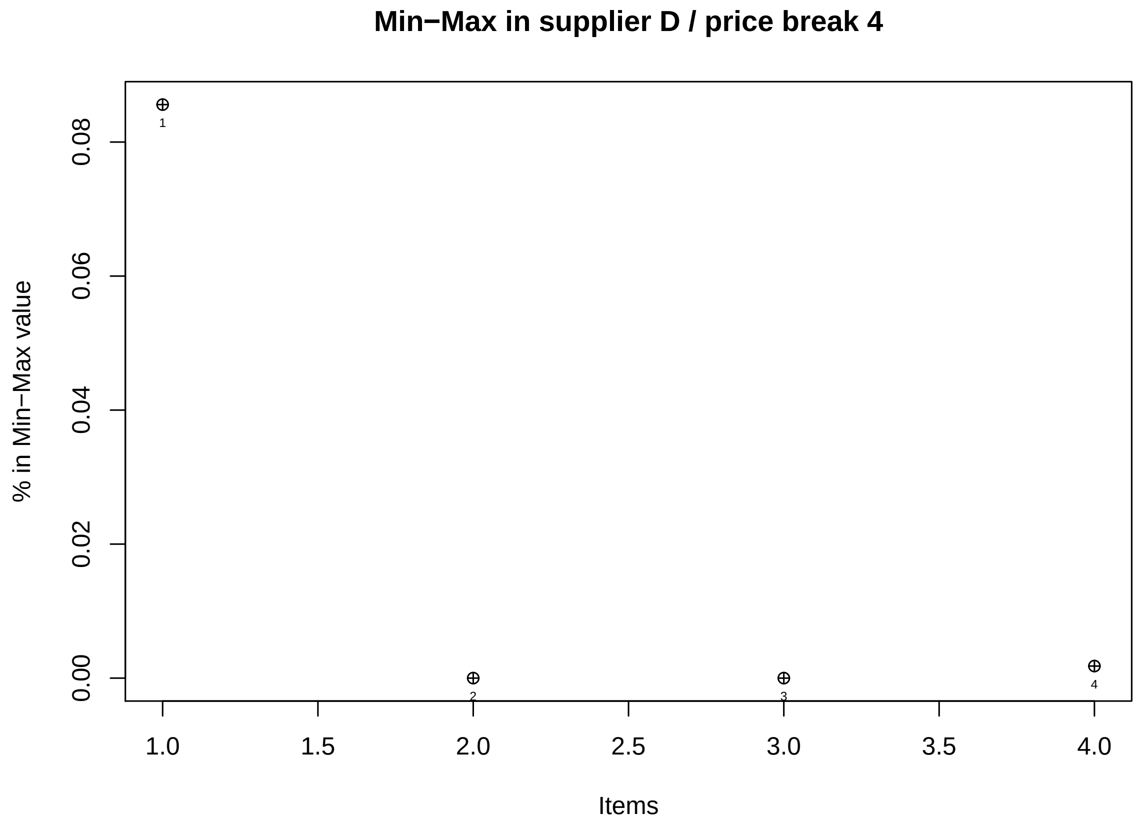


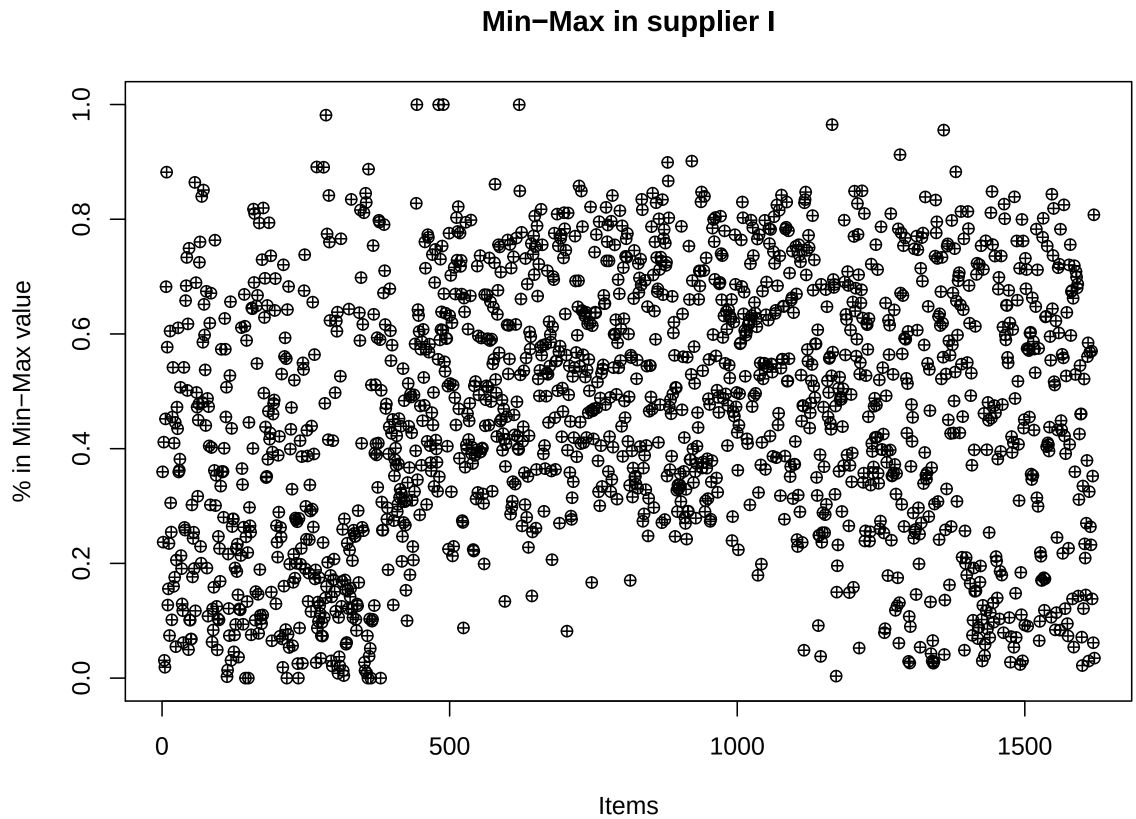

Min−Max in supplier D / price break 1
% in Min−Max value
0.00
0.01
0.02
0.03
0.04
Items
5
10
15
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Min−Max in supplier D / price break 2
% in Min−Max value
0.00
0.05
0.10
0.15
0.20
0.25
5
10
15
20
Items
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20



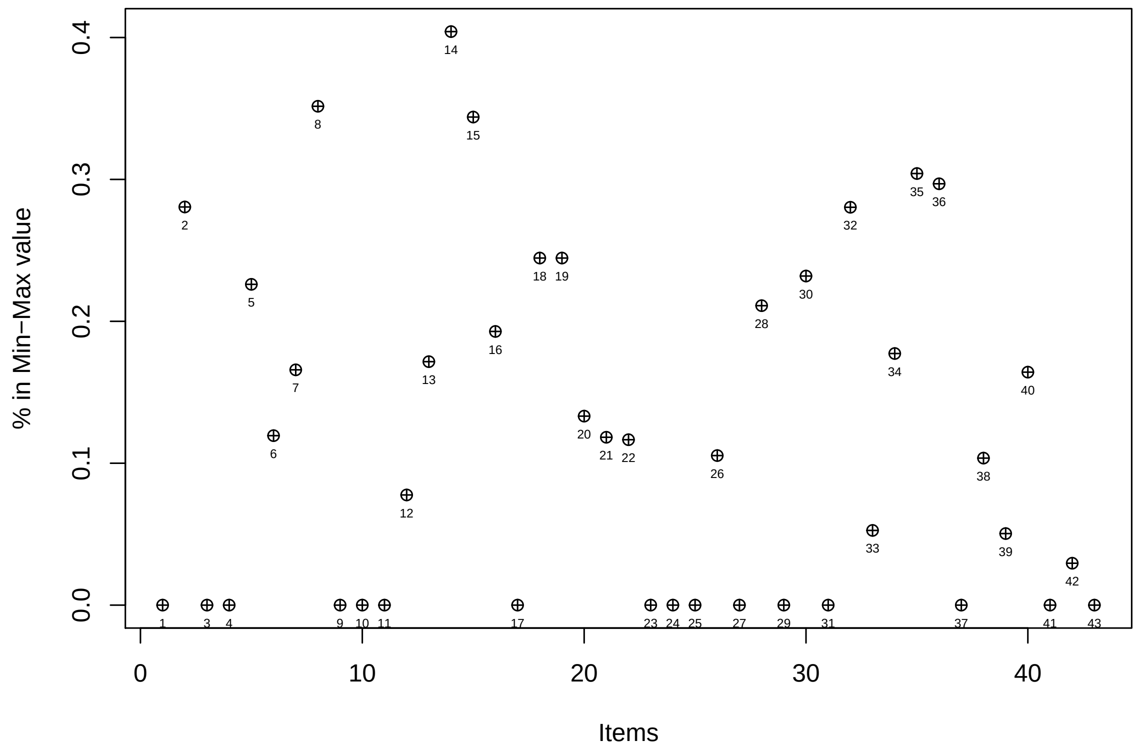
Min−Max in supplier D / price break 3
% in Min−Max value
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Items
2
4
6
8
10
12
14
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

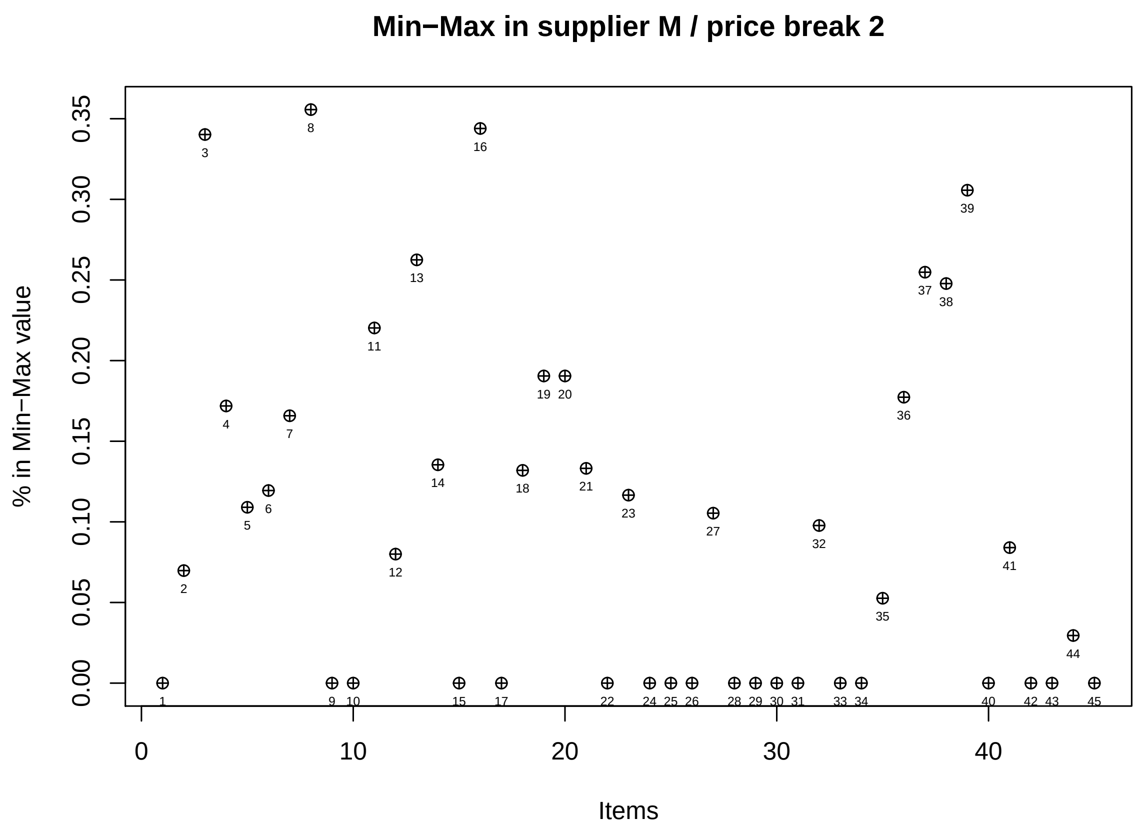

Min−Max in supplier D / price break 4
% in Min−Max value
0.00
0.02
0.04
0.06
0.08
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Items
1
2
3
4



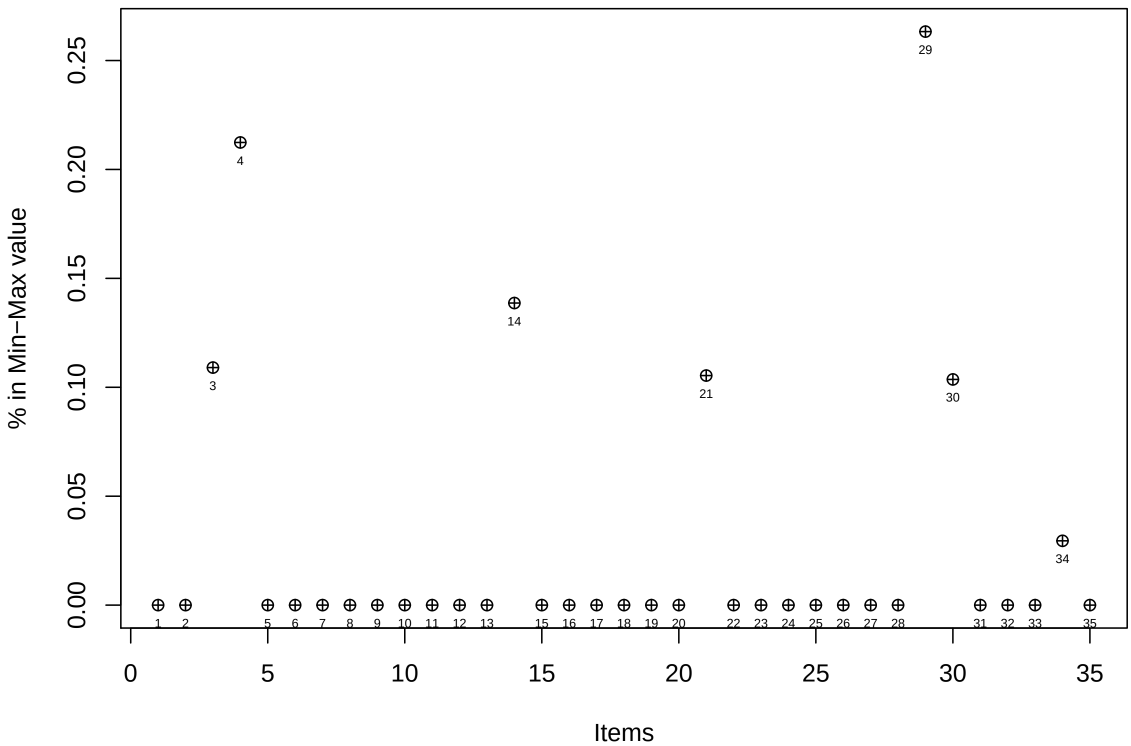
Min−Max in supplier I
% in Min−Max value
Items
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500

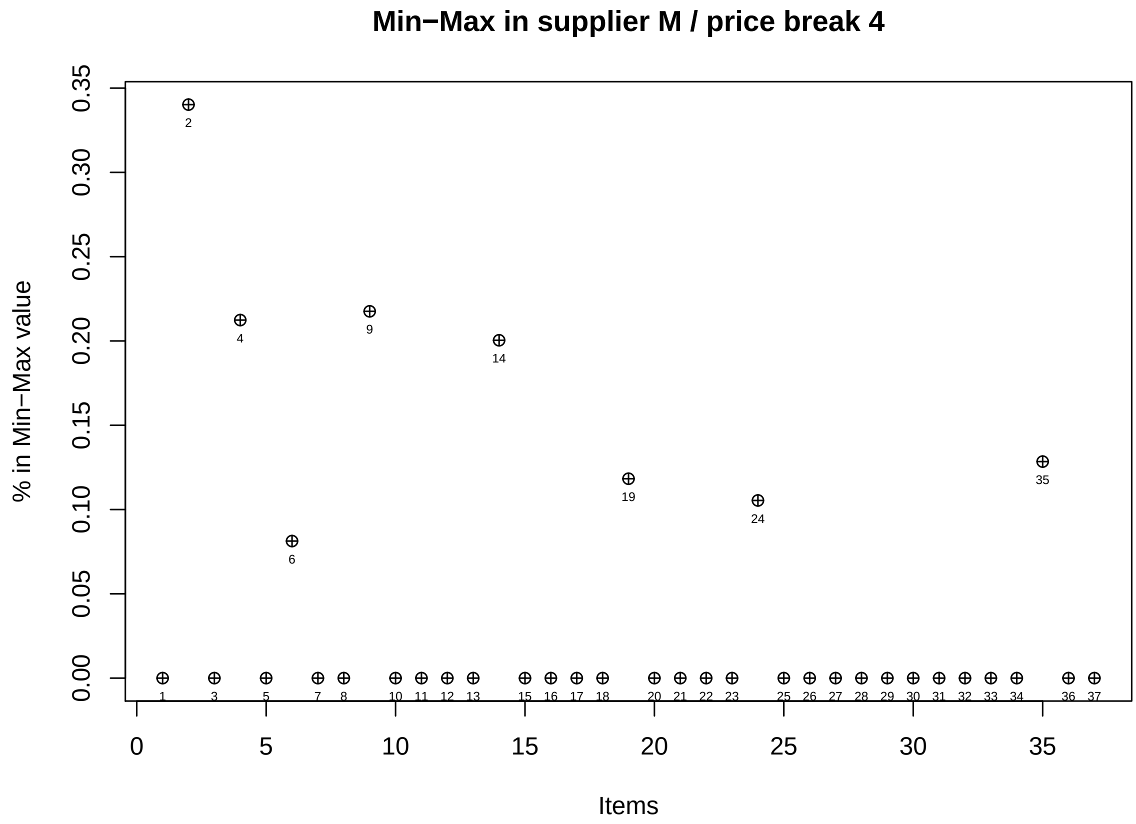
Min−Max in supplier M / price break 1




Min−Max in supplier M / price break 2
% in Min−Max value
Items
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40

Min−Max in supplier M / price break 3





Min−Max in supplier M / price break 4
% in Min−Max value
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
5
10
15
20
25
30
35
Items

The combined metric Figures

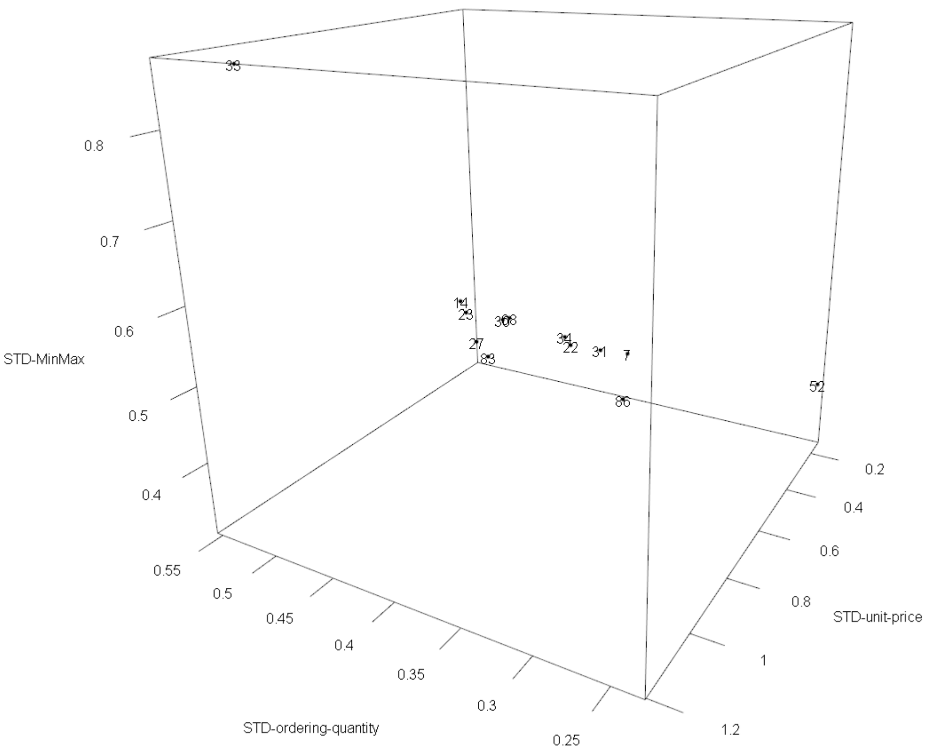


Figure 15: Three dimensional combined metric photo of the items for supplier C.

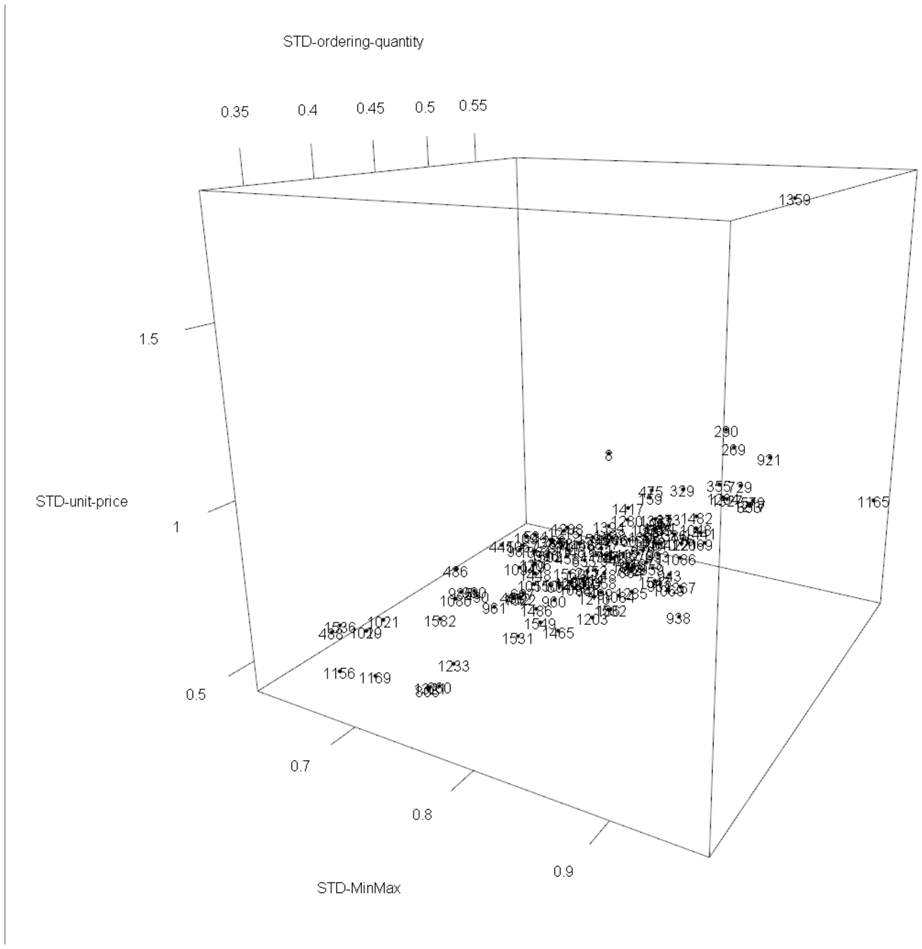


Figure 16: Three dimensional combined metric photo of the items for supplier I.

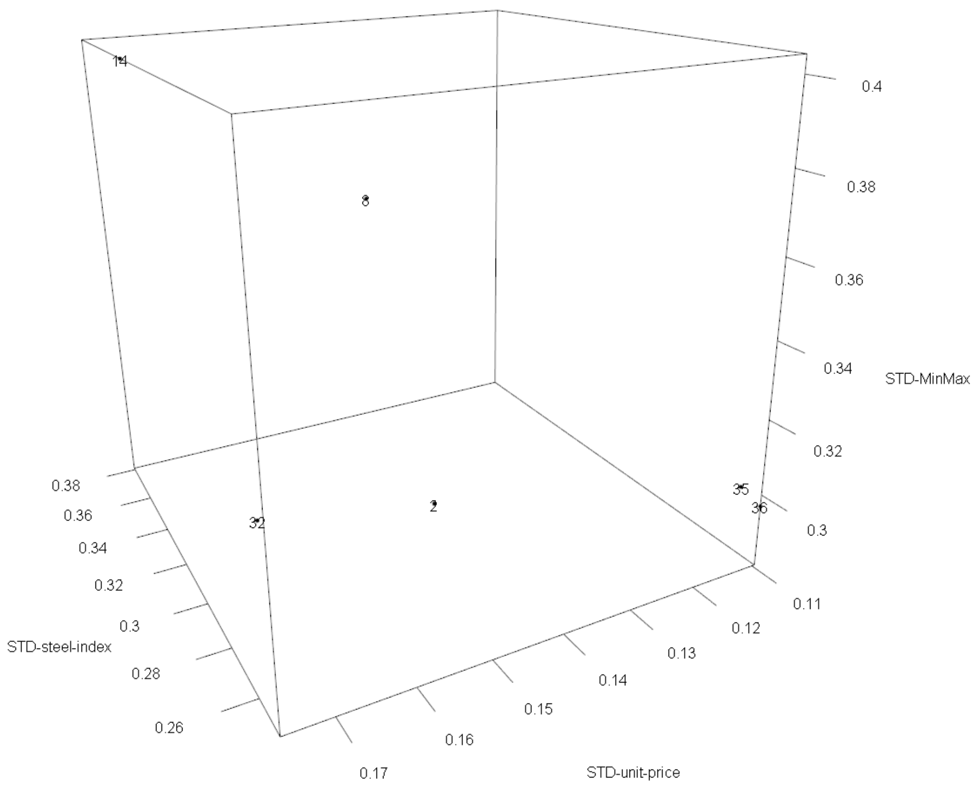


Figure 17: Three dimensional combined metric photo of the first break point of the items for supplier M.

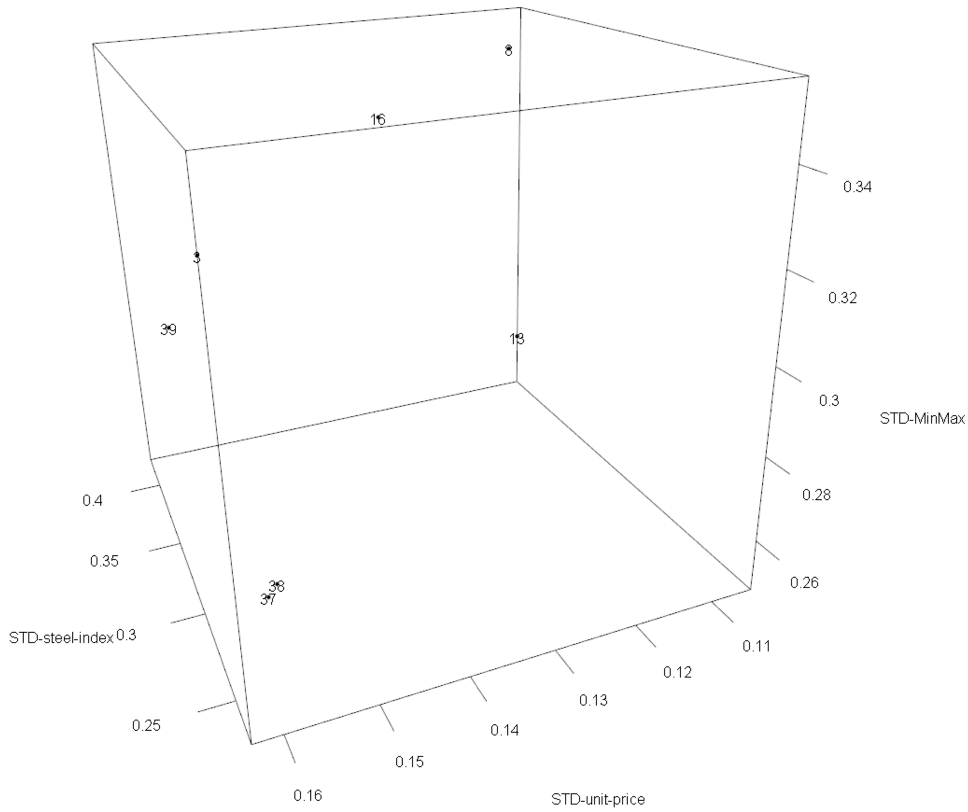


Figure 18: Three dimensional combined metric photo of the second break point of the items for supplier M.

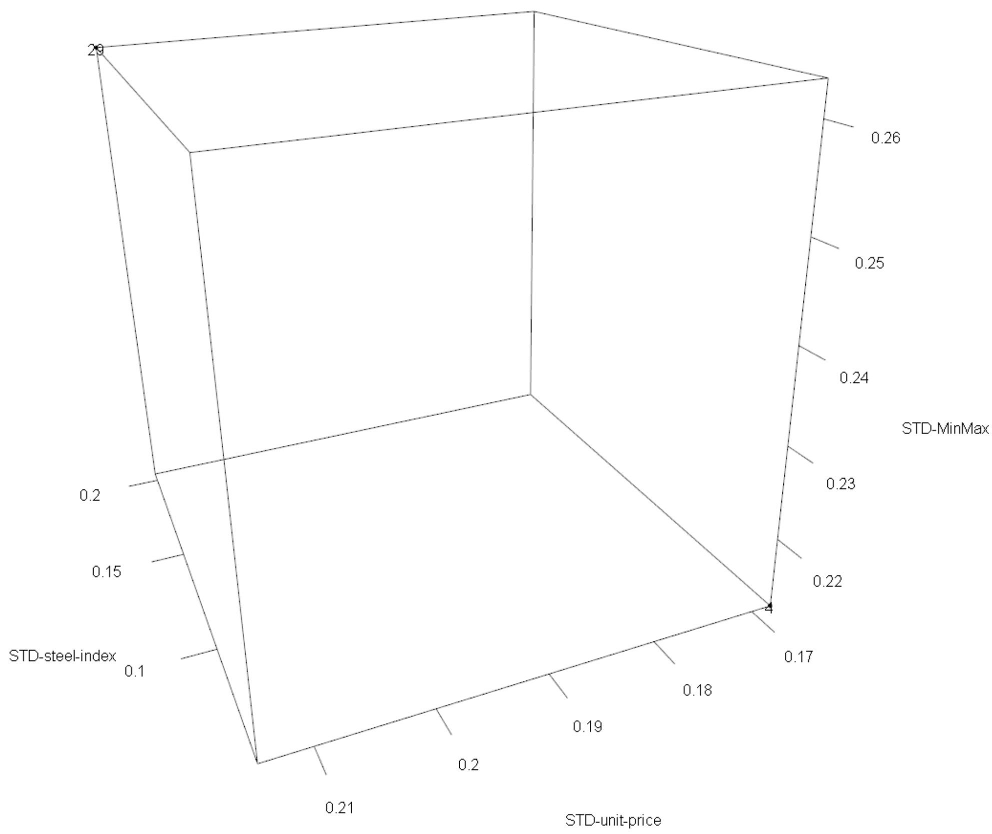


Figure 19: Three dimensional combined metric photo of the thrid break point of the items for supplier M.

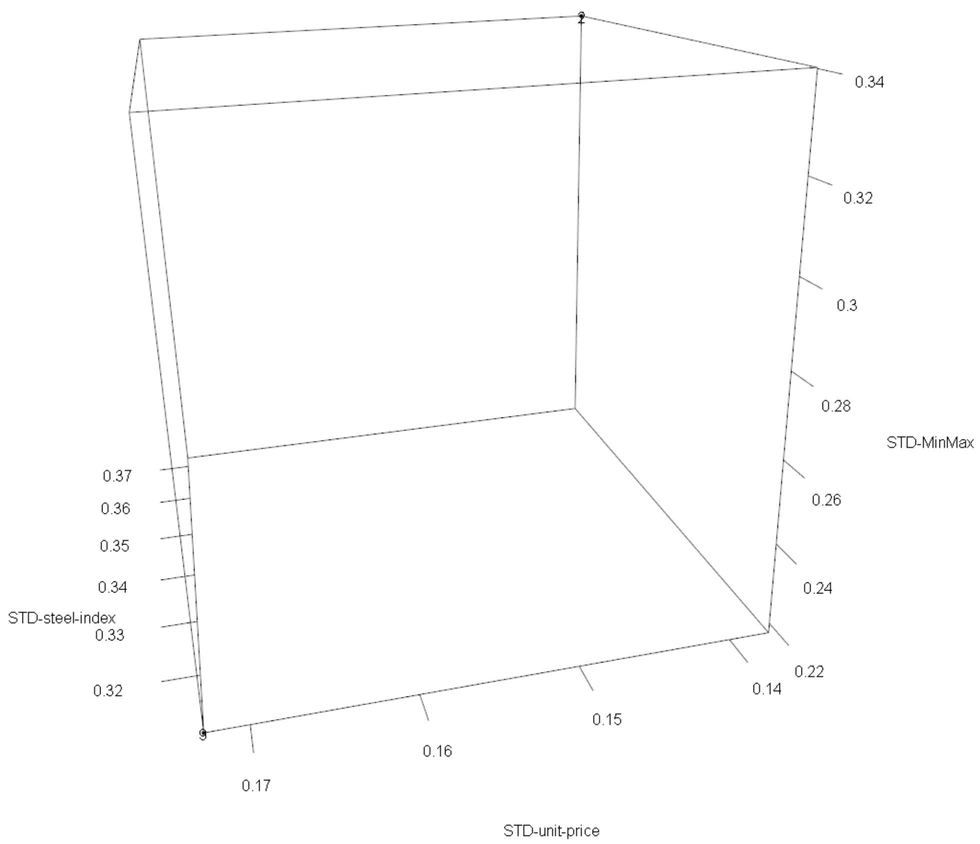


Figure 20: Three dimensional combined metric photo of the fourth break point of the items for supplier M.

Varying the thresholds of supplier I

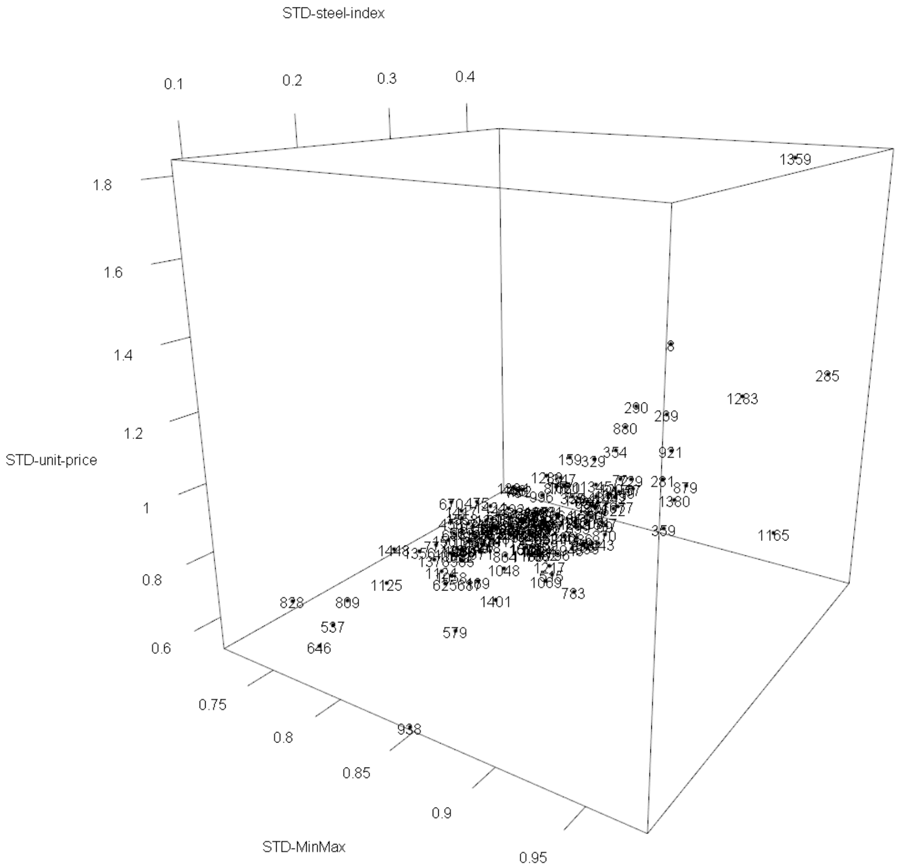


Figure 21: Varying the min-max metric from the threshold value of higher than 0.8 to higher than 0.7 in supplier C.

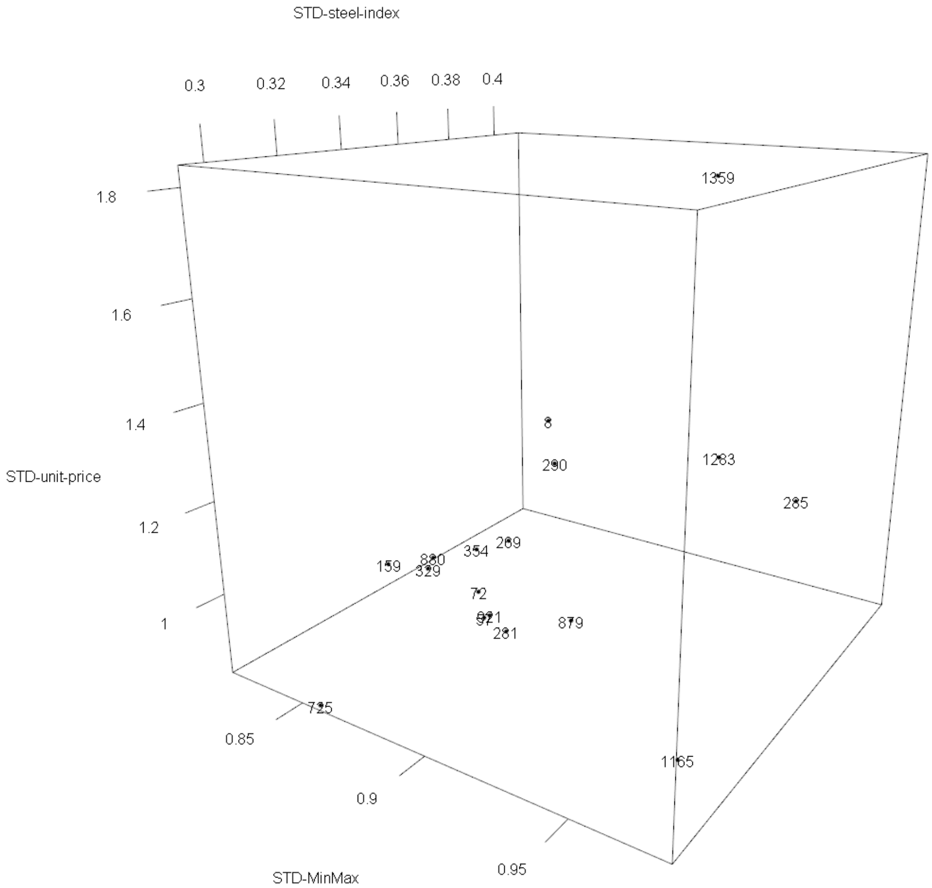


Figure 22: Varying the deviation metric from the threshold value of higher than 0.5 to higher than 0.8 in supplier C.

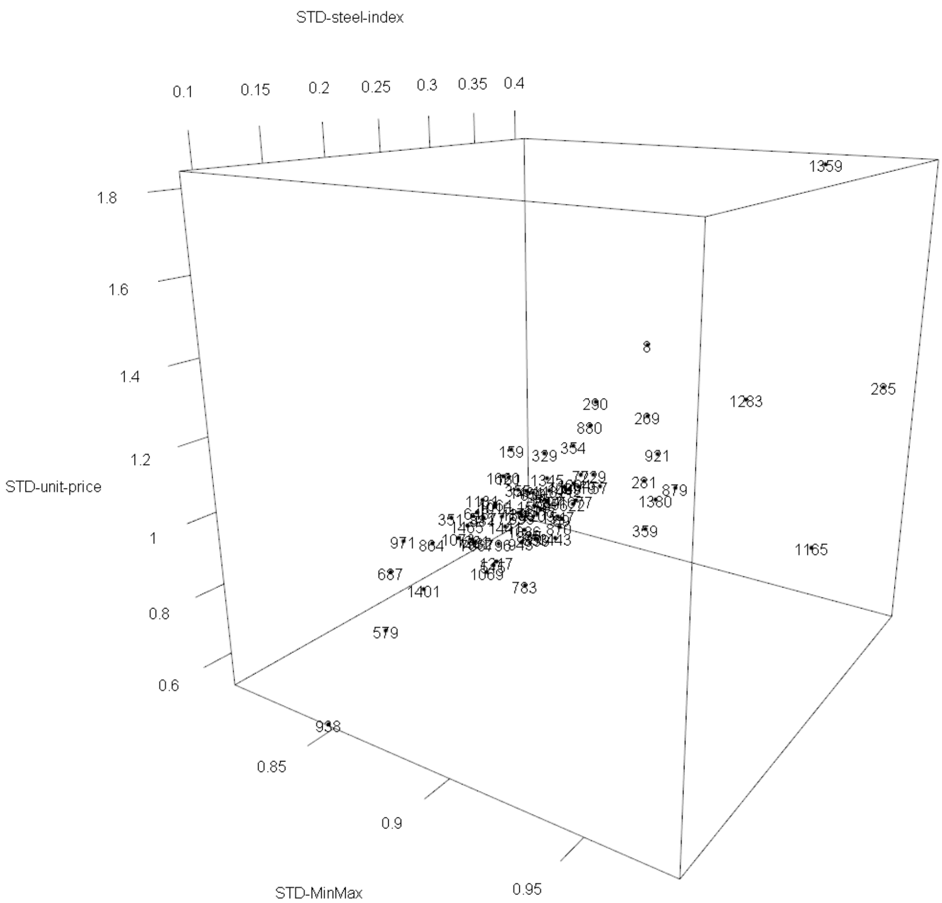


Figure 23: Varying the deviation of steel index metric from the threshold value of lower than 0.8 to lower than 0.5 in supplier C.

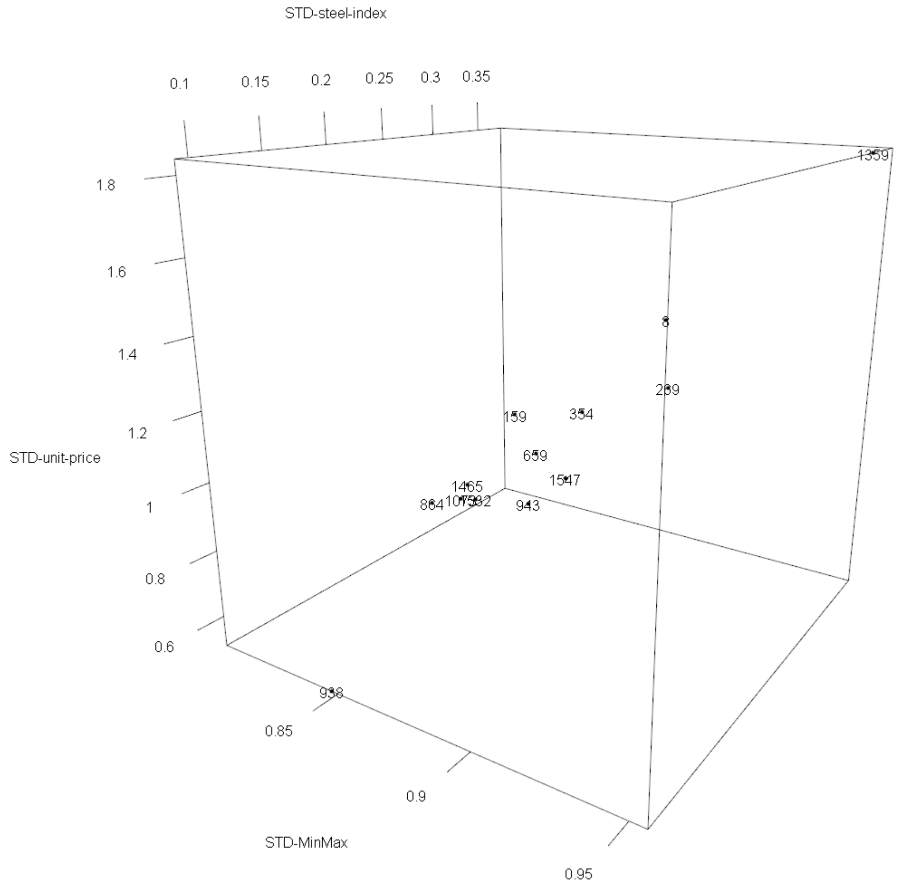


Figure 24: Varying the deviation of quantity metric from the threshold value of higher than 0.2 to lower than 0.5 in supplier C.

Correlation analysis for supplier C

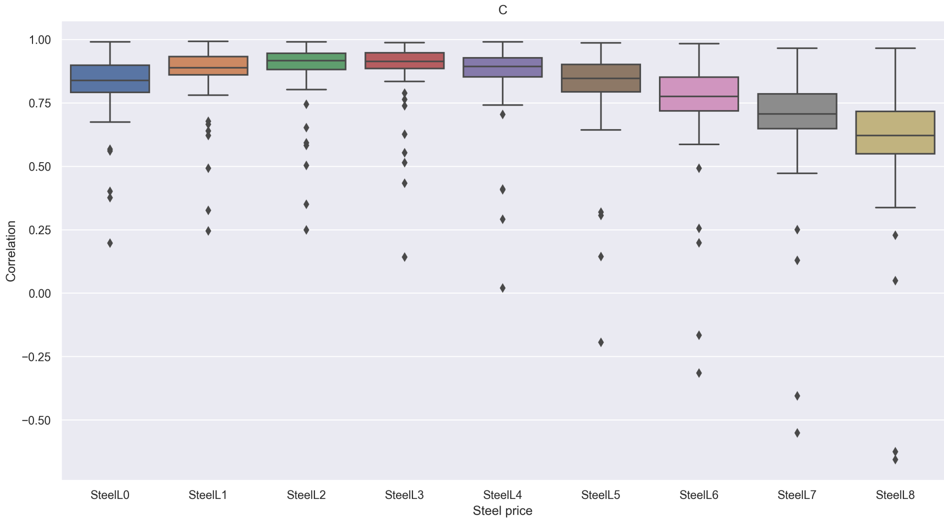


Figure 25: Boxplots of correlation between unit price and steel index with lags of 0-8 months for supplier C at individual item level.

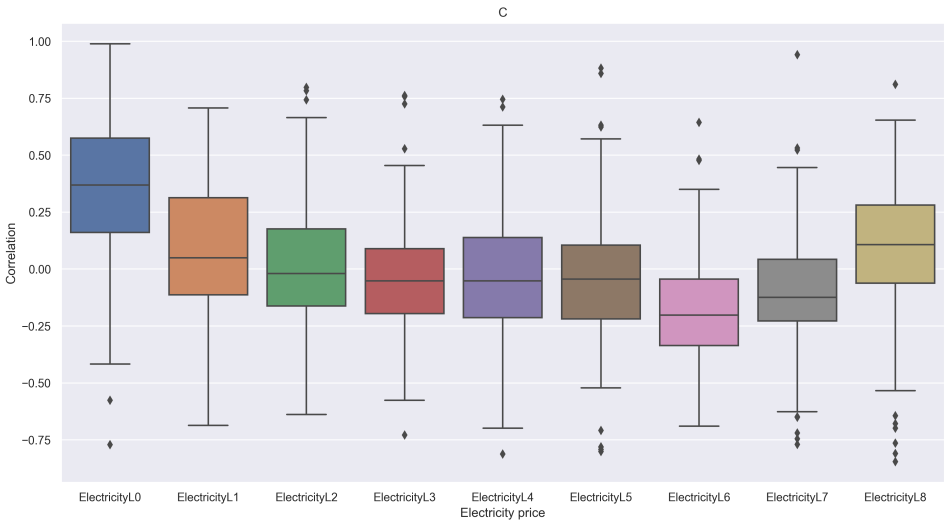


Figure 26: Boxplots of correlation between unit price and electricity price with lags of 0-8 months for supplier C at individual item level.

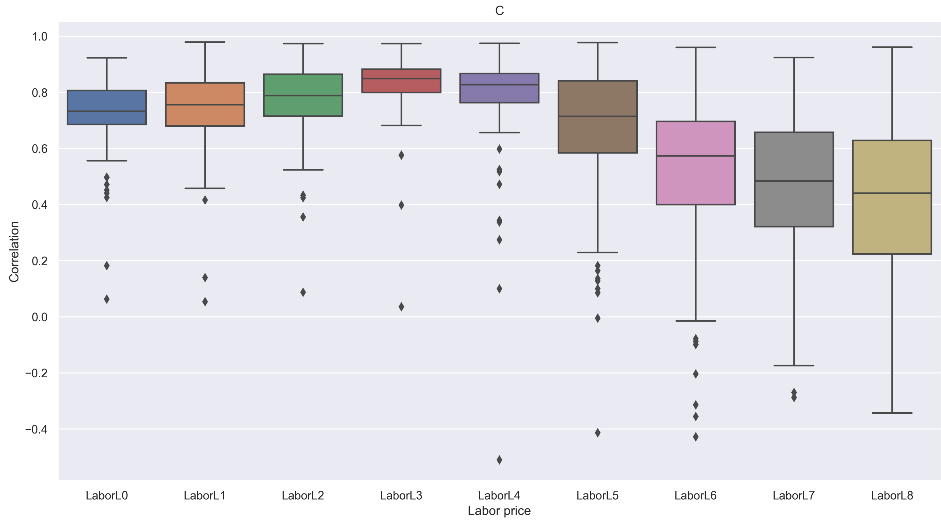


Figure 27: Boxplots of correlation between unit price and labor index with lags of 0-8 months for supplier C at individual item level.

Correlation analysis for supplier D

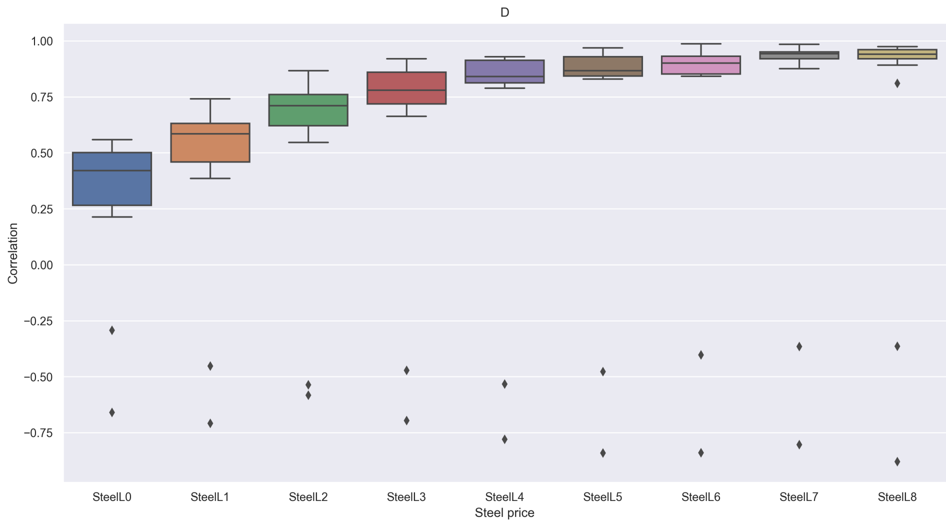


Figure 28: Boxplots of correlation between unit price and steel index with lags of 0-8 months for supplier D at individual item level.

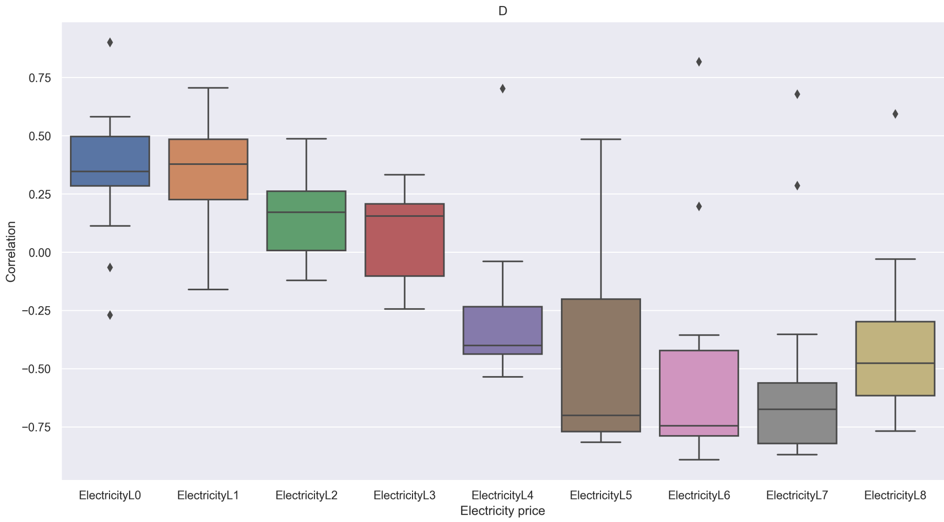


Figure 29: Boxplots of correlation between unit price and electricity price with lags of 0-8 months for supplier D at individual item level.

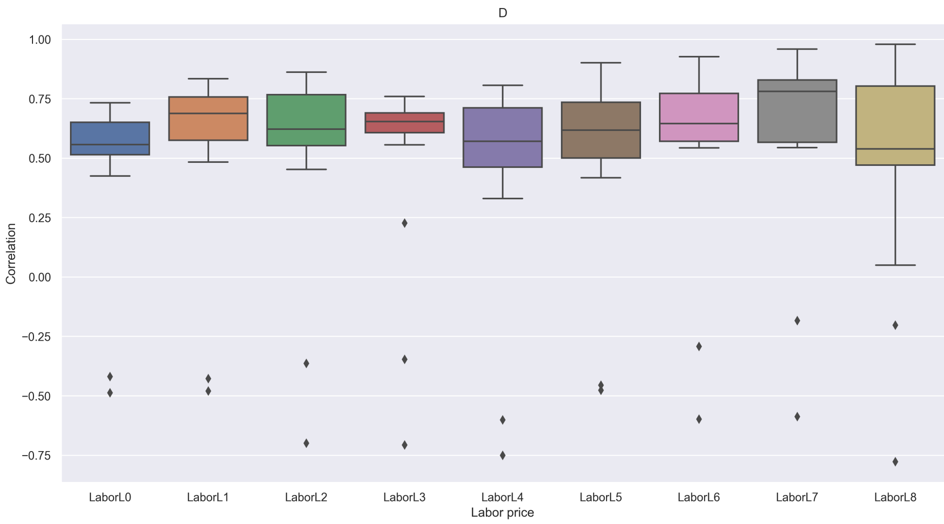


Figure 30: Boxplots of correlation between unit price and labor index with lags of 0-8 months for supplier D at individual item level.

Correlation analysis for supplier M

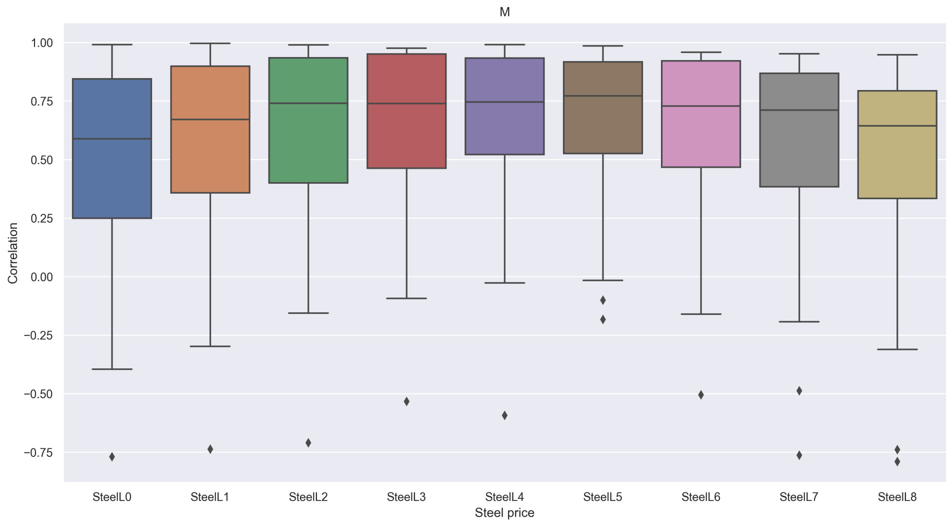


Figure 31: Boxplots of correlation between unit price and steel index with lags of 0-8 months for supplier M at individual item level.

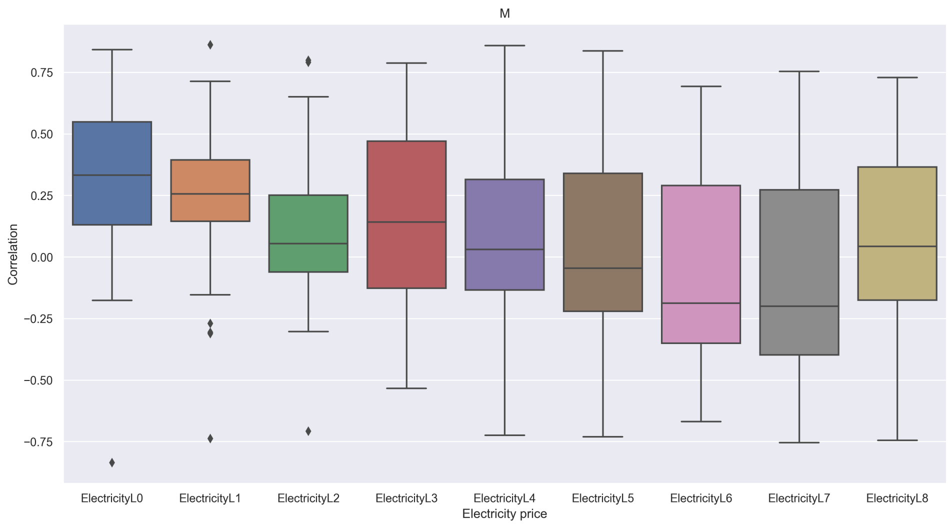


Figure 32: Boxplots of correlation between unit price and electricity price with lags of 0-8 months for supplier M at individual item level.

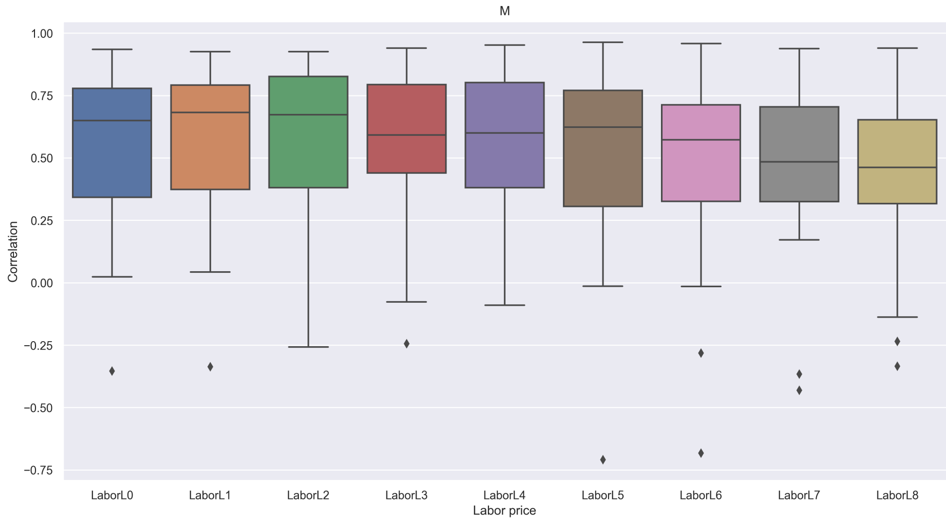


Figure 33: Boxplots of correlation between unit price and labor index with lags of 0-8 months for supplier M at individual item level.

Correlation analysis for supplier I

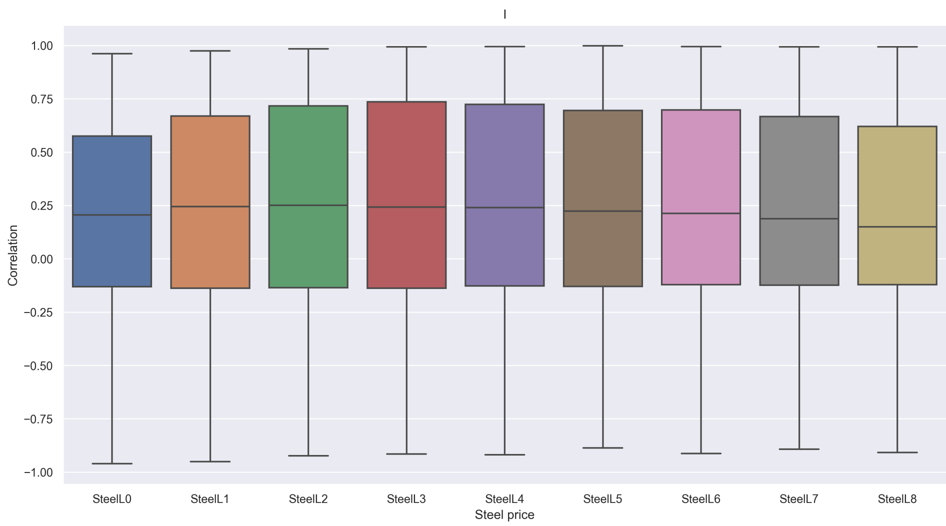


Figure 34: Boxplots of correlation between unit price and steel index with lags of 0-8 months for supplier I at individual item level.

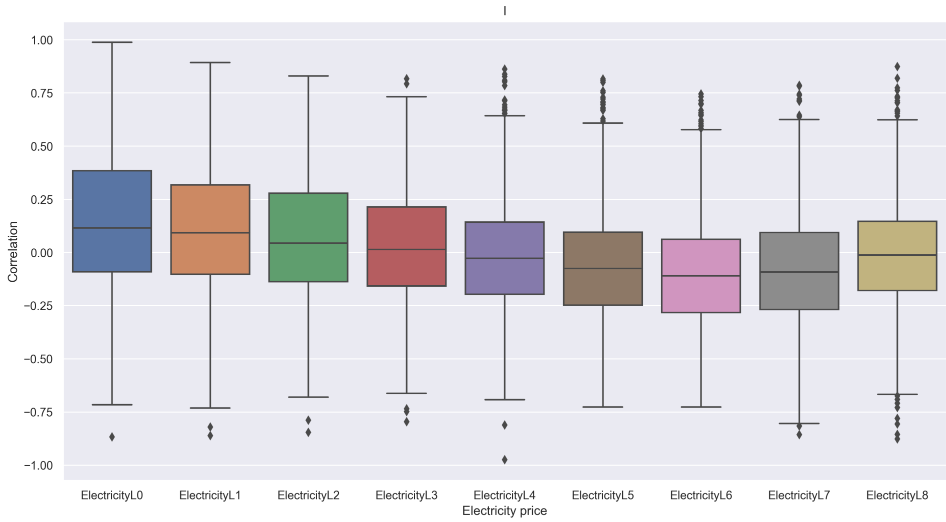


Figure 35: Boxplots of correlation between unit price and electricity price with lags of 0-8 months for supplier I at individual item level.

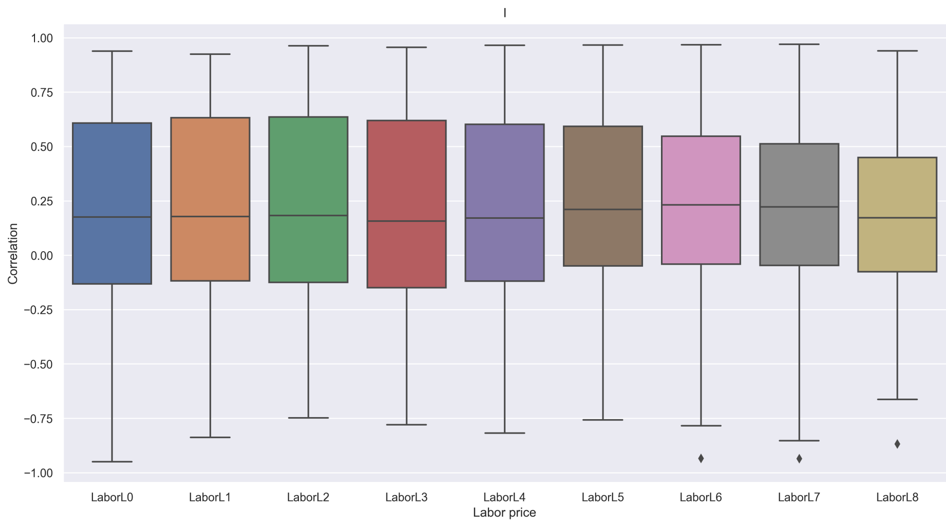


Figure 36: Boxplots of correlation between unit price and labor index with lags of 0-8 months for supplier I at individual item level.

7 Self assessment

7.1 Project progression

Our project followed relatively well the initial project plan and the main phases originally planned were all conducted. However, their duration varied from what was originally planned, resulting into slightly postponed overall schedule. The main departure from the plan was that the original objective was divided to two parts.

Exploratory data analysis took longer than expected, as getting insights from the data on how we should conduct the modelling turned out to be more challenging than expected. Besides, the challenges of analyzing a vast amount of relatively complex data was a bit underestimated, resulting into implementation of the scripts taking longer than expected.

The original objective was divided to two, as it allowed for gathering results with a wider approach than indicated in the original. We believe that this change improved the results of our project, as we could now provide also standalone methods (functioning without the cost model) to detect items with abnormal cost development in the past. This also mitigated the risk of not being able to build the model due to low information value of the data, which partially realized as the models were built only for four suppliers.

7.2 Success of the project

We think that our project was successful given the challenging circumstances. The objectives were achieved, as we were able to build relatively accurate cost models for four suppliers and part of their items as well as determined methods to detect items with abnormal cost development, which both were demonstrated to work in practice. The data had low information value in terms of explaining the unit prices, and thus excluding part of suppliers and items is arguable. Besides, the scope would have been very large, had the models been built for all suppliers and items with such different kinds cost behaviors and probably different cost drivers.

A setback from the perspective of Normet is that the cost model was built at the individual supplier-item level, because it makes implementation more difficult. However, with our limited knowledge of Normet's context and clustering methods, grouping the items and building a model with corresponding aggregation level was not realistic. Technical Item Group level was explored but it was not feasible. Despite the low level of aggregation, our model can be used to estimate the overall cost development and the magnitude of cost changes as a result of steel index changes - which is of huge value with current unstable steel markets. Regarding the second objective, we believe that our methods are valuable even though they are not so practical to implement as they require manual iteration of the thresholds. With the groundwork we have done, we think that Normet's analysts can improve the methods so that the limits can be set more in a more automated way.

7.3 Improvement areas

In terms of our internal working, we could have kept more consistent pipeline for conducting the analyses. Even though we had agreed roles and responsibilities for conducting analyses, team members were conducting analyses with their own preferred methods, lacking integration. Had we used single coding language and done all scripts in an integrated fashion, we could have saved lot of time and effort in the final phases of the project.

In our exploratory data analysis, we think that we could have stayed at a more aggregate levels longer, before diving deep to understand cost behaviors of individual items. This would have enabled us to develop our understanding gradually, instead of directly jumping to analyze individual items. Had we gone slower and gradually from high levels of aggregation to item level of aggregation, we could have been better able to understand the dynamics for grouping items.

One thing that could be improved is to get all relevant information in the early phases of the project. We got the purchase order and steel index data in the beginning but information about the pricing logic only later, when exploratory data analysis had been ongoing for long time already. As a result, part of our analyses needed to be adjusted based on this information. We as a team could have better thought through in the beginning what information is relevant, and asked the right questions faster. However, it was certainly better to get information piece by piece later than not at all, the weekly meetings enabling this continuous learning process. But most importantly it was the great communication and support from Normet's side which helped us to improve our understanding and steer our project to right direction. We also hope that witnessing closely and being part of our analysis process has been a learning experience for Normet!