Demand Forecasting and Inventory Control in Volatile Markets using Machine Learning

Zur Erlangung des akademischen Grades einer

DOKTORIN DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.A. M.Sc. Miriam Kobold

Tag der mündlichen Prüfung: 05.12.2025

Hauptreferent: Prof. Dr. Kai Furmans
Korreferentin: Prof. Dr. Anne Meyer

Determination of the Future Order Variability in Volatile Supply Chains via Behavior-Based Segment Forecasting – Applications from the Automotive Aftermarket

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als Doktorandin bei der Robert Bosch GmbH. Ich möchte mich an dieser Stelle bei allen Personen bedanken, ohne deren Mithilfe die Anfertigung dieser Promotionsschrift niemals zustande gekommen wäre:

Mein Dank gilt zunächst Herrn Prof. Dr. Kai Furmans, Leiter des Instituts für Fördertechnik und Logistiksysteme (IFL) und gleichzeitig Referent für diese Arbeit, sowie Dr. Welf Schneider von der Robert Bosch GmbH, für die Betreuung dieser Arbeit.

Ihre bzw. deine freundliche Hilfe und die mannigfache Ideengebung haben mir einen kritischen Zugang zu dieser Thematik eröffnet. Die zahlreichen Gespräche auf intellektueller und persönlicher Ebene werden mir immer als bereichernder und konstruktiver Austausch in Erinnerung bleiben.

Ich danke Frau Prof. Dr. Meyer für die hilfsbereite und wissenschaftliche Betreuung als Zweitgutachterin. Ich bedanke mich sehr für Ihre wertschätzende Art der Prüfungsdurchführung und die geteilten Ratschläge.

Danken möchte ich außerdem meinen Kollegen am IFL sowie bei der Robert Bosch GmbH, die mich fachlich und moralisch begleitet und unterstützt haben. Mein besonderer Dank gilt Dr. Najdan Vukovic. Dein breites Wissen, der konstante Austausch und deine Freundschaft haben mich nicht nur motiviert, sondern mich mit einer noch größeren Freude und Spaß an dieser Dissertation arbeiten lassen.

Am tiefsten verbunden und dankbar bin ich meinen Freunden und meiner Familie für ihre immerwährende Unterstützung. Mein besonderer Dank gilt meinen Eltern, Martin und Theresia Kobold, die mir meinen bisherigen Lebens- und Karriereweg ermöglichten. Ihnen, insbesondere meiner Mama, widme ich diese Arbeit. Das Beste in mir – ich hab’s von dir.

Vilsbiburg, Dezember 2025 Miriam Kobold

Kurzzusammenfassung

Bedarfsplanung und Bestandsmanagement bei volatilen Bedarfen mit Verfahren des maschinellen Lernens

Bislang galt: Lieferketten sollen die Anforderungen der Kunden zuverlässig und kostengünstig erfüllen. Jedoch ist das Umfeld heute zunehmend unbeständig, geprägt durch Schwerpunktverlagerungen in Produktion und Vertrieb, dem Streben nach mehr Nachhaltigkeit, Handelsschranken, Finanzkrisen und Krieg. Dies erfordert eine Neuausrichtung der Lieferketten im Hinblick auf drei Konzepte: Flexibilität, Robustheit und *szenarienbasierte Optimierung*. Ein Schwerpunkt ist hierbei die Prognose von Kundenbedarfen (Glatzel 2021: 7). Sie bildet die Grundlage für die meisten Planungs- und Steuerungsaktivitäten in jedem Unternehmen und trägt damit maßgeblich zu einem erfolgreichen Supply Chain Management bei. (Danese und Kalchschmidt 2011) Solange keine Prognose der künftigen Nachfrage vorliegt, können Unternehmen nur unter geringer Transparenz Produktionspläne, Aufträge zur Bestandsauffüllung, Transportvereinbarungen und Personalausstattung festlegen. Es ist die Nachfrageprognose, die die gesamte Lieferkette in Bewegung setzt. (Boylan und Syntetos 2021:1)

Dabei ist der Prozess der Prognoseerstellung mit mehreren komplexen Aspekten verknüpft, welche untrennbar miteinander verbunden sind, und welche auch charakteristisch für den vorliegenden Use Case, den *Automotive Aftermarket*, sind:

- Globalisierung und weltweite modulare Beschaffungs- und Vertriebsnetze führen zu steigenden Produktionsvolumina, Variantenreichtum sowie individualisierten statt generalisierbaren Kundenwünschen.
- Nachfragen besitzen einen stochastischen Charakter. Eine Schätzung bezüglich des Eintreffens einer Bestellung und der jeweiligen Ausprägung ist schwierig und die Genauigkeit hängt nicht nur von historischen Absatzdaten sowie den zugrunde liegenden Prognosemethoden ab, sondern auch von der Berücksichtigung einer Vielzahl von Einflussfaktoren.
- Entscheidend ist hierbei nicht nur die Kennzeichnung der Nachfragetrigger und weiterer Signale als wertvoll bzw. irreführend, sondern auch deren datentechnische Repräsentierbarkeit und Kombinationsfähigkeit.

Die Integration und Beachtung dieser Faktoren erfordert den Einsatz von Methoden, die als sogenannte *Advanced Analytics*, d.h. auf Künstlicher Intelligenz basierende Methoden, klassifiziert werden können.

Künstliche Intelligenz (KI) bzw. deren Teilbereich Maschinelles Lernen (ML) wurde in den letzten Jahren national und international als wesentliche Schlüsseltechnologie der Wirtschaftsleistung gesehen und gefördert. Auch in der Logistikbranche nimmt ihre Bedeutung rasant zu. Dies liegt vor allem daran, dass ML-basierte Algorithmen im Gegensatz zu ihren klassisch-statistischen Pendants multivariat arbeiten, Informationen über verschiedene Aggregate hinweg (Zeitpunkte, Produkte und Produktfamilien, Kunden, Regionen) kombinieren und speichern können, lokal und global einsetzbar und mehrstufig prognosefähig sind.

Im Zuge dieser Untersuchung werden multivariate Regressionsmodelle wie Gradient Boosting, Neuronale Netze und Support Vector Regression auf der Grundlage einschlägiger Literatur miteinander verglichen und im Hinblick auf ihre Anwendbarkeit für Zeitreihenprognosen im Allgemeinen und für den vorliegenden Anwendungsfall im Besonderen bewertet. Obwohl im Verlauf dieser Arbeit eine Vielzahl an Methoden evaluiert wird, ist der Fokus ein anderer. Der Dateninput, d. h. Features, welche das Domänenwissen und die Marktspezifika repräsentieren sowie damit einhergehende Feature Engineering Prozesse, werden als Kernelement der gesamten Entwicklung betrachtet. Dieser Ansatz, der gemeinhin als *datenzentrierte KI* bezeichnet wird, benennt die systematische Gestaltung, Verbesserung und Selektion von Datensätzen, um die Prognosefähigkeit der jeweiligen ML-basierten Anwendungen zu optimieren. (Ng 2023)

In der Dissertation gehen die Studien zur Erstellung der Prognosemodelle von der reinen Erkundung zur Erklärung und schließlich zur Bestätigung über. Der gewählte Ansatz umfasst hierbei Elemente aus der qualitativen und quantitativen Forschung und resultiert bei letzterer in einer Vielzahl von Experimenten. Die Kernergebnisse lassen sich anschließend in vier Bereiche untergliedern: *Dateninput und Features*, *ML-basierte Modelle*, *Zeitreihen-Cluster* und *Evaluierungsmetriken*.

Neben dem Training von Modellen mittels transaktionaler Features auf verschiedenen zeit- und produktbezogenen Aggregationsstufen erweist sich auch die Integration von Domänenwissen und Marktfaktoren als vorteilhaft. Merkmale können nicht auf einen bestimmten Typ beschränkt werden: auch sog. *Dynamic Future Unknowns* sind von essenzieller Bedeutung, insbesondere wenn sie in Bezug zu Produktfamilien und Kunden bzw. Kundensegmenten stehen. Neben diesen Domänen sind es vor allem die Population an Primärprodukten einschließlich ihrer Ausfallwahrscheinlichkeiten, Kundenservice-Levels, Preisinformationen sowie Indizes für Saisonalitäten, welche die Forecast-Qualität erhöhen. Abhängigkeiten zu Produktkategorie und Modell bestehen gleichermaßen.

Nach den vorliegenden Ergebnissen sind ML-basierte *semi-globale* Modelle, die sich auf das einzelne Produkt konzentrieren, aber gleichzeitig Informationen aus dem Portfolio verarbeiten können, am besten für die Nachfrageprognose in volatilen Märkten geeignet. Globale Modelle verallgemeinern zu stark; spezifische, d.h. charakteristische Entwicklungen einzelner Zeitreihen werden verkannt. Lokale Modelle erfordern unverhältnismäßig hohe Rechenzeiten und Rechenleistungen. Die Prognosegenauigkeiten liegen auf einem ähnlichen und segmentbezogen auch auf einem leicht besseren Niveau als das traditionelle statistische Ensemble, welches als Benchmark verwendet wird.

Die erfolgreiche Implementierung der vorgestellten *verhaltensbasierten,* d.h. datenzentrierten, semi-globalen Modelle setzt eine Zeitreihensegmentierung als Vorverarbeitungsschritt voraus. Auf der Grundlage der Größe des Portfolios und der abgeleiteten Zeitreihenmerkmale erweist sich das attributsbasierte Clustering im Vergleich zum rein verhaltensbasierten Clustering als überlegen. Die verwendeten und nachweislich sinnvollen Klassifikatoren werden auf verschiedenen Ebenen (Produkt-Markt-Kombination (PMC), Stock-Keeping-Unit (SKU) und Markt) definiert und sind übergeordneter als auch statistischer Natur.

Abstract

Demand Forecasting and Inventory Control in Volatile Markets using Machine Learning

Until now, there was the rule that supply chains should meet customer requirements in a stable and cost-effective manner. Though, as of today the environment is increasingly volatile, characterized by focus shifts in production and sales, the drive for greater sustainability, trade barriers, financial crises and war. This requires a realignment of supply chains with regards to three concepts: flexibility, resilience and scenario-based optimization. One focal point in here is *Demand Forecasting* (Glatzel 2021: 7). It forms the basis for most planning and control activities in any organization (Danese and Kalchschmidt 2011) and hence for successful supply chain management. Unless a forecast of future demand is available, organizations cannot commit to production schedules, inventory replenishment, transportation arrangements and staffing levels. “It is demand forecasting that sets the entire supply chain into motion.” (Boylan and Syntetos 2021:1)

Demand Forecasting itself is entailed with several complexities, which are inextricably linked, and which also heavily concern the market that is the subject of attention in this dissertation: the Automotive Aftermarket.

- Firstly, globalization and worldwide modular sourcing and distribution networks lead to increasing production volumes, more variants and fast-changing as well as individualized customer demands.
- Secondly, the (future) demand for individual products is of a stochastic nature and its estimate is always subject to three uncertainties: primarily, if respectively when demand occurs and, secondly, with regards to the respective size of the order quantity. Hence, accuracy not solely depends on underlying forecasting methodologies seizing historical sales data, but also on the integration of a wide range of influencing factors, that give indications for those uncertainties.
- Thirdly, the decisive factor here is not only the labeling of demand triggers and other signals as valuable or misleading, but also their representability via data as well as their combinability.

Taking these factors into account requires the implementation of methods that can be classified as so-called *Advanced Analytics* i.e., Machine Learning based methods.

Artificial Intelligence (AI) and its subfield Machine Learning (ML) has been seen and promoted nationally and internationally as the essential key technology of economic performance in recent years. And its importance is also rapidly increasing in the logistics industry. This is commonly due to the fact that ML-based algorithms, unlike their classical statistical time series forecasting counterparts, act multi-variate, are able to combine and memorize information across different aggregates (as e.g., time, products and product families, customers, and regions), can be used locally and globally, and are multi-step forecasting compatible.

For this research, multivariate ML-based regression models such as gradient boosting, neural networks and support vector regression are compared on the basis of relevant literature and evaluated with respect to their applicability for time series forecasting in general and for the present use case in particular. Although a variety of models are

evaluated throughout the thesis, the focus compared to past and currently ongoing research projects is different: Input data, i.e., features, representing domain knowledge and market specifics as well as related engineering processes are regarded the core element of the overall implementation. This approach, commonly referred to as *data-centric* AI (Ng 2023), also involves the systematic integration i.e., improvement and selection of datasets to increase the accuracy of one's ML-based applications.

In the thesis, studies to build-up the forecasting models went from pure exploration to being explanatory, and finally, confirmatory, comprising elements from both qualitative and quantitative research resulting in a tremendous number of experiments for the latter. The core findings can be broken down into four areas: *Input Data and Features*, *ML-based Forecasting Models*, *Time Series Segmentation and Clusters* and *Evaluation Metrics*:

Besides training models via transactional features with concomitant aggregation across time and product hierarchies, the integration of domain knowledge and market factors also proofs beneficial. Features cannot be restricted to a specific type: *Dynamic future unknowns* are of essential importance, especially when related to product families and customers respectively customer segments. Besides these domains, it is particularly the populations of the primary products including failure probabilities, stockout indicators, pricing information and indications of seasonality. Dependencies with regards to product category and model exist equally.

According to the present results, ML-based *semi-global* models which are enabled to focus on the single product but can simultaneously seize information from the portfolio are best suited for demand forecasting in volatile markets. A one-fits-all approach proves not to be performant, with global and local models exposing two main downsides: Global models generalize too strongly i.e., characteristic developments of individual time series are not recognized. Local models require high computational time and computing power. The performance in terms of accuracy of the latter is at a similar and sometimes slightly better level than the traditional statistical ensemble used as a benchmark.

The successful implementation of the introduced *behavior-based* i.e. data-centric, semi-global models entail time-series segmentation as a preprocessing step. Based on the size of the portfolio and derived time series characteristics, attribute-based clustering proofs superior compared to strictly behavior-based clustering. Classifiers used and seized are built on different levels (PMC, SKU and market) and of meta and statistical nature.

Outline

List of Figures

List of Tables

List of Abbreviations and Acronyms

General

AA	Automotive Aftermarket
AI	Artificial Intelligence
ATV	All-Terrain Vehicle
BU	Business Unit
BVL e.V.	Bundesvereinigung Logistik e.V.
B2B	Business-to-Business
B2C	Business-to-Consumer
CSL	Customer Service Level
COM	Commodity
GC	Global Currency
GPP	Gross Price Point
HCV	Heavy Commerical Vehicle
HPP	High Price Point Segment
IAM	Independent Aftermarket
ICPR	Index Catalogue Price Point
IPP	Invoiced Price Point
ITG	International Trading Groups
LC	Local Currency
LCPR	Local Catalogue Price Point
LCV	Light Commercial Vehicle
LPP	Low Price Point Segment
MECE	Mutually Exclusive – Collectively Exhaustive
MPP	Mid Price Point Segment
MOQ	Minimum Order Quantity
ML	Machine Learning
NPP	Net Price Point
OEM/OES	Original Equipment Manufacturer / Original Equipment Services
PG	Product Group – product family from a commercial perspective
PH	Product Hierarchy – product family from a technical perspective
PMC	Product-Market-Combination – level of the forecasting process and time-series identifier
PLC	Product Life Cycle

POS	Point-of-Sale
PPC	Product Production Costs
PPP	Pocket Price Point
PV	Passenger Car
RQ	Research Question
SKU	Stock-Keeping Unit
TP	Technical Part
TrP	Transfer Price
TTF	Total Time of Failure
UHPP	Ultra-high Price Point Segment
VAS	Value-added Services
VDA	Verband der Automobilindustrie e.V.
VIO	Vehicles in Operation
WT	Wear-and-tear parts respectively Commodities
XAI	Explainable AI
YoP	Year of Publication

Algorithms and Features

AC	Autocorrelation
ANN	Artificial Neural Network
AR	Autoregression
ARCing	Adaptive Reweighting and Combining
B-LSTM	Bidirectional Long-Short-Term-Memory Network
B-RNN	Bidirectional Recurrent Neural Network
BIRCH	Balanced Iterative Reducing and Clustering using Hierachies
CART	Classification and Regression Tree
CAST	Clustering Affinity Search Technique
CNN	Convolutional Neural Network
CDF	Cumulative Distribution Function
DTW	Continuous Dynamic Time Warping
ED	Euclidian Distance
ETS	Exponential Smoothing
FI	Feature Importance
FNN	Feedforward Neural Network
FSD	Feature Selection Direction

GBR-XGB	Gradient Boosting Regressor
GRN	Gated Residual Network
GRU	Gated Recurrent Units
$H0$	Null Hypothesis
$H1$	Alternative Hypothesis
HF	Helper Feature
IQR	Inter-Quantile Range
KNN	k-Nearest Neighbour
LSTM	Long-Short-Term-Memory Network
MA	Moving Average
MAD	Median Absolute Deviation
MF	Main Feature
MLP	Multi-Layer-Perceptron
NaN	Not a Number
NR	Not Relevant (i.e. the feature is disregarded due to • no data representing the feature being available • being superseded by (an)other feature(s) • not being characteristic for specific demand behavior
PCA	Principal Component Analysis
PDF	Probability Density Function
PI	Prediction Importance
RI	Relevance Indication of MF c for all PMCs i across all models m
RIM	Relevance Indication of MF c for all PMCs i per model m
RIMM	Relevance Indication of MF c per PMC i and model m
RNN	Recurrent Neural Network
RW	Random Walk
SBA	Syntetos-Boylan-Approximation for Croston
SBJ	Shale-Boylan-Johnston Approximation for Croston
Seq2Seq	Sequence-to-Sequence
SF	Segmentation Feature
SVR	Support Vector Regression
TSB	Teunter–Syntetos–Babai Approximation for Croston
WRI	Weighted Relevance Indication of MF c for all PMCs i across all models m
ZF	Zero Forecast

Error Metrics and Business KPIs

AAPE	Arctangent Absolute Percentage Error
AE	Absolute Error
APE	Absolute Percentage Error
FA	Forecast Accuracy
FAI	Forecast Accuracy Index
GMRAE	Geometric Mean Relative Absolute Error
GMSE	Geometric Mean of Squared Error
GRMSE	Geometric Mean of Root Mean Squared Error
NAE	Normalized Absolute Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MBA	Mittlere Bestandsabweichung – Aggregated Forecast Deviation
MdAPE	Median Absolute Percentage Error
MdRAE	Median Relative Absolute Error
MRAE	Mean Relative Absolute Error
MSE	Mean Squared Error
PE	Percentage Error
RAE	Relative Absolute Error
RMSE	Root Mean Squared Error
sAPE	Symmetric Absolute Percentage Error
sMAPE	Symmetric Mean Absolute Percentage Error
TS	Tracking Signal
wFAI	Weighted FAI

Glossary of Notations

T	total historical time period T consisting of individual bins t
D	dataset where $D = \{Y_1, Y_2, \dots, Y_I\}$
$y_{\mathrm{i,t}}$	temporal characteristic of Product-Market-Combination i at time t
$\tilde{y}_i$	median of the temporal characteristics of Product-Market-Combination i
$f_{i,t}$	forecast f for Product-Market-Combination i at time t
$e_{i,t}$	forecast error e for Product-Market-Combination i at time t
c_{t}	covariate (main feature) c at time t
s	forecast origin s
H	overall forecast horizon with $h \epsilon H$
$\widetilde{H}$	MBA forecast horizon with $h \epsilon \widetilde{H}$
$\tilde{Q}$	number of MBA periods $\tilde{q}$ of length $\widetilde{H}$
w^*	lookback window with $w^* \epsilon T$
L	number of lags l for forecast $f_{i,t}$
a	activation function for NNs
α	alpha level to test for statistical significance
b	shape parameter for the Weibull distribution
β	coefficients in a regression model
cd	critical distance
ε	level of precision
ε_t^*	random error in a regression model
e	Euler‘s number
m^*	number of bins in a time series sequence
k^*	integer part of $(h-1)/m^*$
I	number of objects respectively Product-Market-Combination i
M	number of models m
ϑ	step size in gradient descent optimization
$\tilde{d}_i$	median absolute deviation of the temporal characteristics of Product-Market-Combination i
N	overall population size
n	sample size of a population N
Z	abscissa of the Gaussian curve that cuts off an area α
p	proportion in population N
var	coefficient of variance

var^2	squared coefficient of variance
sk	skewness
φ	disperison parameter of the Tweedie distribution
p^*	shape parameter of the Tweedie distribution
r	Bravais-Pearson Correlation Coefficient
v	identifier for a specific vehicle (~RBK)
lt	life time of an object
ToE	time at which the object breaks down (*time-of-event*)
ToO	time at which the life-span of an object starts (*time-of-origin*)
DtE	distance after which the object breaks down (*distance-to-event*)
TtE	time after which the object breaks down (*time-to-event*)
μ_{DoE_i}	expected life span in [km]
σ_{DoE_i}	standard deviation for the life span in [km]
μ_{ToE_i}	expected life span in [years]
FP	expected (static) failure probability in [%]
$FP(lt)$	failure probability being dependent on an object's lifetime lt
$\widehat{FP}(lt)$	expected failure probability being dependent on an object's lifetime lt
$RP(lt)$	reliability probability being dependent on an object's lifetime lt
$mo_{\tilde{k}}$	$\tilde{k}$-th sample moment mo to be used for parameter estimation
d^v	annual mileage per VIO sector-segment-fuel type class in a specific country
$\lambda_i(lt)$	dynamic failure rate for object i being dependent on an object's lifetime lt
η_i	price elasticity of Product-Market-Combination i
$q_{i,t}$	quantity demanded of Product-Market-Combination i at time t
$pr_{i,t}$	price of Product-Market-Combination i at time t
G_k	cluster G_k consisting of a homogenous set of time series $\{Y_1, Y_2, ..., Y_i\}$
$R_{k,l}$	ratio of intra- and inter-cluster dispersion for clusters G_k and G_l

1 Introduction

"What I'm finding is that for a lot of problems, it'd be useful to shift our mindset toward not just improving the code but in a more systematic way of improving the data […]. In other words, companies need to move from a model-centric approach to a data-centric approach."

Andrew Ng, in "Big Data to Good Data"

1.1 Problem Description and Motivation for the Research

Today´s world is modelled by a vast number of zeros and ones. Those numbers, which are called binaries, and which are based on the Boolean logic from 1847, are the simplest form of programming. They enable the conversion of any information by translating words, computer processor instructions, or other data into two-symbol sequences – a process commonly referred to as digitization. Being developed further for decades in the areas of philosophy, mathematics, electrical engineering, and information technology, digitization has now become a buzzword for recent advancements in the context of Industrie 4.0 and also Logistics 4.0 for allowing data processing of all kinds in all formats to be carried intermingled, and with the same efficiency.

The Bundesvereinigung Logistik (BVL e.V.) also addresses the topic of digitization in logistics in its 2021 publication and identifies six focal points (Glatzel 2021: 7) to enable the next level of maturity:

- (Greater) visibility across the supply chain, including upstream and downstream information flows
- Forecasting of customer demands
- Adjustment of production and distribution capacities
- Optimization of inventory
- Securement of logistics capacities
- Management of liquidity and working capital

The decisive point here is that *Demand Forecasting*, which is directly dependent on the first focal point, forms the basis for most planning and control activities in any organization (Danese and Kalchschmidt 2011) and thus for the last four focal points listed above.

From a contextual point of view there are even more reasons for the enduring importance of *Demand Forecasting*.

The first one is the increasing market volatility (Treiblmaier 2014) in combination with the special challenges in the Automotive Aftermarket (AA) distribution network (Pfohl 2018: 238, Hecker et. al. 2012: 47ff) as for example

- globalization, internationalization and increasing competitive intensity (Bousonville 2016: 12f, Klug 2010: 42ff)
- the richness of variants of primary products, product generations and post-purchase-obligations and the thereof derived enlarging scope of the aftermarket portfolio (Klug 2010: 498f.)
- the multi-levelled customer structure as well as the

- financial importance of the sector: despite a decline in order intake (Hecker et. al. 2012: 22) the AA business is, on a value basis, a growing major industrial and economic force.[1]

Another *motivator* is the ever-present conflict of objectives within logistics, in terms of reliability, responsiveness and flexibility on the one side and a perpetual reduction of costs and assets on the other intensifies (Kilger and Wagner 2008, Kilimci et. al. 2019).

The third *enabling* factor is (*big*) data and advanced analytics (Frowein et. al. 2014). Each of us is producer and consumer of data at the same time: details of each transaction are stored - timestamps, products, quantities purchased, amounts invoiced. Based on these data, customers seek products and services being tailored to them: they shall be delivered respectively provided in the right quantity and the right condition, to the right place at the right time at the right price. (Swamidass 2000: 684) To enable industry and companies to meet these requirements, and to make customer behavior correspondingly predictable, continuous efforts are made to extract massive and complex data from disparate sources and to combine them with advanced models and algorithms for data and time-series analysis. (Alpaydin 2019: 1ff) The goal is not to completely identify the process, but to construct a good and useful approximation and to optimize a certain performance indicator. The model is defined with certain parameters and learning in this case means to improve the parameters by using more and more training data from the past and indications from the future. In this case, the model may be descriptive i.e., gaining knowledge based on data as well as predictive i.e., making predictions for the future.

Fourth, methodological developments, meaning advancements for models and algorithms for time-series forecasting, are finally to be seized due to increasing computational power and improved infrastructures. (Murray et. al. 2018)

Delimitation of Terms

The official beginning of Artificial Intelligence is set in Dartmouth in 1956, when John McCarthy organized a workshop comprising seven different topics together with Marvin Minsky, Claude Shannon and Nathaniel Rochester. (Taulli 2019: 6f) Interesting here is the working hypothesis, which now also serves as a general definition of AI: "Every aspect of learning or intelligence in general can be described so precisely that it can be simulated by a machine." As a result of this meeting, a research area emerged that has been offering a very broad field of activity for many scientists until today. (Russel and Norvig 2012: 39ff., Buxmann and Schmidt 2021: 3ff)

In the 1950s and 1960s, the approaches of symbolic AI were strongly promoted alongside the sub-symbolic respectively connectionist techniques that had already emerged earlier. In symbolic AI, explicitly existing knowledge is processed via mathematical logic. Explicit means that there are possibilities of explanation and that we as humans understand how certain decisions have been made or events have been triggered.

In sub-symbolic AI, as for example in ANNs (Artificial Neural Networks), knowledge is represented implicitly. This means that there are no explicit rules about knowledge representation. In academia as well as in operational practice, sub-symbolic systems are

[1] The global automotive aftermarket size was valued at 390.10 billion US Dollar in 2020. It is expected to expand at a compound annual growth rate of 3.8% from 2021 to 2028. (Grand View Research 2020)

ascribed a considerable capacity of performance for two reasons, namely new algorithms with the ability to comprise a close-to-entire relation structure of a certain domain and increased computing power.

Hence, one could say, that symbolic AI is a paradigm with high explainability but limited accuracy performance, whilst sub-symbolic AI is a paradigm with low explainability and high accuracy performance. Most recent research areas now try to combine components from these areas, either being referred to as Explainable AI (XAI), or Neural-Symbolic Computing, focusing on models with both high explainability and high accuracy performance.

A sub-area of AI is machine learning, also referred to as inductive learning, which aims to generate knowledge from experience, i.e., to identify patterns in data and to gain insights that are not immediately recognizable by humans. Thus, it is not possible to develop explicit programs with hard coded logic, because the inner structure of the program to be created is not or at least not completely known.

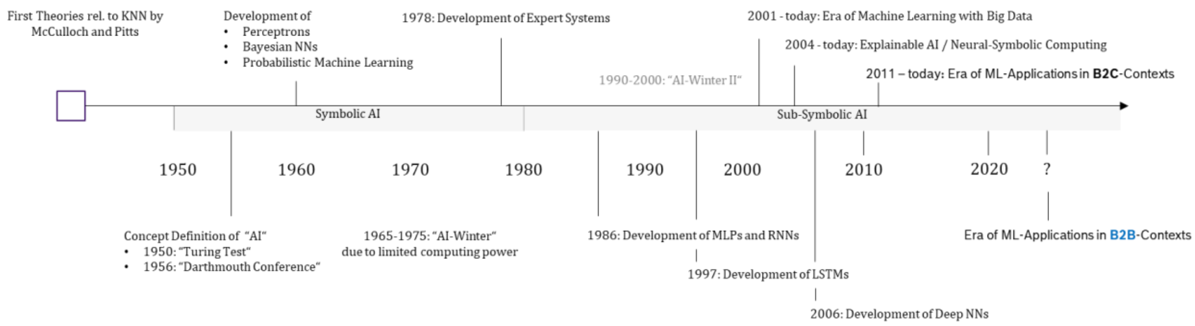


Figure 1.1 History of AI

A differentiation of individual methods in ML depends on the way the model learns. Mainly, *unsupervised*, *supervised* and *reinforcement learning* are distinguished.

- *unsupervised learning*: only input data without desired outputs are available to the algorithm. The goal is that it finds a distribution of the data and thus provides new insights.
- *supervised learning:* in addition to the input data, a desired result is also transmitted to the algorithm. If new, unlearned examples are provided, the model should find an answer analogous to the learned examples.
- *reinforcement learning*: a software agent, i.e., a computer program with a certain degree of self-dynamic behavior, is enabled to learn strategies so that the reward achieved is maximized.

In order to make a forecast, the question which is to be answered by the *supervised ML algorithm* is the following: Based on t consecutive periods of demand and certain relevant factors, what will the demand be during the period(s) $T + 1 \ldots T + H$?

Whereas a traditional statistical model will use a predefined relationship to forecast the demand, a *supervised ML algorithm* will not assume a priori a particular relationship like seasonality or a linear trend; it will learn these patterns directly from the historical demand and the factors and then apply these relationships on new data.

1.1.1. Business Point of View

How does forecasting and forecast accuracy impact logistics? In general, there are two types of demand forecasts that a company can model: short-range demand forecasts, and medium respectively long-range forecasts. The latter are of strategic use. In this scenario, companies typically use the data for budgeting and investment purposes as for trucks or ships, warehouses, distribution centers, and building new hub facilities or production lines. The most useful and at the same time more challenging type of demand forecast though typically comes from the short-range type. This highly influences the operational planning. In here, Kilimci et. al. (2019) differentiate between the two phenomena *excessive stock* (i.e., overstock) as a result of overforecasting and *out-of-stock* as a result of underforecasting. Both phenomena in turn influence different logistics processes.

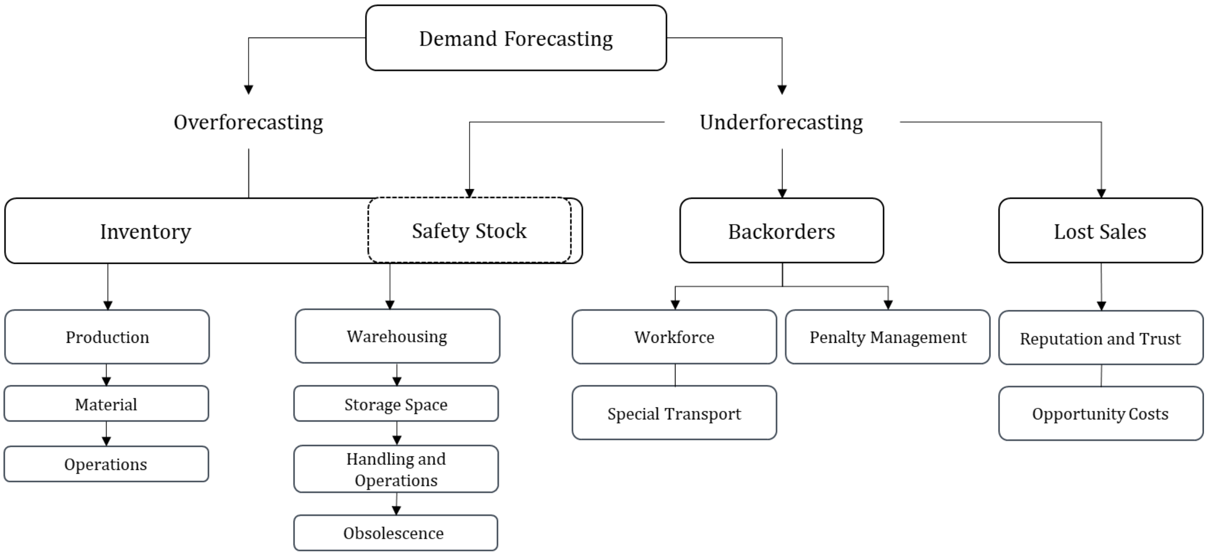


Figure 1.2 Impact of Forecast Accuracy on Logistics Processes

The overview in Figure 1.2 represents the aspects that Cohen et. al. (2006: 131ff), Kilger and Wagner (2008), Klug (2010: 505), Huang et. al. (2014) and Kilimci et. al. (2019) refer to as *areas of impact* in their publications. On the left side, all logistics processes are listed that are associated with overforecasting and thus with increased capital tie-up and operational costs due to stock surplus, increased storage, labor, and insurance costs, as well as with quality reduction and degradation depending on the type of product up to the point of scrapping. (Kilger and Wagner 2008) Underforecasting, on the other side, generally leads to non-satisfaction of demand and consequently to reduced customer satisfaction and loyalty, loss of sales (Cohen et. al. 2006, Huang et. al. 2014), service level fines (Huang et. al. 2014), decreased competitiveness as well as to higher target stock levels and lower reorder points (Nemati Amirkolaii et. al. 2017).

1.1.2. Scientific Point of View

Approach

The initial overarching research problem, how can aftermarket demand planning benefit from ML-based data analytics and forecasting, prompted a thorough literature research regarding data gathering and processing in an aftermarket demand planning context (Chapter 5) and aftermarket demand planning methods (Chapter 2). The search made use of different keywords (I – IV), being used as a general concept but also domain-specific.

- **Keyword I**, the superordinate concept, is *Forecasting*.
- **Keywords II** relate *Forecasting* to the specified concepts *Demand*, *Sales*, as well as *Supply Chain*.
- **Keyword III** adds another specifier by relating *Forecasting* to *business-to-business* (B2B) transactions.
- **Keywords IV** specify the input data and its characteristics for the present use case via using the search keys *Time Series* and *Intermittent*.
- **Keyword V** limits down the scope of methods classified as *Machine Learning*.

The domains defined are directly related to the application areas. Domain (A) limits the scope to the concepts *auto**[2] and *car* whereas domain (B) addresses the *aftermarket* respectively *spare parts* business.[3]

Table 1.1: Keywords and Domains for the Literature Review

KEYWORD					**DOMAIN**		
I	II	III	IV	V	(A)	(B)	(A+B)
super- ordinate concept	specified concept	specified concept	input data	methods			
forecasting					auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2 A.2 + B.2
	demand				auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2
	sales				auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2
	supply chain				auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2
		B2B			auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2
			time series		auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2
			intermittent		auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2
				machine learning	auto* car	aftermarket spare parts	A.1 + B.1 A.1 + B.2

[2] *auto** signifies either the concept *automobile* or *automotive.*

[3] The selection is based on four limitations:

- [year]: 2007 to 2022; exceptions are made for explicit references in publications from 2007 to 2022
- [language]: English only
- [scholarly and peer-review]: in peer-reviewed scientific journals or by renowned institutions
- [type]: digitalized; sole exception is Hecker et. al. 2012

Findings

With demand forecasting being one of the core operation activities in any customer-facing business, the current scope of theoretical and empirical research is vast. There are many works related to both the conventional statistical forecasting methods like ARIMA, exponential smoothing (ETS), or Croston and its adapted versions for intermittent demands, and to the lately advancing ML-based methods like neural nets (ANN), Gradient Boosting (GBR), or Support Vector Regression (SVR) (Carbonneau et. al. 2007, Romeijnders et al. 2012, Nikolopoulos et. al. 2016, Vairagade et. al. 2019, Kilimci et. al. 2019). The latter ones have the advantage that, in addition to point forecasts, also quantile forecasts or so-called probability density functions (PDFs) i.e., complete probability density functions, can be generated and amongst others be used for scenario estimation and inventory management.

Up to now, forecasting results are predominantly compared according to their underlying methodological approaches (i.e., statistical approach 1 vs. statistical approach 2, statistical approaches vs. ML approaches, or ML approach 1 vs. ML approach 2). This is confirmed in different review and research papers as of Kumar and Kumar 2019, Makridakis et. al. 2019, Benidis et. al. 2020, or Pinçe et. al. 2021. No experiments, though, are related to the systematic incorporation of supportive i.e., contextual respectively business knowledge and the respective data – a phenomenon which is called *forecasting inefficiency*. (Leitner and Leopold-Wildburger 2011, Syntetos et. al. 2016, Benidis et. al. 2020)

Overlooking all application areas, there are publications dealing with the integration of endogenous and exogenous demand drivers like macroeconomic (Günay 2016, Saegert et. al. 2018), pricing (Nijs et. al. 2000, Lee and Cho 2009, Ali et. al. 2009, Ma et. al. 2015) or calender data (Vairagade et. al. 2019) into mathematical models for improved demand forecasting. Yet, there is no holistic identification and analysis of those factors and their respective importance. Rather, randomly available ones are included without further justification for their selection. Also, the extent to which a combination of certain factors has a balancing or a reinforcing effect on demand has not yet been sufficiently investigated or proven – neither scientifically nor in practice. According to the literature research conducted (cf. Chapter 2 and Chapter 5.1.1), there is also no known use case from the aftermarket sector or related industries with oligopolistic structures. (Bacchetti and Saccani, 2012; Romeijnders et al.: 2012)

Table 1.2 serves as an overview to relate this PhD project to the existing literature. At the same time, it highlights the aspects of the topic that have not yet been investigated at all or only insufficiently. If the publication's result and identified gaps appear to be relevant for the research and the use case, it is printed in **bold**. Based thereon, three scientific questions (RQs) can be derived:

Table 1.2: Review Papers on Statistical and ML-based Forecasting

	AUTHOR	TITLE	YoP	CONTEXT / DOMAIN	OBJETCIVE	METHODOLOGY, RESULTS, AND IDENTIFIED GAP
0.	Treiblmaier	A Classification Framework for Supply Chain Forecasting Literature	2014	Supply Chain Forecasting	Classification framework for forecasting literature	Results: Criteria to classify supply chain forecasting literature • focus, approach, and method triangulation • data generation • range and timeline • theoretical background and target group

1.	Adya and Collopy	How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation	1998	Forecasting applications in businesses	Analysis and performance evaluation of 48 studies applying ANN in the years 1988 to 1994.	Results: Guidelines to evaluate potential(s) of ANN
						• *Effectiveness of Validation*:
						  • comparison with well-accepted models
						  • use of ex-ante validations
						  • use of a reasonable sample of forecasts
						• *Effectiveness of Implementation*
						  • convergence: in-sample performance
						  • generalization: recognition of patterns outside of the training sample
						  • stability: consistency of results

2.	Leitner and Leopold-Wildburger	Experiments on forecasting behavior with several sources of information – A review of the literature	2011	Forecasting in businesses	Categorization of human forecasting behavior and related approaches • hypothesis testing and forecast building • adjustment of forecasts • revision of forecasts	Results and Gaps: • forecasts are generally compared according to their underlying methodological approaches • **minority of experiments is related to second-source data and integration of knowledge** • judgmental forecasting ability is overweighted compared to statistical forecasting which is mainly due to the inability to incorporate supportive information (*forecasting inefficiency*) • **integration of expertise, domain knowledge, and any contextual knowledge is identified as the major challenge**.
3.	Längkvist, Karlsson, and Loutfi	A Review of Unsupervised Feature Learning and Deep Learning for Time-Series Modeling	2014	Forecasting related to Stock markets, Speech and Music recognition, Videos, Motion Capture Data and Physiology, Electronics	Review of different generative and discriminative deep learning approaches for time-series modeling	Results: Questions to be answered for model selection • Generative or discriminative model? • What are the properties of the data? • How large is the input size?
4.	Syntetos et. al.	Supply Chain Forecasting: Theory, Practice, their Gap and the Future	2016	Demand Forecasting - general	Review paper to define the state of the art in supply chain forecasting with focus on classical time series approaches	Gap: Research has been predominantly **theoretical** in nature and subject to very specific assumptions and demand process formulations. Open topics from scientific and practical perspective comprise the following: • **analysis regarding slow or intermittently moving items** • quantification of information sharing benefits related to both forecast accuracy and inventory reduction • isolation of effects of information sharing • forecasting based on non-predefined hierarchies • **effects of temporal aggregation on** • **mean demand and demand variance estimation**

						• **volatility** • **data collection frequencies** • **forecast horizons** • distinguishing in between alternative demand patterns • quantification of benefits of group-based forecast reconciliation for real-world use cases • quantification of singe-person-based forecast reconciliation for real-world use cases • effectiveness of judgmental forecasting
5.	Tealab	Time series forecasting using artificial neural networks methodologies: A systematic review	2018	-	Literature review of forecasting methods based on ANN from years 2006 to 2016	Results: Quality Assessment of ANN models comprising the following: • explicit mathematical formulation of the model • feature selection • hyperparameter tuning and complexity determination • performance evaluation • application to a real-world use case • definition of training procedures
6.	Kumar and Kumar	The demand forecasting: A comparative review of conventional and non-conventional techniques	2019	Demand Forecasting - general	Literature review of demand forecasting techniques from year 2009 to 2019	Gap: **Increasing complexity of real-world problems has increased potential for non-conventional demand forecasting techniques** with the following challenges arising: • **inputs must be in data format** • **black box phenomenon and rule/relationship determination** • **model selection and model complexity** • hybrid resp. ensemble learning • **computational expensiveness**
7.	Benidis et. al.	Neural forecasting: Introduction and literature overview	2020	-	Literature Review of neural forecasting with four-fold aim: overview of • the advances that have permitted the resurgence of ANN in ML • ANN architectures and state-of-the-art solutions • Literature review with focus on DL • as-is and future application areas	Gap: **Further developments related to NN needed regarding** • hybridizing existing time series techniques • **feature engineering** • integration of state-space-models and renewal or point processes • **bringing innovations from other related areas e.g., sequence-models, attention-based mechanisms, transfer learning** • general purpose techniques **General challenges for ML especially in the B2B scenario:** • **data effectiveness** plus **interpretability, explainability, causality** • **contrast between global and local models**
8.	Pinçe, Turrini and Meissner	Intermittent demand forecasting for spare parts: A critical review	2021	Intermittent Demand Forecasting for Spare Parts	Literature review and quantitative analysis of forecasting methods for spare parts demand	Results: Benchmarking of classical parametric and non-parametric approaches (e.g., Croston, SBA, ETS, TSB, MA, Naïve) with and without aggregation • SBA majorly better if percentage errors are used • an aggregation strategy generates more accurate forecasts than not using such a strategy in 66% of the comparisons
	PhD project	Demand Forecasting and Inventory Control in Volatile Markets using Machine Learning	2023	Intermittent Demand Forecasting for Spare Parts	Literature review and quantitative analysis of multivariate ML-based forecasting methods for spare parts demand wholistically seizing domain knowledge and factors of variance.	RQ1 – RQ2 – RQ3

- **Research Question 1**

  Which factors are most likely to affect individual demand behavior and its forecasting in the Automotive Aftermarket and how can they be categorized and prioritized?

- **Research Question 2**

  How to represent respectively model qualitatively prioritized factors via data and how to determine their relevance from a quantitative perspective?

- **Research Question 3**

  In conjunction with portfolio- and behavioral-based forecasting, which ML-based algorithms promise to lead to sustainable improvements in demand planning in the Automotive Aftermarket?

RQ1 addresses the identification and categorization of factors affecting demand according to their *Context*, their *Type* and *Level of Influence* and the *Demand Variability* caused (cf. Chapter 2.1.2). Experts from different areas of responsibility (Sales, Logistics, Business Units respectively Production) support the qualitative assessment and prioritization. Thereby the gap identified in Leitner and Leopold-Wildburger (2011) as well as in Benidis et. al. (2020) is referenced.

RQ2 elaborates on which factors can be represented by data (*general availability*) and at the same time meet the following predefined criteria (Heinrich and Klier 2011):

- design quality: the data provided matches the data required in terms of granularity, historization and notice in advance
- conformity: the data collected represents the corresponding real-world information. It is multi-dimensional and can be rated according to the four different criteria *currency*, *consistency*, *completeness*, and *correctness*.

In order to prove respectively disprove the factors' relevance also quantitatively, a multi-step approach is used: By relying on the validity of different regression models and metrics, it is to be investigated whether and how often and subsequently how strongly in terms of its effect size (elasticity) the factor influences demand behavior. Likewise, it needs to be examined how much variability in demand is explained by the factor alone and in combination with other factors. Answering these questions is directly related to the relationship determination and black box phenomenon described in Kumar and Kumar (2019), as well as to the interpretability, explainability, and causality aspect for B2B (business-to-business) transactions in Benidis et. al. (2020).

RQ3 is concerned with the implementation of ML-based models that are able to forecast predominantly sporadic i.e., intermittent-lumpy time series more accurately than the conventional benchmarking model by systematically incorporating the pre-selected endo- and exogenous factors. Besides defining the models' type, architecture and parameters, one also has to think about two strategic questions:

- What do experiments look like to identify optimized feature-model-combinations?
- What does the backtesting and evaluation strategy look like, and which error metrics are to be applied?

The ultimate goal is then to imply as much information content as necessary, but at the same time to make it as generalizable as possible.

1.2 Conceptual Framework and Structure

Goal and Goal Achievement

For this reason, the underlying research for this PhD project comprises the following sub goals:

- the systematic identification of determinants of demand in oligopolistic markets (e.g., the spare parts business) as well as representative data sets. The determinants will be denoted *Factors of Variance* in the following.
- the investigation of the influence of those factors on individual and aggregated demand over time both as a stand-alone factor and also in combination. This means that dependencies and interactions will also be considered. The scope for these factors is defined as follows: portfolio information, macroeconomic and microeconomic factors, amongst them sales promoting instruments, customer-, and product-related factors, as well as time-specific determinants.
- the development and implementation of multivariate, multi-horizon time series forecasting models that seize relevant factors in addition to sales historical data, considering that demands and hence the time series used are majorly characterized as follows:
  - intermittent-lumpy (i.e., volatile: chaotic, highly sporadic, and low-volume)
  - non-intermittent-erratic
  - immediate demand does not coincide with end customer demand
- benchmarking of the aforementioned.

Based on this, the hypothesis is formulated that the forecasting accuracy for the total demand of a SKU in an oligopolistic market can be significantly improved, also for longer horizons, by systematically integrating and combining a wide variety of region, market-/customer- and product-specific factors into an ML-based model. Consequently, this is to be considered a contribution to data-centric, meaning behavior- and hence factor-based forecasting rather than to model-centric AI development.

Data and Methods

The data for the research comprises the sales historical records i.e., transactional and meta data[4] of an automotive supplier on SKU-customer level. The starting point for the time series analyses is the year 2014. The forecasting horizon is limited to twelve months and the pre-Covid year 2019. For the selected market, this results in a training period of four years, with another 12 months of lagged information. The validation split is set model specific.

To test the hypotheses, time series data which comprise the prioritized factors are also required. In order to construct the time series, a cumulative data set in long format must

[4] In the context of data management, transactional data is the information recorded directly from transactions. They include the parties involved in the transaction, in the following denoted as *PMC*, the sales quantity as well as all information related to pricing. Another essential information from the transaction is the timestamp *Date*.
Definition 1: The feature *PMC* is specified by the level at which the demand is predicted e.g., *item i* and *customer* or *item i* and the *sales market*, denoting a specific country or region. Hence, the abbreviation *PMC* refers to each possible product and market combination.

be created from the individual cross-sectional data sets and tested for both reliability and validity.

With regards to the forecasting methods, an initial prioritization of possible ML-based methods is carried out by means of literature research. Methods that achieved good forecasting results are also examined more closely according to the following criteria:

- The result is representative
- Input data is time series based and of valid length
- The forecast horizon is of sufficient length i.e., 'one-step-ahead' results are not sufficient
- Methods have achieved good results in the relevant industry, related industries, or using time series being subject to similar characterization

Accordingly, a proof-of-concept based on sequential selection is performed using the following ML-based methods:

- BI-DC-LSTM: local bidirectional stacked long short-term memory networks
- CNN-LSTM: local LSTM with convolutional layers
- (semi-)global LSTM-GRN with attention-mechanism
- (semi-)global GBR-XGB: regularized gradient boosting

A valid comparison of the above ML-based methods requires the assessment of their accuracy also in relation to an established benchmark method, considering the given data constellation. Hence, results are compared to traditional statistical methods such as

- a Naïve model (Random Walk - RW) with lookback window 12 [months]
- a Naïve model (RW) with lookback window 1 [quarter]
- a Moving Average model (MA) with window size 12 [months]
- an ensemble method consisting of naïve and traditional regression algorithms.

Limitation of the Topic[5]

The restriction of the topic occurs from three reference points: channel, type of customer and product life cycle. Focus one is the independent aftermarket (IAM) business with its multi-level and heterogeneous customer structure. OEM/OES (Original Equipment Manufacturer / Original Equipment Services) business is excluded. Nevertheless, learnings and conclusions from the IAM can be transferred to OES demand forecasting with a corresponding adaption of features. Within the IAM, another focus will be laid on B2B transactions, denoting the so-called first level customers, i.e., customers from the segments wholesale, retail, and e-tail who act as distributors and occasionally also as users. With increasing importance of supply chain collaboration and the fact that the failure-related demand only emerges in the workshops, it is of course also necessary to consider the demand behavior of the so-called second-level customers. It is precisely this analysis that J. Greitens is concerned with in his master's thesis 'Feature Extraction and

[5] The master theses listed in the subsequent section were supervised by the author but are not subject of this dissertation.

Engineering from Point-of-Sale Data in the Automotive Aftermarket' (2022) Using local XGB-Regressor models, the key finding is that forecasts relying on POS data are more accurate than forecasts from the same models or traditional ensembles being based solely on pasts demand from first-level customers.

The last restriction is made according to the product life cycle and the available sales history. Forecasts for so-called cold-start products cannot be generated by supervised learning methods (neither ML-based nor traditional ones) due to no or very short history but are rather based on qualitative methods. An alternative approach is implemented and evaluated in the master thesis by A. Goebel. 'Time Series Forecasting for Cold-Start Items from the Spare Parts Sector by Learning from Related Items using Cluster and Classification Algorithms'. (2022) Via using cluster and classification algorithms, similarities between existing and cold-start products are identified. For existing products, both statistical and SKU-specific features are derived, whereas for cold-start products, it is only the latter group, since only these are available before release. SKU-specific features include categorical, metric, and fuzzy-fixed features. The best clustering method is k-means, the best approach amongst classifiers is a Random Forest model. Finally, forecasts for the cold-start products are derived based on the demands of existing products in related clusters.

Outline of the Thesis

The remainder of this thesis is organized as follows:

In section two, theoretical concepts and the state of the art for supply chain forecasting algorithms are reviewed. This enables us to see the main contribution of the thesis. In section three existing premises and approaches for the AA business are provided. Corporate structures, market actors, the product portfolio and methodologies in use define the implications for the development of a behavior-based ML model. Section four describes the general concept behind the behavior-based ML model, introduces the algorithms in scope and presents the major findings from the explorative data analysis. Section five addresses the so-called *Factors of Variance*, their assessment and prioritization from both a qualitative and quantitative point of view, as well as their transformation from lead to leading factors. Section six demonstrates the experimental design and concludes on the final selection of input features. In section seven, results are presented on diverse levels. Those include views on product family levels, value contribution classes and demand profiles as well the forecasting horizons. The last section draws conclusions on the value of a behavior-based ML model being implemented in volatile markets and discusses directions for future research.

2 Theoretical Concepts and State of the Art

This chapter aims to present the supply chain management side as well as the algorithmic contexts from a theoretical perspective. For this purpose, it will summarize the aspects of integrated supply chain optimization and how these are interrelated with data management and data analytics. The structure is as follows:

Chapter 2.1 depicts the characteristics of a *volatile* market by relying on the example of the automotive aftermarket. Generally, such a characterization can be made according to a market's size, its potentials as well as its dynamics, i.e., forces that impact the behaviors of producers and customers, and more specifically according to its participants, competitive structures and the product portfolio. The characterization will facilitate a generalization and a subsequent transfer of the results.

Subsection 2.1.2 highlights theoretical concepts for identifying factors that are to influence demands in these volatile markets. The last section discusses appropriate forecasting concepts (Chapter 2.2) and methods (Chapter 2.3) to improve the overall forecasting performance.

2.1 Supply Chain Complexity in Volatile Markets

In economic theory, a market is generally defined as the convergence of supply and demand and the resulting exchange processes, typically affected by the prevailing competition. (Wöhe and Döring 2012: 416) This definition also applies to the aftermarket, although an extension is needed with two respects:

Firstly, it is a so-called *secondary* market comprising all goods and services contributing to the assembly of a final product and being supplied or rendered after the purchase of the originally purchased primary good in the context of its maintenance or repair. Accordingly, it is the sale of parts (and services) that are directly related to the previously sold good. Therefore, one can also refer to it as *derived* demand. (Yeung 1972)

Secondly, the convergence between supply and demand takes place on multiple levels. B2B transactions occur between manufacturers and first-level customers denoting OEMs, wholesalers, retailers and e-tailers for the OES and IAM business respectively. B2C (business-to-consumer) transactions take place when goods are subsequently re-sold by first-level to second-level customers denoting the workshops, and thus the end user, i.e., the workshop customer.

Regarding the market size and its potential, one may also observe strong divergences as soon as the market is viewed at from a purely regional perspective. These are due to cultural and economic realities. Moderate growth takes place in industrialized countries, growth reserves are related to the increasing number of passenger car inventories in emerging markets. Overall, the secondary market is a higher-yielding market than the business area that handles the actual sale of the primary-product. Market dynamics though do not just result from inherent qualities of the business and its customer base but are also due to exogenous factors with legislation and ecologically motivated trends to remain the biggest disruptors. (Wöhe and Döring 2012: 409, Hecker et. al. 2012)

The phenomena that meet at this point lead to demand patterns that can be described as predominantly *volatile*.[6] Here, the concept of *volatility* is associated with three demand uncertainties (Angkiriwang et. al. 2014):

- Demand uncertainty one: *Does* demand occur?
- Demand uncertainty two: *When* does demand occur?
- Demand uncertainty three: In case that demand occurs, what is the *extent* of this demand?

The probabilistic nature of demand, caused by these uncertainties, can cause a number of problems for the business, especially in managing orders and inventory levels, with effects magnifying through the supply chain. A detailed description of these effects would exceed the scope of this thesis. Therefore, reference is made to Figure 1.2 and to Simangunsong et. al. 2012, who analyze publications on supply chain uncertainty as well as its causes and effects.

2.1.1 Integrated Supply Chain Management and Demand Planning

Supply chains are complex (Vermorel 2018: 3): There is no universal supply chain organization, they differ from one vertical to the other, and, if within the same vertical, differ from company to company due to different strategies, supplier constraints, ERP systems and so on. Within the company, the supply chain keeps evolving due to a changing number of countries, production sites, warehouses, suppliers, and customers. Hence, there is the need for an approach that embraces the very essence of supply chains, the need to be flexible and to adapt, and to cope with the uncertainty of things.

Integrated i.e., connected respectively digitally enabled *Supply Chain Management*, and its sub-concept *Integrated Demand and Supply Planning*, sometimes also referred to as *Sales and Operations Planning*, represents this approach, grasping the overall physiology of supply chains with its three dimensions *length*, *depth* and *time*. (Syntetos et. al. 2016) The implications derived thereof are the following:

- Length: The variance of demand is amplified as it progresses upstream, making it more difficult to forecast accurately. (Cohen et. al. 2006) There are potential gains in accuracy which may be achieved by different forms of collaboration including sharing of demand information between different departments and levels The practice of collaboration has resulted in some major initiatives like *Collaborative Planning, Forecasting and Replenishment* and *Vendor Managed Inventory*.
- Depth: Supply chains are often geographically dispersed, and they are sometimes referred to as supply nets rather than chains. Key factors leading to these hierarchical silos include markets, products, suppliers as well as industrial and final customers, company-owned production sites and distribution centers.
- *Time* denotes the temporal integration and hence reconciliation of all forecasting activities across horizons:
    - Operational planning involves decisions affecting the short-term execution of the company's business.

[6] More details on the two phenomena can be found in chapters *Market Participants and Competition* (cf. Chapter 3.2)as well as *Product Portfolio and Classification* (cf. Chapter 3.3).

- Tactical planning involves resource allocation decisions over medium-term planning horizons and
- strategic planning involves resource acquisition decisions to be taken over long-term planning horizons.

2.1.2 Demand Variability and Factors of Variance

The determinants of demand, in the following denoted as *factors of variance*[7]. (Goodfellow et. al. 2018: 5), are factors that lead to the previously described uncertainties in the economic demand for a product or a service. Here, a categorization can be made based on four dimensions:

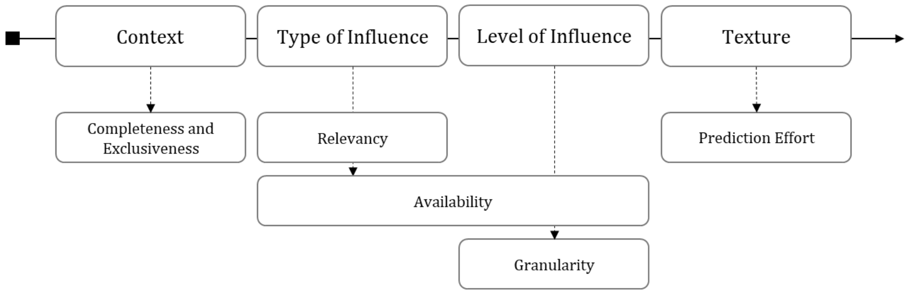


Figure 2.1 Factor Categorization

Literature does not represent the four dimensions in this entirety. Therefore, the individual sections are based on individual research. However, it is worth considering all four areas together to ensure a holistic assessment of the criteria listed for each section.

Context

The *Law of Demand* (Samuelson 1998) is a fundamental principle in economics that states that the quantity demanded varies inversely with price. It acts as an indication of the effect of one economic variable, the price, on another, the demand, provided all other variables are held constant (lat. 'ceteris paribus'). (Lynham 2018: 129) In reality though, demand is not to be defined as a sole relationship between price and quantity. It rather relates to the number of products respectively services a customer is willing and able to purchase at a certain point and within a given period of time. The *willingness to purchase* suggests a desire, based on what economists like Wöhe and Döring (2010: 381) call needs and preferences. Hence, besides pricing, there are other, additional, dynamic, and occasionally mutually dependent factors which affect customer behavior (lat. 'mutatis mutandis'). Those are often referred to as underlying determinants or non-price determinants.

A distinction of non-price determinants is commonly made between macro- and microeconomic factors. The former are influential fiscal, political, or natural events that broadly affect a regional, national, or even international economy. They tend to impact wide swaths of populations, rather than just a few select individuals, and are hence often

[7] Goodfellow et. al.'s concept does not refer to any specific domain such as demand forecasting. Rather, the term can be used for any influencing variables and features in classification or regression problems: „In diesem Kontext bezieht sich das Wort Faktor rein auf die unterschiedlichen Einflussquellen. […] Diese […] sind häufig keine direkt beobachteten Größen, sondern entweder unbeobachtete Objekte oder unbeobachtete Kräfte in der physischen Welt, die ihrerseits die einer Beobachtung zugänglichen Größen beeinflussen. Es kann sich auch um Konstrukte, des menschlichen Denkens handeln, die nützliche verallgemeinerte Erklärungen oder vermutete Ursachen für die beobachteten Daten liefern. (2018: 5f)

closely monitored by governments, businesses, and consumers alike. Whether the impact is neutral, positive i.e., events that foster prosperity and economic growth, or negative i.e., jeopardizing national or international economies, is determined by the intent of the action. While macroeconomics focuses on the behavior of national aggregates, microeconomics narrows its realm of study to individual agents, such as consumers and businesses, or business and business, and their respective economic behaviors and decision-making patterns. Hence, one could say that it seeks to understand how these individuals make decisions; the factors that shape these decisions; and how these decisions affect others.

The context i.e., the content-related reference, is important in order to group the factors and relate them to each other in such a way that no overlaps occur and at the same time a certain degree of completeness is achieved (cf. Chapter 5.1.3).

Type of Influence

The factors' type of influence is directly derived from the components of a time series and helps in assessing relevance and, if positive, general data availability, availability in the training and prediction period as well as the factor's notice in advance. Caniato et. al. (2005) distinguish between *systematic*, *managerial,* and *random* variability. The component they miss to list is related to a time series' level, being denoted as *effective in the long-term* in the following.

Systematic variability is due to structural or cyclical phenomena. Those can either be related to seasonality or to irregularly recurring events. In case of being structural, the phenomena are easy to measure and manage through historical data. For systematic but non-structural recurring events, exogenous information about specific products and/or preferences of the customers is needed to capture the time of occurrence and also its effect. *Managerial* variability refers to fluctuations caused proactively by the business. They refer to manageable influences like marketing campaigns or buying policies. Caniato et. al. (2005) claim, that "[t]his kind of variability needs [both] historical data [on customer level, as] each one has its own behavior, and […] additional [company internal] information [in order to correctly estimate timing and size] of fluctuations". *Random* variability denotes unpredictable variability. It is the one either caused by ad-hoc events or "not explained by any known reason […]" (Caniato et. al. 2005). It can concern individual demands of a single customer as well as overall demand. While random variability often represents a one-time effect, the fourth category *effective in the long-term* is associated with a more sustainable development. This can take the form of a rising or falling trend or even a shift in the nominal values of the process from one level to another.

Level of Influence

A third category 'Level of Influence' is derived from Charles Spearman's *Two-factor Theory of Intelligence* (1904) and subdivides the indicators into *g*- and *s*-factors. The idea remains the same; only the concept of *intelligence* is replaced by the concept of *demand* for this purpose.

The *general* factor represents customer- and product-independent factors, which affect total demand. They are said to be involved in all kinds of happenings, being mainly related to the factor *Time*. For this reason, they are most often to be represented by company-independent, external data, whose aggregation level majorly differs from that of the forecasting problem.

In Spearman's model, this *g-factor* is supplemented by a *specific* factor. It represents an indeterminate number of sub-indicators and features which may not be connected with each other, and which are responsible for individual i.e., product- and/or customer-dependent demand developments. They are related to any of the transactional features *item*, *item category* or *customer* respectively *sales market* and are hence most often to be represented by company internal data.

Texture of the Features

Furthermore, there is the differentiation by whether features reflect static or dynamic (i.e., time-dependent) aspects of the identified factors and whether the developments of the latter are known both during the training period and in the future.

Static-known variables carry time-invariant information. When the model is built with common parameters to forecast multiple time series, these variables allow sharing information within groups of time series with similar static variable levels. Examples include designators or identifiers (i.e., strings) of groups of products, customers, regions, etc., that can be used to mark agglomerates of series that demonstrate similar behaviors. A second example are float or integer features which specify the information content of other factors like e.g., average replacement quantities.

As for the time-dependent covariates, there are two subtypes. *Time-varying-knowns* are useful to identify time-dependent patterns and special events outside the window lookback periods. They can also help the models to react to sharp contrasts induced by discrete events or apparent changepoints. Examples comprise calendar information to identify months and years, or holidays and special events, but also float or integer variables like production costs or the life cycle age. The second group, *time_varying_unknowns*, are Product-Market-Combination-(PMC)-specific covariates, which can be unique to each time series. These variables are usually known to affect the target variable fundamentally and need to be incorporated by models to produce accurate forecasts. However, they do not only bring additional information, but also additional uncertainty to the model, since their usage depends on a forecast made for themselves.

2.2 Demand Forecasting Concepts

Forecasting is a key input for the planning process of most manufacturing and distributing firms. Literature has devoted significant efforts to this issue. Traditionally, three basic approaches have been considered (Dolnicar et. al. 2018, Murray et. al. 2018):

- Hierarchical forecasting
- Attribute-based forecasting
- Behavior-based forecasting

2.2.1. Hierarchical Forecasting

Hierarchical forecasting, which is more specifically known as Top-down, Bottom-up, or Middle-out Forecasting, depending on the strategy chosen, always indicates an aggregation or disaggregation of historical data over time (Athanasopoulos et. al. 2017) or item, i.e., product or customer hierarchies, which are then commonly referred to as family components. (Dangerfield and Morris 1992, Murray et. al. 2018) Top-down implies the application of a prediction model to an upper or the top segment, meaning the aggregated total of an item, and then the distribution of those predictions down to lower levels or even individual items. The performance of bottom-up forecasting, referring to the opposite approach by directly employing the prediction model on a very granular level within a family, is generally rated better in academic literature: Although noise and intermittency are reduced when the data is aggregated, it is regarded to produce more accurate forecasts for individual items or smaller bins as the challenge of disaggregation and the loss of information are ceased to exist. (Athanasopoulos et. al. 2009; Chase 2013; Athanasopoulos et. al. 2017) Murray et. al. (2018) also state that correlations between item groups and the relative frequency of individual items in a group are a general precondition for hierarchical forecasting. However, this is barely fulfilled for hierarchically organized item clusters, e.g., when using legal or organizational structures as a segmentation variable.

2.2.2. Attribute-based Forecasting

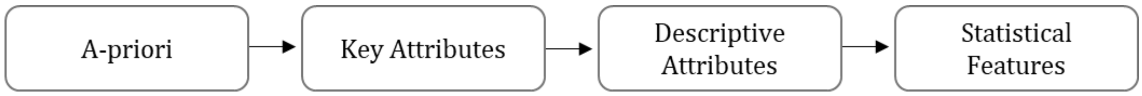


Figure 2.2 Feature Categories for Time Series Segmentation

With attribute-based segment forecasting, the overall population is divided into analogous segments based on attributes associated with either the products or the customers, or both. Murray et. al. (2017) list four categories to define the various types of attributes, “ordering them according to the level of rigor and analytical effort”:

A-priori designates qualitative or categorical segmentation variables, which are solely derived from business knowledge or experience. *Key* as well as *descriptive* attributes are related to Wind’s and Cardozo’s ‘Multi-Stage Segmentation’ from 1974 as well as to Shapiro’s and Bonoma’s ‘Nested Approach’ from 1984. They further divide these attributes into demo- respectively firmographics (e.g., size, location), operating variables (e.g., technology, product, and brand use status) and purchasing approaches (e.g., purchasing policies, buyer-seller relationships) on a macro segmentation level and into situational factors (e.g., order quantities, frequency) and buyers’ personal characteristics (e.g., degree of loyalty, risk aversiveness) on a micro segmentation level. Key attributes can be implemented without any further effort, whereas descriptive variables are often transformed from qualitative to quantitative features and built of various attributes. *Rudimentary* clustering, as Böttcher et. al. (2009) call the sole use of criteria from categories one to three, is not recommended.

2.2.3. Behavior-based Forecasting

Statistical features, the fourth category, are extracted from historical raw data and comprise well-established global features like measures for trend, seasonality, serial correlation, or entropy, but also time specific values like mean, median, variance, kurtosis, sum or peak size and location. (Bala 2012, Hyndman et. al. 2015) Defining and implementing these feature vectors to find and cluster the most similar time series is also known as *transaction-based* segmentation, a variant of *behavior-based* segmentation. (Böttcher et. al. 2009). The similarity between the time series is then a function of these features.

Similar to attribute-based clustering, its effectiveness not only depends on the availability of variables but also on whether the variables used effectively describe the underlying purchasing behaviors. To counteract this disadvantage i.e., to reduce the focus and dependencies from human-selected and manually crafted features to the time series itself and its characteristics, Liao (2005) continues to identify two more clustering approaches within behavior-based forecasting. He distinguishes whether they process directly on raw data, or indirectly with models built from the raw data:

- *Raw data-based* methods match data points between the overall sequence or subsequences of individual time series and determine the points' similarity using various distance definitions (Keogh and Ratanamahatana 2005). According to Kotsakos et. al. (2014: 24-34) as well Aghabozorgi et. al. (2015), one needs to differentiate in-between time- and shape-based methods:
    - In a time-based respectively correlation-based analysis, the correlations among the different time-series data streams are tracked over time to create clusters. Such methods are especially useful when streams exhibit lag correlations within clustered patterns.
    - In shape-based analysis, the time series objects are analyzed in offline manner, and the specific details about when a particular time-series was created is not important. Sequences exhibiting similar patterns are placed into the same cluster based on their shape similarity, regardless of differences in amplitude and phase.

  Results of raw-data-based methods are often proven to be superior to attribute-based segmentation for two reasons: PMCs with different key or descriptive attributes may exhibit similar behavior and they are able to detect similarities that are not expected. Still, researchers are reluctant to use them for their high computational expensiveness. (Murray et. al. 2017)

- *Model-based* approaches replace each time series with a mathematical model of itself, such as a linear regression, a Markov Chain, a Monte Carlo distribution or ARMA, and then determine similarities based on the extracted parameters of those models (Aghabozorgi et. al. 2015). Compared to the previous mentioned approaches, it is shown in Vlachos et. al. (2004) and in Mitsa (2010) that model-based methods have scalability problems and reduced performance when clusters are close to each other.

2.3 Demand Forecasting Methods

Besides forecasting concepts (cf. Chapter 2.2), it is also the computational methods, which determine individual forecasting results. In academic literature, quantitative approaches – mainly based on time-series data – are prevalent, as is shown by the meta-study from Fildes et al. (2008) who analyzed a total of 558 publications in forecasting research.

Time-series data are quite common in all forms of temporal tracking. Those reach from pathological brain activity exploration in medicine, consumption patterns in the energy sector, to stock market analysis in finance, but also to industrial areas like process and quality control or inventory management. (Liao 2005, Zhang 2012, Aghabozorgi et. al. 2015) Their ubiquity has generated a substantial interest in querying, indexing, clustering, classifying, and modeling of such data and their attributes, and much effort has been devoted over the past three decades to the development and improvement of time-series forecasting methods. (Zhang, Patuwo and Hu 1998, Keogh 2006, Zhang 2012, Benidids et. al. 2020, Boylan and Syntetos 2021)

Definition 2: A time-series Y = {y_1, y_2, ..., y_T} is an ordered set of numbers that indicate the temporal characteristics of objects at any time t of the total time period T. Specifically, they contain a contextual attribute (time) and a behavioral attribute (data value). In addition, the single values are temporarily dependent on one another resulting in the following characteristics: linearity and trend, seasonality as well as stationarity.

Using the time series approach for forecasting, forecasters collect and analyze historical observations to determine a model to capture the underlying data-generating process. The model, either local or global[8], is then extrapolated to forecast future values.

Hyndman (2006) proposes two different ways to generate forecasts F where $F = F_{I,(S),H}$ = {$f_{1,(1),\mathrm{T}+1}$, … , $f_{I,(S+J),H}$} from data set D = {Y_1, Y_2, ..., Y_I} and time-series Y_i = {y_1, y_2, ..., y_T} as input data via a particular method m. Those involve

- static forecast simulations with constant origins s, and
- dynamic i.e., rolling forecast simulations with varying origins $s + j$ where $j = 1, \ldots, J$.

Besides the range of origins, it is also the range of horizons h with $h \in H$ which defines the type of the forecasting model:

- One-step forecasting models only predict the next time step of time series Y_i (Benidis et. al. 2020).
- Multi-step forecasting in contrast refers to generating predictions for longer horizons. According to Makridakis et. al. (2018), the three alternative ways comprise the following:
    - *direct* forecasting: For each out-of-sample prediction a separate model is developed.
    - *recursive* forecasting: The recursive strategy involves using a one-step model multiple times, iteratively using the prediction for the prior time step as an input for making a prediction for the following time step. Hence, each prediction of the future horizon depends on the predicted values of the previous time steps.
    - *multi-output forecasting*: One model that is capable of predicting the entire forecast sequence in a one-shot manner is developed. Given an input sequence and generating an output sequence those models are often referred to as Sequence-to-Sequence (*seq2seq*) models. (Benidis et. al. 2020)

[8] *Local* model: With local models, the parameters of the model are learned individually for single time series in a collection of time series. Hence, information retrieval is limited to this series and cannot happen portfolio-wise.
Global model: With global models, parameters and patterns are learned jointly on every series from a collection of time series.

In addition to quantitative or statistical forecasting systems, there are also examples of papers like Forge (2009) or Caniato et. al. (2011), which combine quantitative approaches with qualitative ones, or which solely rely on a qualitative approach.

In the following, a literature review on forecasting methods for intermittent and non-intermittent demands, demand indicators, and accuracy measures is provided. More than 140 articles on spare part and non-spare part demand forecasting are reviewed and three tables of best performing and frequently applied forecasting methods (Chapter 2), input features (Chapter 5.1.1) and evaluation metrics (Chapter 2.3.3) are compiled.

Some of these specialized forecasting models for spare parts, but also general time series prediction approaches applicable to spare part demand, will subsequently be discussed with regards to underlying logics, individual advantages as well as downsides.[9] The scope of ML-based forecasting methods is defined through the literature-based prioritization in Chapter 2.

2.3.1 ML-based Demand Forecasting Methods

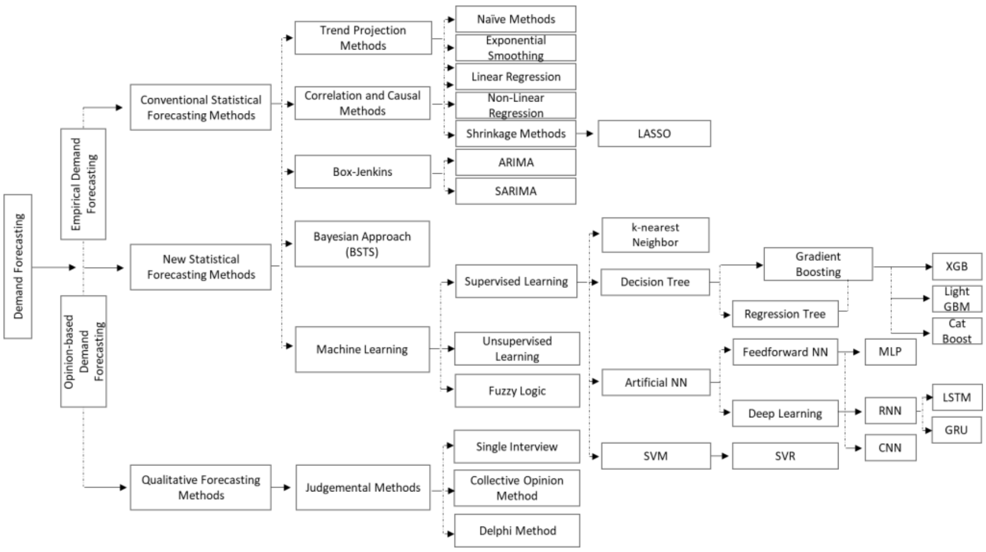


Figure 2.3 Literature based Prioritization of Demand Forecasting Methods

Besides the model classes, and the prediction origins and horizons, it is also the natures of forecasts, which need to be differentiated: Point forecast models (i.e., deterministic models) generate point estimates of the future value, which are either based on the expected value of the time series or some quantile.

Probabilistic forecasting comprises *interval* (i.e., quantile) forecasting, providing an estimate of the possible future range of demand, and *distributional* forecasting. Here, the forecast models predict the conditional distribution of the future values of the time series given past values of the time series (discriminative models) and/or relevant covariates (generative models).

[9] As of Adya and Collopy (1998), a *new* approach must be evaluated in 'terms of alternatives', meaning, that in order to understand what this new approach really contributes to a forecaster's ability, its performance and results need to be contrasted to other methods.

For stochastic, stationary demands, future values can be estimated by means of a distribution assumption.

Fundamentals of Supervised Machine Learning Algorithms

Artificial Neural Networks (ANN)

According to Zhang (2012), ANN are biologically inspired mathematical, massively parallel, supervised learning models, containing layer-wise processing units, so-called neurons or nodes, which are connected. The neurons in the input layer are used to receive information from the data. For a time-series forecasting problem, past observations from the dependent variable (y_1, ..., y_T) and from covariates (c_1, ..., c_T) are used as inputs. The input is processed from the input layer to the output layer via several optional hidden layers[10]. Each neuron calculates its output by a non-linear transfer function and an assigned weight, and passes the result to the next neuron, until the output layer is reached. The output from the network is then used to predict the future value(s) of a time series.

The number of successive layers is called the depth of the network. The overall structure including the depth, width, types of layers or units and the way they are connected is defined by the topology, or architecture of the ANN, with the input layer being the most decisive factor according to Zhang (2012).

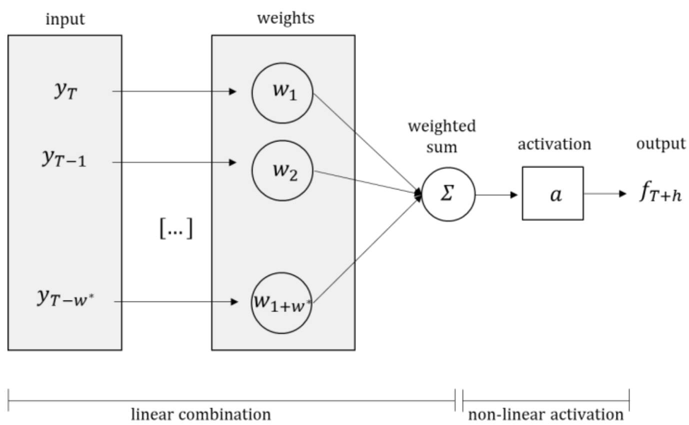


Figure 2.4 Topology of a Perceptron

Figure 2.4 shows the model of a single neuron, a so-called perceptron. The perceptron, originating from a paper by Rosenblatt from 1957, is one of the most widely used basic units of an ANN. The outcome of the perceptron is calculated according to Equation (1).

$$f = a\left(\sum_{1}^{T} w_t \cdot y_t + w_0\right) \tag{1}$$

[10] Number of hidden layers:

- 0: only capable of representing linear separable functions or decisions.
- 1: can approximate any function that contains a continuous mapping from one finite space to another.
- 2: can represent an arbitrary decision boundary to arbitrary accuracy with rational activation functions and can approximate any smooth mapping to any accuracy

The weighted inputs ($w_t \cdot y_t$) and a bias w_0, representing a threshold value, are used to calculate the output of the summing element. The result of the non-linear element is subsequently generated by the unit step function defined in Equation (2). This means, that the perceptron is only activated if $w_t \cdot y_t + w_0 \geq 0$, which is controlled by the learned weights and the bias.

$$f(w_t \cdot y_t + w_0) = \begin{cases} 0 \text{ for } w_t \cdot y_t + w_0 < 0 \\ 1 \text{ for } w_t \cdot y_t + w_0 > 0 \end{cases} \quad (2)$$

While ANN architectures for modern applications have increased in complexity, they continue to be composed of combinations of basic structures, such as MLPs, RNNs and CNNs which have been well-researched and explored for decades (cf. Figure 1.1). (Zhang 2012, Benidids et. al. 2020) Despite very different advantages and downside, Zhang et. al. (1997) generally attribute four advantages to them:

- ANNs are data-driven, self-adaptive methods with few a priori assumptions which learn from “functional relationships among the data even if […] [those] are unknown or hard to describe.
- ANNs can generalize i.e., they can learn from historical samples and subsequently extrapolate the information.
- ANNs are universal approximators, i.e., they can approximate any continuous function and make use of it for estimating future values.
- ANNs are non-linear, i.e., they are capable of performing non-linear modelling.

Feedforward-NN (FNN)

Zhang (2012) states that FNNs, in particular Multi-Layer-Perceptrons (MLP), are amongst the most widely used models for time series forecasting. They propagate information one-directional from the input nodes through hidden layers to the output nodes. Training the NNs refers to the estimation of connection weights w_t. Here, literature proposes various algorithms. Still, the most influential one remains the backpropagation algorithm as of Werbos (1974), which suggests using a gradient descent approach to adjust and determine weights to minimize an overall error function e.g., the sum of squared errors. (Alpaydin 2019: 255ff.)

$$\Delta w_t = -\vartheta \frac{\partial E}{\partial w_t}, \forall t \quad (3)$$

$$w_t = w_t + \Delta w_t \quad (4)$$

One of the main limitations of simple FNN, in particular MLPs, is that they only remember the most current input information. Moreover, the number of inputs and outputs is fixed making them inapplicable to problems with varying input and output sizes as in time series forecasting. (Elman 1990).

Recurrent Neural Networks (RNN)

Due to the shortcomings of classical FNNs, Elman (1990) proposed so-called recurrent connections. These attempt to model time or sequential dependent behavior, i.e., that hidden units learn some kind of feature representation of the raw input, feeding the output of the hidden units from time step $t-1$ back to themselves in each time step t, providing the network with a dynamic memory. Also, at each time step t, the network receives external input y for time t. One crucial detail here is that the same network is used for all time steps, i.e., the weights of the network are shared across time steps. This weight-sharing idea is similar to that of convolutional neural networks where the same filter is used across different parts of the input (cf. subsequent section). This allows the RNNs to handle sequences of varying length during training and, more importantly, to generalize to sequence lengths not seen during training. (Alpaydin 2019: 313f.) Although RNNs have been widely used in practice, a common issue while training them by gradient-based methods is that of so-called *vanishing* or *exploding gradients*. (Hochreiter 1991) This is because small gradients i.e., weights are multiplied several times over the multiple time steps, with the gradients shrinking asymptotically to zero. This means that the weights of the early layers will not be changed significantly preventing the network from learning long-term dependencies. (Alpaydin 2019: 316)

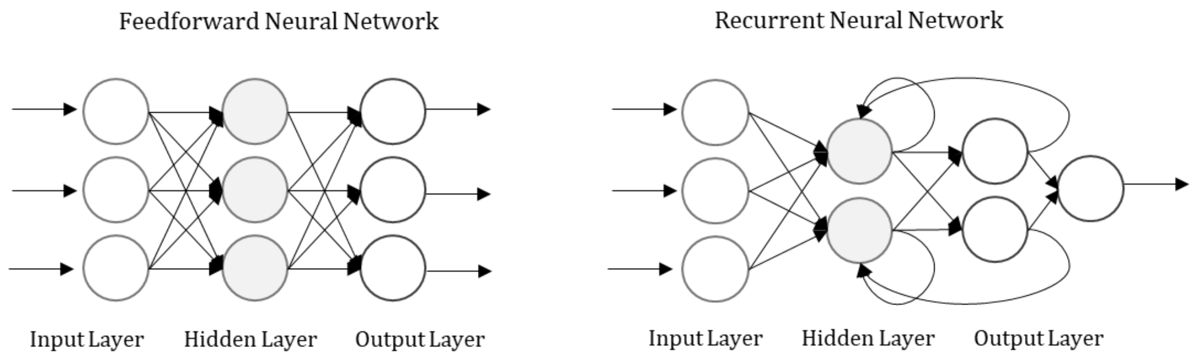


Figure 2.5 Topologies of and Knowledge Transfer in FNNs and RNNs

LSTMs were proposed by Hochreiter and Schmidhuber (1997) to address this problem. Instead of using a simple network at each time step, LSTMs use a more complicated architecture composed of so-called LSTM cell blocks. These blocks again have various components to control the flow of the input to the cell as well as to decide on which kind of information should be kept and which should be propagated to the next time step.

- The input gate decides how much information from the current input flows to the cell state[11]. Relying on an activation function that normally approaches results close to zero or one for inputs, it serves as a forget logic, which removes information from the current input that is not needed.

[11] The cell state represents the internal memory of the cell which stores both short and long-term memories.

- The forget gate is a layer that is filtering i.e., removing information from previous cell states based on current inputs and previous hidden states[12]. This also happens via activation functions.[13]
- The output gate decides how much information from the current cell state flows into the hidden state to enable the LSTM to either pick long-term memories or short- and long-term memories.

With LSTMs being a variant of RNNs, there are again several variants of LSTMs that are widely used in practice. Gated recurrent units (GRU) are a simplification of LSTMs that do not use a separate memory cell and are hence computationally more efficient while still being comparable to LSTMs. (Chung et. al. 2014)

Bidirectional RNN (B-RNN) and **Bidirectional LSTM (B-LSTM)**

In 1997, Schuster and Paliwal propose another variant of RNNs respectively LSTMs, the so-called bidirectional recurrent neural networks (B-RNN). They can be trained using all available input information from both the past and the future of a specific time frame. For this purpose, two RNNs respectively LSTMs are trained – one on the as-is input sequence and one on a reversed copy of it.

Providing the sequence bi-directionally was initially justified in the domain of speech recognition because there is evidence that the context of the whole utterance is used to interpret what is being said rather than a linear interpretation. (Goodfellow et. al. 2018: 436f.) This holds true for time series as well: Forward and backward propagation of the input information can provide additional context to the network and result in faster and even fuller learning of the problem. (Siami-Namini et. al. 2019)

Convolutional Neural Networks (CNN)

CNNs or ConvNets are locally connected feed-forward neural nets that use convolutional layers to exploit ordinal structures i.e., neighboring information, present in three- to one-dimensional input data as used for image recognition, computer vision applications, and time series. (Goodfellow et. al. 2018: 369ff.)

A convolutional layer applies a linear operation, the so-called *convolution*, to smaller patches of the input data. Convolution here refers to the process of computing moving weighted sums by sliding a so-called *filter* or *kernel* matrix over different parts of the input data (cf. Figure 2.6; as in Goodfellow et. al. 2018: 373). The size of these patches as well as how the filter is slid across the input are part of the hyperparameters of the models. The aim is that the weights, the filter is containing, are adjusted in such a way that it is able to extract relevant features from the raw input data. Since the given input data might have various useful features relevant to the problem, more than one filter is typically learned in a convolutional layer, with each filter being used to convolve different parts of the input to extract a single feature.

In addition to convolutional layers, CNNs also use a pooling layer to reduce the size of the feature representation as well as to make the features extracted from the convolutional

[12] The hidden state represents the calculated output information with respect to the current input, previous hidden states, and the current cell input. Also, it can decide to only retrieve short, long-term or both types of information stored in the cell state to make the next prediction.

[13] For more information on activation functions please refer to Appendix 1.

layer more robust. Similarly, to the convolution operation, the pooling operation is applied to smaller neighborhoods by sliding the corresponding filter over the input. A pooling layer, however, does not have any learnable weights and hence both the convolutional and the pooling layer are counted as one layer in CNNs. (Goodfellow et. al. 2018: 379ff.)

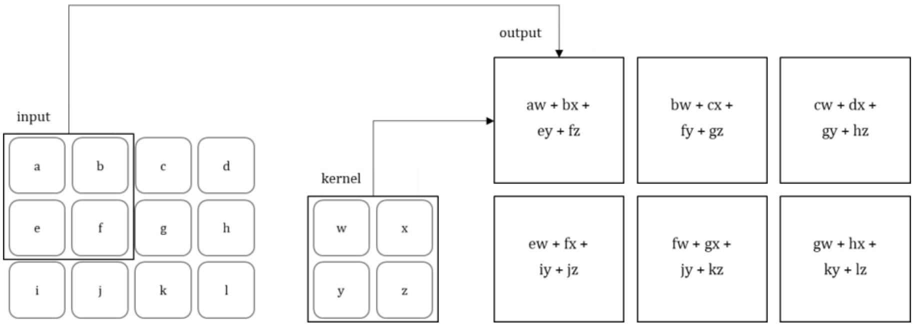


Figure 2.6 Convolution for a 2D Tensor without Kernel Flipping

Encoder-Decoder Architectures and Attention Mechanisms

An autoencoder is a three-component NN designed for the domain of representation learning, i.e., it reconstructs the original input and tries to represent it in the output layer. The inherent advantage is, that through its two functions, the encoding and decoding, one can impose a bottleneck in the NN which forces a compressed and meaningful knowledge representation of the originally inserted features. (Goodfellow et. al. 2018: 563ff.) Accordingly, the ideal autoencoder model balances the following:

- Sensitive enough to the inputs to accurately build a reconstruction and
- insensitive enough to the inputs so that the model does not simply memorize or overfit the training data.

The main challenge derived from this is composed in the question of how many neurons are needed for the hidden layer? If the number of hidden neurons is equal to the number of input neurons, it is likely that the information will be passed through the NN in an unchanged way. The obvious solution are so-called *sparse* autoencoders, which consist of significantly fewer hidden neurons than the input neurons. (Goodfellow et. al. 2018: 566)

A model that makes use of the encoder-decoder functionality is the Sequence-to-Sequence (Seq2Seq) model introduced by Google researchers Sutskever et al. (2014) within the natural language processing (NLP) domain. Here, the significance of a word is often identified in relation to its context. Correspondingly, "[i]n times series data, points of significance are often identified in relation to their surrounding values – such as […] change-points or cyclical patterns." (Lim et. al. 2020) The way it works is as follows: Each instance of the input sequence is captured by a cell, propagated through the encoder model, and then summarized within a fixed-length context vector. The decoder model takes this as an initial input with recurrent units and activation functions estimating the most probable output. This serves as input of the next unit respectively layer.

Moreover, the attention mechanism proposed by Bahdanau et al. (2015) emerged as an improvement over the encoder decoder-based neural machine translation system in NLP. Later, this mechanism, and its variants, were used in other applications, including computer vision, speech processing, and recently also in demand forecasting. (Lim et. al. 2020). The reason for it is its power to overcome the bottleneck phenomenon caused by the fixed-length vector respectively the so-called long-range dependency. In contrast to the basic version that processes and encodes an entire sequence at once, the attention-based approach calculates a context vector for each output instance, the overall context and allowing the decoder to look at the sequences selectively by reacting to specified relative importance values. (Bahdanau et al. (2015), Lim et. al. 2020)

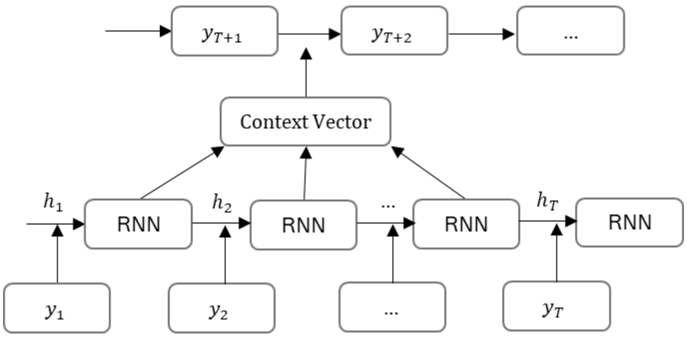


Figure 2.7 Encoder-Decoder Architecture

Gradient Boosting Regressor

Originally designed for classification problems, the concept of boosting came out of the idea of whether a weak learner respectively a weak model can be modified to become better. These thoughts built upon Leslie Valiant's work on *Probably Approximately Correct Learning* from 1984, a framework for investigating the complexity of ML problems where she states that, "[t]he idea is to use the weak learning method several times to get a succession of hypotheses, each one refocused on the examples that the previous ones found difficult and misclassified." (Valiant 2014:152) Hence, boosting could be named a procedure that combines many weak models respectively the outputs of many weak models to produce a powerful *committee* (Hastie et. al. 2017: 337ff.). The most suited and therefore most common algorithm used with boosting are decision trees. In decision trees, the goal is to predict the value of a target variable f_{T+h} by learning simple decision rules inferred from the data features $(y_1, \ldots, y_T)$ and $(c_1, \ldots, c_T)$. All data points are divided into two groups so that similar data points are adjacent; these groups are then further split binary, resulting in fewer and more and more homogeneous data points being related specifically to one node. This process is called recursive partitioning, the models are referred to as *ARCing* (Adaptive Reweighting and Combining) or *Stage-wise Additive Models* (Breimann 1997).

While *Random Forest*, one type of boosting, relies on a random combination of binary questions to split the data points into increasingly homogeneous subgroups, *Gradient Boosting* chooses splits such that the prediction accuracy improves with each iteration. The objective function consists of a loss function and a regularization term, the procedure applied is gradient descent. The result is the weighted average of the predictions of all trees.

Amongst Gradient Boosting, *LightGBM* and *XGB* are optimized distributed i.e., parallelized gradient boosting algorithms. (Chen and Guestrin 2016) For the latter, trees grow depth-wise while in LightGBM, trees grow leaf-wise which is the fundamental difference between the two frameworks. Compared to other versions, XGB is the only algorithm that uses the second order Taylor approximation in the loss function. Hence, the minimization of errors is optimized as more information about direction and magnitude of the descent is provided. Another inherent advantage is its regularization mechanism: XGB has in-built L1 (Lasso Regression) and L2 (Ridge Regression) regularization which prevents the model from overfitting. That is why XGB is also called the *regularized* form of Gradient Boosting.

Ensemble Learning

One of the major developments in time series forecasting is model combining or ensemble modeling. (Qiu et. al. 2014; Galicia et. al. 2019; Masini et. al. 2020) Sharkey (1996) state that "[a]n ensemble can be formed by multiple […] architectures, the same architecture trained with different algorithms […], or even different methods." Based on this, the basic idea of this multi-model approach is the use of each component model's unique capability to better capture different patterns in the time series. Both theoretical and empirical findings have suggested that combining different models can be an effective way to improve the predictive performance of each individual model, especially when the models in the ensemble are quite different. (Brown 2010: 312ff.)

Fundamentals of Unsupervised Machine Learning Algorithms – Clustering

Clustering is one of the most popular data mining methods, not only as a preprocessing step or subroutine for other techniques or as part of a complex system, but also due to its exploratory power in three respects:

- Performance: Identify patterns and correlations in an unlabeled data set, that can either be frequent or rare, and that help to understand and process data further in a faster and more structured way. (Keogh and Lin 2005; Liao 2005; Aghabozorgi et. al. 2015)
- Transparency and Prioritization: Make smarter decisions on which PMCs to focus on. Achieve a better trade-off with regards to forecast accuracy and computational expensiveness and decide on how to best allocate time, budget, and resources by identifying high-value PMC segments and initiatives with the greatest potential business impact.
- Personalization: Understand how different groups of time series should be targeted with different models at the most appropriate parameters. (Keogh et. al. 2004)

According to Wunsch and Xu (2008) as well as Hennig and Maila (2015: 3) clustering methods can be hard respectively crisp, soft respectively fuzzy, or hierarchical. Hard clustering means that the object is included in exactly one cluster, while in soft clustering, time series are members of multiple clusters to varying degrees. Hierarchical clustering also assigns every data point respectively time series to multiple clusters, but at different levels of aggregation.

Definition 3: Given a data set $D = \{Y_1, Y_2, \ldots, Y_I\}$ where Y_i is a time series, the unsupervised partitioning process of D into $G = \{G_1, G_2, \ldots, G_K\}$ occurs such that homogenous time series data are grouped together based on similarity i.e., the intra-cluster distance is minimized and the inter-cluster distance, i.e., dissimilarity across subsets, is maximized. (Kotsakos et. al. 2014: 26, Hastie et. al. 2009: 502). The grouping is called time series clustering. G_k is then called a cluster where $D = \cup_{i=1}^{k} G_i$ and $G_i \cap G_j = \emptyset$ for i $\neq j$.

Algorithms

Whatever the categorization is though, for any time-series clustering approach, the main issues to be considered for effective usage is the choice and design of the clustering algorithm together with the inherent parameter setting (Liao 2005). Reasons are as follows:

- accuracy, as every method expresses homogeneity and separation of clusters differently,
- efficiency and scalability, as the computational costs differ from one method to another, and
- domain dependency, as the approaches have often been developed for particular application areas

As illustrated in Figure 2.8, the approaches can generally be divided into five families (Murtagh 2015: 23-29):

- connectivity-based i.e., hierarchical,
- point-based i.e., partitioning,
- density-based,
- model-based i.e., distribution-based, and
- graph-based clustering.

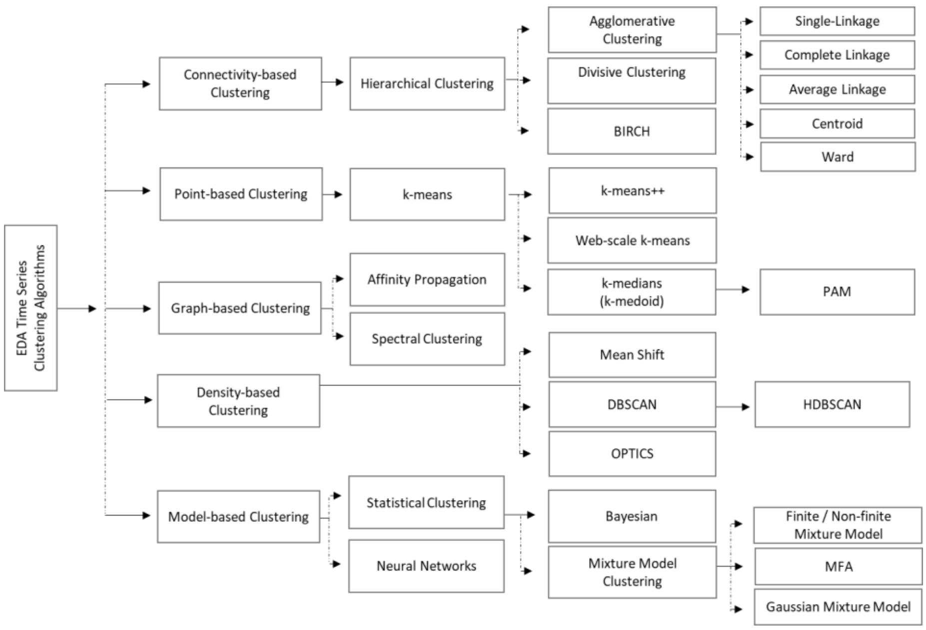


Figure 2.8 EDA Time Series Clustering Algorithms

Liao (2005) also states that the majority of use cases relies on the 'classic' i.e., hierarchical-agglomerative and point-based methods. Commonly applied are also specific adaptions of either the time series themselves by transforming them into flat data or by replacing the default distance measure.

For hierarchical clustering, agglomerative or divisive algorithms are used, with the agglomerative algorithm considering each item as a separate segment and then gradually merging the clusters. The divisive algorithm starts with perceiving all objects as one cluster and then moving top-down to split the objects into single segments. BIRCH, as a third category, has recently started receiving attention due to its success over large data sets. (Murtagh 2015: 23-29) It abstracts smaller, more compact representations from large datasets commonly denoted as *clustering feature entries*. Formally, a clustering feature entry is defined as an ordered triple being composed of the number of data points, the sum of data points and the squared sum of data points in the cluster. This already reveals the major disadvantage of BIRCH: It is only able to process metric features. (Zhang et. al. 1996)

A point-based clustering method makes k groups from i unlabeled objects in the way that each group contains at least one object. Amongst them, *k-means* and *k-medoids* are the most representative examples. From these, *k-medoids* is usually preferred: unlike *k-means*, *k-medoids* computes the dissimilarity matrix of all data sequences and uses actual sequences as cluster centroids. In contrast, *k-means* requires the computation of artificial sequences as centroids, which hinders the easy adaptation of distance measures other than the *Euclidian Distance* (ED). However, from all these methods, only the *k-means* class of algorithms can scale linearly with the size of the datasets.

Table 2.1: Categories of Cluster Algorithms with Strengths and Weaknesses

ALGORITHM	**STRENGTHS**	**DRAWBACKS**
connectivity-based	• no prior knowledge regarding the number of clusters as an initial parameter required • transparency via dendrogram	• quadratic computational complexity • poor scalability • weak in terms of quality as no adjustment of clusters after merging or splitting is possible
point-based	• wide range of distance metrics • centroid is easy to determine	• number of clusters needs to be determined in advance • random choice of clusters and cluster centers • globular clusters
density-based	• extraction of dense clusters leaving sparse background classified as 'noise' • no globular clusters, • no noise lumping in • no pre-setting of cluster number • stable across runs • performance: few clustering algorithms can tackle datasets as large as DBSCAN	• low performance for varying density • distance threshold ε becomes challenging to estimate especially for high-dimensional data
model-based	• support of mixed membership • models account for variance and return the probability that a data point belongs to each of the clusters G_k	• computationally expensive
graph-based	• deterministic over runs • no purely globular cluster assumption	• computationally expensive • number of clusters needs to be determined in advance • random choice of clusters and cluster centers

Aside from all of these conventional approaches, some new papers emphasize the enhancement of algorithms and present customized models, typically denoted as hybrid methods, for time series data clustering. (Zhang et. al. 2010, Aghaborzorgi et. al. 2015, Kumar et. al. 2015)

Distance Measures

The use and efficiency of these algorithms are dependent on the determination of a measure of similarity respectively dissimilarity. ED and its generalizations are the most commonly used metrics (Agrawal et al., 1993; Keogh et al. 2001).

$$\mathrm{ED} = \sqrt{\sum\nolimits_{1}^{T} \left(y_{i,t} - y_{j,t}\right)^2} \tag{5}$$

Its drawback though is its sensitivity towards slight shifts across the time axis, which limits it in terms of identifying time series that are similar in time. (Aghaborzorgi et. al. 2015) Hence, a variety of other measures has been proposed in recent years. Those included autocorrelation (Wang and Wang, 2000), cross-correlation (Agrawal et al., 1993), piecewise probabilistic measures (Keogh and Smyth, 1997), cosine wavelets (Huhtala et al., 1999), piecewise normalization (Indyk et al. 2000), and *cepstrum* (Kalpakis et al. 2001). However, empirical comparisons in Keogh and Kasetty (2002) revealed that ED still performs favorably compared to others when tested on the same datasets.

Continuous Dynamic Time Warping (DTW), an elastic i.e., one-to-many/one-to-none metric, has also been applied in time series mining to resolve the difficulty in clustering time series even when they are shifted or temporally stretched (Kruskal and Liberman 1983, Keogh and Ratanamahatana 2005, Murray et. al. 2018). According to Paparizzos and Gravano (2015) it has been proven to be more appropriate for many time series tasks, the reasons for it lying in its inherent nature of offering invariances to both scaling and shifting.[14] The downside of DTW is mainly related to scalability problems as quadratic computation is required. Hence, in practice it is only applicable if preprocessing through sub-clustering takes place.

Evaluation Metrics for Clusters

In general, the evaluation of extracted clusters is difficult, and it is still an open problem in scientific research (Murray et. al. 2017, Dolnicar et. al. 2018). The dominant reason is the absence of data labels.[15] Moreover, results are highly dependent on the domain as well as on the algorithms, parameters, cluster numbers and size, outliers, and the similarity measures. They are all concepts which are to be referred to individually.

Nonetheless, Keogh and Kasetty (2002) conclude in their paper 'On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration' that the

[14] Time-series distortions and their invariances amongst others comprise amplitude and offset invariance, sampling invariance, phase respectively shift invariance, occlusion invariance as well as complexity invariance. For a more detailed review refer to Batista, Wang and Keogh 2011.

[15] A so-called external index is used to measure the similarity of formed clusters to the externally supplied class labels or ground truth and is the most popular clustering evaluation method. In literature, this index is also known as external criterion, external validation, or extrinsic validation as the ground truth is available. The internal index, on the contrary, is used to measure the goodness of a clustering structure without respect to external information. This index is denoted internal criterion, internal validation, intrinsic and unsupervised metrics. (Lossio-Ventura et. al. 2021)

evaluation of time series clustering should follow certain principles which are defined as follows:

- the implementation bias needs to be avoided by the careful design of experiments,
- the validation of algorithms should be performed on various ranges of data sets unless the algorithm is created only for a specific set and
- complex similarity measures should always be applied in combination with simple and stable metrics.

It is also Palacio-Nino and Berzal (2019) who discuss the desirable marks of *good* clustering results, recalling Kleinberg's *impossibility theorem*, and describing a taxonomy of evaluation criteria for unsupervised machine learning, specifically focusing on the concepts of *cohesion* (intra-cluster distance) and *separation* (inter-cluster distance):

$$Coh(G_k) = \sum_{y_i \in G_k,\ y_j \in G_k} prox(y_i, y_j) \tag{6}$$

$$Sep(G_k, G_l) = \sum_{y_i \in G_k,\ y_j \in G_l, G_k \neq G_l} prox(y_i, y_j) \tag{7}$$

Both are based on a proximity function that determines how similar a pair of time series is. Similarity, dissimilarity and distance functions can be used for this purpose. When determining intra-cluster distances, small values indicate a good cluster quality; for inter-cluster distances, values should be maximized.

The *silhouette coefficient* (SC) as of Rousseeuw (1987) is the most common way to combine the concepts of cohesion and separation in a single metric. It determines for each object whether it is closer to the objects of its own cluster than to those of other clusters. The coefficients move in the interval [-1; 1]. If it approaches the value of 1, the object is closest to the points of its own cluster, if it approaches -1 it is closer to the objects of the nearest cluster. Scores around zero indicate overlapping clusters. Being defined as

$$SC = \frac{\frac{1}{|G_l|} Sep(G_k, G_l) - \frac{1}{|G_k|} Coh(G_k)}{\max\left(\frac{1}{|G_k|} Coh(G_k), \frac{1}{|G_l|} Sep(G_k, G_l)\right)} \tag{8}$$

$\frac{1}{|G_k|} Coh(G_k)$ represents the average distance of a cluster member to all other members within a cluster; $\frac{1}{|G_l|} Sep(G_k, G_l)$ on the other is the average distance of a cluster member to all other members of the nearest cluster. (Liu et. al. 2010)

The *Calinski-Harabasz* (CH) coefficient, proposed as second metric, is also known as the variance-ratio-criterion. It is a measure based on the internal dispersion of clusters and the dispersion between clusters. In this case, dispersion is equal to the sum of squared distances (Caliński and Harabasz 1974).

$$CH = \frac{\sum_1^G prox(g_k, g)^2 \cdot \frac{1}{K-1}}{\sum_{y_i \in G_k} prox(g_k, y_j)^2 \cdot \frac{1}{K}} \tag{9}$$

g_k denotes the cluster representative (e.g., the mean over all cluster members of a specified cluster G_k respectively the centroid), g the overall mean and K the number of clusters. A higher value of the CH index suggests more compact clusters with better separation from each other.

The *Davies-Bouldin* (DB) index is defined as the average similarity of each cluster with a cluster most similar to it. Here, similarity is defined as the ratio of the clusters' size and distance to each other. It is formed by specifying

$$R_{kl} = \frac{\frac{1}{|G_k|} Coh(G_k) + \frac{1}{|G_l|} Coh(G_l)}{prox(g_k, g_l)} \tag{10}$$

$\frac{1}{|G_k|} Coh(G_k)$ respectively $\frac{1}{|G_l|} Coh(G_l)$ denote the averaged distances between the individual objects in cluster G_k respectively G_l; $prox(g_k, g_l)$ represents the distance between centroids g_k and g_l. Basically, here cluster similarity is equal to the sum of two intra-cluster dispersions divided by the separation measure.

The actual DB index is subsequently defined as:

$$DB = \frac{1}{K} \sum_{1}^{K} \max_{k \neq l} R_{kl} \tag{11}$$

Like SC, the DB index does not require a-priori knowledge of the ground-truth labels. The minimum score is zero and – differently from most performance metrics – the lower the resulting values the farther apart and less dispersed are clusters.

According to literature, there are more respectively other metrics with individual advantages and disadvantages. However, the previously defined ones, especially SC and CH, both require similar prerequisites and expose similar reaction patterns (e.g., being higher for convex clusters), although the information used differs. (Liu et. al. 2010)

Literature Review on Conventional Statistical and ML-based Forecasting Methods

In the following some works that make use of conventional statistical and ML-based approaches for time series prediction are discussed to underline the importance of this approach.

As in Aburto and Weber (2007), Carbonneau et. al. (2007/2012) or Makridakis et. al. (2020), predominant methods used are to be found in the primer group, focusing on regular demand patterns with no specialization for spare part characteristics. Those are Exponential Smoothing (ETS) models, Autoregressive (AR) models or Moving Average (MA) approaches, as well as combinations and modifications of the same like models of the Autoregressive Moving-Average (ARMA) family. Bartezzaghi et al. (1999) as well as Makridakis et. al. (2020) note that all these methods assume a certain degree of stability in the nature of the time series, which is often not given for (automotive) spare parts demand. Boylan and Syntetos (2010) continue to argue that the ignorance of such properties could lead to substantial over- or underestimation of future demands.

So, Croston (1972) was the first who presented an exponential smoothing method for the predominantly intermittent demands in this domain, proposing to use separate estimates for the two unknowns *demand size* and *interval*. Syntetos and Boylan (2001) show that Croston's method is biased and suggest an adjustment to overcome this issue in their follow-up publication from 2005 (SBA). Other variants of Croston's method are suggested in literature as well, commonly referred to as TSB and SBJ methods (Kourentzes 2014). In a comparative study though, Teunter and Sani (2009) show that the variants of Syntetos and Boylan (2001/2005) are the most promising ones. Other studies compare variants of Croston with other conventional statistical methods; for this see e.g., Willemain et al., (1994), Ghobbar and Friend (2003), Eaves and Kingman (2004), or Syntetos et. al. 2015. The studies show that most Croston variants outperform these methods on average, but not for all possible situations. An example is Romeijnders et. al. (2012). Relying on top-down based approaches, they develop a two-step forecasting method for the aviation industry that seizes the advantages of exponential smoothing. Results confirm the accuracy of the proposed approach and indicate that in contrary to the methods described previously, it could use information on repair operations and planned maintenance on component level to significantly mitigate forecasts errors. Another example is Regattieri et. al. (2005). They investigate the behavior of 20 forecasting methods when dealing with lumpy demand for high priced aircraft spare parts, relying solely on historical data. The results showed that the best approaches in terms of accuracy are the weighted MA, exponentially weighted MA and Croston.

Gutierrez et al. used an ANN for forecasting of lumpy spare part demands in 2008. According to the authors this is the first time this kind of model was applied to lumpy spare part demand forecasting. They used a three layered MLP. Both the current demand and the period between the last two successive demand occurrences were taken as input variables. Gutierrez et al. compared the performance of the ANN with the conventional approaches Croston, SBA and ETS. Despite the simple topology of the NN, it was found to outperform the other models. Lolli et al. (2017) also published a paper about ANNs for the prediction of intermittent spare parts demand. Different NNs with varying inputs and hyperparameters were tested and compared with Crostons' method and SBA. In an expensive statistical evaluation, it is shown that the ANN models outperform both Croston and SBA.

Arslankaya and Öz (2018) estimate upcoming demands for passenger cars likewise relying on ANNs. Besides the model architecture it is also the human selected factors from macro- and microeconomic domains that contribute to higher accuracies than the ones achieved by simpler benchmarking models as e.g., MA and ETS. Shmueli et. al. (2017) share this opinion: They state that NNs are used for time series forecasting because they incorporate multivariate input into the forecast, and for Yang and Zhang (2019), forecasts that consider exogenous factors are generally better than those solely relying on univariate modeling. Besides incorporating additional information to the forecasting process, Murray et. al. (2017) regard clustering an obligatory task: They benchmark different attribute-, connectivity- and point-based methods prior to implementing ANNs for generating forecasts within the retail sector. So do Nikolopoulos et. al. (2016) and Murray et. al. (2018).

Because of their specialty to capture time dependent patterns, RNNs have also been heavily used for time series forecasting. Bianchi et al. (2017) performed a comparative study and evaluated several recurrent networks, including RNNs, LSTMs, GRNs, Non-linear Autoregressive Exogenous models and Echo State Networks on synthetic and real world data sets. They found that there is no general solution and that each task has specific requirements to the model. They also found Elman RNNs to outperform the more complex gated RNNs on some time series problems, whereas the LSTMs outperformed the other tested NNs in case the time series were non-linear. Amin-Naseri and Tabar (2008) also employed MLPs, RNNs and generalized regression NNs for forecasting spare parts from lumpy demand class. They gathered real world data to examine the performance of the proposed approaches and compared them to Croston and SBA. The results confirmed the superiority of the ML-based methods with the RNN being the best. Ma et al. (2015) published a LSTM model for speed prediction. The model was evaluated on travel speed data collected by sensors on an expressway. An extensive benchmarking against conventional statistical models and other types of ANNs and SVR was accomplished, and the LSTM was found to outperform all of them in terms of accuracy and stability. Even though the application area is different from this thesis, the authors' result enables an important conclusion: LSTMs can capture characteristics of the time series, like seasonality and trend.

In 2017, Hsu proposed a LSTM model augmented by an autoencoder. He argues that the LSTM can capture long-term dependencies of time series, but has difficulties to capture short-term relations correctly, which he tries to overcome by implementing an autoencoder functionality. Experimental evaluation on four data sets, including chaotic time series, shows that the proposed model is superior to other state of the art time series prediction approaches. Salinas et. al. 2020 recommend an autoregressive LSTM (*DeepAR*) based on negative binomial likelihood for probabilistic demand forecasting, stating that their architecture is capable of providing forecasts even for items that have "little or no history available" by learning from globally used time series history from similar SKUs. Comparisons with conventional statistical models and baseline RNNs show that their approach achieves highest point forecast accuracies. Lim et. al. 2020 develop an attention-based deep NN architecture for multi-horizon forecasting. The logic for this mechanism enables the model to flexibly decide which of the input variables are most relevant for each instance of the output sequence. LSTMs take over the function of the encoder and decoder network, gating mechanisms, so called GRNs, "help to skip over any unused components of the architecture." Besides the advantage of generating quantile predictions, Lim et. al. (2020) also point to its capability of seizing short- and long-term dependencies from static, dynamic known and dynamic unknown variables. Also, respective importance scores are returned after each iteration to guarantee interpretability. A benchmarking shows the superiority of the approach compared to conventional statistical and more complex ML-based ones as proposed by Salinas et. al. (2020).

Just recently, Zhuang et. al. (2022) introduced a two-stage method for intermittent demand forecasting in the automotive spare parts sector based on boosting algorithms and compare it to SBA and bootstrapping methods. The forecasting problem is divided into two sub-problems, "going first to predict whether there is demand and then to predict how much demand there is." (Zhuang et. al. 2022) The booster used for regression, a LightGBM model, is enabled to also seize discrete time features, statistical features, and product-related factors. Across different horizons and metrics, the combination of two LightGBMs with an accompanying classification threshold estimation proves most useful.

Table 2.2: Literature Review on Clustering Methods

	AUTHOR	TITLE	YoP	CONTEXT / DOMAIN	METHOD / MODEL	INPUT VARIABLE(S), SAMPLE SIZE	METRICS and RESULTS
1.	Dangerfield, Morris	Top-down or Bottom-up: Aggregate vs disaggregate extrapolations	1992	M-Competition	attribute-based clustering ETS	Input: time series data Scope: 15,753 time series 66 months history with 48 months for training	Distance Metrics: - Results: • bottom-up achieves better results compared to top-down approach (MAPE is reduced by ~ 30%)
2.	Wang et. al.	Characteristic-based Clustering for Time Series Data	2006	Synthetic Data Set	attribute-based clustering: k-means, fuzzy c-means, ViscoverySOMine software, hierarchical clustering	Input: 13 characteristic-based (statistical) features Scope: 50 time series with 4097 observations	Distance Metrics: - Results: • usage of all 13 features give the lowest possible result • best result with 7 features • results of cluster algorithms are quite similar with k-means and fuzzy c-means delivering the best • number and choice of features may vary from data set to data set but results strongly support claim that the results learned on one dataset will generalize to future datasets of the same type.
3.	Nimineț	Automotive Market - From a General to a Market Segmentation Approach	2013	Automotive Industry – B2C	attribute-based clustering	Input: - geographics B2C psychographics Scope: -	Distance Metrics: - Results: • additional classifiers besides geographic aspects for clusters proof beneficial • customer-related attributes are from the functional, sportive and social functional area
4.	Weinstein	Segmenting B2B technology markets via psychographics: an exploratory study	2014	B2B Industry	attribute-based clustering	Input: B2B psychographics: motivation, relationship, risk Scope: -	Distance Metrics: - Results: • demonstration of advantage of attribute-based segmentation • classifiers are derived from the customers' behavior
5.	Aghabozorgi et. al.	A Hybrid Algorithm for Clustering of Time Series Data Based on Affinity Search Technique	2014	synthetic data set real world data sets	raw-data-based clustering: k-medoids k-means hybrid: sub-cluster via CAST prototype definition via calculation of affinity	Input: time series data Scope: 8000 time series	Distance Metrics: Piecewise Aggregate Approximation, ED, DTW (only for synthetic data set) Results: • TTC outperforms benchmarking methods. • challenges in determining number of sub-clusters and in evaluating approaches w/o data labels

					final TTC (Clustering of sub-cluster prototypes via DTW)		
6.	Murray et. al.	Forecasting Supply Chain Demand by Clustering Customers	2015	Bulk Materials	raw-data-based clustering: k-means	Input: sales history industry type label location seasonality (additional ones not named) Scope: 3,500 time series	Distance Metrics: ED Results: removement of outliers in the data set is more decisive than the selection of cluster algorithms
7.	Hyndman, Wang and Laptev	Large-Scale Unusual Time Series Detection	2015	Yahoo Server Metrics	attribute-based clustering PCA	Input: Statistical Features (mean, variance, ACF, trend, linearity, curvature, seasonality. peaks., entropy, lumpiness, etc.) Scope: 1500 time series	Distance Metrics: - Results: • PCA in combination with multi-dimensional outlier detection is used to identify unusual time series • Best performance by α -hull method (generalization of the convex hull)
8.	Rizand-Mure et. al.	Unsupervised Time-Series Clustering of Distorted and Asynchronous Temporal Patterns	2016	synthetic data set	spatiotemporal mean-shift DTW	Input: time series Scope: 176x234x36 time points	Distance Metrics: DTW Results: combination of mean-shift and DTW outperforms benchmark methods
9.	Murray et. al.	Market segmentation through data mining: A method to extract behaviors from a noisy dataset	2017	synthetic data set industrial data set (bulk liquid materials)	raw-data-based clustering hierarchical clustering, k-means ANN	Input: time series Scope: -	Distance Metrics: DTW and ED Results: • improved insights into customer behavior when data base is clustered via DTW • number of cluster segments still to be optimized
10.	Murray et. al.	Forecast of individual customer`s demand from a large and noisy dataset	2018	industrial data set (bulk liquid materials)	raw-data-based: DTW hierarchical agglomerative clustering ARIMA on segment level	Input: time series Scope: -	Distance Metrics: DTW Results: • improved clusters when data base is clustered via DTW instead of attribute-based methods • overfitting when top-down forecasting is applied

Table 2.3: Literature Review on Forecasting Methods

	AUTHOR	TITLE	YoP	CONTEXT / DOMAIN	METHOD / MODEL	INPUT VARIABLE(S), SAMPLE SIZE, LEVEL and FORECAST HORIZON	METRICS and RESULTS
1.	Aburto Weber	Improved supply chain management based on hybrid demand forecasts	2007	Retail (six out of top 50 SKUs)	(Seasonal) Naïve Unconditional Average (S)ARIMA MLPs, Ensemble of ARIMA and MLP	Input: Sales History 13 input features Scope: 5,000 SKUs	Metrics: MAPE, NMSE Results: hybrid approach and MLPs are best
2.	Carbonneau et. al.	Machine Learning-Based Demand Forecasting in Supply Chains Forecasting Supply Chain Demand Using Machine Learning Algorithms	2007 2012	Retail (top 100 SKUs)	ETS ARMA Theta Local/Global MLP Local/Global RNN Local/Global SVR	Input: sales history Scope: 148 monthly bins with 124 months for training	Metric: AE, NAE, NMAE, rank derivation acc. to accuracy metrics Results: • ML-based models are best acc. to accuracy • ETS and SVR are best if rank is considered as well
3.	Kourentzes	On Intermittent Demand Model Optimization and Selection	2014	Spare Parts – Automotive Synthetic data set	Croston SBA SBS TSB ETS	Input: sales history Scope: real: 3000 time series (~ 1,200 time series being lumpy and clumped) synthetic: 5000 time series	Metric: MAE, MSE, MASE; rank derivation acc. to accuracy metrics Results: • TSB and ETS are best acc. to accuracy • optimizing the non-zero demand and inter-demand interval estimates separately found be beneficial for Croston and SBA
4.	de Rego de Mesquita	Demand forecasting and inventory control: A simulation study on automotive spare parts	2015	Spare Parts – Automotive	MA SBA Single Demand Approach Bootstrapping six models for demand distribution	Input: sales history Scope: 10,032 SKUs 72 months of history with 42 months for training	Metric: Target Fill Rate, Realized Fill Rate Results: • main findings are related to spare parts inventory management • configurations and policies depend on demand profile • no best model
5.	Nikolopoulos Babai Zied Bozos	Forecasting supply chain sporadic demand with nearest neighbor approaches	2016	Automotive Industry Synthetic data set	ETS Croston SBA TSB KNN	Input: sales history Scope: 3,000 time series monthly bins	Metrics: MAE, MSE Results: • acc. to overall accuracy NN are en par with top parametric methods when clustering is applied • without any pattern, SBA or TSB are recommended
6.	Lolli et. al.	Single-hidden layer neural networks	2017	Spare Parts – Automotive	Croston SBA	Input:	Metrics: MAPE, bias Results:

		for forecasting intermittent demand			FNN with different settings RNN Time-delay neural network	sales history Scope: 24 time series weekly bins	• no significant difference with regards to MAPE • FNN proposed as in Mukhopadhyay et al. (2012) performs best with regards to bias
7.	Makridakis et. al.	Statistical and Machine Learning forecasting methods: Concerns and ways forward	2018	Data from M3 competition: six domains (limited to high-volume and non-intermittent time series)	Naïve ETS Box-Jenkins MLP Bayesian NN Radial Basis Functions Generalized Regression NN KNN CART SVR Gaussian Processes RNN LSTM	Input: time series Scope: 3003 monthly time series characterized by considerable seasonality, some trend and a fair amount of randomness	Metrics: MAPE, MASE Results: • value of ML method is still to be empirically proven in an objective, indisputable manner • ML methods need to become more accurate, require less computer time, and be less of a black box. • traditional statistical methods are more accurate than ML ones • best is ETS • best ML model is Bayesian NN • worst is MLP
8.	Salinas et. al.	DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks	2020	Five domains: automotive manufacturing e-commerce (2) traffic electricity	Croston ETS negative-binomial autoregressive method state space model RNN with different settings DeepAR (RNN)	Input: sales history covariates Scope: 1046 time series (automotive) 50 months of history (automotive)	Metrics: NAE, NRMSE Results: • DeepAR outperforms benchmarks • automotive: results from RNN with negative binomial are comparable with DeepAR worst is RNN with Gaussian distribution
9.	Lim et. al.	Temporal Fusion Transformer for Interpretable Multi-horizon Time Series Forecasting	2020	Electricity Traffic Retail Stocks	ARIMA ETS Temporal Regularized Matrix Factorization Model DeepAR Deep Structured Semantic Model Transformer Model with local convolutional processing Sequence-to-Sequence Model Multi-horizon Quantile RNN TFT (RNN)	Input: transactional time series data related master data calendar data Scope: 135,000 time series (retail) 120 days of history with 90 days for training	Metrics: quantile loss Results: • benchmark methods are limited for retail and stocks • TFT outperforms competing methods across all experiments • Second best is Multi-horizon Quantile RNN
10.	Zhuang et. al.	A combined forecasting method for intermittent demand using the automotive aftermarket data	2022	Spare Parts – Automotive	SBA Bootstrapping LightGBM Adjusted TrAdaboost	Input: sales history statistical and product features (retention data for the car type, part type, FM/SM) Scope: 3,089 SKUs 120 weeks of history with 116, 111 and 108 weeks for training	Metrics. AUC, MASE, MAAPE Results: • Combination of boosting models to estimate the probability of demand occurrence (via classification) and demand size (via regression) • Best is the combination of LightGBM for classification, Best threshold searching and LightGBM for regression

2.3.2 Feature Engineering and Lead Order Determination

[F]eature engineering is another topic which doesn't seem to merit any review papers or books, or even chapters in books, but it is absolutely vital to […] [machine learning] success. […] Much of the success of machine learning is actually success in engineering features that a learner can understand.

Scott Locklin, in "Neglected machine learning ideas"

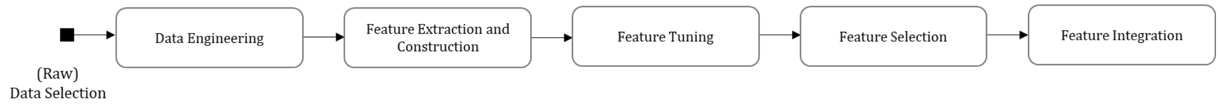


Figure 2.9 Process Steps within Feature Engineering

Preprocessing the data involves both data engineering and feature engineering. *Engineered Features* is a concept which refers to the dataset comprising the tuned features expected by the model—that implies performing certain operations on the information represented by those features. The process behind the concept *Feature Engineering*, sometimes also referred to as *Applied Machine Learning*, is an iterative process and can be done manually or automatically e.g., PCA-based or via multiple layers in a NN, or in combination of both. Applying it in the sense of crafting the features manually, it requires domain knowledge and in-depth analysis of different sample data as well as thinking about how best to expose them to predictive modeling algorithms.

Besides feature engineering, the need for feature selection often arises in ML-based forecasting problems. With high-dimensional data, typically many features are irrelevant and/or redundant for a given learning task, having harmful consequences in terms of performance and/or computational cost. Moreover, a large number of features requires a large amount of memory or storage space. In this context, selection techniques play an important role in reducing the data's dimensionality. (Liu and Motoda 2000; Ferreira and Figueiredo 2012) In general, there are three classes when metaheuristics are applied. Those comprise *filter* methods, *wrapper* methods and regularization. Filter methods evaluate the importance of the factors (John et. al. 1994) outside of the predictive models applying a statistical measure to assign a score to each feature (e.g., *information gain* or correlation respectively regression coefficients). The features are ranked by the score and either selected to be kept or removed from the dataset. The methods may be univariate, considering the features independently and only with regards to the dependent variable, or multivariate. (Kuhn and Johnson 2013: 490) Wrapper methods consider the selection of a set of features as a search problem, where different combinations are prepared, evaluated and compared to other combinations. A predictive model is used to evaluate different combinations of features and assign scores to find the optimal set. It may be methodical, stochastic or heuristic with determining and varying the *Feature Selection Direction* (FSD)[16] being an integral part of it. Regularization respectively embedded models enable the learning of which features best contribute to the accuracy of the model while the model is being created. Another option is the application of *greedy* algorithms. Here, algorithms incorporate features one by one. Besides deciding on a selection technique, there is also the challenge that finding all relevant features is NP-hard.

[16] Forward Selection works as follows: Start with empty feature set $C_0 = \emptyset$; Incrementally add features c_i such that $c_i \not\perp\!\!\!\perp Y_i \mid C_{i-1}$
Backward Selection works as follows: Start with full feature set C_0; Incrementally remove features c_i such that $c_i \perp\!\!\!\perp Y_i \mid C_{i-1} \backslash c_i$

2.3.3 Backtesting and Evaluation Metrics

The benefit out of a forecasting method arises from its forecast quality measured by both the forecast accuracy and the robustness of the forecast. With regard to accuracy, there are two types that equally represent the success of a model:

- *ex-post* forecast accuracy: How well can a model approximate, i.e., reproduce, the historical time series?
- *ex-ante* forecast accuracy: How well can a model forecast the future, i.e., approximate out-of-sample data?

The most basic form of evaluation e is defined as follows

$$e_{it} = y_{it} - f_{it} \tag{12}$$

determining the deviation between the actual value y_{it} and prediction f_{it}.

In literature a variety of additional accuracy metrics, sometimes also denoted as *error* or *evaluation metrics*, have been applied. Amongst others, it is Armstrong and Collopy (1992), Makridakis (1993), as well as Hyndman (2006) and Hyndman and Koehler (2006) who are highlighting their individual characteristics, commonly distinguishing in-between *scale-dependent* and *scale-independent* errors. The latter are to be applied for assessing the forecast accuracy for a portfolio of time series respectively PMCs i. Additional criteria, amongst others listed by Martin et. al. (2020) are also valid for the present use case:

- levering the zero property of division, enabling the following scenarios:
  - $y_{it} = 0;\ f_{it} > 0$
  - $y_{it} > 0;\ f_{it} = 0$
  - $y_{it} > 0;\ f_{it} > 0$
  - $y_{it} = 0;\ f_{it} = 0$
- outlier insensitivity
- symmetry i.e., even penalization of under- and over-estimation
- assessing the variance-bias-trade-off
- interpretability as well as
- economic considerations i.e., the proposed metric should be zero for optimal storage by a perfect prediction and if the deviation from the perfect prediction increases, the metric should be greater than zero, also representing the costs of the misprediction for the specific PMC i.

They are used to correctly assess capabilities of the specific statistical and business-related metrics in the subsequent sections and to make a final selection (cf. Table E 2).

Besides the metric, it is also the *process* of evaluating a forecasting method against existing historical data, which determines how realistically the degree of robustness can be defined. This is referred to as *hindcasting* or more commonly *backtesting*. In the context of time series forecasting, random splits or cross validation is not possible. A simple train-/validation-/test-split only provides a one-time assessment. Hence, Hyndman and Athanasopoulos (2018) recommend using walk-forward validation (*Rolling Window*

Forecast) relying on expanding windows by varying the forecast origins for the model.[17] A more detailed description of the approach and its practical application are found in Chapter 6.1.

Statistical Forecast Error Metrics for Point Forecasts

The most commonly used scale-dependent metrics, i.e., measures for which the scale depends on the scale of the data, are based on absolute errors (AE) or squared errors (SE). (Fildes and Makridakis 1988) They are useful when comparing different forecasting methods that are applied to single time series or data with the same scale. (Kim and Kim 2016)

Being not unit-free, the authors Armstrong and Collopy (1992: 69ff) as well as Chen and Yang (2004) state for MSE and RMSE that besides being incapable of enabling a comparison of different forecasting methods based on a portfolio of time series, those are also highly sensitive against outliers.

$$MAE_i = \frac{\sum_{t=1}^{T} |y_{it} - f_{it}|}{T} \tag{13}$$

$$MSE_i = \frac{\sum_{t=1}^{T} (y_{it} - f_{it})^2}{T} \tag{14}$$

$$RMSE_i = \sqrt{\frac{\sum_{t=1}^{T} (y_{it} - f_{it})^2}{T}} \tag{15}$$

For intermittent demand data, Syntetos and Boylan (2005) recommend the use of the geometric mean of the squared error (GMSE) respectively the geometric mean of the root mean squared error (GRMSE). They point out, though, that GMSE and GRMSE have the flaw of being equal to zero when any error is zero, a problem which will occur when both the actual and forecasted demands are zero – a reason for Hyndman and Koehler (2006) to consider them an inappropriate measure for assessing accuracy on intermittent demand data.

$$GMSE_i = \sqrt[T]{\prod_{t=1}^{T} (y_{it} - f_{it})^2} \tag{16}$$

$$GRMSE_i = \sqrt[2T]{\prod_{t=1}^{T} (y_{it} - f_{it})^2} \tag{17}$$

Percentage errors have the advantage of being scale independent, a key characteristic for a good measure according to Makridakis (1993), making them be frequently used to compare forecast performance between different time series. (Kim and Kim 2016) The absolute percentage forecast error (APE) is defined as in (18) with the two available sub-variants symmetric absolute percentage error (sAPE) and arctangent absolute percentage error (AAPE).

$$APE_i = \left| \frac{y_{it} - f_{it}}{y_{it}} \right| \tag{18}$$

[17] A variation of this approach focuses on either a single or a multi-step forecast horizon for each test set.

$$sAPE_i = \left|\frac{y_{it} - f_{it}}{y_{it} + f_{it}}\right| \tag{19}$$

$$AAPE_i = \arctan\left(\left|\frac{y_{it} - f_{it}}{y_{it}}\right|\right) \tag{20}$$

The most commonly used percentage error metrics are the mean absolute percentage error (MAPE) and the median absolute percentage error (MdAPE).

$$MAPE_i = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{it} - f_{it}}{y_{it}}\right| \tag{21}$$

$$MdAPE_i = \text{median}\ (\mathrm{APE_{i1}}, \mathrm{APE_{i2}}, \ldots \mathrm{APE}_{iT}) \tag{22}$$

Measurements based on percentage errors have two main disadvantages. They are placing a heavier penalty on positive errors (i.e., underestimation) than on negative errors (i.e., over-estimation) which leads to asymmetry. (Makridakis 1993) Moreover, they are having an extremely skewed distribution when actual values are close to zero or are infinite respectively undefined if there are zeros in the series, either for the actual or the predicted values, as it is frequent for intermittent data. (Kim and Kim 2016)

sMAPE, proposed first by Armstrong in 1985 and later by Makridakis (1993), resolves the issue of producing infinite results for y_t being equal to 0 by defining the denominator as half of the sum of the actual and the forecasted value. It is defined as in (23), and in practice and for more comprehensive interpretability as in (24).

$$sMAPE_i^{200} = \frac{200}{T}\sum_{t=1}^{T}\left|\frac{y_{it} - f_{it}}{|y_{it}| + |f_{it}|}\right| \tag{23}$$

$$sMAPE_i^{100} = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{y_{it} - f_{it}}{|y_{it}| + |f_{it}|}\right| \tag{24}$$

While fixing the asymmetry of boundlessness, sMAPE introduces another kind of delicate asymmetry caused by the denominator of the formula, in that under-forecasting ($y_{it} > f_{it}$) is penalized more severely than over-forecasting ($y_{it} < f_{it}$). (Ord 2001) Results are also unstable when both f_{it} and y_{it} are very close to zero and account for being undefined, when both f_{it} and y_{it} are equal to zero. Hence, a popular and commonly accepted approach is to add minor absolute values to both terms in the denominator.

Moreover, whenever either f_{it} or y_{it} is zero, $sMAPE_i$ will automatically hit the upper boundary value 200 or 100 for having equal values in the numerator and the denominator. For demand occurring infrequently, there are countless such cases.

$MAAPE_i$calculates the mean arctangent percentage error between eventual outcomes and the forecast.

$$MAAPE_i = \frac{1}{T}\sum_{t=1}^{T}\arctan\left(\left|\frac{y_{it} - f_{it}}{y_{it}}\right|\right) \tag{25}$$

The function *arctan* APE_{it}, is the inverse function for the bijective tangent of APE_{it}, in the range $]-\frac{\pi}{2};\frac{\pi}{2}[$. Hence, it is defined for all real values from negative infinity to infinity (i.e., $D=\mathbb{R}$) with $W=]-\frac{\pi}{2};\frac{\pi}{2}[$.[18] With a slight manipulation of notations, for the range of $[0,\infty]$ of APE_{it}, the corresponding range of $AAPE_{it}$ is $\left[0,\frac{\pi}{2}\right]$ with zero being the best possible result and $lim_{APE_{it}\to\infty} tan^{-1} APE_{it}=\frac{\pi}{2}$. Unlike MAPE and MdAPE, MAAPE is thus finite when the true value y_t is close to zero i.e., in this case the arctangent absolute error approaches $\frac{\pi}{2}$. (Kim and Kim 2016) A further advantage of MAAPE is that it is scale-free as its values are expressed in radians.

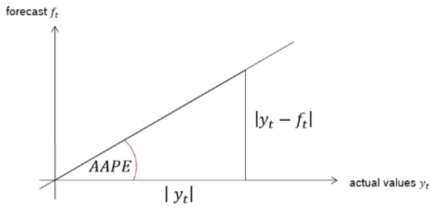


Figure 2.10 arctangent of APE_{it}

The remaining disadvantage is that AAPE is undefined when y_{it}, or both f_{it} and y_{it} are equal to zero. Like with sMAPE, true values y_t which equal to zero need to be disregarded or a minor value needs to be added. Kim and Kim (2016) propose 0.01. Also, MAAPE is not appropriate, if "extremely large forecast errors are considered as genuine variations that might have some important business implications, rather than being due to mistaken or incorrect measurements [...]." (Kim and Kim 2016)

$$R_i^2{}^{[19]} = 1 - \frac{\sum_t^T (y_{it} - f_{it})^2}{\sum_t^T \left(y_{it} - E(y_{it})\right)^2} \qquad (26)$$

R^2 is calculated by dividing the sum of squared prediction errors and the total sum of the squared deviations of y_{it} and the expected value of y_{it}. Results are in the range zero and one. The higher the values, the better the fit between predictions and actual values.

The final category are scale-independent relative metrics, which use reference forecasting values to scale errors. Let

$$RAE_{it} = \left|\frac{y_{it}-f_{it}}{y_{it}-f_{it}^*}\right| \qquad (27)$$

denote the relative error where $y_{it}-f_{it}^*$ is the forecast error obtained from a benchmark method. Then the following metrics can then be defined as follows:

$$MRAE_{it} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{it}-f_{it}}{y_{it}-f_{it}^*}\right| \qquad (28)$$

[18] $\tan(\alpha) = \tan(APE_t) = \left|\frac{y_{it}-f_{it}}{y_{it}}\right|$ and $\alpha = \arctan\left(\left|\frac{y_{it}-f_{it}}{y_{it}}\right|\right) = AAPE_{it}$

[19] where $E(y_{it}) = \frac{1}{T}\sum_t^T y_{it}$

$$MdRAE_{it} = \text{median}\ |RAE_{it}| \tag{29}$$

$$GMRAE_{it} = \sqrt[m]{\prod_{t=1}^{m} \left| \frac{y_{it} - f_{it}}{y_{it} - f_{it}^{*}} \right|} \tag{30}$$

Like MAPE and MdAPE, relative error metrics become unstable, if there are several periods with zero demands, resulting in undefined values.

As of Hyndman and Koehler (2006), the mean absolute scaled error (MASE) is proposed as the best and also the generally applicable measurement of forecast accuracy without the problems seen in the other measurements. It is obtained by scaling the forecast error based on the in-sample mean absolute error using a naïve forecasting method. Hence, it is defined as follows:

$$MASE_i^{non\_seas\ 20} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{|y_{it} - f_{it}|}{\frac{1}{N-1} \sum_{n=2}^{N} |y_{in} - y_{in-1}|_B} \right| \tag{31}$$

$$MASE_i^{seas} = \frac{1}{T-M} \sum_{t=1}^{T} \left| \frac{|y_{it} - f_{it}|}{\frac{1}{N-1} \sum_{n=M+1}^{N} |y_{in} - y_{in-M}|_B} \right| \tag{32}$$

The result, a scaled error, is less than one if it arises from a better forecast than the average one-step, naïve forecast computed in-sample. Conversely, it is greater than one if the forecast is worse than the average one-step, naïve forecast computed in-sample. (Hyndman 2006)

The MASE can be used to compare forecast methods on a single series, and, since it is scale-free, to compare forecast accuracy across series. Also, it penalizes positive and negative forecast errors as well as errors for forecasts related to high quantities and low quantities equally. On the other hand, as a relative measure it relies strongly on the selection of the benchmark method and on how good the time series can be forecasted by the selected method. Another disadvantage is that the MASE is infinite or undefined if all historical observations are equal.

Business KPIs for Point Forecasts

The forecast accuracy index (FAI) as defined by VDA 9000 (VDA 2010) and LG07 (Odette 2012) is a performance indicator ranging from 0% to 100%. 0% represents the worst value of forecast accuracy, while 100% means that the forecast values f_{tl} show no (weighted) deviation from the reference value y_{it}. A special characteristic of the metric is that not a single, but different lag forecasts f_{itl} for time t with lag $l \in [1; L]$ are evaluated. Weights α_l are chosen according to business needs with $\alpha_1; \dots;\ \alpha_l \geq 0$ and $\sum_{l=1}^{L} \alpha_l = 1$.

The formula is as follows:

$$FAI_i = \frac{1}{T} \sum_{t=1}^{T} FAI_{it}\ \text{with} \tag{33}$$

[20] where $\frac{1}{N-1} \sum_{n=2}^{N} |y_{in} - y_{in-1}|_B$ is the within-sample MAE_{it} of the naïve benchmark method and n = 1 … N is the set of forecasting sample periods.

$$FAI_{it} = \begin{cases} \sum_{l=1}^{L} \alpha_l \cdot \max\{0\,;\,(1 - |APE_{itl}|)\}, & \text{if } y_{it} \neq 0 \\ 0, & \text{if } y_{it} = f_{itl} = 0 \\ \infty, & \text{if } y_{it} = 0 \cap f_{itl} \neq 0 \end{cases} \tag{34}$$

In order to also consider the relevance of different products, it is recommended to apply a weighted aggregation with weights w_i either representing production costs, purchase, or sales prices.

$$wFAI_t = \frac{\sum_{\mathrm{i}}^{I} \max\{y_{it}; \sum_{l=1}^{L} \alpha_l \cdot f_{itl}\} \cdot w_i \cdot FAI_{it}}{\sum_{\mathrm{i}}^{I} \max\{y_{it}; \sum_{l=1}^{L} \alpha_l \cdot f_{itl}\} \cdot w_i} \tag{35}$$

The second business-related metric, the MBA (Mittlere Bestandsabweichung), ensures interpretability. MBA is an acronym for the German translation of *Aggregated Forecast Deviation*. Its calculation is based on an ex-post-comparison between cumulated forecasts $\frac{1}{\widetilde{H}-t}\Sigma_t^{\widetilde{H}} f_{it}$ and true values $\frac{1}{\widetilde{H}-t}\Sigma_t^{\widetilde{H}} y_{it}$ for a specified period $\widetilde{H}$ with $\widetilde{H} \epsilon H$ and with $\widetilde{H} - t$ being equivalent to the lead time of PMC i. Hence, the MBA is defined as

$$MBA^{21} = \frac{1}{\widetilde{Q}} \sum_{\tilde{\mathrm{q}}=1}^{\widetilde{Q}} MBA_{\tilde{\mathrm{q}}} \text{ with} \tag{36}$$

$$MBA_{\tilde{q}} = \frac{\frac{1}{\widetilde{H}-t} \left| \sum_{t}^{\widetilde{H}} (y_{it} - f_{it}) \right|}{\frac{1}{\widetilde{H}-t} \sum_{t}^{\widetilde{H}} y_{it}} \tag{37}$$

An estimate of the forecast quality is based on the following empirical values:

$$MBA = \begin{cases} [0; 1] = good \text{ forecast} \\ ]1; 2.5] = medium \text{ forecast} \\ ]2.5;\ \infty[= bad \text{ forecast} \end{cases}$$

A decomposition of the forecast error e_{it}, as described in more detail in Appendix 2, proves that it is not only the deviation itself, but also the model's tendency to over- or underfitting that must be evaluated.

The tracking signal (TS) is a performance indicator with values ranging from -1 to +1. In case of positive values, it indicates a (weighted) over-forecasting tendency. The lower the values, the stronger the under-forecasting tendency.

$$TS_i = \frac{\sum_{t=1}^{T} (f_{itl} - y_{itl}) \cdot \alpha_l}{\sum_{t=1}^{T} \left|(f_{itl} - y_{itl})\right| \cdot \alpha_l} \tag{38}$$

[21] where the number $\tilde{q}$ of MBA periods of length $\widetilde{H}$ is defined as $\tilde{q} = \frac{H}{\widetilde{H}-t}$.

2.4 Sample Size Definition

A challenge arises due to the high volume of data in combination with the computational expensiveness of the models. As in statistics, information is often inferred about a population by studying a finite number of individuals from that population, i.e., the population is sampled, and it is assumed that characteristics of the sample are representative of the overall population, the very same approach will be applied for all quantitative experiments.

According to Cochran (1963:72ff), an adequate sample size n is determined taking into consideration the population size N, the margin of error ε as well as the confidence level denoting the reliability of the estimation procedure like in Equation (39):

$$n = \frac{Z^2 \cdot p \cdot (1-p)}{\varepsilon^2} \tag{39}$$

Z^2 is the abscissa of the Gaussian curve that cuts off an area a at the tails (where $1-a$ equals the desired confidence level), p is the estimated proportion of an attribute that is present in the population. Yamane (1967) relies on the same approach but provides a simplified formula demonstrated in Equation (40):

$$n = \frac{N}{1 + N(\varepsilon)^2} \tag{40}$$

Given a large population with unknown variability in the proportion, a maximum variability p = 0.5 is assumed[22]. Furthermore, suppose a 95% confidence level as well as a rather high precision rate ε with $\varepsilon_1 = \pm 3\%$ respectively $\varepsilon_2 = \pm 5\%$ is required. Applying Yamane‘s formula with the above values, sample sizes in Table 2.4 result:

Table 2.4: Sample Sizes for Different Precision Levels

SIZE OF POPULATION	**SAMPLE SIZE n FOR PRECISION ε OF:**			
	±3%	**±5%**	±7%	±10%
1,000	*use all*	286	169	91
2,000	714	333	185	95
5,000	909	370	196	98
10,000	1,000	385	200	99
20,000	1,053	392	204	100
50,000	1,087	397	204	100
100,000	1,099	398	204	100
> 100,000	1,111	400	204	100

For very large population sizes (> 5,000), an upper bound of $n = 1{,}000$ is hit. In order to represent the composition of the original product portfolio in the sample when applying global models, the share of the individual value contribution classes (A/7.61; B/12.75; C/79.64) must be considered. Consequently, the product class with the lowest share in the portfolio is fixed at the (sub)sample size $n_A = 1{,}000$. This results in a (sub)sample size $n_B = 1{,}675$ and a (sub)sample size $n_C = 10{,}461$. The total sample size n results in $n =$ 13,136. In addition, consideration is given to a fair distribution of the different regions and BUs.

[22] The use of the level of maximum variability (p =0.5) in the calculation of the sample size for the proportion generally will produce a more conservative sample size (i.e., a larger one) n.

2.5 Conclusion

Various ML methods have been proposed in academic literature as alternatives to conventional statistical ones for time series forecasting. The same is valid for clustering approaches. The results though clearly show that no single model is the best and that the best performer is dependent on the data and other conditions. (Murray et. al. 2017; Dolnicar et. al. 2018; Makridakis et. al. 2018)

However, within the area of ML-based models, there are some indications (cf. Table 2.5) which were originally listed by Hastie et. al. (2009: 351) and that are now adapted according to current developments. The green triangle is synonymously used for the key *good*, the grey box for *fair* and the red triangle for *poor*. Those and the fact that it is not completely new models, but rather new combinations of existing models and components that contribute to the state-of-the-art (cf. Figure 1.1) enable a prioritization for this thesis.

In addition to the discussion of forecast accuracy, Hansmann (1983:141) also calls for an evaluation of computational costs as well as the complexity or ease of use of the methods available. This would also require individual assessment of the utility of appropriate models and procedures per use case and data set. As this thesis compares multivariate behavior-based ML methods with conventional statistical forecasting methods detached from the investment decision of their implementation, an assessment of the forecasting quality comes to the fore and results in the selection of models displayed in Chapters 1.2. and 4.3.

Table 2.5: Characteristics of Different ML-based Methods

CHARACTERISTIC	NEURAL NETS	SVM	DECISION TREES	k-NN, KERNELS
Natural handling of data of *mixed* type	▽	▽	△	▽
Handling of missing values	▽	▽	△	△
Robustness to outliers in input space	▽	▽	△	△
Computational scalability (large D and high number of C)	□	▽	△	▽
Ability to deal with irrelevant inputs	□	▽	△	▽
Ability to extract linear combinations of features	△	△	▽	□
Interpretability	□	▽	□	▽

In accordance with the preceding explanations of the evaluation metrics, a single metric will not meet all corresponding requirements. Instead, an attempt is made to achieve the best possible coverage for the criteria defined on page 41 by selecting several statistical and business-related KPIs. For the former, it is $sMAPE^{200}$ and $MAAPE^{arctan2}$; for business KPIs $wFAI$, MBA as well as TS are selected.

To enable the scenarios listed below, the following adaptions apply:

- if: $y_{it} = 0$; f_{it} = 0 then: APE_{it} = 0 (originally undefined)
- if: $y_{it} = 0$; f_{it} > 0 then: APE_{it} = 1 (originally infite)
- if: $y_{it} = 0$; f_{it} = 0 then: $MAAPE_{it}$ = $MAAPE_{it}^{arctan2}$ = $\frac{1}{N}\sum_{t=1}^{N} \text{arctan2}\left(\left|\frac{y_{it}-f_{it}}{y_{it}}\right|\right)$

as arctan2 $\left(\left|\frac{y_{it}-f_{it}}{y_{it}}\right|\right)$ with $y_{it} = 0$ returns a finite value in the range of [0; $\frac{\pi}{2}$].

In order to objectively evaluate the relative performance of the ML methods, it is never the sole averaged error value plus a 0.95-confidence interval that is considered, but rather their distribution.

3 Analysis and Evaluation of Existing Premises and Approaches

This chapter intends to provide a basic understanding of existing premises and the methodology in use. It takes up the findings from Chapter 2.1 and presents the challenges related to corporate structures, market participants and the product portfolio in more detail.

3.1 Corporate Structures and their Implications in a Volatile Market

In a multinational automotive company with a wide product portfolio, for which it needs to offer aftermarket service, an extensive range of spare parts needs to be provided. The complexity that arises from this variety is multiplied by the prevailing regional and customer-driven conditions. Bode and Wagner (2015) call this *spatial* and *vertical supply chain complexity.* On the *horizontal*, the competition in the market affects the supply and demand behavior of the individual players.

3.2 Market Participants and Competition

Automotive parts and components – and thus also spare parts – are manufactured by car makers and, above all, by automotive suppliers who already produce these parts for new car production. In addition, copy manufacturers are entering the spare parts market, reproducing original parts. In general, the above manufacturers sell into two different sales channels. The first is the traditional sales channel through the auto makers' service organizations (OES), which serves the car manufacturers' authorized repair shops. In the second sales channel, the IAM, independent formats, such as wholesalers, retail and internet providers as well as independent workshops, are served. Traditionally, the IAM has been a fragmented market with a limited number of first level customers. Hence, one could speak of a demand oligopoly. Continued consolidation increases market power on the customer side and supports economies of scope, in the sense that targets can be negotiated more effectively and at the same time achieved more easily. Another special feature is that there are no clear hierarchies or predefined sales paths anymore. Rather, transactions can be made to and from all channels and formats (cf. Figure 3.1).

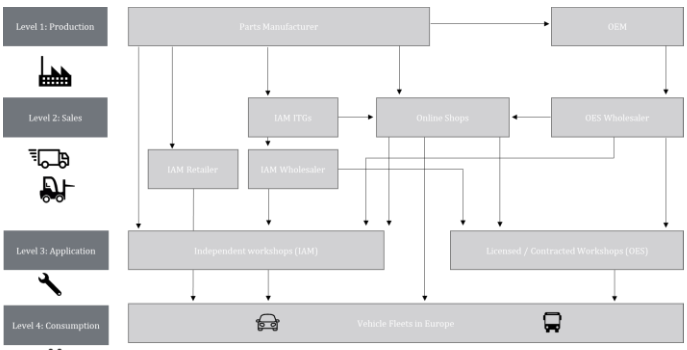


Figure 3.1 Sales Structure in the Automotive Aftermarket

A differentiation of the supplier market and the prevailing conditions can be made by classifying the products into Technical Parts (TP) and Commodities (COM). While the number of TP suppliers is low to medium and market entry is difficult, there is a large number of suppliers for COM products. Here, high intensity of competition with concomitant increasing pressure on prices and margins is prevailing.

In general, there are fixed and well-established business relationships between suppliers and customers. Annually updated contracts and target agreements aim to strengthen long-term ties, fix quotas, generate post-purchase consistency and avoid post-purchase dissonance.

3.3 Product Portfolio and Classification

The product portfolio of a supplier in the aftermarket is broad and deep. This is due to two reasons: The first is the Principle of *One-Stop-Shopping*. It is based on an agglomeration of several related or unrelated product groups and enables customers to satisfy complementary needs (and goals) at a single point of contact. The second reason relates to the large number of product generations and the existing follow-up delivery obligations with a fixed term of more than ten years. This means, that even though primary products are phased out and replaced by new product generations, spare parts must be kept in stock for the previous primary products as well as for the new ones. This leads to very long product life cycles (cf. Figure 3.2) and a constant expansion of the product range, resulting in demands that can be characterized by small volumes, low frequencies, C-parts status and a highly volatile and intermittent nature. (Nemati Amirkolaii et. al. 2017)

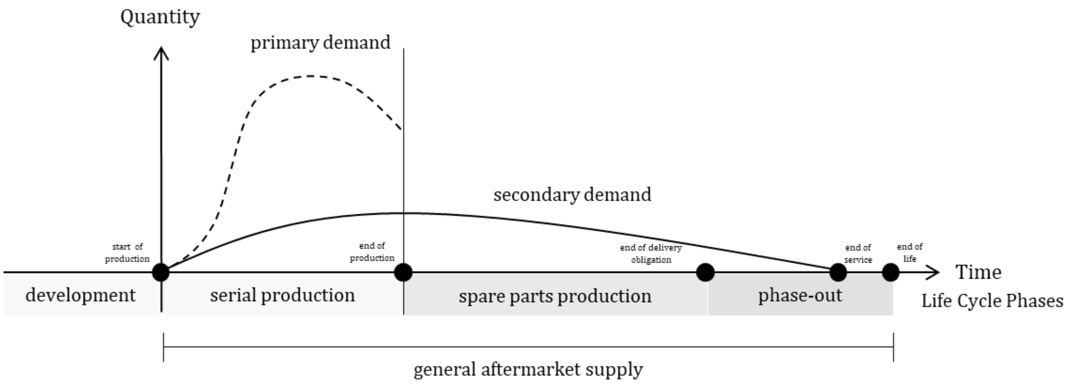


Figure 3.2 Product Life Cycles of Spare Parts (acc. to Klug 2010)

In order to apply the best intervention and to support planning and inventory management of these parts, a classification framework that considers the context related to each case is needed. (Jouni et al. 2011). For this the following criteria apply:

- the degree of complexity and a corresponding differentiation according to the previously described product type TP and COM
- the product category with elements *service*, *wear-and-tear* and *accident* part that enables a reference to the product's technical life cycle
- the economic product life cycle which comprises six different phases from prototyping to maturity phase to phase out

- the product status with corresponding turn rates i.e., fast movers and slow movers and its value contribution

In general, parts with highly sporadic demands and a low value contribution (C-part) are often labeled as *Make to Order* products and are accordingly without requirement for a forecast.

3.4 Analysis of the Methodology in Use

The methods in use within the business are key figure (FAI, CSL)- and extrapolation-based forecasting methods.

For obtaining the system prognosis, historical sales data comprising 36 months, are projected into the future using an ensemble of Holt's exponential smoothing, MA, linear regression and seasonal trend decomposition, with the linear regression being solely used to forecast the trend component within the seasonal trend decomposition on a higher aggregated level.[23] The indices are subsequently applied to the individual cluster members. For time series that exceed a pre-defined threshold of trailing zeros, zero forecasting (ZF) is applied. In case of time series that are shorter than the required minimum length, results are based on a simple MA algorithm. Thresholds for the outlier correction and parameters for the forecasting models are estimated via a grid search method per cluster[24].

Although the algorithm uses the merits of different methods, a common drawback becomes evident: It is univariate, i.e., lead factors are neither integrated nor seized. Rather, an adjustment is made subsequently via the subjective assessments of experts. These experts are usually from interface departments between sales and logistics and rely on continuous information transfer with regards to estimated market development, empirical values, internal business knowledge and experience.

3.5 Conclusion

The previous chapter shows how vast and diverse market participants, sales channels, products and associated complexities are. Hence, the precondition to build and implement a successful *behavior-based* benchmark model is to correctly identify and map the demand triggers in an individualized manner.

[23] The clusters are formed attribute-based, seizing regional and product hierarchy (PH) labels. The trend is determined on PH3-, seasonality on PH2-level. (cf. Appendix 3) The clusters need to exceed a minimum size of 25 PMCs with at least 48 orderlines each. In case that the minimum size is not reached, a fallback scenario is applied by choosing the next higher product hierarchy.

[24] The clusters are formed attribute-based, seizing regional, product hierarchy, value contribution (based on orderlines) and RIS labels. The product hierarchy chosen is the BU.

4 Development of a Behavior-based ML Model for Demand Planning

The following chapter deals with what *behavior-based* actually means, how a *behavior-based* model is defined, which concept is considered valid for it (Chapter 4.1), and which challenges should be paid attention to with respect to data and model selection.

In general, *behavior-based* means that the behavior of a time series, which represents the evolution of demand over time, is to be explained and extrapolated by selecting and combining empirical values (data) in such a way that the highest possible transparency regarding the future behavior is achieved. Whilst in past scientific studies, progress was to be achieved mainly through model improvement and extension, this approach's focus lies on the process of creating representations of knowledge that increase the effectiveness of the model. This is also referred to as *data-centric* machine learning.

It is to be noted that the models' output does not improve by using all data that is available. It rather means to learn from experience (i.e., past sales volumes), to pay attention to business specifics and future rule-based and ad-hoc phenomena and to combine it with smart feature engineering and selection.

It is equally important to analyze the portfolio's characteristics and make the model(s) react to them in the best possible way – globally or cluster wise. Accordingly, Chapter 4.2 represents the results of a profound EDA and highlights realities from a statistical as well as from a business point of view.

4.1 Concept

The concept points back to the result targets listed in Chapter 1.2 and seeks the implementation of a multivariate, multi-step ML model to provide monthly forecasts for all active PMCs for a specified horizon. In order to make the model behavior-based, lead factors from which demand i.e., consumption behaviors can be extracted, interpreted and predicted need to be identified and in a second step, reduced to the most relevant ones without losing information. Appropriately, Chen and Guestrin (2016) state, that "there are two important factors that drive these successful applications: [first of all, the] usage of effective models that capture the complex data dependencies and [secondly] scalable learning systems […]."

With regards to the former, a one-fits-all approach i.e., the implementation of a purely global model is not recommended. (Hu et. al.: 2018) At the same time, Syntetos and Boylan (2021: 250) argue that complex classification schemes can indeed offer greater flexibility but at the cost of a more challenging and expensive implementation. As forecasting is still required on a non-global level, the approach will follow Hu et. al.'s (2018) recommendation to cluster according to a multi-criteria approach.

4.2 Data Management and Analytics

Data Management incorporates all activities for transforming data into decision-relevant information for the forecasting process. It comprises the model-independent sub-steps identification, acquisition and storage, pre-processing as well as loading. *Explorative Data Analytics* (EDA) then refers to the subsequent application of algorithms and software to acquire intelligence from the previously gathered data. (Kwon et. al. 2014)

In Table 4.4, the scope of the research subject is illustrated from a data point of view and compares the extent to which it represents the overall portfolio. The sample's number of time series exceeds 70,000 comprising thirteen BUs and represents demands from three different regional markets and four customer segments. There are no restrictions for part types or product families.

Within preprocessing, a distinction must be made between model-agnostic or model-specific adaptions. For model-agnostic ones, another distinction is made into error-based and case-based corrections[25]. The former primarily focuses on missing value identification and imputation for dates and feature columns. Besides this, there may be *outliers* both in the sales history of the respective products and in the data representing the factors affecting demand. According to Osborne and Overbay (2004) an outlier is an observation that lies "outside the norm" from other values in a random sample from a population. In a sense, this definition leaves it up to the business respectively the analyst to decide what is to be considered *outside the norm*.

Outliers can arise from several different mechanisms or causes. As Table 4.1 illustrates, these are either due to errors in the data (Osborne and Overbay 2004) or business induced.

Table 4.1: Causes for Outliers in Time Series Data

TECHNICAL OUTLIERS	OUTLIERS IN DEMAND HISTORY
• error during manual input • transfer-related errors	• one-off business • second source • initial coverage

Depending on the cause, it is worthwhile to use different methods for outlier correction. For technical outliers, a correction via inter-quantile ranges (IQR), for business-related outliers Median Absolute Deviation (MAD) is applied.

[25] Model-agnostic case-based corrections comprise the following:

- aggregation across four dimensions: time, product, customer, market
- normalization of the dependent variable y_t and covariates c_t via *min-max* where $y_{scaled} = \frac{y-y_{min}}{y_{max}-y_{min}}$ and $c_{scaled} = \frac{c-c_{min}}{c_{max}-c_{min}}$. By doing so, all data values will be transformed into the range [0, 1] meaning that the minimum and maximum value of a variable or feature is going to be 0 and 1, respectively.
- encoding of categorical features
- feature engineering and selection
- categorization of forecastable PMCs: Definition of the number of leading and trailing zeros and respective thresholds to identify cold-start products and PMCs approaching phase-out.

In general, MAD is a common and robust way to detect outlying values in univariate time series, comprising two properties that have made many statisticians like Huber (1981) describe it as the "single most useful ancillary estimate of scale". These properties are the following:

The median $\tilde{y}$ is, like the mean, a measure of central tendency but offers the advantage of being very insensitive to the presence of outliers. One indicator of this insensitivity is the so-called breakdown point (Donoho and Huber: 1983). The estimator 's breakdown point is the maximum proportion of observations that can be *contaminated* (i.e., set to infinity) without forcing the estimator to result in a false value (infinite or null in the case of an estimator of scale). For example, when a single observation has an infinite value, the mean of all observations becomes infinite; hence the mean's breakdown point is 0. By contrast, the median value remains unchanged. The median becomes absurd only when more than 50% of the observations are infinite. With a breakdown point of 0.5, the median is the location estimator that has the highest breakdown point. The same can be said about the MAD as an estimator of scale. Moreover, the MAD is totally immune to the sample size.
For defining if y_t is anomalous, the following applies: $\frac{|y_t-\tilde{y}|}{\tilde{d}} > threshold$[26] where

$$\tilde{d} = \text{median}(|y_1 - \tilde{y}|, |y_2 - \tilde{y}|, \ldots, |y_T - \tilde{y}|) \tag{41}$$

Applying MAD for AA-specific time series, a challenge arises due to their inherent characteristics, which are described more precisely in the upcoming section *demand profile identification and time series classification.* It is especially the demand's sporadicity and volatility, which lead to majorly identifying values as outliers, that are indeed different from other data points in the time series, but not considered outliers from a business perspective (column 1 in Figure 4.1)

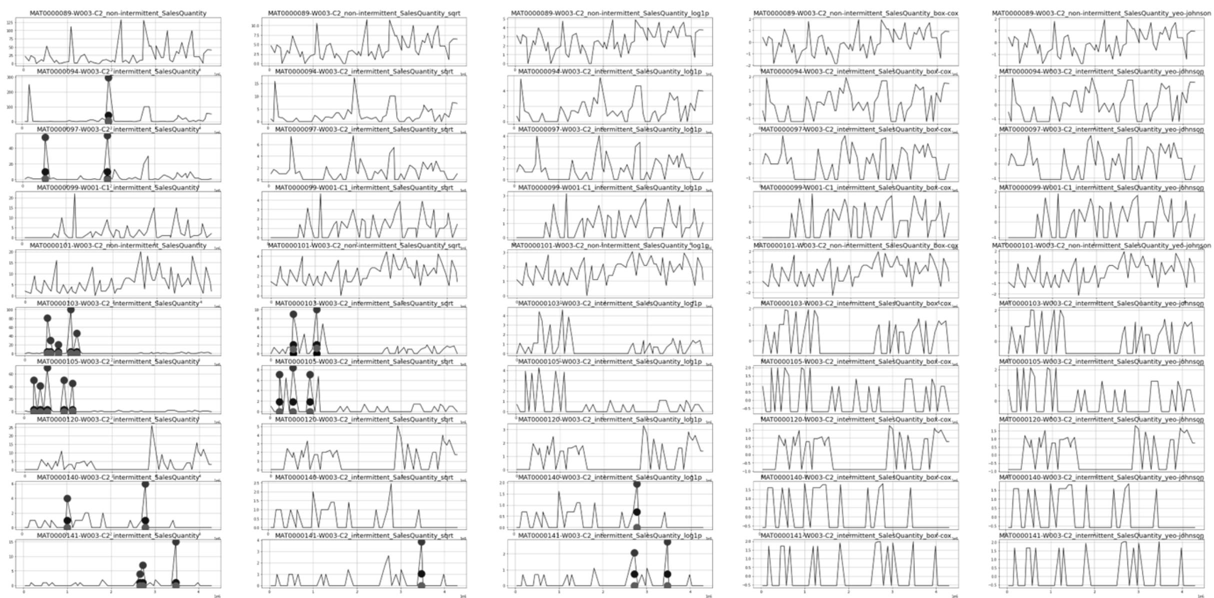


Figure 4.1 MAD Outlier Detection in Raw and Transformed Format (Example)

[26] The threshold is an analyst-defined cut-off value. For the use case it is defined as follows:
IQR = time series quantile(.90)-time series quantile(.10)
cut-off coefficient = 2.

To counteract this phenomenon, different non-linear transformations, amongst them square root (cf. column 2 in Figure 4.1), log (cf. column 3 in Figure 4.1), Box-Cox (cf. column 4 in Figure 4.1) and Yeo-Johnson (cf. column 5 in Figure 4.1) have been implemented, finally resulting in an outlier detection identifying only values which are significantly off from the main pattern. Upper bubbles label outliers per PMC and transformation, centered and lower markers highlight corrected volumes based on the 0.9- respectively the 0.5-quantile. To achieve a sound level of tolerance and to limit the computational effort for transformation and back-transformation, an application of square root and log is preferable in comparison to Box-Cox and Yeo-Johnson. Another downside of square root though is that back-transformation induces bias and by definition underestimates the transformed label. Hence, a correction is needed. Also, to what extent transformation improves accuracy is to be evaluated separately in the forecasting procedure.

Model-specific preprocessing respectively EDA always aims at verifying and potentially fulfilling statistical, numerical and algorithmic concerns in order to validly apply suitable models. This comprises testing for stationarity of time series, autocorrelation and autoregression, demand profile identification, pareto analysis and value contribution, as well as distributional fitting. In the following, the most important results from these analyses are presented.

Stationarity and Non-Stationarity of Time Series

Time series stationarity refers to a time series for which statistical properties do not vary over time. This is not to be confused with a more general sense of 'change', as many stationary time series do in fact exhibit some level of visual change. Instead, it is the way in which the series itself changes that does not change over time. To be more precise: the mean, variance, and autocorrelation structure of a given time series must remain constant.

The Augmented Dickey-Fuller test (ADF) is arguably the most popular test for identifying stationarity. It works by attempting to test for the presence of a *unit root*[27], resulting in the following hypotheses:

- $H0$: the time series has a *unit root*, hence is non-stationary
- $H1$: time series is stationary

For AA, the ADF reveals that 71% of all COM-related time series and 83% of all TP-related time series are stationary. This is incredibly important in the realm of time series statistics as many statistical tests and processes only work with stationary data. Thus, any non-stationary data must first be transformed to be stationary via a process known as differencing.

Decomposition of Time Series to identify Seasonal and Trend Indices

On the other hand, stationary time-series do rarely have predictable patterns, amongst them trend-cycle components or seasonality.

[27] A unit root is a feature of a stochastic process. Removing the unit root is achieved by differencing and renders the time series stationary. If a time series has x unit-roots, it must be differenced x times until the series is stationary.

For most AA time series, neither indices for seasonality nor trend can be derived on item level. Hence, the decomposition is again applied on different product aggregation levels, with PG3 resulting in higher but still no significant shares of time series revealing trend and seasonality. Only on PG2 level, stable values are to be derived. Those are eventually related to the item from the respective PG2. Doing so, 42.03% of the overall time series reveal seasonal patterns on the aggregated level. The largest portion of non-seasonal time series can be assigned to the TP type. The highest share of time series with seasonal patterns is related to two BU units from the COM sector with shares of 60%. Their seasonality results from both weather-related circumstances and demonstrably from business induced marketing activities.

Table 4.2: Share of Items revealing Seasonality on PG2-level across Countries

	BU	**IS_SEASONAL**	
		0_PROPERTY [%]	**1_PROPERTY [%]**
1.	TP_1	65.91	34.08
2.	TP_2	60.50	39.50
3.	TP_3	52.12	47.87
4.	TP_4	69.66	30.33
5.	TP_5	63.57	36.42
6.	TP_6	50.64	49.35
7.	TP_7	85.34	14.65
8.	TP_8	65.97	34.02
9.	COM_1	50.34	49.65
10.	COM_2	31.33	**68.66**
11.	COM_3	46.15	53.84
12.	COM_4	40.71	59.28
13.	COM_5	38.21	*61.78*
	All	57.96	42.03

Autocorrelation and Autoregression

Autocorrelation (ACF) respectively *serial correlation* means 'correlated with itself', i.e., different observations of a specified variable y_{i} are correlated with each other. For such a pattern to be interpretable, the order of observations needs to obey a logical order, as it is the case with time series, for example. In the simplest case, each value of y_{it} is correlated with the value of the previous period. Of course, values that are further apart can also be correlated with each other. An ACF analysis can give an answer, of whether the observed time series is to be modeled with an MA-model, and if yes, what the order is.

An *autoregressive* (AR) process is a stochastic process used in statistical calculations in which future values $y_{\mathrm{i,t+1}}$are estimated based on a weighted sum of past values, the so-called *lags*. It operates under the premise that past values have an effect on current values. A process considered AR(1) is the first order process, meaning that the current value y_{it} is based on the immediately preceding value $y_{\mathrm{it-1}}$. An AR(2) process has the current value y_{it} based on the previous two values $y_{\mathrm{it-1}}$ and $y_{\mathrm{it-2}}$. The order of an AR-model is normally derived from the partial autocorrelation (PACF) analysis.

Two assumptions are made prior to autoregression modeling:

- The preceding time steps are useful in predicting the value at the next time step. Hence, one may assume a certain dependance between values.
- The time series is stationary.

From both analyses applied to AA time series, one can make the following observations:

- There are hardly any autocorrelations that are significantly non-zero if the analysis is applied on PMC (SKU)-level. This is expected as for a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly. Therefore, the majority of time series is considered random.

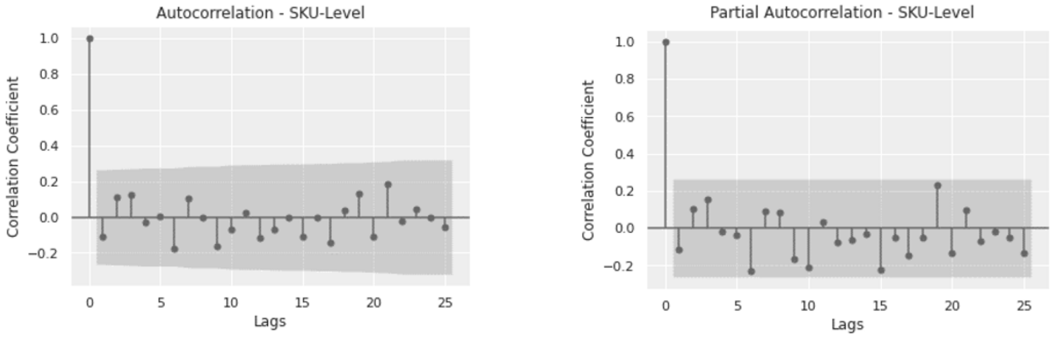


Figure 4.2 ACF and PACF on SKU-Level (Example)

- There are several autocorrelations that are significantly non-zero if the analysis is applied on PCM (PG2)-level. Therefore, the time series is non-random.
- There is a high degree of autocorrelation between adjacent (lag 1) and near-adjacent (lag 2 and lag 3) observations in ACF plots if the analysis is applied on PG2-level. A strong correlation can also be seen at a lag of 12.

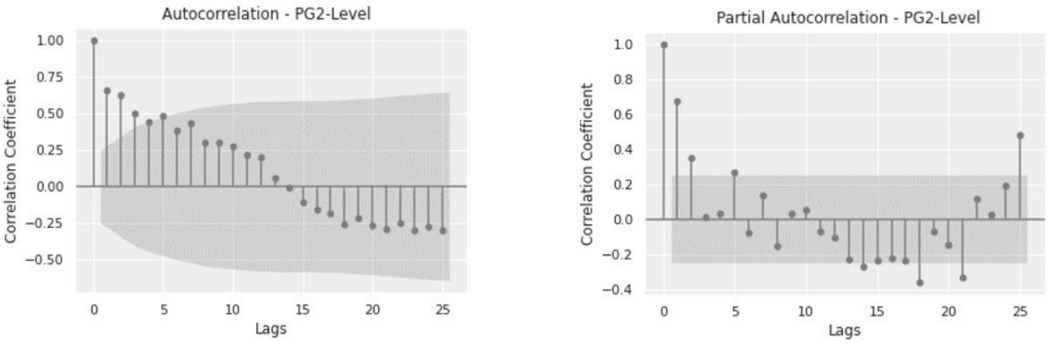


Figure 4.3 ACF and PACF on PG2-Level (Example)

- To test results for significance, a two-step verification is applied in setting the threshold for the $correlation\ coefficient > 0.5$ and by depicting a 95%-confidence interval: That means, anything within the grey area is statistically close to zero and anything outside the grey area is statistically non-zero. This results in the following plots.

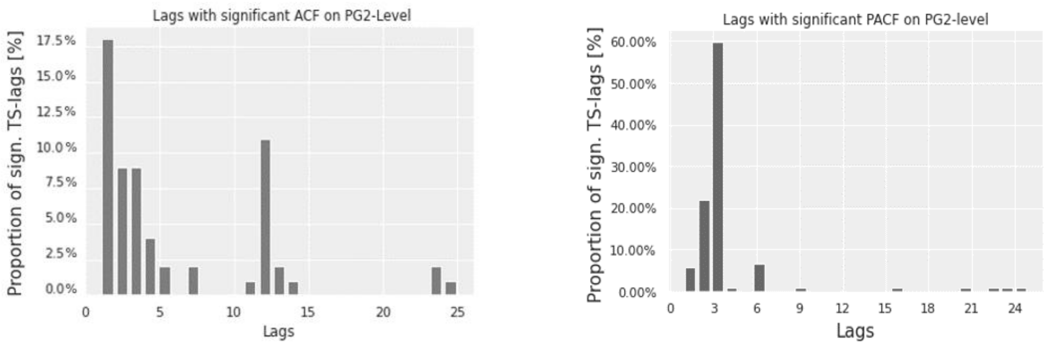


Figure 4.5 Lags with Significant Correlations in AA Time Series (Example)

- Modeling random time series (*white noise*) on SKU-level is difficult because no parameters from the ACF and PACF plots can be derived. Fitting autoregressive models results in models of many different orders. A distribution of these can be derived from the figure below.

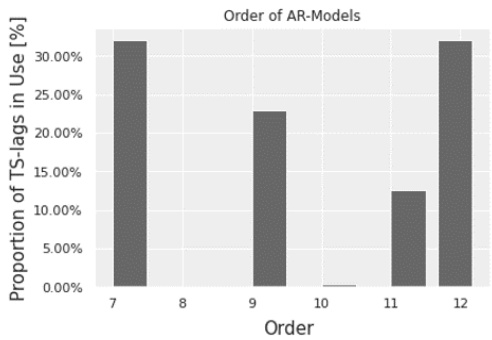


Figure 4.4 Order of AR-Models

Demand Profiles and Time Series Classification

There has been a considerable amount of research on classifying and analyzing forecasting methods (Hu et. al. 2018, Januschowski et. al. 2019), but not on distinguishing between different demand patterns to guide forecasting and inventory control. Notable exceptions are the papers by Gelders and Van Looy (1978), Williams (1984), Eaves and Kingsman (2004), Syntetos et. al. (2005), Syntetos and Boylan (2006) as well as Boylan et. al. (2008). The latter build on the ideas from their predecessors and propose to use the following classifiers: the mean inter-demand interval (ADI) to estimate the variability in demand timing with

$$ADI_i = \frac{\text{number of total demand periods of PMC}_\text{i}}{\text{number of non−zero demand periods of PMC}_\text{i}} \qquad (42)$$

[time periods]

and the squared coefficient of variance (var_i^2) to determine the variation of demand sizes.

Reasons to do so are found in the different PMCs i which are associated with different underlying demand structures, which in turn require majorly different methods for forecasting. The way this task is performed has significant implications in terms of stock control and customer satisfaction (CSLs). Therefore, categorization rules constitute a vital element of intelligent inventory management systems. Applying KIPs 1 (42) and 2 results in the categorization scheme illustrated in Figure 4.6. Time series assigned to the left side of the matrix represent demands with a very homogeneous demand quantity, whereas lumpy and erratic time series are defined by high volatilities. Products, designated as intermittent or lumpy, share the common characteristic that they are sold very infrequently, i.e., sporadically, whilst smooth and erratic ones do not or hardly comprise so-called zero-demands.

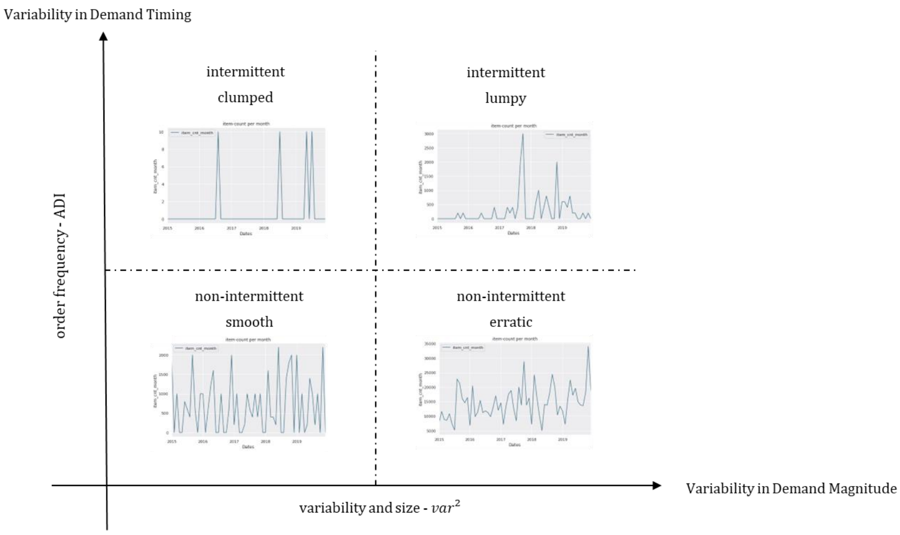


Figure 4.6 Demand Profiles based on Demand Timing and Magnitude

The scheme meets various theoretical and practical requirements, though, the breakpoints need to be chosen empirically i.e., case-study based, so that they make sense for the situation that is analyzed. (Eaves and Kingsman 2004) For AA, the thresholds were defined across all BUs with ADI at 1.33[28] and var^2 at 0.36[29]. In addition to these threshold values, the length of the time series is also an influencing variable that significantly affects the categorization. In order to confirm the clear dominance of the intermittent-lumpy demand profile (Table 4.4), which results from the analysis being applied on the overall time series length, the analysis is repeated at annual level and transition probabilities are derived (Figure 4.7).

[28] The calculation of the cut-off value for ADI is based on business internal assumptions for non-intermittency: for a maximum of one fourth of the total number of bins, demands may be equal to zero: This results in demand quantities exceeding zero in 75% of the total number of bins. An example calculation: total number of bins = 12 (months); bins with zero demand = 3; bins where demand exceeds zero: 12-3 = 9; threshold derived from this calculation: 12/9 = 1.333

[29] The calculation of the cut-off value for var^2 is based on business internal RIS classifications with the subsequent threshold values: R (regular): $var \leq 0.6$; I (irregular): $var > 0.6$.

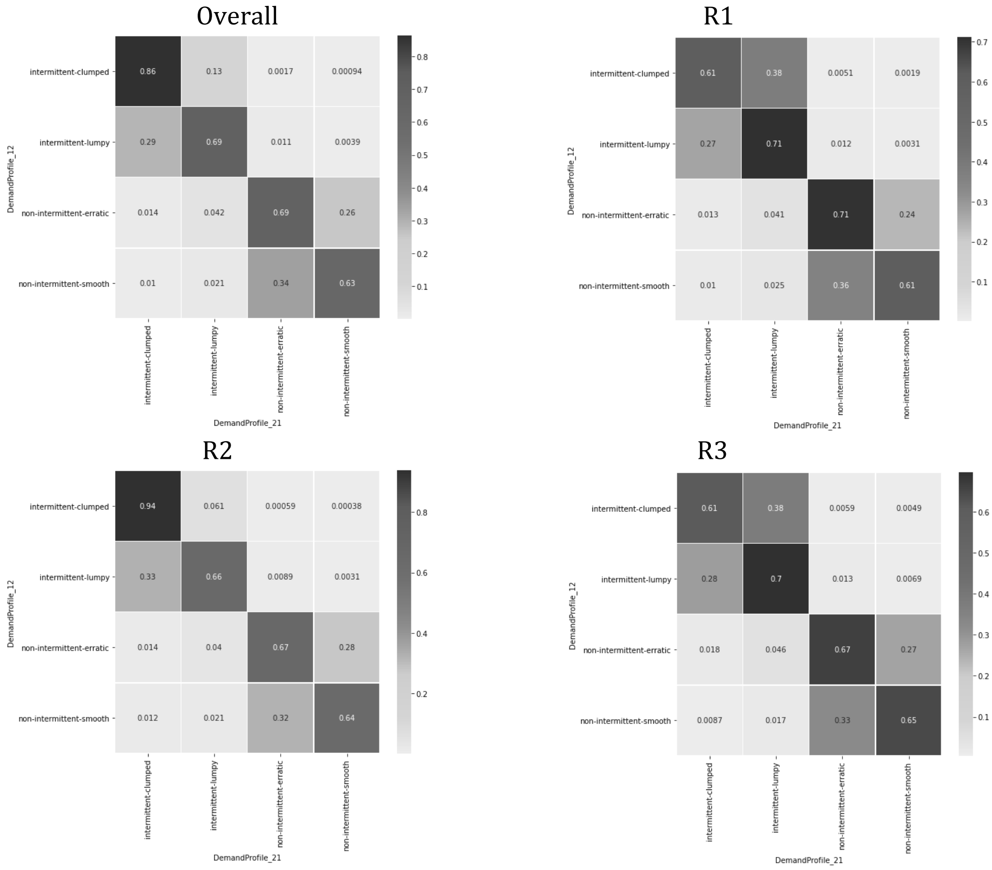


Figure 4.7 Markovian Transition Matrix for Demand Profiles

Pareto Analysis and Value Contribution of Time Series Clusters

In order to understand the relation between demand profile assignments and high-value products, a pareto analysis based on the single SKU's turnover is performed. As one can see in Table 4.3 and in Table 4.4, the dominant type of series is intermittent-lumpy with a share of approximately 80% when determined with monthly bins, while non-intermittent erratic ones are the runner-up with almost 15%. Non-intermittent smooth series are represented by approximately 5%. Furthermore, it becomes evident that the most valuable and simultaneously best-to-predict products can be ascribed to non-intermittent erratic (i.e., heterogeneous order size but homogenous order frequency) as well as to non-intermittent smooth (i.e., homogenous order size and order frequency) time series on a monthly level, amounting to approximately 80% of the turnover.

Table 4.3: Value Contribution and Demand Profiles per Region

	R1			R2			R3		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
	6.65	13.13	80.22	7.47	12.98	79.55	9.69	11.67	78.64
intermittent-lumpy	2.23	6.92	90.85	2.35	7.55	90.10	3.62	7.49	88.89
non-intermittent-erratic	18.63	37.79	43.58	28.60	39.28	32.12	24.81	22.65	52.54
non-intermittent-smooth	47.72	43.59	8.69	44.85	43.09	12.06	36.51	28.48	35.01

Table 4.4: Characteristics of the Sample Data Set

SCOPE	REGIONS	SUB-REGIONS	CHANNEL	SEGMENT	PRODUCT TYPE	BU	PG1	PG2	PG3	SKU	VALUE CONTRIB. [%]	TURN RATE [%]	RIS [%]	DEMAND PROFILE [%]
OVERALL	7	9	OES IAM	OEM Wholesale Retail Others Workshop	TP COM	15	18	133	451	>550k	threshold [80-95-100]		threshold [0.6-1.2-∞]	
SAMPLE	1	3	IAM	Wholesale Retail Others Workshop	TP COM	13	18	114	371	71622	A: 7.60 B: 12.74 C: 79.64			I-L: 80.27 NI-E: 14.72 NI-S: 5.00
R1	1	1	IAM	Wholesale Retail Others Workshop	TP COM	13	18	102	304	66438	A: 6.65 B: 13.13 C: 80.22	FM: 40 SM: 60	R: 3.99 I: 16.74 S: 79.27	I-L: 80.71 NI-E: 14.97 NI-S: 4.32
R2	1	1	IAM	Wholesale/ Retail Others Workshop	TP COM	13	18	100	322	38321	A: 7.47 B: 12.98 C: 79.55	FM: 33 SM: 67	R: 4.44 I: 13.79 S: 81.77	I-L: 83.45 NI-E: 11.79 NI-S: 4.76
R3	1	1	IAM	Wholesale/ Retail Others Workshop	TP COM	13	18	97	286	27757	A: 9.69 B: 11.67 C: 78.64	FM: 46 SM: 54	R: 6.33 I: 20.48 S: 73.19	I-L: 75.09 NI-E: 18.25 NI-S: 6.66

4.3 Algorithms

In line with the criteria mentioned in Chapter 1.2 and the results from the literature review presented in Chapter 2, the following four models – three neural networks and one decision tree – are selected:

BI-DC-LSTM:

- The deeply connected LSTM remembers certain long-term dependencies as well as more current experiences.
- The bidirectional layer is used as a wrapper layer consisting of a forward and backward layer and serves to simultaneously seize information from the past and the future for one specified data point.

CNN-LSTM: It is helpful to think of this architecture as a hybrid model defining two-sub-models.

- The CNN model is used as an encoder to interpret subsequences for feature extraction and determining the relationship between different observations, being then provided as time steps to an LSTM
- The LSTM model serves as the backend for interpreting the features across time. Hence, it supports sequence prediction and prevents the model from treating observations independently.
- A time-distributed layer serves as a wrapper layer. It is used around the output layer so that one value per time step can be predicted given the full sequence provided as input

LSTM-GRN:

- The LSTM-GRN (TFT) uses recurrent layers i.e., LSTMs for local processing and interpretable self-attention layers for long-term dependencies.
- In addition, it utilizes specialized components to select relevant features and a series of gating layers to suppress unnecessary components.
- The model is also capable of probabilistic forecasts by using quantile regression. However, this capability is disregarded in this dissertation.

GBR-XGB:

- GBR-XGB uses a pre-sorted and histogram-based algorithm for computing the best split. It works as follows: For each node, all input variables and features are enumerated over. For every feature, instances are sorted by the feature value. Using a linear scan, the split is made along with the feature basis information gain. The best-split solution is picked along with all the features.
- Another advantage is the application of the so-called Tweedie loss, which is particularly suitable for intermittent time series due to its characterization (cf. Chapter 4.2). Setting it as the objective function, it basically forces the model to maximize the likelihood of the distribution and to thus approximate the correct number of zeros for its predictions.
- Besides performance, one of the most important reasons for its usage is its scalability. According to Chen and Guestrin (2016), it runs more than ten times

faster than other existing ML-based solutions on a single machine with LightGBM being even faster than GBR-XGB.

4.4 Conclusion

The vastness und diversity seen from the automotive aftermarket's sales channels, market participants and products (cf. Chapter 3) is reflected in the time series. Standard analyses from statistics hardly lead to reliable results:

- Outlier correction is hampered by the sporadic nature of the time series. A route-causes analysis thereto is human-dependent and not to be verified via data.
- A decomposition of the time series only leads to an identification of seasonality and trend on aggregated level.
- Autocorrelation and autoregression is likewise only to be detected at PG2 level.
- This is contrasted by the classification according to demand profiles and value contribution: A clear majority assignment is made to the lumpy cluster and C-parts.
- The lumpiness characteristic also fits to the result of distributional fitting: Tweedie, characterized by a high point mass being concentrated at zero, may represent the majority of aftermarket demand best (cf. Appendix 4).

Consequently, the requirements to the data input, especially the covariates, but also to the ML-based model are immense: it needs to identify and learn from local context as well as from meta data and it needs to detect and seize short-term as well as long-term dependencies. The comparison of the prioritized methods based on the analysis of existing literature with the above requirements limits the scope of the methods for the present dissertation to three NNs and one boosting method.

5 Factors of Variance in a Behavior-based ML Model for Demand Planning

In the context of the thesis, various methods are used to determine and prioritize the different factors influencing customer demand.

Stage one, a descriptive-empirical investigation, is based

- on a comprehensive review of journals, books, and studies from general areas but also from the very specific automotive respectively spare parts environment, as well as
- on perceptions and insights gained through expert interviews on customer-specific order behavior, product-specific trends, and their root-causes.

Herein, the goal was not to primarily collect quantitative data, but to create understanding and in-depth knowledge about how the market works. (Mack et. al. 2005)

Stage two, a quantitative-empirical investigation helps to compare the human-aided against a statistical selection of indicators, and to finally select appropriate lead factors from the complete set of potential ones.

5.1 Identification and Categorization

The literature analysis is of importance for obtaining a complete framework of potential indicators, and also for designing the guideline and questions for the subsequent interviews. In this way, relevance and focus are guaranteed and the frequently expressed criticism that results from personal discussions are exclusively based on subjective experience and insider knowledge is counteracted. (Kromrey 2002: 353) Thus, one could also say, that in combination with the interview of various focus groups and experts, it provides the descriptive-empirical foundation of the overall study.

The questionnaires, the focus groups were to answer, are not a static construct. They are rather defined in an iterative and interactive, i.e., deductive process, as new aspects that arise during an interview can be relevant for follow-up surveys as well as for previous discussions. (Friedrichs 1990: 22) The content is neither product-dependent nor related to the area of responsibility of the focus groups, and is structured as follows:

- Personal information
- Question block 1: General Questions
- Question block 2: Indicator Categories Influencing Demand
- Question block 3: Single Indicators Influencing Demand
- Optional Section: Time Series Analysis[30]

The personal information section will be used to demonstrate compliance with the criteria listed in the following, describing the participants' expert status. Furthermore,

[30] The optional section 'Time Series Analysis', which illustrates past sales developments on different product aggregation levels via time series, is asking for specific interpretation of patterns, anomalies or conspicuous features.

they enable a differentiated evaluation on the basis of regional- or even focus group-based affiliations.

It is regarded important

- that all interviewees can give answers within in their own reference system,
- that results are dependent on attitude differences rather than on verbalization skills, and
- that at least a certain level of comparability is achieved.

Hence, section one provides open, sections two and three rank-scaled closed questions, with responses being rated on six different levels. (Porst 2014: 69ff)

In choosing focus groups and experts, it was first necessary to clarify which persons qualify as specialists and can thus represent the entire research area. This status was ascribed equally with regards to methodological-relational and social-representative aspects (Bogner et. al. 2014: 10f). On the one hand, persons were identified who could be assumed to possess relevant knowledge related to the research area (cf. Table 5.1). On the other hand, persons, who are made experts socially and/or within the enterprise and have task-related, precise but also contextual knowledge, were regarded highly relevant. (Schütz and Luckmann 2017: 358ff, Bogner et. al. 2014: 9)

Table 5.1: Expert Status Definition for Focus Groups

FOCUS GROUP	FOCUS	CRITERIA
Key Account Managers / Customer Relationship Managers	customer	customer proximity or strong customer-bond with insights regarding their behavior sales-related targets: • increase of sales • high service levels
Sales Logistic Interface Managers / Customer Supply Chain Service Managers / Logistics Planning Managers	customer with logistics focus	logistics-related targets: • high service levels • efficient inventory management
Product Specialists / Business Unit (BU) Managers	product	proven know-how in a specific product segment sales-related targets: • ensuring competitiveness • strengthening of market positions

Moreover, focus group and expert rounds were to be arranged in such a way that reference points

- to different medium- and large-sized companies from the automotive sector being active in the market as manufacturers and/or suppliers on a global level, i.e., comprising different sales regions and markets,
- to the top 20% of high-revenue customers as well as
- to products from both the COM and the TP sector

were enabled.

5.1.1 Lead Factors from Scientific Literature

Historical demand data is normally used to establish a baseline for forecasting future demand. It consists of the three variables

- sales quantity shipped and invoiced, the so-called history orders,
- lost sales as well as
- back orders.

However, its result, the actual historical demand is not always the best indicator of final demand. In other words, only considering what happened in the previous years and neglecting everything that is taking place in the current market, a business‘ demand forecast will be incomplete.

As already stated in Chapter 2.1.2, one can generally differentiate between *movement* and *shift* factors. The price of the focal product is within the first group. Amongst the second, there are six more factors from the macro- and microeconomic domain, that may result in a shift of the demand curve: future developments and expectations, the number and income of customers, the customers‘ preferences and tastes as well as the price of substitute products and the price of complementary products. (Lynham 2018: 128ff.)

Table 5.2: Influencing Factors on Automotive Spare Parts Demand

B2C	B2B – AUTOMOTIVE SPARE PARTS
price of the focal product	pricing of and promotions for the focal product
future developments and expectations	legislation (Crone 2010, Hecker et. al. 2012)
	technological advancements and trends (Crone 2010, Hecker et. al. 2017)
	ecology (Klug 2010)
number of customers	market volume – number of primary products and failure rates (Voss 2006; Pfohl 2009; Klug 2010; Hecker et. al. 2012/2017)
	number of channels (Hecker et. al. 2017)
	number of customers (Klug 2010; Hecker et. al. 2012)
income of customers	size of first level customer (Hecker et. al. 2012)
	purchasing and negotiating power of first level customer (Hecker et. al. 2012)
preferences and tastes of customers	intensity of use and maintenance strategy of second level customers (Loukmidis and Luczak 2006; Pfohl 2009/2018; Klug 2010)
	purchasing strategies, order and inventory policies of first level customers (Pfohl 2009/2018; Klug 2010)
	quality and brand (Hecker et. al. 2012)
price of substitute products	number of competitors (Hecker et. al. 2012)
	number of alternative goods and their prices
price of complementary products	number of complementary goods and their prices
	artificial linkage via target agreements

The works of Pfohl (2009/2018), Klug (2010), Hecker et. al. (2012/2017) and the VDA (2014) are suitable for transferring these factors from B2C to B2B and even more specifically to the spare parts use case (cf. Table 5.2).

In his chapter ‘Ersatzteillogistik’, Pfohl (2009: 241) lists a comprehensive number of factors influencing the demand for spare parts and hence also its prognosis: Besides the number of primary products in use, the need for spare parts and its forecast quality is also influenced by

- planned future sales of primary products,
- the intensity of use and operating conditions,
- the maintenance strategy of the user, as well as
- information on product life cycles, failure curves and wear and tear.

In his later edition, ‘inventory policies of first level customers’ is cited as another decisive factor. (Pfohl 2018: 241) Klug (2010: 497 ff.) relates Pfohl's depiction to the automotive aftermarket and adds the item ‘market and environmental conditions’. In contrast to the already cited works, Hecker et. al. (2012/2017) elaborate on how the automotive aftermarket progresses in response to different influences from the economic, political and technological environment and how participants – competitors, suppliers and customers – interact.

It should be noted that all of these publications are exclusively qualitative in nature and are solely based on continuously updated market analysis. A first exception is the framework from the VDA, entitled ‘Aufbau eines Frühindikator-Systems für das Automobil-Ersatzteile-Geschäft im Independent Aftermarket als Grundlage der Mittelfristprognose.’ The aim was to develop a theoretical indicator model by means of literature research and expert interviews, which was then to be evaluated quantitatively by applying correlation and regression methods across different lags. The result is a ten factor[31] comprising framework which is to be used for forecasting on category level. A detailed description of the approach and its results can be found in Figure 5.1.

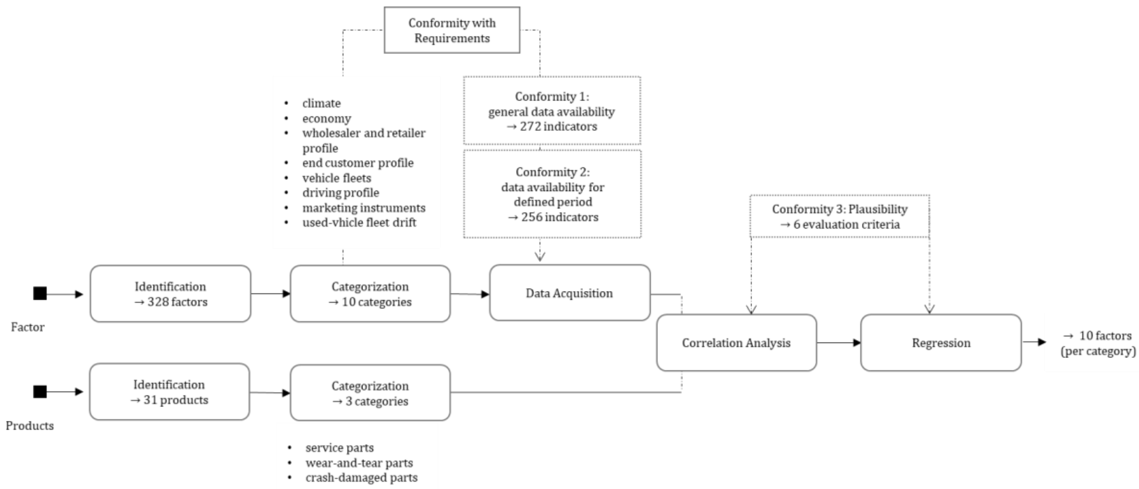


Figure 5.1 VDA Approach to Identify Lead Factors for Aftermarket Demand

[31] The factors are prioritized according to their explanatory power (R^2): (1) domestic sales [€] of vehicles and engines; (2) total sales volume [€] of the automotive industry; (3) production volume of motor vehicles and engines with 1.0 - 1.5 liters cubic capacity in Germany; (4) production volume of motor vehicles and engines with 1.5 - 2.0 liters cubic capacity in Germany; (5) production volume of motor vehicles and engines with 2.0 - 3.0 liters cubic capacity in Germany; (6) new registrations of passenger cars in Germany; (7) wages and salaries in mining and manufacturing in Germany; (8) short-time work in Germany; (9) fuel types of vehicles in use in Germany; (10) accident statistics and fatalities due to car accidents in Germany

Positive aspects of the methodology are the clearly defined requirements for the potential factors, the data collection and the data engineering. However, it is equally critical that the correlation and causality analysis is conducted solely with highly aggregated product data i.e., sales figures of several companies are accumulated across a broad range of product hierarchies. Another limitation, resulting in the fact that the present findings can only be used as a starting point for this study, is the short observation period of six months and the regional filter.

Scientific works dealing with the strategic integration of factors into mathematical models for improved demand forecasting are very scarce. The majority is not based on a holistic identification and analysis of the aforementioned. Rather, single, quantitatively available factors are integrated without further justification for their specific selection, though their integration is traceably useful as demonstrated in Trapero et. al. (2012), Ma et. al. (2015), Nemati Amirkolaii et. al. (2017), and Sagaert et. al. (2017).

The majority of the publications can be assigned to the domain *retail*, the majority of the factors to marketing instruments. Examples are Nijs et. al. (2000), Caniato et. al. (2005), Aburto and Weber (2007) as well as Kilimci et. al. (2019), who propose to exploit information on promotional activities to estimate both the timing and the size of demand fluctuations. The same focus is set by Ailawadi et. al. (2006) who analyze artificially inflated demand caused through campaigns. Lang et. al. (2015) apply a hierarchical Bayesian semi-parametric approach to seize pricing information for predicting sales and identifying the best pricing for products in the retail sector. Like Ma et. al. (2015) they integrate both item price effects and inter-category promotional information. Occasionally, factors from more than one domain are integrated. Huang et. al (2014) apply stepwise selection via LASSO to improve demand forecasting for retail operations. Their set of factors comprise data on pricing and promotions as well as competitive information. Ali et. al. (2009) successfully seizes promotional data next to information about the products' life cycle phases. This idea is comforted by Syntetos and Boylan (2021: 269): They state that "[o]bsolescence is a natural issue to consider in a spare parts context […]. […] After its launch to the market, a product will typically experience increasing sales (growth), which will then stabilize during the maturity phase, before falling off in the final phase (decline) of its life cycle. […] The final phase is extremely important from an inventory perspective, given that enough needs to be stocked to satisfy future customer demand but not too much because, at some (unknown) future point in time, sales will be discontinued, and any leftover quantities can only be discarded." Despite the importance of this issue, they continue to explain that "obsolescence has not been appropriately addressed when developing forecasting methods for spare parts demand." More ideas for customer-related factors can be derived from their chapter 'Demand generation'. (2021: 254-269) Herein they emphasize the value of considering the number and heterogeneity of customers, as well as the variety and frequency of their requests. Crone (2010) concludes in his dissertation that in addition to microeconomic, internal company data, it is also worthwhile to search for external sources from the political, socio-cultural and technological environment. Especially in the case of a long-term forecast horizon, those may contribute additional benefit. Bai and Ng (2008) as well as Forge (2009) confirm this assumption by seizing macroeconomic respectively micro-, meso- and macroeconomic factors. Sagaert et. al (2018) also uses the latter but identifies two challenges. The first is limited data availability in general. The second is its granularity which complicates its use in industry and means that any model has to be reformulated for each bin, using only appropriate lagged realizations of the variables.

Table 5.3: Literature Review on Seizing Lead Factors for Demand Forecasting

	AUTHOR	TITLE	YoP	CONTEXT / DOMAIN	METHOD / MODEL	INPUT VARIABLE(S), SAMPLE SIZE, LEVEL and FORECAST HORIZON	METRICS and RESULTS
1.	Nijs et. al.	Category Demand Effects of Price Promotions	2000	Retail – Supermarket Supplies	auto-regressive models	Input: sales history marketing mix variables Scope: 560 SKUs (fast movers) 208 weeks	Results: Promotional intensity, advertising intensity, competitive structure and covariates affect effectiveness of pricing information to actually determine its importance. • bias is caused through aggregation • frequent promotions have effects on short-run consumer sensitivity, but no effect in the long run • the more oligopolistic the market, the stronger the promotional effect • promotional effects are higher for perishable products and lower for products with frequent new-product introductions
3.	Caniato et. al.	Clustering customers to forecast demand	2005	Retail – Groceries	clustering linear regression	Input: sales history incl. seasonal indices variables related to managerial variability (i.e., promotions)	Result: • investigation of the relationship between sources of demand variability and the possibility to adopt a clustering approach to forecasting • the higher the incidence of systematic and managerial variability, the more the clustering approach is suited for forecasting • the higher the incidence of random variability, the less the approach can be adopted
4.	Ali et. al.	SKU demand forecasting in the presence of promotions	2009	Retail - Groceries	ETS SVR Regression Tree	Input: sales history pricing information Scope: 35 products from four stores 71 weeks with 51 weeks for training	Metrics: MAE, MAPE Results: advanced multivariate methods (regression Tree and SVR with RBF or polynomial kernel are best for forecasting promotional sales but at the expense of increased data manipulation and technique complexity
5.	Lee and Cho	Demand forecasting of diesel passenger cars considering consumer preference and government regulation in South Korea	2009	Passenger Cars	qualitative and quantitative methods: scope of the survey: 500 adults	Input: sales history government regulations customer preferences regarding pricing and car attributes Scope: -	Metrics: t-value, p-value Results: forecasts are generated based on the outcomes of the qualitative research and by testing scenarios. strong dependance on the model, the type of powertrain, pricing of cars and fuels, market penetration and share.
6.	Trapero et. al.	Impact of Information Exchange on Supplier Forecasting Performance	2012	Retail	ETS ARIMA	Input: sales history POS data	Result: Information Sharing significantly improves accuracy.
7.	Romeijnders, Teunter and Jaarsveld	A two-step method for forecasting spare parts demand using	2012	Spare Parts – Aviation	ZF, Naïve, MA, ETS	Input: sales history	Metrics: MAE, MSE, MAD Results: two-step multivariate component level forecast and ETS are the best performing algorithms

		information on component repairs			Croston, SBA, TSB 2S	repair history Scope: 17,012 SKUs ten years	
8.	Williams et. al.	Predicting retailer orders with POS and order data: The inventory balance effect.	2014	Retail - Consumables	Holt-Winters	Input: - Scope: 110 weeks with 104 weeks for training	Metric: MAPE Result: forecast quality is improved via POS data sharing and integration
9.	Huang et. al.	The value of competitive information in forecasting FMCG retail product sales and the variable selection problem	2014		Stepwise selection and LASSO ETS base-times-lift approach Autoreg. Distributed Lag model	Input: sales history promotional information Scope: 120 SKUs 200 weeks	Metric: MAE, MAPE, sMAPE, MASE Result: • autoregressive distributed lag model with pre-selected promotional variables exceeds accuracy of benchmark models • building diffusion indices via factor analysis is regarded the superior pre-selection method
10.	Lang et. al.	Accommodating heterogeneity and non-linearity in price effects for predicting brand sales and profits	2014	Retail	Bayesian semi-parametric approach	Input: sales history 11 characteristics from the socio demographic domain Scope: 8 brands in 81 stores 89 weeks	Result: • development of additive regression models where own- and cross-price response is estimated flexibly using P-splines • heterogeneity in price response across stores is accommodated by multiplicative store-specific random effects that scale the nonlinear price curves
11.	Ma, Fildes, and Huang	Demand Forecasting with high-dimensional data: the case of SKU retails sales with intra- and inter- category promotional information	2015	Retail – Food and Beverages	Granger Causality three-stage LASSO ETS	Input: Sales history Promotional history of focal product Inter-categorial promotional information Intra-categorial promotional information Scope: best-selling SKUs	LASSO as regularization technique for variable selection advantage: • shrinkage w/o information loss • selection of predictors out of a group of correlated predictors Results: • the extent of promotional interactions among individual SKUs are unstable and dynamic across time periods. • similar purchasing patterns for non-promoted products due to promotions for complements (portfolio importance)
12.	Sagaert, et. al.	Tactical Sales Forecasting using a very large set of macroeconomic indicators	2017	Raw Material - Global Tire Industry	PCA sNaïve, Linear Regression, ETS, Holt-Winters, ARIMA,	Input: sales history macroeconomic variables (consumption, feedstock, financial, housing, labor, import/export, industrial, transport) Scope:	Metric: MAPE Results: • LASSO outperforms other methods • Improvement of FA by 18.8 % • expert opinion on integration of indicators improves FA

					ARI LASSO	six and a half years with five and a half years for training; monthly bins	
13.	Nemati Amirkolaii et. al.	Demand Forecasting for Irregular Demands in Business Aircraft Spare Parts Supply Chains by using Artificial Intelligence	2017	Spare Parts – Aviation	Croston, SBJ, TSB MA SES Statistical method from industrial partner NN with concomitant attribute-based clustering (ABC)	Input: sales history 16 statistical features (min, max, and mean across different periods, sum, STD, CV^2, ADI, demand at the end of the immediately preceding target period, number of periods separating the last two non-zero transactions) price Scope: 30 SKUs from all four demand profile categories (intermittent, erratic, lumpy and smooth) four years with three years for training; monthly bins	Comparison of four scenarios Metrics: positive and negative errors Results: MLP-NNs with higher number of features improve demand forecasts significantly for intermittent demands along with reduction in associated financial implications. • Best results: when all features are used with all parts. Indication of strong correlation with input features not considered in existing forecast methods. • Worst results for the univariate scenario. Results are expected as forecasting accuracy is likely dependent on specific features; hence, when one feature is used for all parts, the worst results emerge.
14.	Kilimci et. al.	An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain	2019	Retail (faster movers)	k-means PCA 9 traditional methods SVR DL Approach	Input: sales and stock history promotional data weather conditions Scope: 1500 SKUs from 5500 stores 106 weeks	Metrics: MAPE, MAD Results: • 20 clusters with different numbers of products for each store • 20 different models per store • best results with ensemble, especially ensemble with Deep Learning and integration of additional features • results from ensemble w/o Deep Learning better than application of single methods
15.	Vairagade, et. al.	Demand Forecasting Using Random Forest and Artificial Neural Network for Supply Chain Management	2019	Retail – Grocery (top 10 SKUs)	Random Forest NN	Input: sales history oil price holidays Scope: 4100 SKUs from 54 stores 56 months history	Metrics: MAE, R^2 Result: slightly better results via Random Forrest compared to NN
16.	Salais-Fierro et. al.	Demand Prediction Using a Soft-Computing Approach: A Case Study of Automotive Industry	2020	Light Vehicles	Delphi Holt-Winters fuzzy logic ANN	Input: sales history environmental factors (GDP, inflation, etc.) Scope: 168 months history	Metrics: MAD, MSE, MAPE Result: nonlinear autoregressive network with exogenous inputs exceeds accuracy of nonlinear autoregressive network and Holt-Winters' model

5.1.2 Lead Factors based on Business Knowledge

In total, 41 discussions and interviews have been conducted in groups or individually from July 2020 to February 2021. As of Gaskin et. al. (2010: 5) and Guest (2016), sample sizes of six to fifteen seem appropriate for qualitative studies using in-depth interviews. This range also proofed well-suited for the present use-case to obtain profound answers both per focus group and market. The groups of interviewees are composed as follows:

Table 5.4: Composition of Interviewees

FOCUS GROUP and EXPERTS	MARKETS[32]	COMPANY
41.5% from sales 39% from logistics 19.5% from business units	58.5% from EU 14.6% from NA 14.6% from LA	Six different medium- and large sized enterprises

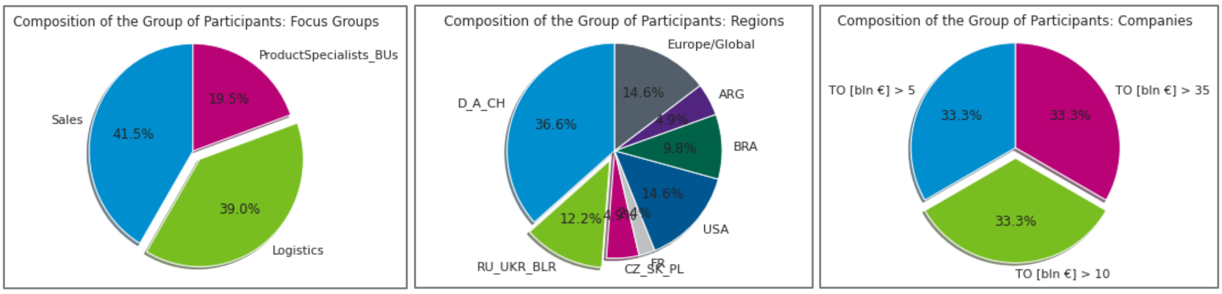


Figure 5.2 Composition of the Interviewees

The individual factors, subsequently referred to as *items*, which are examined for their significance and impact, are subdivided according to the categories (*item battery*) named in Appendix 6.[33] Moreover, the influence of the category *Time*, which is directly but super-ordinately related to the other categories and to demand behaviors, is examined.

There will also be a distinction as to whether and how much a factor influences demand is dependent on the respective product type.[34]

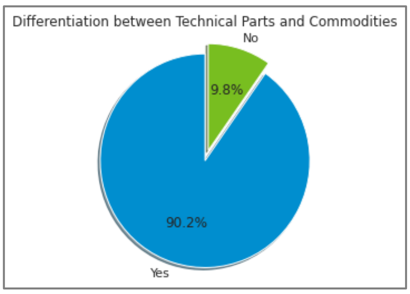


Figure 5.3 Differentiation of Products in Participating Companies

[32] Missing 12.2%: Global View due to organizational structures in the company

[33] The interviews are evaluated using the so-called item battery / item method.

[34] The more granular distinction made by VDA between accident and failure, wear-and-tear as well as service parts is not adopted here. Instead, in accordance with the definition in Chapter 3.3, accident and failure parts are assigned to the *TP* category, while wear and tear as well as service parts are assigned to *COM*.

In addition to the frequency distribution and an associated prioritization of the items, the attributes managerial (proactive), random (reactive), structural (seasonal or cyclical) and/or effective in the long term are also assigned in order to obtain indications of the respective relevance as well as data representation capabilities of the indicators. (Caniato et. al. 2005)

Results from Open Questions

A total of 88 single indicators were named. 55 are attributable to factors that influence the demand of a specific customer for a specific product. These tend to have a long-term effect if they are assigned to the item battery *market participants*, i.e., suppliers and competitors as well as customers. If the factor is product-specific and at the same time part of the sales promoting instruments, it is attributed a highly volatile nature.

33 factors are listed which are assumed to have an influence on total demand. They both reference the macro and the industry specific environment. If they are of a political-legal, or economical nature, a long-term effect is assigned. Exceptions are currency fluctuations as well as weather and climatic conditions. They are attributed a selective, recurring to highly dynamic effect.

A detailed overview of all factors named is provided in Table E 6 and Table E 7.

Results from Closed Questions

Macroeconomics

In general, experts consider economic, political, and regulatory macroeconomic factors to be moderately to highly influential on the demand for *TP* and *COM* with medians of 3, 3.5 and 4.0. Inconsistencies in the individual scoring can primarily be attributed to regional settings.

High evaluations are mainly given by participants with responsibilities in import-dependent countries (e.g., Russia, USA, and Argentina). Here, the customers' purchasing behavior is majorly influenced by the exchange rate differences between foreign and domestic currencies. In each focus group, the reasoning is as follows: "aggressive sometimes also speculative ordering due to exchange rate fluctuations". In contrast, participants from countries with a high level of domestic production and/or a common or rather stable currency rate, the importance is regarded lower.

According to focus groups 1 and 2, the influence of import duties or embargos is considered less significant than currency fluctuations for two reasons:

- They are known in advance, depending on the amount of time in-between announcement and enforcement, and
- Their effect is mitigatable by means of free trade agreements, price adaptations or incoterms (~"paid by departure land").

Economic stimuli like the eco-rebate ('*Abwrackprämie*') for cars in Germany in 2009, or the pandemic-motivated one-off payment to US-citizens have a demonstrably positive effect on demand. However, the measures are rated rather low. One reason is their irregular and very sporadic nature. The second one is, that they are either indirectly taken into account by considering the vehicle population with a certain time lag, or not at all, as the effect of industry-independent measures is hardly ascertainable to specific products.

Evaluations are high if technical and ecological mobility trends are politically motivated. Accordingly, there is a strong correlation between the ratings for political and technological respectively ecological indicators.

High fluctuations in the ratings for the macroeconomic item 'Technology' again are to be attributed to regionally prevailing conditions. To make mobility as resource-friendly and emission-free as possible, the powertrain mix is not only promoted politically and by environmental associations, but also by manufacturers in forms of a broad portfolio comprising combustion engines, electric drives, and fuel cells. Hence, the increasing use of alternative drive systems whether being ecologically motivated and/or financially incentivized, and the advancing developments in the portfolio, result in an increasingly diverse vehicle population.

Thus, experts for mid-European markets give high ratings for TP with medians of 4, stating, that the indicator gains in importance with the increasing heterogeneity of the vehicle population. This trend is reinforced by the fact that the maximum average age of a car with 11.3 years is much lower than in Eastern European or South American countries, i.e., that the fleet changes even faster. (Hecker et. al. 2012). For COM, there are divergent opinions as the term includes both products that are exclusively installed in combustion-powered vehicles like spark plugs or fuel filters, being therefore highly affected, and products that can be installed in any vehicle regardless of the drive system like wipers or lighting.

Lower ratings from South America are vindicated with an existing limited powertrain mix, as e.g., in Brazil, where passenger cars are solely run on bioethanol or on gasoline, but not on Diesel, and also with the fact, that e- or hybrid drives are currently not affordable by the vast majority of the population.

The future of mobility though is not only characterized by the disruptive trends in drive systems, but also by the demand for advanced connectivity and the associated integration of smart hard- and software components. This trend is also leading to an expanded portfolio and a steadily growing importance of the item 'Technology'.

The item 'Culture' comprises different trends: ongoing urbanization, new urban mobility concepts encouraging shifts to more sustainable and diverse modes of transport, and the sharing culture, all resulting in a changing mobility behavior. Still, its significance is considered rather low. The medians are at 1 in European and South American countries and at 2 in North American countries. The frequency distribution of the ratings and respective reasonings show that at present it is more of a vision, than an actual occurrence. However, statements prove that the influence of the item will develop upwards in the future.

Vacation-related periods are only considered to have a strong effect if they are of sufficient length and are not offset by regional differences.

The ratings for the item 'Ecology' are strongly correlated with the ratings for the 'Technology' item. BUs state that the measures for climate change mitigation are directly related to technological innovations. A low rating is given majorly in Eastern European markets, in Russia and in Brazil since as yet there are no or very tolerant limits and regulations for reducing emissions. An increased rating for both *TP* and *COM* can also be explained by the factor 'Atmospheric Conditions'. These differ from normal seasonal weather and climate conditions in their rather irregular to unpredictable occurrence and in their intensity. Examples are broad temperature ranges and sudden significant peaks or drops, precipitation phases and thunderstorms.

Microeconomics

The ratings for the items 'Market Structures and Dynamics' and 'Industry Structure and Market Participants' in the item category 'Microeconomics' diverge strongly depending on whether the product belongs to *TP* or *COM*. The reason behind this is the nature of the respective market. There are very few suppliers for *TP*. Those are well established and have specific expertise. Their number remains limited, as the entry barriers to the market are high. The positions in negotiations with customers are strong. With *COM*, there is an open predatory competition, which results in steady shifts of competitive power. At the same time, consolidation and concentration processes amongst customers are taking place, particularly in European and North American markets. Thus, the position of customers is strengthened not only by the possibility of purchasing parts from second- or third-source suppliers, but also by the constant expansion of the companies themselves and by their international distributional networks.

Generally, the item 'Market Position' is to be regarded one of the most significant indicators for *TP* as well as *COM* with medians of 4 respectively 5. The slightly lower scores for *TP* can be derived from the suppliers' strong market position, the proven importance of the brand, and the often-existing price leadership. Regarding *COM*, the item is of great importance, as suppliers not only seek to remain in but to strengthen their positions, to increase market shares and to strive for price leadership. Hence, there are also correlations with the acquisition, engagement and development of customers, the consideration of potential growth rates for internal targets as well as customer-sided contractual agreements, and the use and effect of marketing measures.

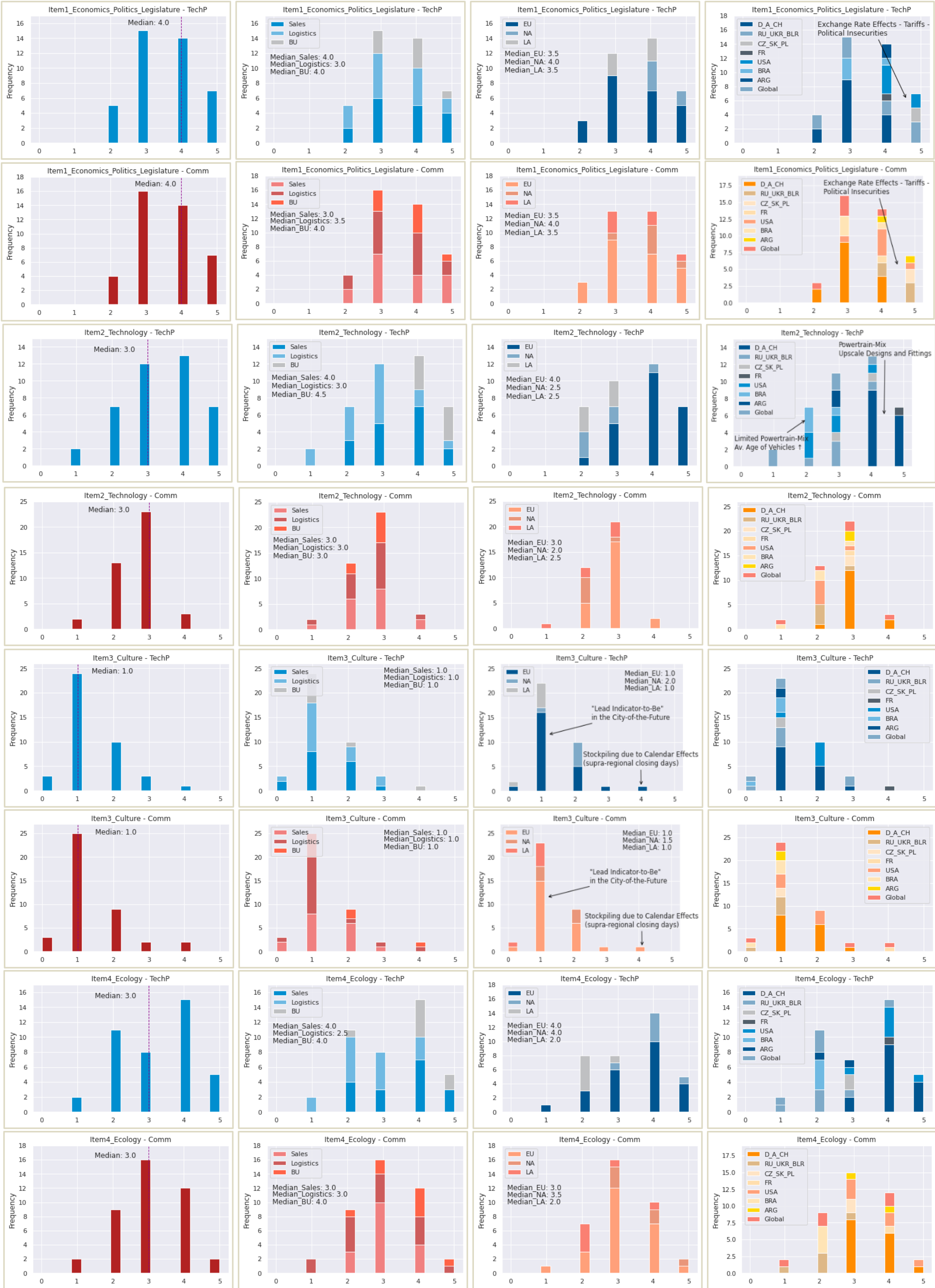


Figure 5.4 Rating of Macroeconomic Factors of Variance

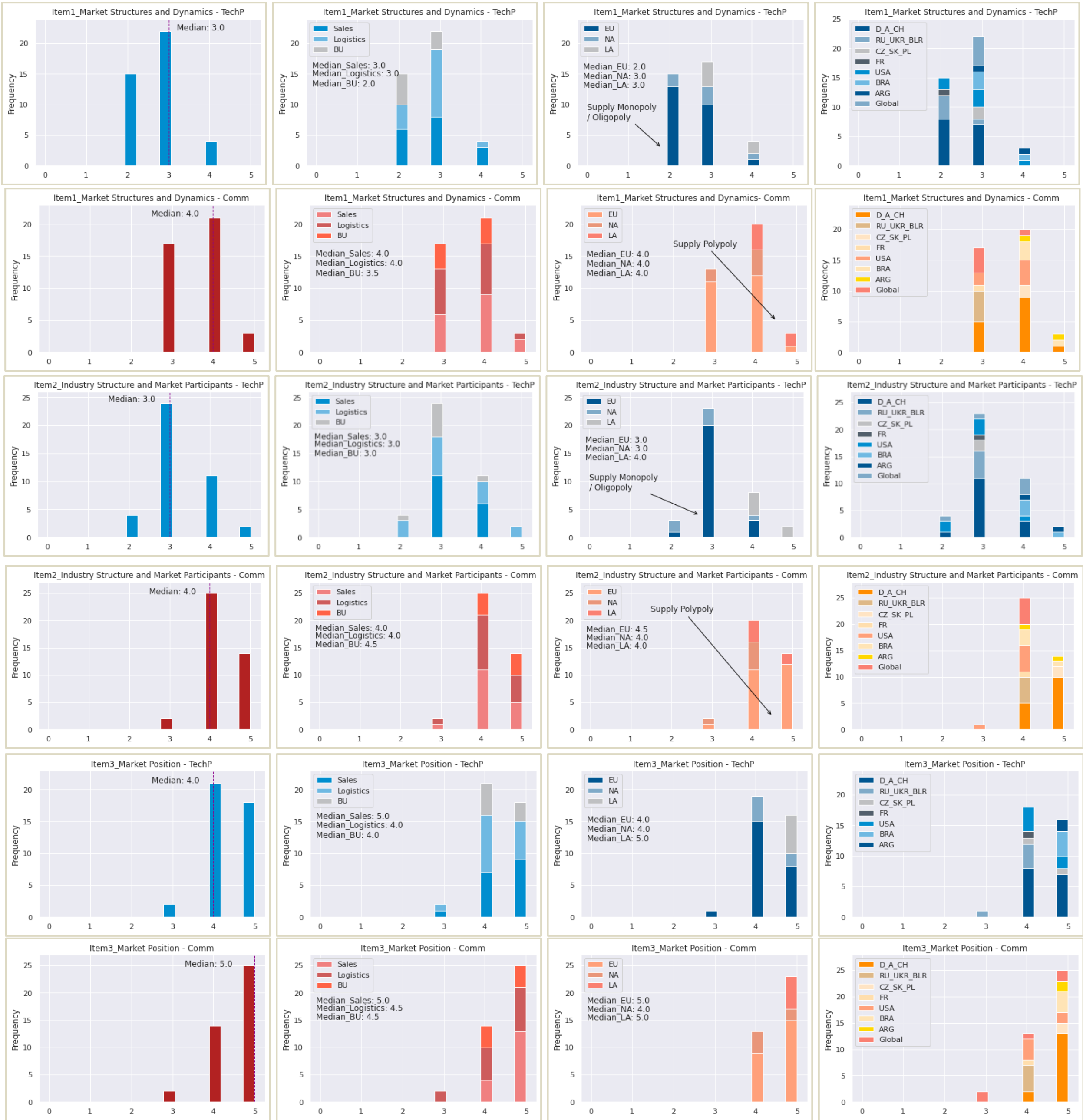


Figure 5.5 Rating of General Microeconomic Factors of Variance

Microeconomics – Customer

Customer-related items are subdivided into the categories *Firmographics* and *Psychographics*.

Firmographic-related item 1 ‘Size, Ownership and Distribution Network’ is currently defined primarily by an increasing consolidation of the market. While the market share of the top three players in the Northern American area amounts to 50%, the European counterpart is still more fragmented with the leading actors having a joint share of about 15%. Current developments, though, suggest that the automotive aftermarket in Europe, especially in the DACH region and France, will follow a similar development path. (Roland Berger GmbH and HSH Nordbank AG 2018)

In addition to the size and network of the customer, negotiating skills and the resulting purchasing behavior are of major importance in determining demand. Negotiations take place on a predefined hierarchy. Depending on the affiliation of the respective subordinate customers, they also apply to them. In case of renegotiations due to takeovers and a redefinition of targets and pricing at a higher level, agreements on national/subordinate levels are superseded.

Focus groups attribute average medians of 4 to all three items. One exception is the item ‘Negotiating Power and Communication’. For *TP* medians are at 2 and 3 due to the supply oligopoly, whereas for COM, there is a strong tendency towards a median of 5, especially in North America and Europe.

Item 2 ‘ERP-System’ plays a medium to high role with a median of 3.

Figure 5.6 Rating of Microeconomic Factors - Firmo- and Psychographics

Microeconomics – Product

Goods can be divided into at least two categories: Fast-moving and slow-moving products or, more casually, fast sellers and slow sellers. The designations stand for the frequency with which the products enter and leave the warehouse. Generally, the status is derived from the demand for the respective product, but at the same time can be an indicator of future developments. Medians for both *TP* and *COM* are at 5. A further distinction is made with the categories *service*, *wear-and-tear,* and *accident* part. Demand for service as well as wear-and-tear parts is mostly triggered through mileage or – as in the DACH region also through fixed intervals. *TP*, on the contrary, are mostly replaced due to randomness, e.g., after an accident. The focus groups regard the differentiation by product category as important, especially the BUs with a median of 4. Experts from South America give slightly lower ratings with a median of 3.

Items 3 and 4 refer to a product's life cycle (PLC). It mostly refers to the economic development of a product over a certain period of time from its introduction to its phase-out from the market on the basis of sales, revenue, and profit curves (cf. item 3 'Economic PLC Phase'). Theoretically there are six different phases: development, cold start/phase-in, growth, maturity, decline and phase-out. In economic reality, though, they cannot always be separated so precisely from one another, and they also extend over different periods of time. While experts from the BUs and sales recommend referencing the development of similar/related products for cold-starts, specific information must be found for the other very individual phases. A consideration of the same is rated as very important for *TP* and *COM* with medians of 4 to 5.

The economic PLC is not to be confused with the technical PLC, subsequently referred to as the *wear-and-tear life cycle*. There is a direct link to the mileage, which explains the different ratings: Wear-and-tear hardly affects *TP* resulting in a median of 2, whereas demand for *COM* is as already stated mainly triggered by it.

In a company-internal marketing research campaign[35] initiated by the s*ales* department of one of the companies asked, item 5 'Brand and Quality' was named amongst the top five reasons for purchasing barriers besides pricing, distribution, and product range. Furthermore, it was concluded that brand positioning (i.e., achieving both awareness and loyalty) is defined through four aspects:

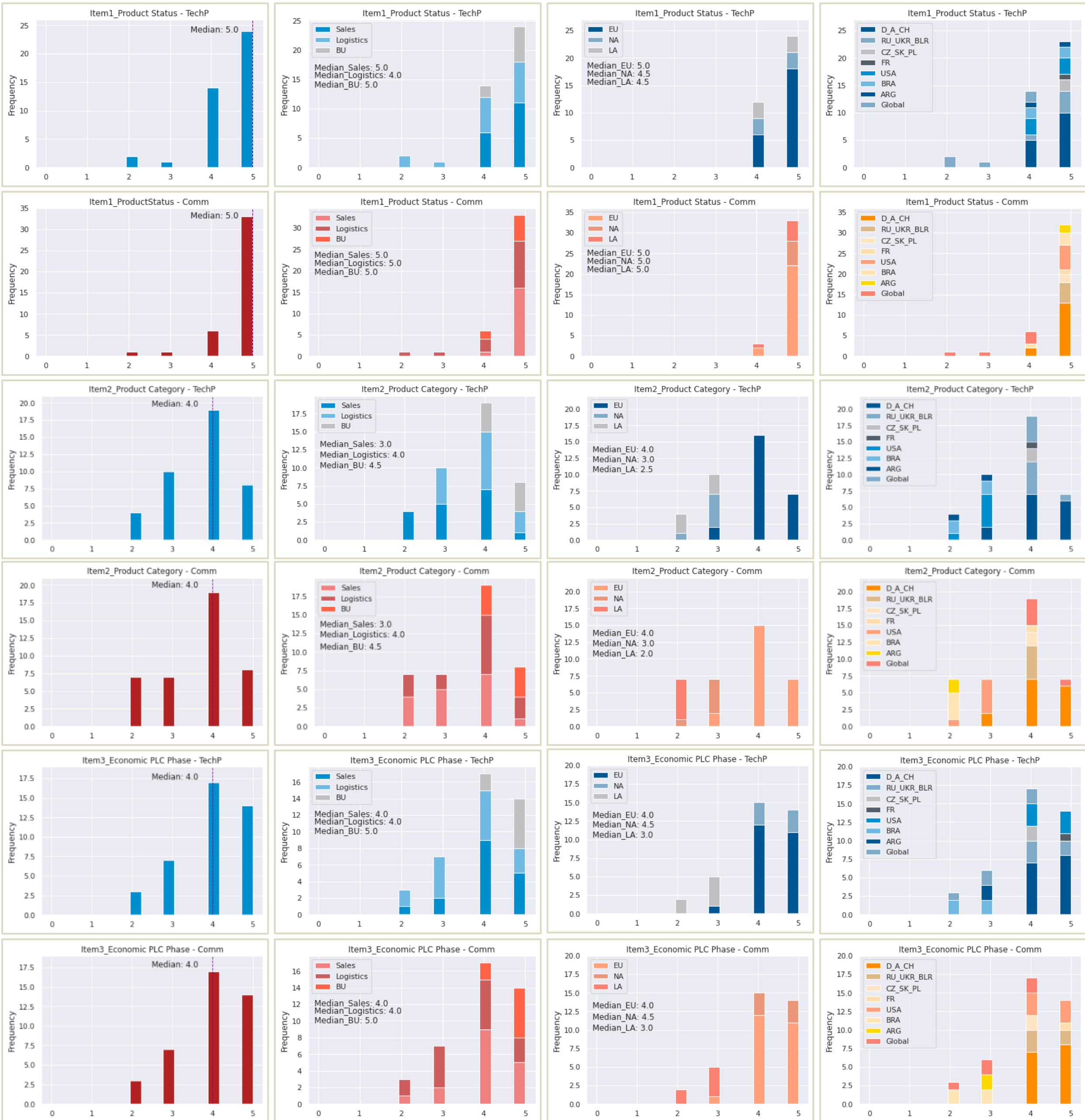
- quality: reliability and after-sales support for 1st and 2nd level customers
- type of solution: convenient, effective, and efficient repairs, fit to the vehicle's market value, connected technology
- responsibility: value-orientation, conservation of resources
- internationalization: worldwide distribution, customer focus on global level.

The results from the study are consistent with the focus group assessments. Due to the respective market position of the manufacturer/supplier and the price, high quality standards are an inherent expectation that must be met for *TP*. Medians of 3, 3.5 and 4 for *COM* are slightly lower, but there is also a significant right skew in the distribution. The

[35] The campaign was conducted in cooperation with an external institute comprising B2B (sample size: 200) and B2C (sample size: 1000) relations.

respective relevancies have a common reason: high quality standards are considered a competitive advantage here.

The information gain from the final item ‘Product Complexity’ is highly correlated with other items according to experts. Sales relates demand for complex products to wealth as well as the level of education in a specific region. BUs cite repair effort as a correlating factor, and both emphasize pricing information. In particular it is the latter which promises more value than mapping the degree of complexity. This results in respective medians of 3 and 1 for *TP* and *COM*.



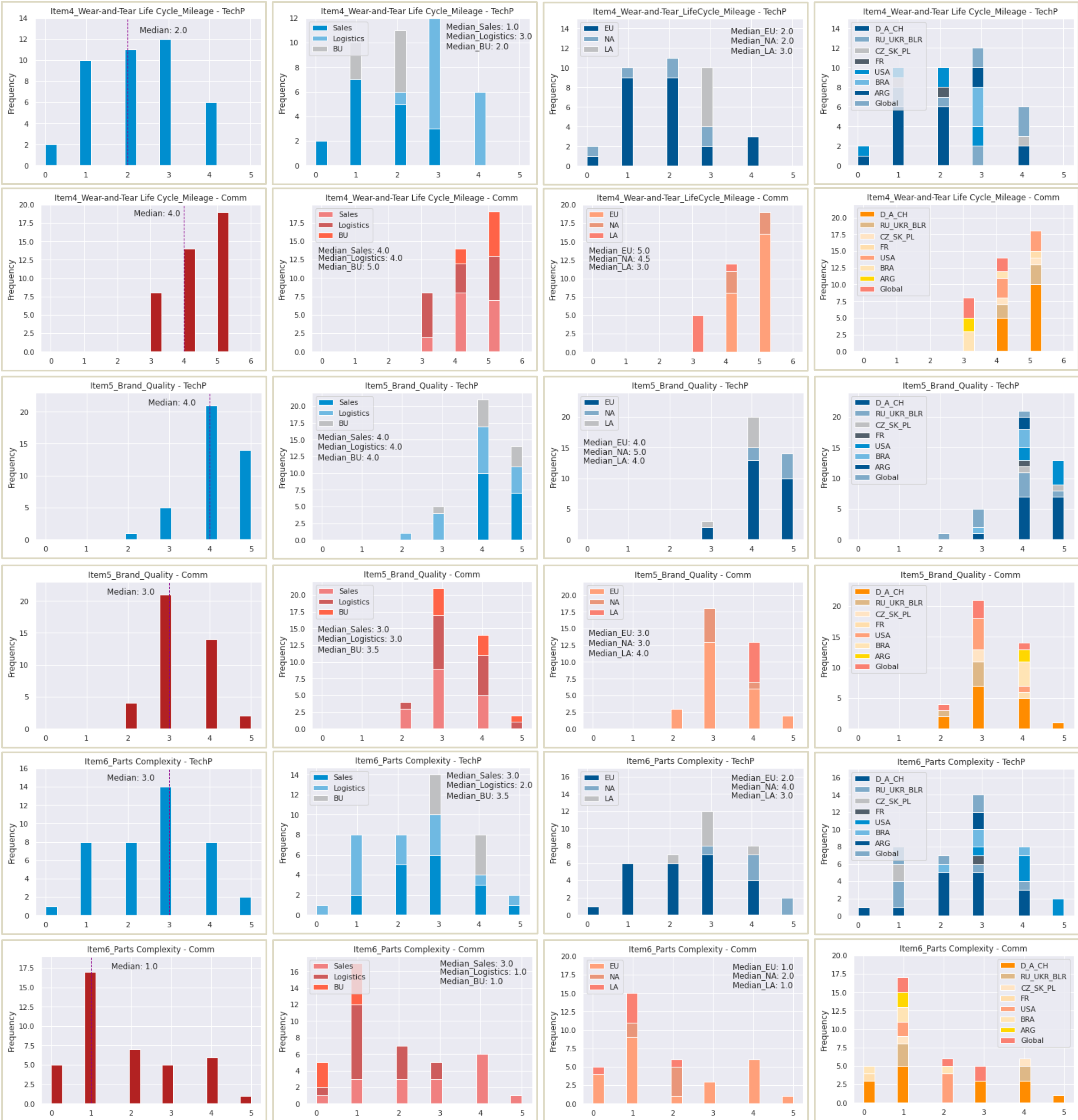


Figure 5.7 Ratings of Product-related Factors of Variance

Microeconomics – Marketing Instruments

The term *pricing* is used to summarize all measures that serve to determine the price at which a product is offered on the market. Various factors, internal ones like costs of production and margins, but also external ones like the market form, competitor information as well as country-specific dependencies play an important role in this process.

Whether a price is competitive and triggers sales or an increase in sales is ultimately determined by two factors:

- the market price of the competitor(s) compared to the own price setting and
- the customers' degree of internationalization and their network

Competitor pricing is particularly important for *COMs*. Both suppliers and consumers are numerous. Suppliers secure their market share through pricing, and in higher segments also through the brand, quality standards and bonus services, so-called value-added services (VAS). The analyses of the expert ratings show that it is not the retail price that is decisive here, but the net price, i.e., the final price offered after all discounts and downstream credits have been deducted. For *TPs*, price pressure from competitors is much lower. This can be seen in a comparison of the medians, which are at 4 and higher for *COMs* and at 3 for *TPs*.

In general, the item 'Concessions' is considered more important than the item ‘Price Segments and Pricing’. The reason for this is that if market orientation needs improvement, i.e., the general market price level has been missed, discounts can be used to correct prices and to achieve compensatory effects. Also, high ratings are due to the fact that concessions can be granted in two different forms:

- type 1 is related to direct discounts on the product demanded
- type 2 is related to complementary discounts, i.e., discounts granted on products from the portfolio, if certain products or certain amounts of certain products are purchased.

For the item 'Campaigns', referring to marketing measures in forms of incentives, there are large variances. They have a demonstrable effect regardless of whether they are offered at distributor- or at workshop level. However, they are often considered a trigger for artificially inflated demand both on supplier and customer side which lead to peak orders and high inventories as long as campaigns are active but prevent levelled subsequent orders in post-campaign months. Hence, there is the motivation to rather offer all-time lower instead of periodically decreased prices and to limit campaigns to specific, majorly emerging markets, to selected products at a very granular level, and in cyclical intervals.

The harmonization of prices across several markets, especially neighboring regions, is important to consider, when it comes to the customers’ degree of internationalization and their network. Otherwise, a cross-regional comparison of prices will offer the possibility of so-called *gray imports* i.e., that goods are legally imported from another country through channels other than the market’s official distribution system. Consequently, it is necessary to find a trade-off between defining local market pricing according to local market conditions in order to remain competitive, and to keep the market orientation supra-regional to prevent gray imports.

Incentivized sales targets are defined across all as well as for individual product groups. Their level is based equally on internal business targets, and on customers' past sales figures, including a continuous upward trend. According to consistent feedback, their influence is rated with medians of 3 respectively 4 for *TPs* and medians of 5 for *COMs*. It is *TPs* which contribute to the achievement of targets through natural demand behavior, whilst major shares are predominantly reached via *COMs*. High influence is only attributed to *TPs* if goals are product specific. In both cases, peak orders are common at the end of respective target periods, independently from whether they are defined quarterly, half-yearly or yearly.

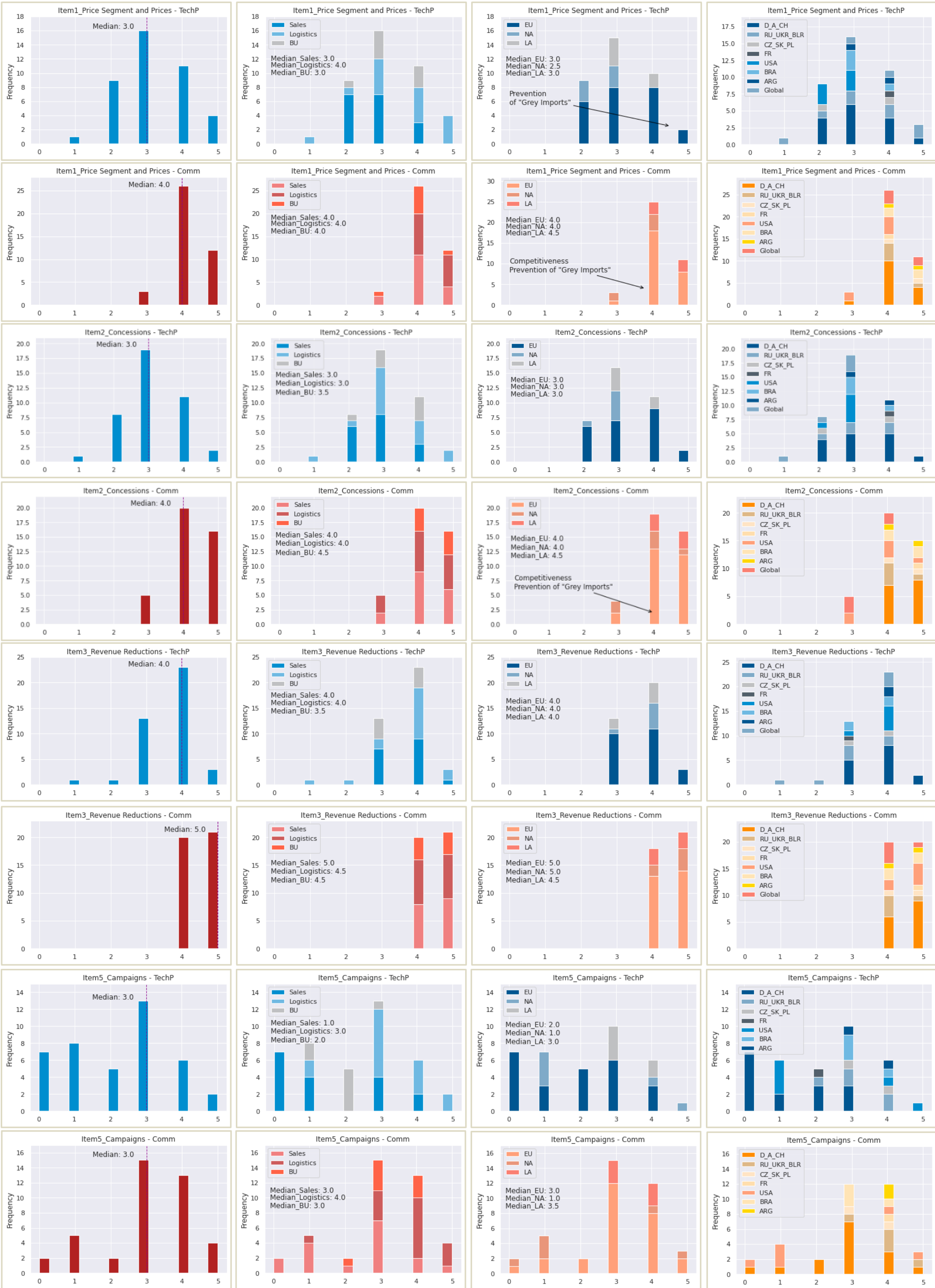


Figure 5.8 Rating of Microeconomic Factors of Variance - Marketing Instruments (1)

For ‘Value Added Services’ there are high ratings for *TPs* with medians at 3, 4 and 5, which can especially be attributed to the North and Latin American markets. While for North American customers, VAS are an inherent expectation especially for products from the HPP and UHPP segments, it is the Latin American markets, which majorly offer and worthship VAS in the form of trainings. Despite the common reasoning that VAS may serve as a USP amongst *COMs*, limited influence is attributed to this product item with medians of 1 to 2. The item ‘Service Level’, here mainly referring to customer service levels (CSL), is a measure, indicating if order lines are dispatched at the right quantity and at the right time. For *COMs*, failing to send the quantity needed or missing delivery dates, the risk of losing customers to a competitor increases. Hence, medians are at 4 within sales and BUs, and at 5 within logistics experts. For *TPs*, low service levels result in reduced customer satisfaction, as it is again an inherent expectation induced through corresponding prices and price segments.

Figure 5.9 Rating of Microeconomic Factors of Variance - Marketing Instruments (2)

The MOQ specifies a minimum order quantity and thus refers to the lowest quantity of products or units that a supplier is willing to produce or deliver at one time. It exists for both *TPs* and *COMs*. The requirement can be a soft constraint if it is linked to preventing too high fixed or setup costs per unit, or a hard constraint, if it is related to business requirements such as packaging units. (Musalem and Dekker 2005) Hence, the MOQ is an alternative way to achieve economies of scales in production and transportation. (Zhao and Katehakis 2006) As it is a product-specific but customer-independent variable, flexibility on part of the customer is restricted. Still, focus groups, also logistics, and BUs regard sales target quantities more influential than MOQs within the item 'Contractual Agreements'.

Time

With regard to the item 'Time', one needs to differentiate, if demand swings up and down due to climate or weather-related seasonality, or if the influence on demand behavior is artificially inflated, i.e., by the company and contractually fixed deadlines.

Compared to *COMs*, demand for *TPs* is less dependent on weather- or climate-related influences as well as on the target agreements discussed in the previous section. Hence, the majority of ratings amount to 0, 1 and 2. For the item 'Target Agreements', more weight is generally attributed to the *quantity* component than to the *time* component with medians of 3 respectively 4. For *COMs* high medians are related to both climate-induced peaks, as for example with pollen filters or wiper blades, and to artificially generated seasonality through periods for target achievement.

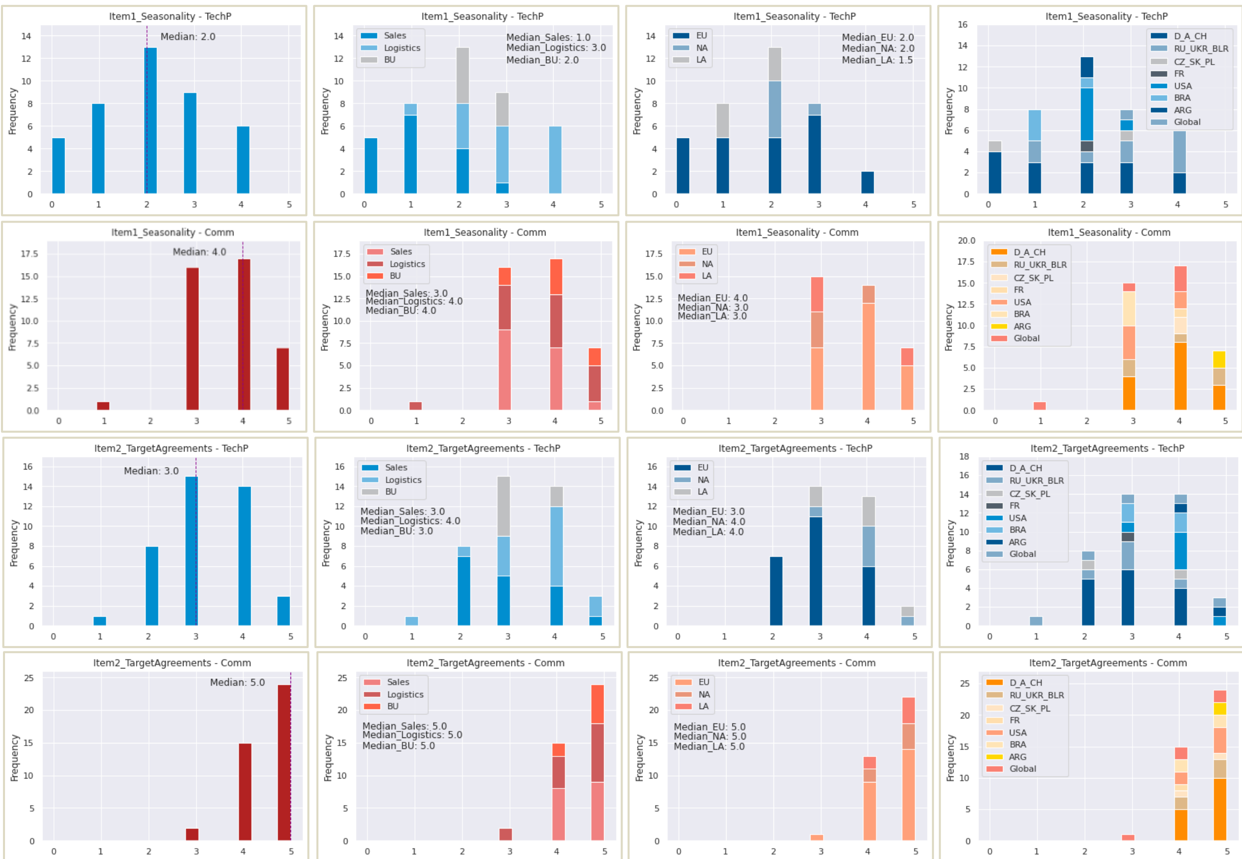


Figure 5.10 Rating of Time-related Factors of Variance

5.1.3 Conclusion

The research from Chapters 5.1.1 and 5.1.2 results in a vast pool of indicators comprising several domains. In order to proceed in a solution-oriented manner, they are subsequently presented via logic trees based on the so-called MECE principle (Minto 2021) MECE means *mutually exclusive, collectively exhaustive* implying that individual indicators do not overlap content wise on a structural level (*ME*) and that the sum of the indicators on one structural level completely define the information gain on the next higher level (*CE*). In this case, splits are hierarchical.[36] In the subsequent pages, two exemplary logic trees, one for TPs and one for COMs, are displayed (cf. Figure 5.11 and Figure 5.12). The corresponding region is Europe[37], the threshold value for the median is 3. If it is exceeded, the respective factors are highlighted with a dark-colored background.

In addition to the content-related reference, the following questions also play a role for the subsequent analyses:

- Which factors can be represented by sales historical data?
- Which factors need to be added additionally?
    - From which sources can the data be extracted?
    - What is their level of granularity? – Monthly? Quarterly? Yearly?
    - How are features, representing those factors, characterized?
- How many leading factors should be regarded?

[36] Even with this approach, only a certain degree of completeness is achieved. Neither the individual logic trees, nor the tables in Appendix 5, nor the final feature space claim to be complete.
[37] The data set used for the quantitative evaluation of the factors and the forecasting models is limited to different markets within Europe.

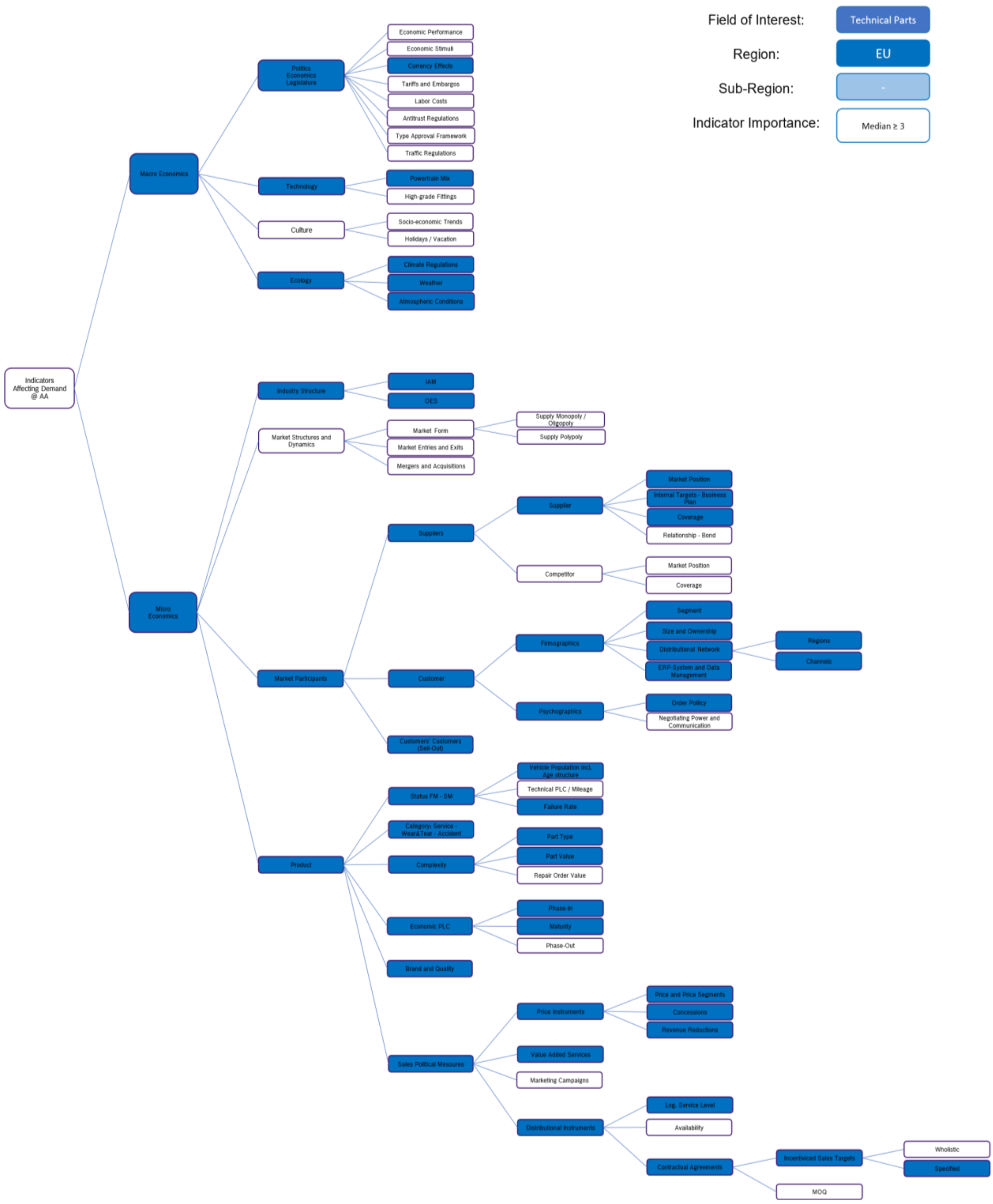


Figure 5.11 Logic Tree for Factors of Variance Affecting COM

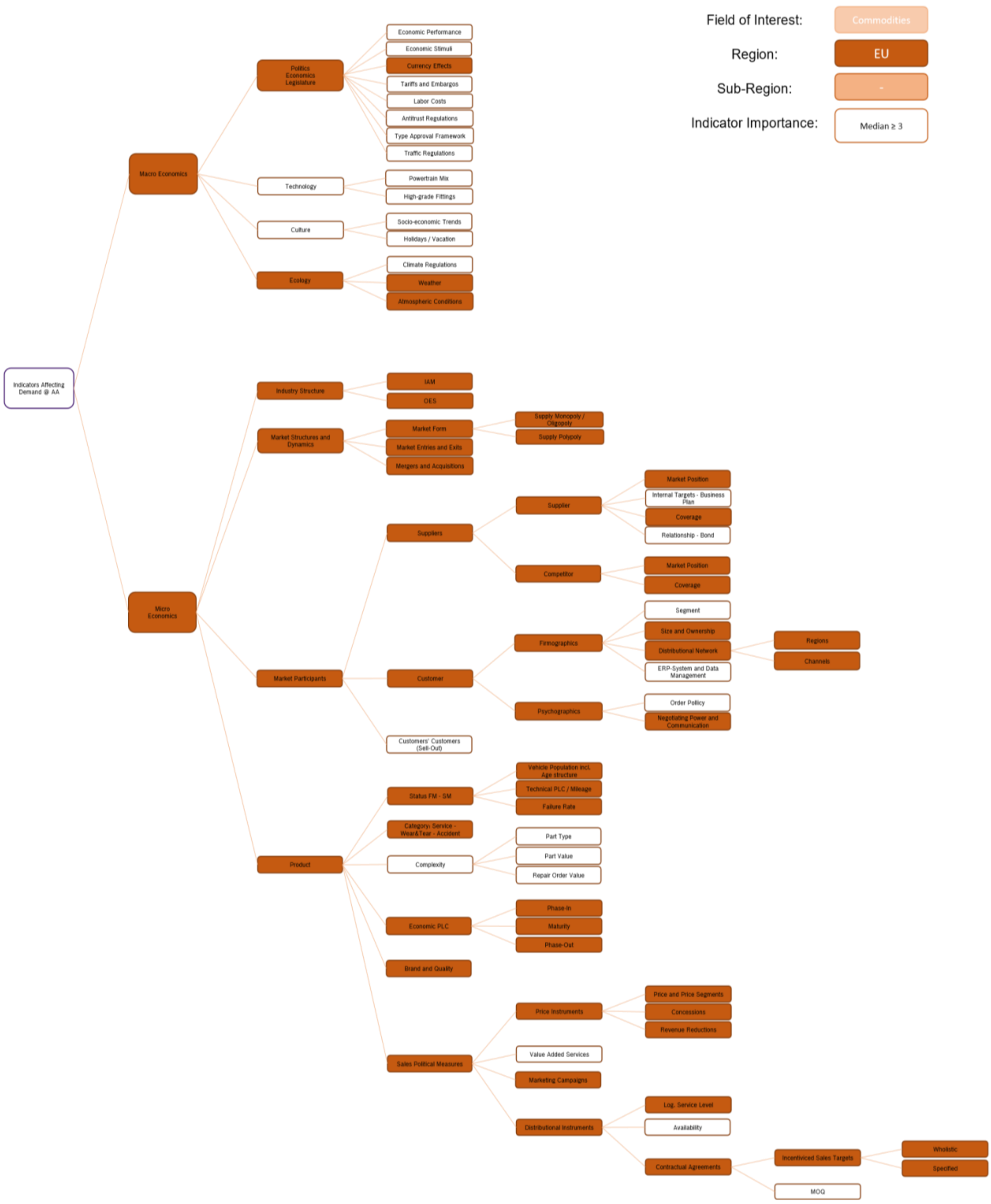


Figure 5.12 Logic Tree for Factors of Variance Affecting TP

5.2 Implementation of a Behavior-based Framework

The tables in Appendix 7 list all features that meet the criteria of general data availability and that do not represent one-off effects. Also excluded are ad-hoc factors i.e., events that occur very rarely and without notice in advance.

The grouping happens context-based according to the MECE categories, and from a data-related perspective according to the level-of-influence, as well as by texture respectively data type (cf. Chapter 2.1.2). Furthermore, there will be the classification into Main Features (MFs), Helper Features (HFs) and Segmentation Features (SFs). MFs are features that are used as input to train the model. They go through steps one to four of the process illustrated below with their importance and ranks being determined in a twofold way to correctly answer RQs 1.2 and 1.3. Subsequently, the prediction importance is determined only for those MFs that also fulfill the criterion of availability in the observation period or that can be extrapolated with a corresponding quality. HFs are not directly used as input features but help in building other HFs or MFs whereas SFs are features which serve as active segmentation variables or as supplementary profiling variables i.e., filters that enable evaluations cluster-wise and on different levels.

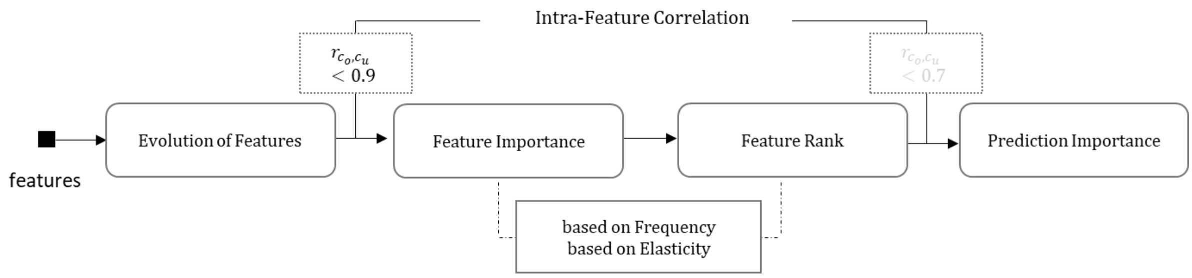


Figure 5.13 Feature Selection Process for a Behavior-based Forecasting Model

5.2.1 Time Series Segmentation

Before selecting a forecasting algorithm and corresponding MFs, it is essential to determine appropriate clustering methods and suitable SF (cf. Chapter 2.2, 2.3 and 5.2). The task is complicated by the fact that clustering assumes the existence of distinct segments in the data. However, the demands of PMCs are often distributed continuously in multivariate space without any distinct, underlying segments. Furthermore, the methods which may lead to positive results are unknown. (Caniato et. al.: 2005) To verify statements for the present use case, two approaches are applied:

- Behavior-based clustering seizing meta and statistical SF as well as unsupervised learning algorithms
- Attribute-based clustering seizing meta and statistical SF derived from business knowledge

The choice to adopt these methodologies was driven by the need to strike a balance between performance and computational demands, as well as the observation that both exhibit less sensitivity towards imcomplete or noisy data.

The procedure for applying behavior-based clustering is as follows:

- Definition of the PMC and collection of historical demand data

- Determination of the identifiers the PMC is formed of e.g., product, customer or sales market, and derivation of corresponding segmentation features SF= $\{SF_1, SF_2, \dots, SF_s\}$ on time series as well as global level (cf. Table 5.5).

Table 5.5: Categories for Segmentation Features

META FEATURES		STATISTICAL FEATURES	
Key Attributes	Descriptive Attributes	Business-induced Statistical Features	Structural Time Series Characteristics[38]

- Encoding of categorical and normalization of metric SF_s.
- Selection and application of the best statistical cluster technique with prioritized SF set via establishing a trade-off between (time series) similarity of cluster members, evaluation metrics and computational effort
- Profiling and interpretation of segments via cluster representatives

In total, 144 SF have been identified and processed. (cf. Table E 14 – Table E 22) As of Han, Kamber, and Pei (2012: 415) though, clustering algorithms do not result in meaningful segmentation for high-dimensional feature sets. This is also evident for the present use case. Hence, dimensionality is reduced and SF_s are concomitantly prioritized via establishing a trade-off between the number of clusters K, the size of clusters (i.e., number of within-cluster members) and evaluation metrics. (cf. Appendix 15)

Within the above categories, the following SF_s can be identified as the ones leading to an improved clustering result.

Table 5.6: Prioritized Feature Set

	META FEATURES		STATISTICAL FEATURES	
PMC	-	-	'GrowthRate_SA_PMC' 'APV_PMC'	'benford_correlation' 'number_peaks__n_10' 'variance_larger_than_standard_deviation' 'linear_trend__attr_slope'
SKU	'PartType'	'PPC-P AA_Mat10'	'Seasonality_PG2'	-
SR	'Dom_CS'	-	'Mean_SA_Year_SR' 'AvAPV_SR'	-

In addition, the following findings emerge from SF prioritization:

- The result for clustering improves when integrating SF_s from different levels and when combining meta- and statistical(/transaction-based) SF_s
- SF_s that have limited or no correlation with other SF_s in the respective category lead to high within-cluster homogeneity and inter-cluster distance
- Normalized SF_s lead to better cluster results than non-normalized ones
- The type of encoding is decisive: integer respectively ordinal encoding of SF_s degrades metrics whereas the same SF_s being one-hot encoded improve cluster results.

[38] Structural time series features are built via the Python based machine learning library *tsfresh.* (Christ et. al. 2018)

- Further insights about theoretical product and customer similarity can be derived:
  - Categoricals representing market-specific behavior as e.g., Dom_CS increase within-cluster homogeneity and inter-cluster distance
  - Product-related categoricals as e.g., technical or commercial item categories increase within-cluster heterogeneity and degrade metrics. Hence, they are not considered suitable to form clusters containing PMCs that are similar in their demand behavior and thus in time series progression.

A benchmarking of different cluster algorithms using the final feature set and ED as distance metric proves that there are distinct segments, but not for all PMCs. Being determined experimentally, i.e., iteratively and in dependance on the algorithms, the number of clusters G_k is on average set at 43 to 61 for a sample of 13,136 time series. Via different evaluation metrics (Table 5.7) and an accompanying random visualization of individual clusters, the BIRCH algorithm demonstrates superiority. Second best is k-means.

Table 5.7: Benchmarking Results of Clustering Algorithms[39]

	K-MEANS	K-MEDOIDS	BIRCH	DBSCAN
SC	0.69 (+/-0.015)	0.26 (+/-0.009)	*0.76* (+/-0.011)	**0.80** (+/-0.027)
CH	*1730.52* (+/-18.73)	583.18 (+/-13.90)	**1984.92** (+/-15.07)	1319.0 (+/-12.85)
DB	*0.87* (+/-0.043)	2.24 (+/-0.056)	**0.48** (+/-0.024)	1.22 (+/-0.054)

The number of within-cluster members is heterogenous with an average cluster size of 248 [BIRCH] and 279 [k-means]. An examination of the distribution shows that there are outliers both on the lower level (size_min = 17 [BIRCH]; size_min = 9 [k-means]) and the upper level (size_max = 958 [BIRCH]; size_max = 795 [k-means]).

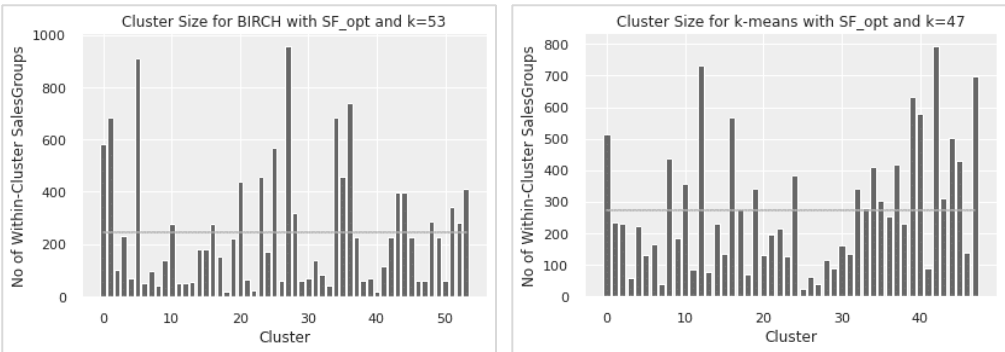


Figure 5.14 Cluster Sizes derived from Clustering Algorithms

A closer segment profiling concludes that members can in general be

- differentiated by their label *COM* and *TP*
- allocated to a specific sales market and
- assigned to three or two and occasionally also to one specific BU or PG1. This can be seen as proof of the one-stop-shopping principle as the congruence of two time series is obviously not determined by the product hierarchy.

[39] Best results are printed in bold; second best results are in italics.

The clusters themselves can be differentiated into three distinct groups. The first category comprises clusters which are considered *good* in terms of their size and the nature of their members. Examples are Cluster 21, Cluster 27 and Cluster 30 in Figure 5.15.

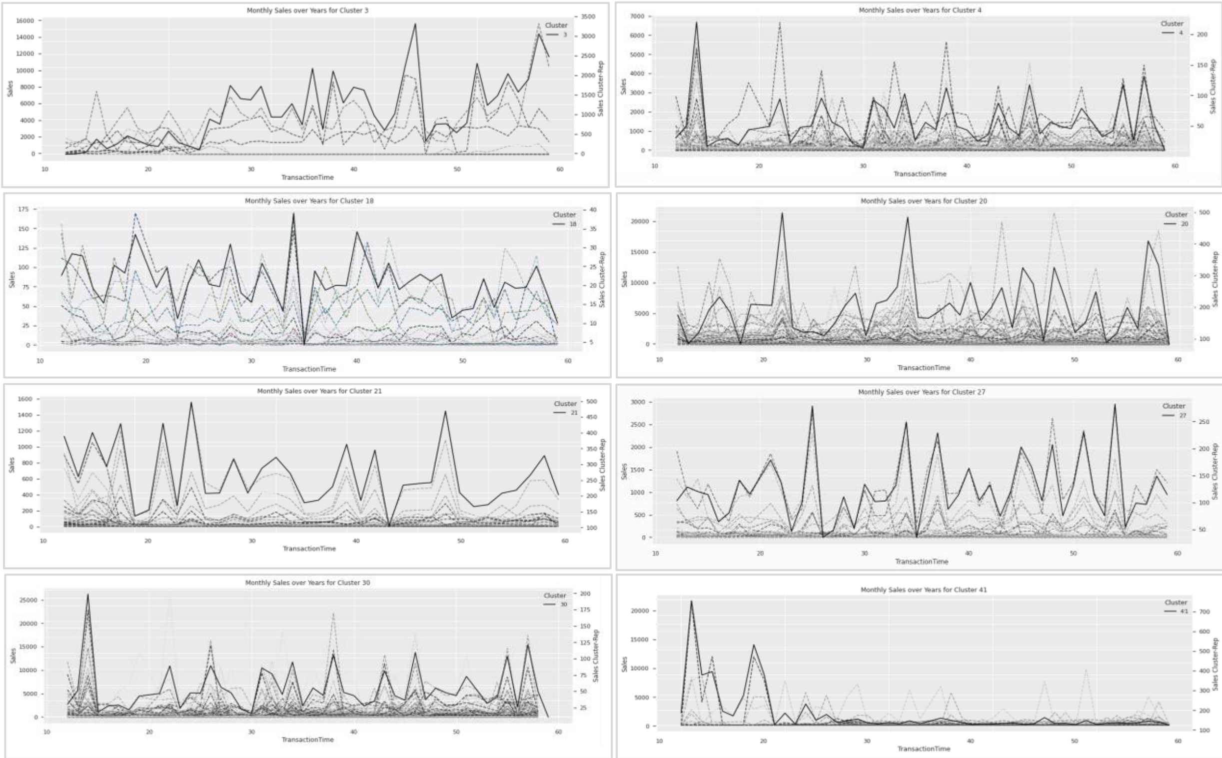


Figure 5.15 Exemplary Cluster Representation from AA Time Series Segmentation

Besides the measures that have been introduced in Chapter 2 and applied beforehand, there are additional, simple ways (Keogh and Kasetty 2002) to verify if the similarity of time series is to be regarded *good*. One approach is to use off-set translation in combination with amplitude scaling i.e., the time series is shifted by its mean and subsequently scaled by its standard deviation. Another option is to apply first differencing and scaling. The second step is to determine the Bravais-Pearson correlation coefficient r between the transformed time series y_i.and y_j. To obtain the upcoming results, the second option is chosen with the cluster members always being compared to each other as well as to the cluster representative. The distinct curves of the empirical cumulative distribution functions (ECDF) formed from the coefficients are represented in Figure 5.16.[40]

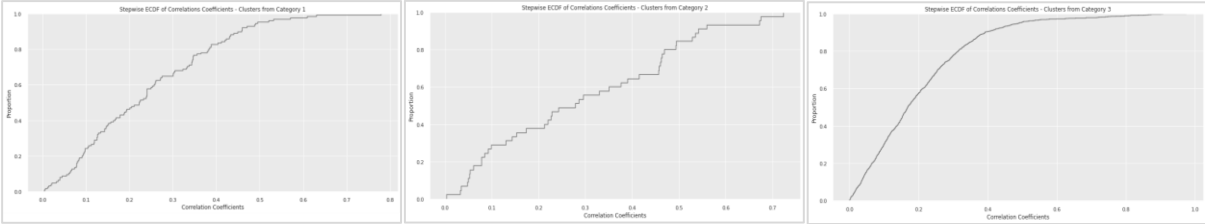


Figure 5.16 Clusterwise Time Series Corrrelation

[40] Category 1: mean: 0.2601; std: 0.1658; min: 0.0013; max: 0.7812; Category 2: mean: 0.2938; std: 0.2042; min: 0.0033; max: 0.7230; Category 3: mean: 0.2028; std: 0.1594; min: 0.0002; max: 0.9748

The second category consists of clusters whose PMCs expose similar behavior. However, the number of members is too low to achieve statistically significant results and train a global model sufficiently well. Examples are Cluster 3, Cluster 18 and Cluster 41 in Figure 5.15. Clusters in the third category do have a significant number of members. A limited number, mostly the dominant ones from a sales quantity perspective, are also similar in terms of their time series behavior, but there is also a significant share of time series which can generally be referred to as noise. Examples are Cluster 4 and Cluster 20 in Figure 5.15.

Several new conclusions result from these findings:

- Individual i.e., cluster specific model selection and parameter configuration is very expensive both with regards to time and computational effort.
- There is a lack of statistical significance due to a partially small number of cluster members and
- noisy time series from category three will affect the selection of models and blur its parameter settings.
- Evident cluster characteristics from individual clusters can be re-used as SF for attribute-based clustering.

For the second approach, nine attributes in four categories are defined (Table 5.8), resulting in 24 standalone clusters. Only those classifiers are listed which both met the criteria of availability and improved the models' performance with regards to cluster homogeneity and forecast quality. Concomitant comes an empirical reasoning for their choice (cf. Table E 56).

Table 5.8: Cluster Attributes Applied for AA Time Series Segmentation

CATEGORY	CLUSTER ATTRIBUTES				
Product Type	COM			TP	
Value Contribution + Demand Profile	A + B_non-intermittent			B_intermittent + C	
Obsolescence	Cold-Start + Phase-in	Growth	Maturity	Decline	Phase-out
Currency Area	R1			R2	

As a first category, the product type signifying the differentiation of products into TP and COM is used. The selection of this category is based on the findings from the previous cluster approach as well as the business-based indications from the preliminary qualitative analysis (cf. Chapter 5.1.2). It is also founded on specific ideas from literature, where amongst others, Syntetos, Boylan and Teunter (2011a) recommend seizing *design and manufacturing process technology* as a cluster attribute.

The second category is likewise proposed in literature and highlighted in Chapter 4.2. It follows the approach of using the degree of *intermittency* as a classifier. In order to focus both on best-to-predict and value-contributing time series, *value contribution* is used as another classifier. This choice supports the business point of view referencing the trade-off between resource utilization and improvement potential. (Triparthi et. al. 2018)

The third category, *Rate of Obsolescence*, denoting a product's lifetime is represented via the economic PLC respectively its individual phases. In the absence of reliable empirical data, the phases are determined using the *Mann-Kendall Trend* test (Mann 1945, Kendall

1975, Gilbert 1987). Its purpose is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward respectively downward trend means that the variable consistently increases respectively decreases over time, but the trend may or may not be linear. It can be used in place of a parametric linear regression analysis, which tests if the slope of the estimated linear regression line is different from zero. The regression analysis requires that residuals from the fitted regression line are Gaussian distributed; an assumption not required by the non-parametric i.e., distribution-free *Mann-Kendall Trend* test. Phase-ins and Phase-out products are labelled via a different approach. [41]

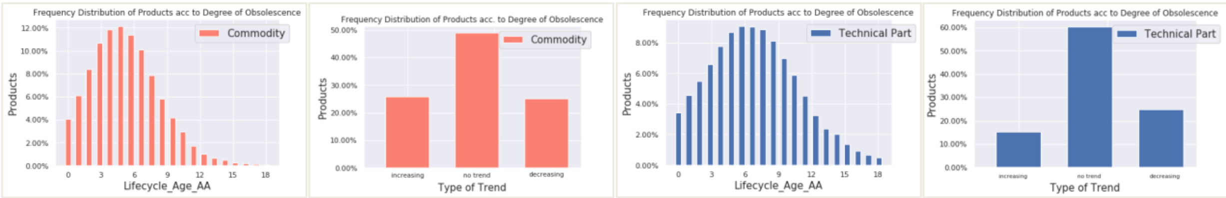


Figure 5.17 Frequency Distribution of Products acc. to Life Cycle Age and Phase

Clustering by product life cycle allows to adapt parameters, especially the quantile used, and reduces the influence of time series with converging behavior e.g., with respect to the trend.

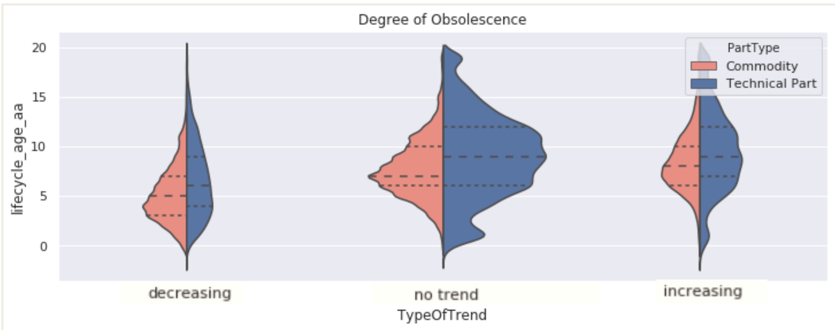


Figure 5.18 Degree of Obsolesence of COM and TP

Fourth, the region respectively currency area is used as the final cluster attribute. The background is that the demand and indicators of the same are country-specific.

[41] Phase-in products are labelled as such if the constraint that the first 24 months of the five years comprising history only represent zero demands is fulfilled. Growing, mature and declining products are labeled by determining the type of trend via the *Mann-Kendall Trend* test. It tests whether to reject $H0$ and accept $H1$, where $H0$: There is no trend present in the data. and $H1$: A trend (positive or negative) is present in the data.
Phase out products are identified by determining the number of trailing zeros with a threshold set at 12. The threshold is based on business implications.

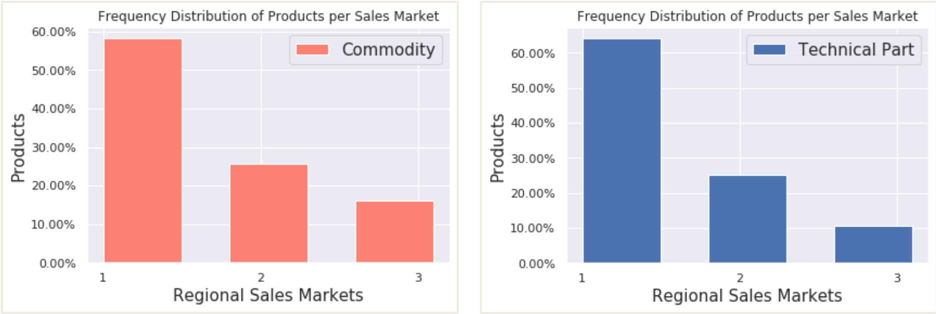


Figure 5.19 Region-specific Frequency Distribution of Products

This is evident in the fact that the portfolio does not have the same level of demand in each region. Figure 5.19 illustrates that the large majority of the products in scope (approximately 60% for COM and 65% for TP) are only demanded in one of the markets considered. Hence, the circumstance that all markets are intra-European obviously has only a minor influence. Approximately 25% of the products are demanded in two of the three markets, and 15% respectively 10% of the products are demanded in all three markets. If there is demand for a product in several regions, this demand varies both regarding frequency and absolute quantities.

Possible reasons for this comprise the country-specific composition of the vehicle population, country-specific pricing as well as currency effects.

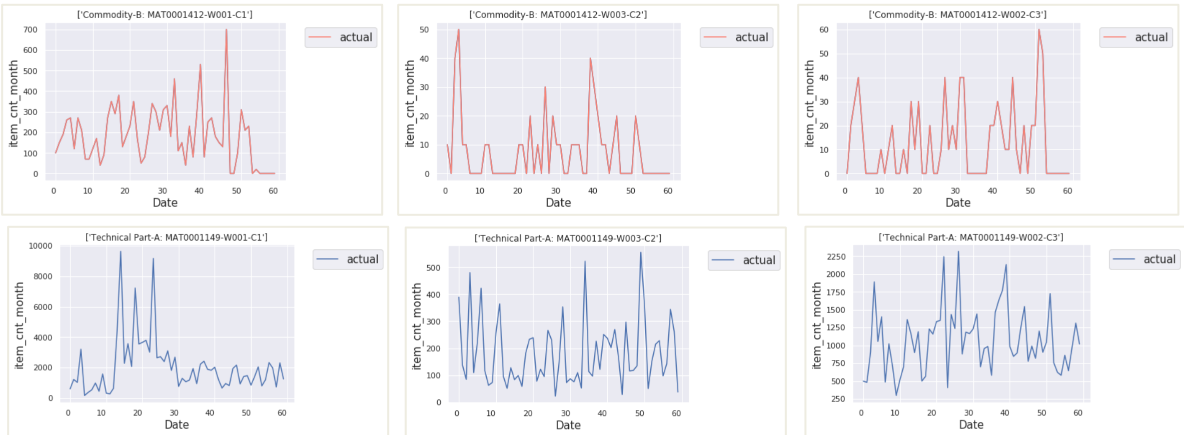


Figure 5.20 Region-specific Demands for Identical Products

5.2.2 Assessment of Lead Factors

There is no best feature selection method – at least not universally, just like there is no best set of input variables or best ML algorithm. Hence, based on the preceding chapters, a procedure of several steps is set up to answer the following questions:

First, how is the MF characterized? This is directly related to the evolution, i.e., the nature of MFs over time and is assessed by means of three statistical tests:

- ADF to test for unit roots i.e., the existence of stationarity
- Shapiro-Wilk test as the test of normality, and
- depending on normality, Levene's test or Bartlett' s test to test homoscedasticity i.e., homogeneity of the data

The nature of the MFs is relevant for two reasons:

- assumptions that apply to various statistical tests are to be met and
- adequate methods need to be derived to forecast MFs that are both dynamic and unknown in the future

Second, is the indicator affecting demand behavior and how much is the indicator affecting demand? Research on this will help identifying the leading factors among the lead factors.

To answer questions one to three, the statistical relevance for the predetermined dependent variable y_i is determined via a *wrapper* method (cf. Chapter 2.3.2).

The data consists of I PMCs with $i = 1, 2, \dots, \mathrm{I}$ and C MFs with $c = 1, 2, \dots, \mathrm{C}$, each of T observations. Different, traditional and ML based regression algorithms m with $m = 1, 2, \dots, M$ (cf. Appendix 8) are applied to compute the following score:

$$RIMM_{im}^{c},$$

the **R**elevance **I**ndication of MF c per PMC (**M**aterial) i and **M**odel m.

$RIMM_{im}^{c}$ is a relative scoring metric with $\sum_{c=1}^{C} RIMM_{im}^{c} = 1$. It indicates how useful respectively valuable each MF is for predicting the dependent variable y_i compared to all other MFs. It can be coefficients, quality measures, or error metrics, with its choice depending on the traditional respectively ML based models applied.

In order to determine the importance of MF c for the overall product portfolio I or a subset J of the product portfolio I with $J \subseteq I$ based on the results of a specific model m, the RIM is determined, by defining

$$RIM_{m}^{c} = \frac{\sum_{i}^{I=\infty} RIMM_{im}^{c}}{I} \qquad (43)$$

and $0 \le RIM_{m}^{c} \le 1$.

It indicates the relevance of MF c per model m where the number of PMCs $i \geq 2$.

To correctly assess the robustness of RIM_{m}^{c} i.e., to either confirm or disprove MF c's relevance via multiple models m, and to enable conclusions about if and how often MF c

is impacting demand (RQ 1.2), the distributions of RIM_m^c over all models m are tested for differences. Here, it is important to note that the scoring value itself is not included in the analysis, but only the information whether MF c has an impact i.e., $RIMM_{im}^c > 0$ or not i.e., $RIMM_{im}^c = 0$. Based on those counts, model-dependent ranks are derived for MFs c and differences in their central tendencies can be detected. (Demšar 2006)[42]

The statistical tests to be used for the current use case are displayed in Table 5.9.

Table 5.9: Test for Robustness of Central Tendencies

for $C > 3$	**HOMOSKEDASTIC**	**HETEROSKEDASTIC**
GAUSSIAN DISTRIBUTED	ANOVA Tukey HSD	Friedman Nemenyi
NON- GAUSSIAN DISTRIBUTED	Friedman Nemenyi	

With $H0$ stating that the mean value for each population is the same and being rejected for $\alpha \leq 0.05$, post-hoc tests are applied to identify MF(s) c that differ significantly from model m to model m with regards to their importance respectively rank.[43]

Provided $H0$ is not to be rejected for $\alpha \leq 0.05$, the importance of MF c can be determined using RI, the **R**elevance **I**ndication for MF c based on multiple models m with

$$RI^c = \frac{\sum_i^{I=\infty} \sum_m^{M=\infty} RIMM_{im}^c}{I * M} \tag{44}$$

with $\sum_{c=1}^{C} RI^c = 1$. Using the actual scoring values for $RIMM_{im}^c$, a score can be derived that

- is demonstrably valid regardless of the models m used and that
- indicates information on how much MF c is impacting demand.

RI^c is also calculated with respect to the overall product portfolio I, or a subset J of the product portfolio I with $J \subseteq I$.

[42] How such testing is done has already been described in the well-known paper by Demšar (2006): Based on the number of populations i.e., the number of MFs c (in here: C ≥ 3) and their nature, one must decide whether the repeated measures ANOVA with Tukey's HSD as posthoc test, or Friedman and Nemenyi's post-hoc test is the suitable statistical framework. Tukey HSD test and the Nemenyi test are applied to infer which differences are significant when comparing the model-dependent ranks of each MF c with the expected rank of MF c. For Tukey HSD, populations for MF c are significantly different if confidence intervals for their mean values are not overlapping. Nemenyi tests if differences are surpassing a specified threshold being commonly known as the critical distance cd. It is defined as $cd = q_{\alpha,M}\sqrt{\frac{M \cdot (M+1)}{6 \cdot C}}$, where α is the confidence level, M is the number of models and C is the number of measurements, here MFs c. To calculate $q_{\alpha,M}$, the *Studentised Range Statistic* for infinite degrees of freedom divided by $\sqrt{2}$ is used. Based on results of Tukey HSD respectively Nemenyi, MFs c showing significant differences in their ranks are eliminated from the dataset.

[43] The most relevant MF is getting the rank of one, the second-best rank two and so on. Here, rank one means that the frequency of MF c impacting demand is highest; rank two means that the frequency of MF c impacting demand is second highest and so on.

The averaged scores $\frac{\sum_{m}^{M=\infty} RIMM_{im}^{c}}{M}$ across models m additionally serve to determine the $ECDF^{c}$ per MF c. The graphical representation of the same allows a comparison between the distributions of the importances across individual products. At the same time, a probability can be derived per MF and its relevance.

RIM_{m}^{c} and RI^{c}enable a cross-product assessment of the factors. Hence, MFs c are prioritized which show effects across the portfolio. A disadvantage, though, is that the relevance may primarily be determined by the numerical dominance of a specific group of products, like e.g., C-Parts. To rather consider the actual value contribution of the same, the final and most essential metric for feature selection will be WRI^{c} with

$$WRI^{c} = \frac{\sum_{i}^{I=\infty} \sum_{m}^{M=\infty} RIMM_{im}^{c} * w_{i}{}^{44}}{I * M} \qquad (45)$$

Based on the WRI^{c}, the lead order i.e., ranks[45] of MFs c can be derived. This gives

$$lim_{RIMM_{im}^{c} \to 0} WRI^{c} = 0 \qquad \text{and}$$

$$lim_{RIMM_{im}^{c} \to 1} WRI^{c} = 1$$

PMCs with a rather low value contribution amplify the first and weaken the second effect. PMCs related to high-selling and/or high-priced products weaken the first and intensify the second effect.

In order to rule out multicollinearity of MFs c, the intra-feature correlation r_{c_o,c_u} according to Bravais-Pearson is determined. It comes down to dividing the covariance and the product of the standard deviations. Results range from -1, denoting a perfect negative correlation to +1, denoting a perfect positive correlation.

It is defined as

$$r_{c_o c_u} = \frac{\sum_{1}^{T} (c_{o,t} - \overline{c_o})(c_{u,t} - \overline{c_u})}{\sqrt{\sum_{1}^{T} (c_{o,t} - \overline{c_o})^2} \sqrt{\sum_{1}^{T} (c_{u,t} - \overline{c_u})^2}} \qquad (46)$$

[44] The averaged value contribution w_i is defined as $w_i = \frac{turnover_i}{\sum_{1}^{I} turnover_i}$.

[45] The most relevant feature is getting the rank of one, the second-best rank two and so on. Here, rank one means that the elasticity of MF c is highest; rank two means that the elasticity of MF c is second highest and so on.

5.2.2.1 Domain D_0: Transactional Data, Portfolio Information and Demand Profiles

Besides the actual transactional information, so-called *Industry Volume Features* based on the individual product families and levels are seized in order to represent portfolio information in a meaningful way. Another group, the *Demand Profiles* (Chapter 4.2) are used for clustering purposes and for a granular evaluation, but not as input features. Lag-Features are not considered in the feature importance analysis as they are only used in the XGB Model to counteract the memoryless-ness. They also expose high intra-feature correlation. (cf. Figure E 4)

As demonstrated in Table E 25, the majority of features in the COM sector are considered stationary; the proportion of stationary time series increases with the hierarchy. On the contrary TP parts are predominantly non-stationary; the proportion of stationary time series decreases further with the hierarchy.

Intra-Feature Correlation

In comparison with the sum-based features, average-based features are highly correlated with the portfolio comprising correlation coefficients being determined as follows:

- $r_{f_5,f_6} = 0.88, r_{f_5,f_7} = 0.75, r_{f_5,f_8} = 0.74$
- $r_{f_6,f_7} = 0.85, r_{f_6,f_8} = 0.84$
- $r_{f_7,f_8} = 0.98.$

The highest correlations for both the accumulated and averaged volumes are identified for MFs 3 and 4, and 7 and 8, respectively. The Pearson-Bravais correlation coefficients for those are as follows: r_{f_3,f_4} = 0.94; r_{f_7,f_8} = 0.98 with the reason being found in the minor differences between the product family label 'item_category1' and the product family label 'BU'. Hence, MF4 and MF8 are eliminated. Likewise correlated with the majority of the other MFs is the average-based level 2 MF with r_{f_2,f_6} = 0.86, r_{f_5,f_6} = 0.88, and r_{f_6,f_7} = 0.85. MF 9, i.e., the sales quantity aggregated over time, reveals the highest correlation with the dependent variable with r_{y,f_9}= 0.81. Accordingly, the feature importance determination based on rank will be conducted for six populations i.e., MFs.

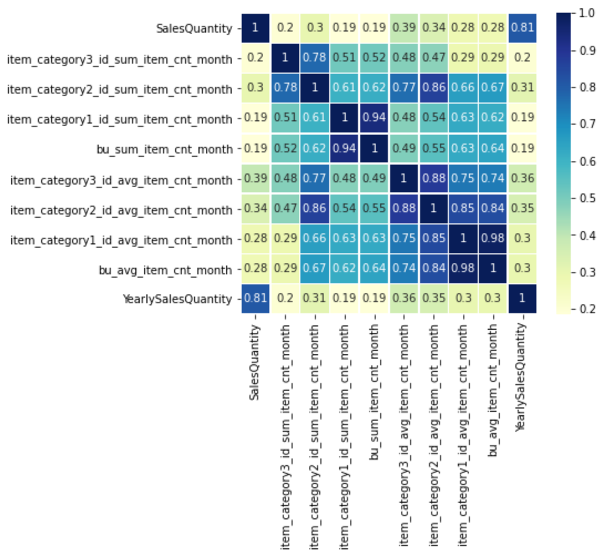


Figure 5.21 Correlation Testing for Industry Volume Features

Feature Importance and Feature Rank

… based on Frequency

For COM, we fail to reject $H0$ of the Shapiro-Wilk test that the population is Gaussian distributed for all populations. The same applies for TP. Therefore, it is assumed that all populations are Gaussian distributed. Based on Bartlett's test for homogeneity, it is assumed that the data related to COM is heteroscedastic, whereas data related to TP is homoscedastic.

Based on the results from the non-parametric Friedman/post-hoc Nemenyi test respectively the repeated measures ANOVA/post-hoc Tukey HSD test, we assume that there is no statistically significant difference between the mean values of the populations across product types and value contribution classes.[46] Hence, for all MFs, the prediction importance scores WRI^f are calculated. On behalf of those, the final prioritization and transfer of the MFs to the ML models are realized.

Table 5.10: Importance Testing for Industry Volume Features - Frequency

min. observed p-values for $C = 6$		**NON- / GAUSSIAN DISTRIBUTED**	**HOMOSKEDASTIC / HETEROSKEDASTIC**	**ROBUSTNESS of CENTRAL TENDENCIES**
COM	A	0.136	0.010	0.067
	B	0.079	0.006	0.294
	C	0.043	0.024	0.089
TP	A	0.135	0.190	0.069
	B	0.297	0.366	0.156
	C	0.295	0.322	0.126

… based on Elasticity

For COM, MF5 and MF1 are ranked highest with WRI^{f5} = 0.210 and WRI^{f1} = 0.205. This is expected as they are related to the product hierarchy being closest to the respective SKUs and hence PMC. MF7 and MF3 are ranked second highest with WRI^{f7} = 0.156 und WRI^{f3} = 0.150; MF6 and MF2 with WRI^{f6} = 0.144 and WRI^{f2} = 0.136 are at rank 5 and 6. The ranking for MFs related to TP is identical, with values for WRI^f differing slightly compared to COM.

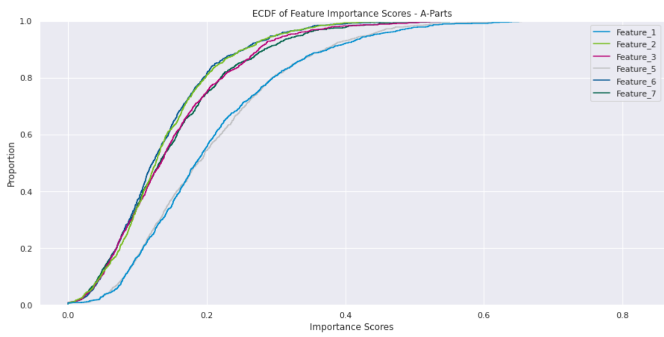


Figure 5.22 ECDF of Importance Scores for Industry Volume Features - COM-A

[46] The ranks derived from the frequency analysis are represented in Table E 31 in the appendix.

Table 5.11: Importance Testing for Industry Volume Features - Elasticity

RI^f (WRI^f)		**MF1** 'item_category3_ID_ sum_item_cnt_month'	**MF2** 'item_category2_ID_ sum_item_cnt_month'	**MF3** 'item_category1_ID_ sum_item_cnt_month'	**MF4** 'BU_ID_ sum_item_cnt_month'
COM	A	0.208	0.141	0.150	eliminated due to r_{f_3,f_4} >0.9
	B	0.181	0.142	0.174	
	C	0.173	0.147	0.175	
	overall	0.180 (0.205)	0.147 (0.136)	0.167 (0.150)	
		MF5 'item_category3_ID_ avg_item_cnt_month'	**MF6** 'item_category2_ID_ avg_item_cnt_month'	**MF7** 'item_category1_ID_ avg_item_cnt_month'	**MF8** 'BU_ID_ avg_item_cnt_month'
COM	A	0.208	0.138	0.151	eliminated due to r_{f_3,f_4} >0.9
	B	0.183	0.143	0.174	
	C	0.172	0.144	0.175	
	overall	0.183 (0.210)	0.145 (0.144)	0.170 (0.156)	
		MF1 'item_category3_ID_ sum_item_cnt_month'	**MF2** 'item_category2_ID_ sum_item_cnt_month'	**MF3** 'item_category1_ID_ sum_item_cnt_month'	**MF4** 'BU_ID_ sum_item_cnt_month'
TP	A	0.191	0.145	0.158	eliminated due to r_{f_7,f_8} >0.9
	B	0.169	0.150	0.173	
	C	0.158	0.152	0.180	
	overall	0.164 (0.194)	0.151 (0.144)	0.175 (0.154)	
		MF5 'item_category3_ID_ avg_item_cnt_month'	**MF6** 'item_category2_ID_ avg_item_cnt_month'	**MF7** 'item_category1_ID_ avg_item_cnt_month'	**MF8** 'BU_ID_ avg_item_cnt_month'
TP	A	0.189	0.148	0.158	eliminated due to r_{f_7,f_8} >0.9
	B	0.168	0.152	0.173	
	C	0.157	0.151	0.179	
	overall	0.163 (0.201)	0.152 (0.153)	0.175 (0.155)	

Besides MF4 and MF8, all six remaining MFs are transferred to the ML models.

5.2.2.2 Domain D_1: Customer-related Features

From one sales market to the next, very heterogeneous behavior can be observed with regard to the share of expenditure for the respective products. The situation is similar for the customer segments. (cf. Figure 5.23)

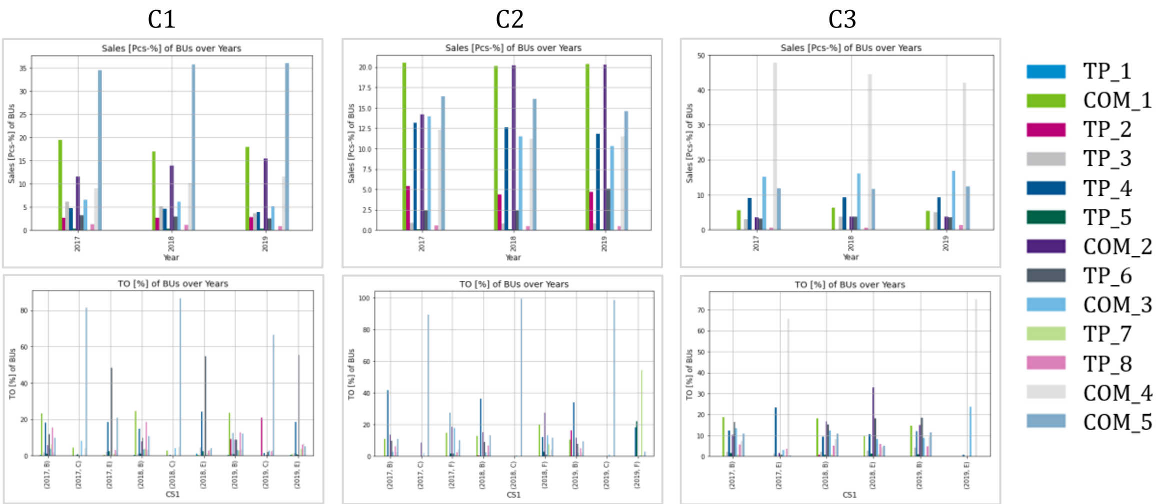


Figure 5.23 Buying Behavior and Expenditure Shares from Market to Market

In general, sales are dominated by products from the product type COM. Looking at it from a turnover perspective, products from product type TP also dominate. It is noteworthy that C1 and C2 are similar in their behavior with COM_1, COM_2 and COM_5 being sold most often, while C3 clearly prioritizes COM_4 and purchases parts from product type TP only sporadically. With regards to customer segments, *b* is the dominant one, although the distinction between wholesale and retail is sometimes a country-specific phenomenon and customers are assigned to segment *b* rather than *c* if no distinction is made. Another reason is that the portfolio is sold in its entire breadth to segment *b*, while sales in segments *c* and *f* are mainly limited to products from product type COM.

In order to accordingly transmit the behavior of the individual customers and customer segments to the ML Model, MFs from three main categories are built: The first category comprises MFs related to sales from the top customers in each sales market. The identification is based on a pareto analysis of customers on CH2-level. The resulting MFs are expressed in absolute (topcus_c, top1_c, top2_c, top3_c) and relative values (topcus_c_pct, top1_c_pct, top2_c_pct, top3_c_pct). The second category represents sales volumes per customer segment (b, c, e, f). The final category focuses on the number of customers and their evolution over the year and from year to year (unique_cus(_month), unique_cus_year).

Intra-Feature Correlation

There are no congruent MFs amongst customer-related MFs, as seen previously with industry volume features.

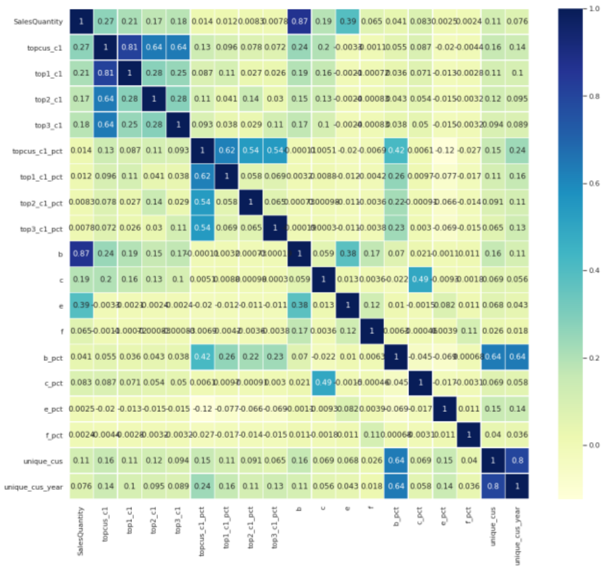


Figure 5.24 Correlation Testing for Customer-related Features

The dominance of segment *b* across all regions is also evident from the correlation analysis with a coefficient of $r_{f_9, f_{DepVar}}$ = 0.87. In addition, the number of customers correlates with the relative MF13: $r_{f_{13}, f_{17}}$ = 0.64 und $r_{f_{13}, f_{18}}$ = 0.64.

Most highly correlated is the monthly number of customers with their yearly counterpart: $r_{f_{17}, f_{18}}$ = 0.8. In addition, the sales of the top customers are highly correlated with the accumulated sales of the top customers both in absolute and relative form with r_{f_1, f_2} = 0.81; r_{f_1, f_3} = 0.64, r_{f_1, f_4} = 0.64, r_{f_5, f_6} = 0.62, r_{f_5, f_7} = 0.54, r_{f_5, f_8} = 0.54.

Importance and Feature Rank

… based on Frequency

For COM, we reject $H0$ of the Shapiro-Wilk test that the population is Gaussian distributed for all MFs. The same applies for TP. Here, too, MFs within each value contribution class are non-Gaussian distributed. Based on these results, the non-parametric Friedman/post-hoc Nemenyi test is applied to determine if there are any significant differences between the median values of the populations. According to its results, it is assumed that this applies to eight MFs (MF2, MF3, MF4; MF6, MF7, MF8, MF12, MF14) from the COM sector and to nine MFs (MF2, MF3, MF4; MF6, MF7, MF8, MF10, MF12, MF14) from the TP sector across all value contribution classes and to a ninth MF (MF10) from the COM-C sector.[47] This means, that close to 50% of all customer-related MFs are used in different frequencies by the regression models. Hence, for ten respectively nine MFs, the importance scores WRI^f are calculated. On behalf of those, the final prioritization and transfer of the MFs to the ML models are realized.

Table 5.12: Importance Testing for Customer-related Features - Frequency

min. observed p-values for $C = 18$		**NON- / GAUSSIAN DISTRIBUTED**	**HOMOSKEDASTIC / HETEROSKEDASTIC**	**ROBUSTNESS of CENTRAL TENDENCIES**
COM	A	0.000	-	0.000
	B	0.000	-	0.000
	C	0.000	-	0.000
TP	A	0.000	-	0.000
	B	0.000	-	0.000
	C	0.000	-	0.000

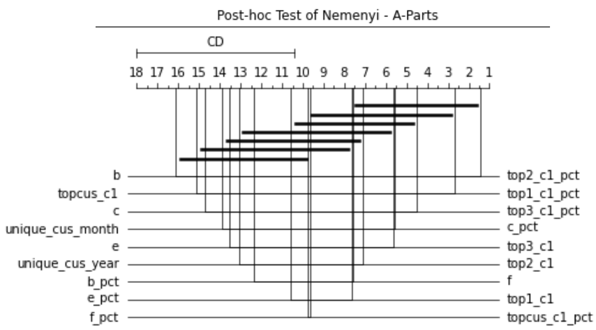


Figure 5.25 CD-Plot for Customer-related Features – COM-A

… based on Elasticity

For COM, MF9, MF1, MF10 and MF17 are ranked highest with WRI^{f9} = 0.592 , WRI^{f1} = 0.267, WRI^{f10} = 0.070 and WRI^{f17} = 0.033. It is notable that MFs were selected from every domain i.e., customers, customer segments and number of customers, with CS b clearly dominating. Relative MFs are an exception – they fall below the threshold for WRI^f = 0.01. The ranking for MFs related to TP is similar with MF10 (CS c) being less prioritized. This is expected as CS b offers both COM and TP parts to second level customers, whereas CS c predominantly sells less complex parts from the COM sector to final customers.

---

[47] The ranks derived from the frequency analysis are represented in Table E 33 in the appendix.

Table 5.13: Importance Testing for Customer-related Features - Elasticity

		MF1	**MF2**	**MF3**	**MF4**
RI^f (WRI^f)		**'topcus_c'**	**'top1_c'**	**'top2_c'**	**'top3_c'**
COM	A	0.234	eliminated due to Δ rank and mean rank $> CD$	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	0.214			
	C	0.181			
	overall	0.241 (0.267)			
		MF5	**MF6**	**MF7**	**MF8**
		'topcus_c_pct'	**'top1_c_pct'**	**'top2_c_pct'**	**'top3_c_pct'**
COM	A	0.008	eliminated due to Δ rank and mean rank $> CD$	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	0.010			
	C	0.021			
	overall	0.015 (0.007)			
		MF9	**MF10**	**MF11**	**MF12**
		'b (wholesale)	**'c'** (retail)	**'e'** (others)	**'f '** (workshops)
COM	A	0.638	0.047	0.024	eliminated due to Δ rank and mean rank > CD
	B	0.652	0.033	0.020	
	C	0.639	eliminated	0.014	
	overall	0.611 (0.593)	0.039 (0.070)	0.019 (0.022)	
		MF13	**MF14**	**MF15**	**MF16**
		'b_pct'	**'c_pct'**	**'e_pct'**	**'f_pct'**
COM	A	0.011	eliminated due to Δ rank and mean rank $> CD$	0.005	0.001
	B	0.012		0.006	0.001
	C	0.033		0.005	0.001
	overall	0.016 (0.019)		0.004 (0.004)	0.001 (0.001)
		MF17	**MF18**		
		'unique_cus(_month) '	**'unique_cus_year'**		
COM	A	0.020	0.001		
	B	0.039	0.011		
	C	0.079	0.016		
	overall	0.039 (0.033)	0.017 (0.007)		
		MF1	**MF2**	**MF3**	**MF4**
		'topcus_c'	**'top1_c'**	**'top2_c'**	**'top3_c'**
TP	A	0.138	eliminated due to Δ rank and mean rank $> CD$	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	0.115			
	C	0.128			
	overall	0.135 (0.218)			
		MF5	**MF6**	**MF7**	**MF8**
		'topcus_c_pct'	**'top1_c_pct'**	**'top2_c_pct'**	**'top3_c_pct'**
TP	A	0.010	eliminated due to Δ rank and mean rank $> CD$	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	0.015			
	C	0.037			
	overall	0.025 (0.0091)			
		MF9	**MF10**	**MF11**	**MF12**
		'b' (wholesale)	**'c'** (retail)	**'e'** (others)	**'f '** (workshops)
TP	A	0.733	eliminated due to Δ rank and mean rank $> CD$	0.026	eliminated due to Δ rank and mean rank > CD
	B	0.666		0.035	
	C	0.598		0.007	
	overall	0.651 (0.672)		0.019 (0.015)	
		MF13	**MF14**	**MF15**	**MF16**
		'b_pct'	**'c_pct'**	**'e_pct'**	**'f_pct'**
TP	A	0.023	eliminated due to Δ rank and mean rank $> CD$	0.006	0.000
	B	0.042		0.007	0.000
	C	0.093		0.002	0.000
	overall	0.052 (0.014)		0.004 (0.004)	0.000 (0.000)
		MF17	**MF18**		
		'unique_cus(_month)'	**'unique_cus_year'**		
TP	A	0.049	0.012		
	B	0.095	0.020		
	C	0.103	0.018		
	overall	0.079 (0.055)	0.025 (0.022)		

The prioritization resulting from the correlation and importance analysis, is as follows: For COM, customer sales-related MF1 (topcus_c), segment-related features MF9 (b), MF13 (b_pct), MF10 (c) and MF11 (e) as well as the average number of customers per month (MF17) are transferred to the ML models. For TP, the same MFs are to be used but without MF10.

5.2.2.3 Domain D_2: Product-related Features

Within the MECE-category *product* there are two distinct types of information:

- The first group, the master data, serves as a precise description of the products and is mostly defined company internal. It comprises details on production costs, quantity constraints (e.g., pallet quantities and lot sizes), lead times, start- and end-of-production times as well as the thereof derived product age and historized customer service levels (cf. Table E 18).
- The second group on the other hand is related to factors that define a product's status and future trends from the outside. Hence, it is specifically the political, ecological and cultural motivators affecting mobility being addressed in this category (cf. Table E 19).

The subsequent chapter and analysis will be structured alike to the data's categorization.

Product-related master data

As demonstrated in Table E 27, it is cost- and age-related MFs that are predominantly non-stationary whereas lead times and CSLs only seem to be subject to minor fluctuations. Pallet and lot sizes are rather static in nature – hence, the test were not to be applied.

Intra-Feature Correlation

Correlation coefficients r between product-related MFs are in general very low. The range of absolute values is [0.001; 0.099]. As expected, only MF6 and the lagged versions of it (MF7 and MF8) are higher correlated with r_{f_6,f_7} = 0.61, r_{f_6,f_8} = 0.52 and r_{f_7,f_8} = 0.56.

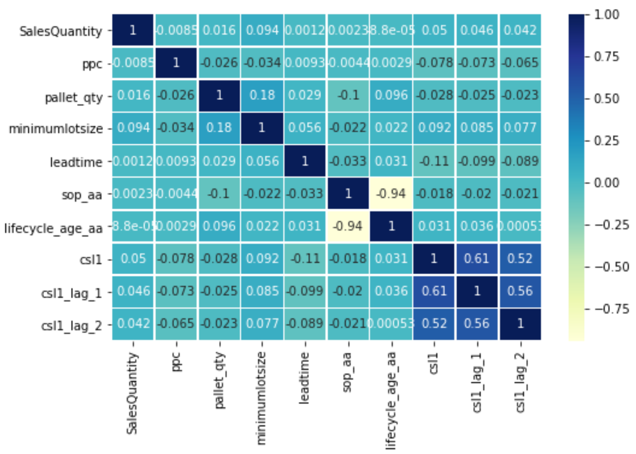


Figure 5.26 Correlation Testing for Product-related Features

Importance and Feature Rank

... based on Frequency

For COM, we reject $H0$ of the Shapiro-Wilk test that the population is Gaussian distributed for MF2 (pallet_qty), MF6 (csl19) and MF7 (csl1_lag_1). For TP-A and TP-B, $H0$ is rejected for MF4 (leadtime). Therefore, we assume that not all populations are Gaussian distributed. For TP-C we fail to reject $H0$ that the population is Gaussian distributed for all populations and reject $H0$ that the data is homoscedastic. Thus, we assume that our data is heteroscedastic.

According to the non-parametric Friedman/post-hoc Nemenyi test, significant differences between the median values of the populations are to be assumed for MF2 and MF3 for all value contribution classes from the COM sector, for MF4 for the COM-A and COM-B sector, for MF2, MF3 and MF8 for all value contribution classes from the TP sector, MF4 from the TP-A and TP-B sector as well for MF1 from the TP-C sector.[48] Hence, for five respectively six MFs, the importance scores WRI^{f} are calculated.

Table 5.14: Importance Testing for Product-related Features - Frequency

min. observed p-values for $C = 8$		NON- / GAUSSIAN DISTRIBUTED	HOMOSKEDASTIC / HETEROSKEDASTIC	ROBUSTNESS of CENTRAL TENDENCIES
COM	A	0.006	-	0.038
	B	0.005	-	0.000
	C	0.000	-	0.000
TP	A	0.005	-	0.000
	B	0.005	-	0.000
	C	0.016	0.000	0.000

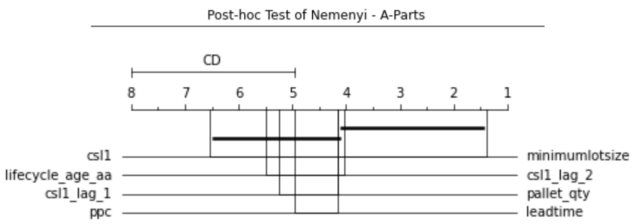


Figure 5.27 CD-Plot for Product-related Features – TP-A

... based on Elasticity

For COM, MF6 and MF7 reveal highest priority with WRI^{f6} = 0.239 and WRI^{f7} = 0.209. They represent the CSL in its actual and lagged version. With a slightly lower WRI^{f1} = 0.207, MF1 (ppc) is the runner up. MF5, the age of a SKU, is ranked fourth place. Within the TP sector, MF1 is assigned highest relevance with WRI^{f1} = 0.306. MF5 is ranked second place with WRI^{f5} = 0.281. MF6 and MF7 also matter whereas quantity constraints (e.g., pallet quantities and lot sizes) as well as the lead time have hardly any effect. The fact that MF6 and MF7 (CSL) is considered more important for COM may be partly due to

[48] The ranks derived from the frequency analysis are represented in Table E 35 in the appendix.

the existing competitive situation and confirms the results from the qualitative research represented in Chapter 5.1.2.

Table 5.15: Importance Testing for Product-related Features - Elasticity

		MF1	MF2	MF3	MF4
RI^f (WRI^f)		**'ppc'**	**'pallet_qty'**	**'minimumlotsize'**	**'leadtime'**
COM	A	0.246	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	0.249			
	C	0.146			0.000
	overall	0.177 (0.207)			0.000 (0.000)
		MF5	**MF6**	**MF7**	**MF8**
		'lifecycle_age_aa'	**'csl1'**	**'csl1_lag1'**	**'csl1_lag2'**
COM	A	0.211	0.227	0.165	eliminated due to Δ rank and mean rank > CD
	B	0.241	0.212	0.159	
	C	0.168	0.449	0.133	
	overall	0.187 (0.176)	0.377 (0.239)	0.144 (0.209)	
		MF1	**MF2**	**MF3**	**MF4**
		'ppc'	**'pallet_qty'**	**'minimumlotsize'**	**'leadtime'**
TP	A	0.246	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	0.285			
	C	eliminated due to Δ rank and mean rank > CD			0.000
	overall	0.246 (0.306)			0.000 (0.000)
		MF5	**MF6**	**MF7**	**MF8**
		'lifecycle_age_aa'	**'csl1'**	**'csl1_lag1'**	**'csl1_lag2'**
TP	A	0.245	0.304	0.202	eliminated due to Δ rank and mean rank > CD
	B	0.195	0.338	0.179	
	C	0.369	0.401	0.218	
	overall	0.149 (0.281)	0.479 (0.221)	0.121 (0.192)	

Based on the analysis it is MF1, MF5, MF6, MF7 and MF8 for the COM sector as well as MF1, MF5, MF6 and MF7 for the TP sector, that are to be used in the forecasting process.

Vehicles in Operation (VIO) and Failure Rates

The underlying demand in the aftermarket is mainly driven by a need for maintenance or repair of the finished product, in this case vehicles. According to Romeijnders et. al. (2012) as well as Kareem and Lawal (2015) maintenance respectively repair activities are either

- planned preventive for wear-and-tear parts, meaning that the replacement is time- or usage-caused, or
- unplanned corrective for both defective and wear-and-tear parts meaning that failure is mostly due to randomness.

However, the typification of a component as a wear-or-tear respectively defective part (cf. Chapter 3.3) does not provide much information on the expected life span from a technological point of view. To approach the answer, life span data, the failure rate, and based thereon failure probabilities need to be considered. These criteria are in turn dependent on other criteria.

VIO Data

For the spare parts, whose demands are to be modeled, a mapping exists that enables the assignment to all vehicles in which the component is or can be installed. The lowest level specifies the vehicles according to brand and model, powertrain system and motorization. Besides this, differentiations can also be made on segment[49] and sector[50] level in nine respectively three categories. Accordingly, the related vehicle population database comprises the number of cars, busses, and trucks on different aggregation levels per country and year. The failure behavior of this population, whether caused by non-recoverable defects, accident-related total losses or exports, and the resulting disappearance of the vehicles from the overall population is considered in the simulation.[51]

Furthermore, the vehicle population is of a different age. This results in the population of relevant vehicles in operation (VIO) including the age structure per country over the course of the years.[52]

Causal Factors

To correctly identify the linkage between a certain product, its sales and the VIO within a specified market and year, different causal factors (cf. Table 5.16) resulting in different scenarios need to be considered. The latter are illustrated in Table 5.17. [53]

Table 5.16: Causal Factors for Failure

	DEFECTIVE TP		WEAR-AND-TEAR COM
CAUSAL FACTOR	Early Failure	Randomness	Wear-and-Tear • material fatigue and age • environmental factors (temperature, humidity) • human factors (mileage, driving behavior)
FAILURE RATE	early-life-failure-rate	random-failure-rate	wear-out-failure-rate
FAILURE PROBABILITY	early-life-failure-probability	random-failure-probability	wear-out-failure-probability

[49] According to the KBA (2021) and the European Commission (2021), vehicle segments are to be categorized as follows: Micro Car, Small Car, Lower Middle Class, Middle Class, Upper Middle Class, Upper Class, Heavy Bus, Medium Truck (6-16t), and Heavy Truck (>16t).
[50] The vehicle sectors considered comprise Passenger Cars (PV), Light Commercial Vehicles (LCV) < 6 t, and Heavy Commercial Vehicles (HCV). Excluded are vehicles related to agriculture and forestry, construction, off-highway / All-Terrain Vehicles (ATV), as well as 2-3-wheelers and power sports.
[51] The simulation of VIO on country and yearly level follows three steps using a Monte Carlo simulation. First, a country's overall fleet size is determined per year. Second, populations within each sector and segment are generated, considering the respective drive system and motorization. Step three comprises the simulation of the respective vehicle life cycle under consideration of various events.
[52] Based on the analyses presented in Appendix 11, the aggregation level chosen is per country and year on *sector_segment_fueltype_age* level. More detailed as e.g. brand, model, etc. that are mentioned above will be neglected.
[53] Scenarios one to three are predominantly defined by the company-specific product portfolio.

Definition of Life Span, Failure Rate and Failure Probability

A part's *life span*[54] refers to the interval that an individual component of a system survives without the failure of its overall or partial functionality and without the need for repair. (Hugh 2022: 293ff) Here, interval can be related to specific time units like years or months, but also to other variables like mileage. *Failure* is when a particular component no longer fulfills its specified function, either due to the sudden total loss of a function or the gradual reduction in the fulfillment of a function. For the latter, the component is replaced condition-dependently, and only when a set limit, the so-called *failure criteria*, is exceeded.

Once the status for failure has sufficiently been defined, the next step is to consider the failure rate. It provides information about the reliability of a component and denotes the share of its failures in a defined period of time. According to Kosky et. al. (2020: 229-257), plotting the failures of a group of identical components for a specified period of time results in curves that could be approximated by one of the illustrations in Figure 5.28.

Type A corresponds to the classic failure rate pattern and can be regarded a combination of several different failure patterns. The so-called *bathtub curve* divides the technical life cycle into

- a phase of *early failures* (I), which are mostly caused by defective components, followed by
- the phase of *random failures* (II). Random failures are mostly caused by external influences and must be expected in any of the three phases. They follow the principle of historicity, i.e., that the life span of a component has no influence on its further failure behavior. With operating time increasing,
- the final phase of *aging failures* (III) begins, resulting in an upward trend for the failure rate. The most important technical, environmental, and human-based stress parameters are type of material respectively its composition, temperature, speed as well as life span. The wear occurred is usually eliminated by the component's replacement, and the life span starts again. (Wilker 2010) Hence, the challenge, which arises in here, is that the vehicle's age is not necessarily equal to its built-in products' age.

Despite being commonly used to model failure rates, several empirical studies have shown that in practice, assumptions based on the bathtub curve are only realistic in a very limited number. (Kareem and Lawal 2015: 69ff) In this respect, those of wear-and-tear parts are mainly attributed towards types A, B and C with sometimes very short phases I and II, whereas defective parts are rather represented by types D, E or F.

[54] In literature, *life span* is often also denoted as *service life*, *total time of failure* or *time-of-event (ToE).*

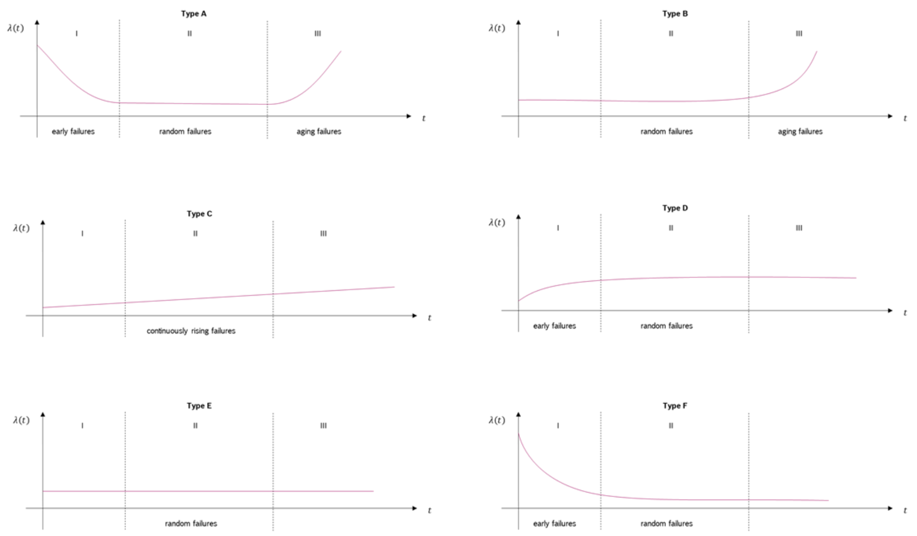


Figure 5.28 Failure Rates Modelling

Since neither the exact stress nor the exact stressability of a specific component are known in advance and can possibly only be determined posteriori, the question of whether a specific component will fail cannot be answered deterministically. One can only estimate the probability of failure; in other words, one can only determine which proportion of the population of identical components is likely to fail, but not whether a specific component will fail.
Data collections comprising stress parameters as well as failure rates or probabilities for spare parts from the automotive sector do currently not exist to the best of our knowledge. If available, measured, and predicted values of failure and reliability are derived from various manuals and differ by up to a factor of 100 due to

- newly evolved spare parts of different characteristics and configurations in regard to quality and reliability, making the underlying data being outdated or the fact, that
- influencing variables are not considered at all or are not considered realistically.
- Additionally, a fundamental objection arises from the fact that these calculations mostly operate with constant failure rates, assuming spare parts failure homogeneity both over time, making it therefore only cover random failures (cf. phase II of the bathtub curve or type E), and products. (Kareem and Lawal 2015: 69ff)

A statistical tool which addresses the question 'How long would it be, before a particular event i.e., the failure, occurs?' by considering the aforementioned shortcomings is the so-called *Survival* respectively *Time-to-Event Analysis*. (Hosmer et. al. 2008) Originally from medical domains, it is now also applied in customer and marketing analytics, actuaries, and predictive maintenance.

There are three main methods to assume the failure respectively survival curve: The *parametric* [55]approach comprises the assumption of a parametric model which is based

[55] Amongst the non-parametric approaches, it is the *Kaplan-Meier Estimator*, which is used to estimate the probability that a certain event will not occur for an object within a certain time interval, as well as the *Cox-*

on a specified distribution, its parameter estimation, and the final formation of the estimator of the survival function. Here, two- and three-parameter Weibull distributions are widely used to represent the respective strength distributions.

The *Weibull* distributional form was first derived through an extreme-value approach by Fisher and Tippett in 1928. As noted by Mann (1968), it became known as the *Fisher-Tippett Type III* distribution of smallest values or as the third asymptotic distribution of smallest (i.e., extreme) values. The findings are briefly outlined in the following:

An object can be divided into various components, with corresponding lifetimes $lt_1 \dots, lt_i$. If a series structure of the single components is assumed, the lifetime lt of the object is equal to the lifetime lt_i of its weakest element:

$$lt = \min(lt_1 \dots, lt_i)$$

The lifetime lt of the object thus corresponds to the lowest failure time (1st rank) of a sample of extent lt_i. If the procedure is repeated for several objects, the 1st rank will always be different – it will scatter. The distribution describing this scattering is called *extreme value distribution*, as the first rank size is an extreme rank. It can be shown that for $lt \rightarrow \infty$, the distribution always corresponds to the form of a *Fisher-Tippett Type III* (i.e., Weibull) distribution, independent of the individual elements' lifetimes' type of distribution.[56]

In its two-parameter form, the Weibull family is represented by PDF $f(lt)$ with

$$f(lt) = b \cdot ToE^{-1} \left(\frac{lt}{ToE}\right)^{b-1} e^{-\left(\frac{lt}{ToE}\right)^b} \tag{47}$$

$$\text{with } b > 0,\ ToE \geq 0$$

where ToE is the *time-of-event* (i.e., location) variable, b is the shape parameter and $\left(\frac{lt}{ToE}\right)$ is the inverse scale parameter. b is also known as the Weibull slope where specific values result in specific distributions:

- b = 1: exponential distribution
- b = 2: Rayleigh distribution
- b = 3.602: distributions with reduced skewness (~ Gaussian distribution)

A more common representation is CDF $F(lt)$ with

$$\begin{aligned} F(lt) &\coloneqq P(ToE < lt) = \int_0^{lt} f(lt)\,\mathrm{d}lt \\ &= 1 - e^{-\left(\frac{lt}{ToE}\right)^b} \end{aligned} \tag{48}$$

Proportional-Hazard Model, a regression model that assumes the behavior of failure rates as a function of different environmental factors.

[56] In 1939, Swedish scientist Waloddi Weibull (1939a) derived the same distribution as Fisher and Tippett for analyzing the strengths of materials. He gave several practical examples over the next decades (1939b, 1951, 1952), firmly establishing the name *Weibull* in the context of reliability analysis. (Evans et. al. 2019: 1)

and the thereof derived failure probability denoted $FP(lt)$ with

$$FP(lt) \coloneqq \int_0^{lt} f(lt)\,\mathrm{d}lt - \int_0^{lt-1} f(lt)\,\mathrm{d}lt. \tag{49}$$

For wear-and-tear parts, a certain time lt^0 must pass before a failure occurs. The three-parameter Weibull distribution, which is obtained after the transformation $lt \rightarrow lt - lt^0$, serves as a model for the statistical description of this phenomenon and is defined as follows:

$$F(lt) = 1 - e^{-\left(\frac{lt-lt^0}{ToE-lt^0}\right)^b} \tag{50}$$

with $F(lt) \cdot N$ giving the proportion p of population N with $ToE < lt$.

The Reliability Function is represented as

$$\begin{aligned} RF(lt) \coloneqq P(ToE \geq lt) &= \int_t^{\infty} f(lt)\,dlt \\ &= 1 - F(lt) \end{aligned} \tag{51}$$

with $RF(lt) \cdot N$ giving the proportion p of the total population N with $ToE > lt$.

Numerous methods of estimating Weibull parameters have been suggested in literature (Evans et. al. 2019: 1-8) and can generally be divided into four categories:

- *Method of Moments* according to Pearson (1894 and 1895)
- *Least Squares for Location- and Scale-dependent Distributions* according to Lloyd (1952 and 1962)
- *Percentile Estimation* according to Dubey (1967) and Zahakis (1979) as well as
- *Maximum-Likelihood Estimation* according to Fisher (1912 and 1921) and Halperin (1952)

Considering that besides the empirically derived variables

- *expected life span* in [km] denoted μ_{DtE_i} along with its standard deviation denoted σ_{DtE_i}, the
- *expected life span* in [years] denoted μ_{ToE_i} with $\mu_{ToE_i} = \frac{\mu_{DtE_i}+\sigma_{DtE_i}}{\frac{1}{V}\sum_{v=1}^{V} d^v}$ [57], also the
- expected failure probability denoted FP, being static, normalized to one and transformed to percentage values

are known for each product on PH1-level, and that completely recorded empirical life data is missing, the subsequent focus will be on the *Method of Moments* method of estimation:

[57] The annual average mileage d is formed over vehicles v, whose sector correspond. Further information on how the annual mileage d^v per vehicle type (sector_segment_fueltype_age) is derived is provided in Appendix 13.

The procedure consists of equating as many population moments $mo_{\tilde{k}} = \int_{-\infty}^{+\infty} lt^{\tilde{k}} \cdot f(lt)\,\mathrm{d}lt$ respectively central moments $mo_{\tilde{k}\ central}$ with $\tilde{k} = 1, 2, \dots$ such as

- for $\tilde{k} = 1$: the expected value or mean $\widehat{FP}(lt)$ denoted $mo_{1\ \mathrm{central}}$
- for $\tilde{k} = 2$: the variance $var(lt) = mo_{2\ \mathrm{central}}$, or
- for $\tilde{k} = 3$: the skewness $sk(lt) = mo_{3\ \mathrm{central}}$

to sample moments as there are parameters to estimate. Mathematical support for the dependence between parameters and moments is discussed in detail by Kendall and Stuart (1969), who state that two distributions which have a finite number of lower moments in common will be approximations of one another. Hence, the distribution of the data is approximated by equating the moments of a distributional form to the empirical moments.

Fitting parameter b via minimizing

$$MSE_i(\widehat{FP}_i, FP_i) = \left(\left(\frac{1}{ToE_i}\sum_{1}^{ToE_i}\widehat{FP}_i(lt)\right) - FP_i\right)^2$$

results in a BU- and PH1-comprising $b_{opt} = 2.96$ for Wear-and-Tear-Parts (COM). Determining the parameter PH1-wise, the range comprises values of [2.03; 5.21].[58]

By means of b_{opt}, dynamic failure rates $\lambda_i(lt)$ with $\lambda(lt) = \frac{f(lt)}{R(lt)} = \frac{\left[\left(\frac{R(lt)-R(lt+\mathrm{d}lt)}{dlt}\right)\right]}{R(lt)}$ are modeled for COM.

To consider the phenomenon that a product's life span starts again after phase III and to correctly model failure probabilities $\widehat{FP}_i(lt)$ over the entire life cycle of the vehicle population, relative probabilities are determined not only for the overall, but also for sub-populations. Accordingly, the combined probability $\widehat{FP}_i^{COM}(lt)$, indicating that for vehicle population N, a certain number of COM products i will be replaced at time t, is calculated via the probabilities for the first failure of product i, denoted $\widehat{FP}_{1i}(lt)$, lag shifts of life span ToE_i for defining the n-th failure of product i in vehicle v, denoted $\widehat{FP}_{ni}(lt)$, and the resulting conditional probabilities. Accordingly,

$$\widehat{FP}_i^{COM}(lt) = \widehat{FP}_{1i}(ToE_{1i}) + \widehat{FP}_{ni}(ToE_{ni}) + \dots + \widehat{FP}_{Ni}(ToE_{Ni}) \qquad (52)$$

results in a specified number of failed wear-and-tear spare parts at vehicle age lt: $P(lt) \cdot \widehat{FP}_i^{COM}(lt)$. For defective parts, the empirically determined constant failure rates $\lambda_i^{TP}$and probabilites $\widehat{FP}_i^{TP}$ are assumed per PH1.

$P(lt)$ is derived from the previously defined aggregation level (sector_segment_fueltype_age).

[58] The variable *time-of-origin* (ToO) and *time-to-event* (TtE) in [years] are derived via the vehicle-specific age and empirical research i.e., $ToE = ToO + TtE$. The maximum value for the vehicle-specific age is set at 19.

The resulting MF is called *vio_replaced_dyn* respectively *vio_replaced_stat.*

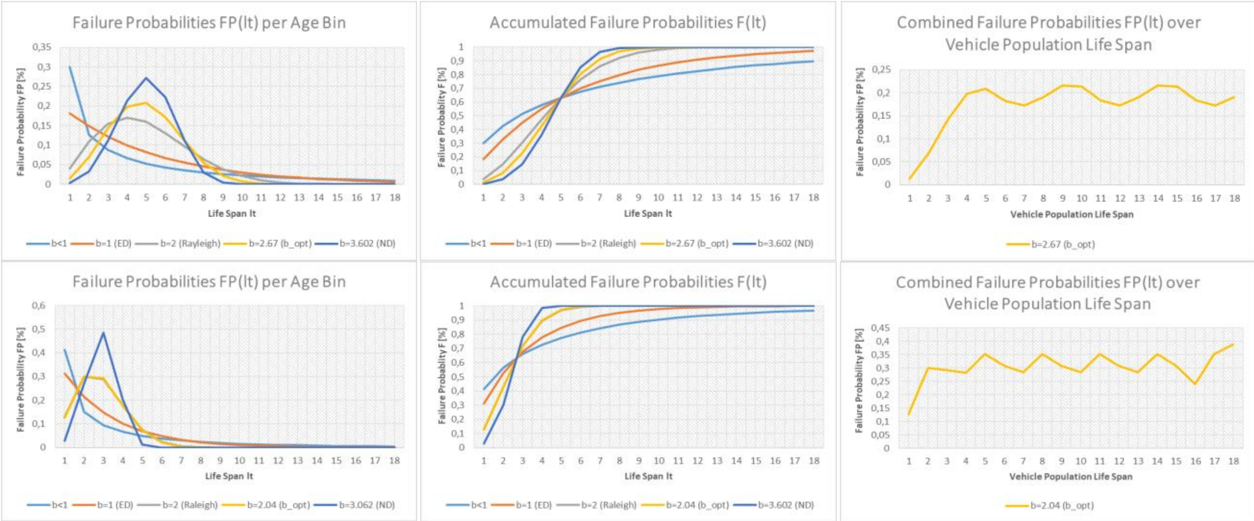


Figure 5.29 Failure Probabilities $\widehat{FP}_i^{COM}$ based on different b-values (Example)

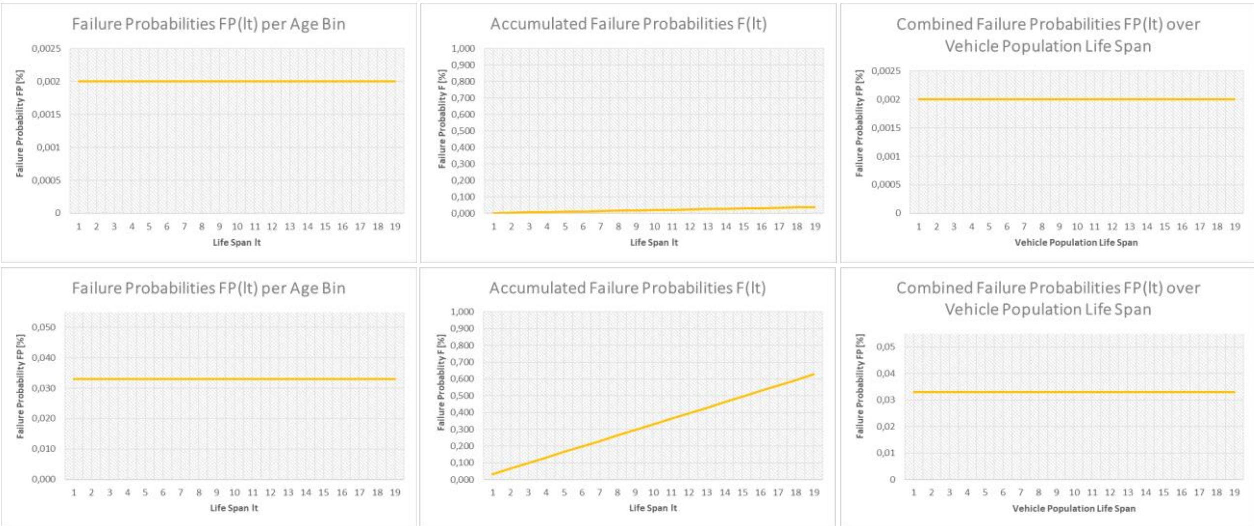
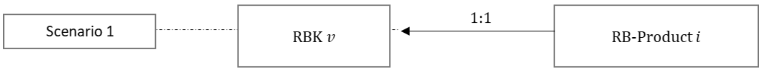


Figure 5.30 Failure Probabilities $\widehat{FP}_i^{TP}$ based on different b-values (Example)

Table 5.17: Scenario Development for Correlation Analysis of VIO and PMCs

INFLUENCING FACTOR 1: Company Portfolio

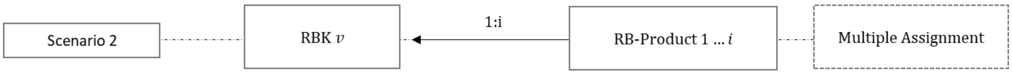
- Scenario 1: A specific product fits into a specific vehicle → 1: 1 relationship



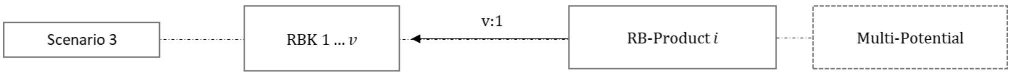
- Scenario 2: Several products fit into a specific vehicle, a phenomenon which is called multiple assignment. It is mostly due to high technical similarity between products and the different price segments available and triggers an internal competition → 1: i relationship.

  The multi-assignment probability is defined as

$$mult\_ass = \frac{1}{\mathrm{n_{unique}prod.\ per\ rbk\ within\ a\ specific\ PH3}} \tag{53}$$

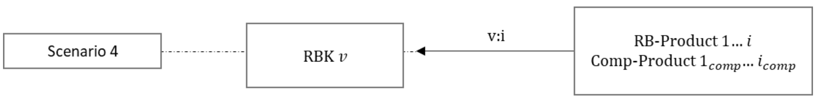


- Scenario 3: A specific product fits into several vehicles. This increases the sales potential for the single product. v: 1 relationship.



INFLUENCING FACTOR 2: Competitor Portfolio and Pricing

- Scenario 4: Besides the company-specific product portfolio, it is also the competitors' product range and pricing which have to be considered, as a certain number of components that are installed in the vehicles, are sourced from other market participants. For antitrust reasons, no or only very limited information is available regarding this subject. Therefore, estimates of market shares are used as an indicator of the relative strength of a supplier to the market.[59] It is defined per BU and country as well as per BU and region due to cross border sales.



- An additional variable to be considered is the so-called *Average Replacement Quantity*, specifying the average number of spare parts exchanged at a replacement process e.g., for spark plugs or brake pads.

VIO-related MFs are derived from the scenarios in Table 5.17 and hence five-fold. MF1 accumulates the number of assigned vehicles per product, split according to their segment, sector, fuel type and age. MF2 to MF5 all imply failure probabilities of the respective parts considering the vehicle type's mileage per country and year with MF3 to MF5 also adding the potential influence of average replacement quantities, multiple assignments and market share.

[59] Market shares need to be seen in the context of the definition of the market. Hence, figures are based on the markets' size on BU-, country- and distribution channel-level, and by expressing a company's revenue as a proportion of that total. As there may also be inaccuracies on an IAM country level due to cross border sales, country-comprising shares which amount for whole regions e.g., the total of Europe, are also provided.

Intra-Feature Correlation

The correlation analysis emphasizes the importance of dynamic failure probabilities: Apparently MF2 has stronger reference with r_{y,f_2}= 0.41 than the pure VIO variable with r_{y,f_1}= 0.32. For TP on the other hand, values for MF1 and M2 are on the same level.

Furthermore, it is striking for COM that the correlations to the dependent variable y and between MFs decrease significantly when taking into account multiple assignment factors as well as market shares, whereas there is no noticeable change for TP.

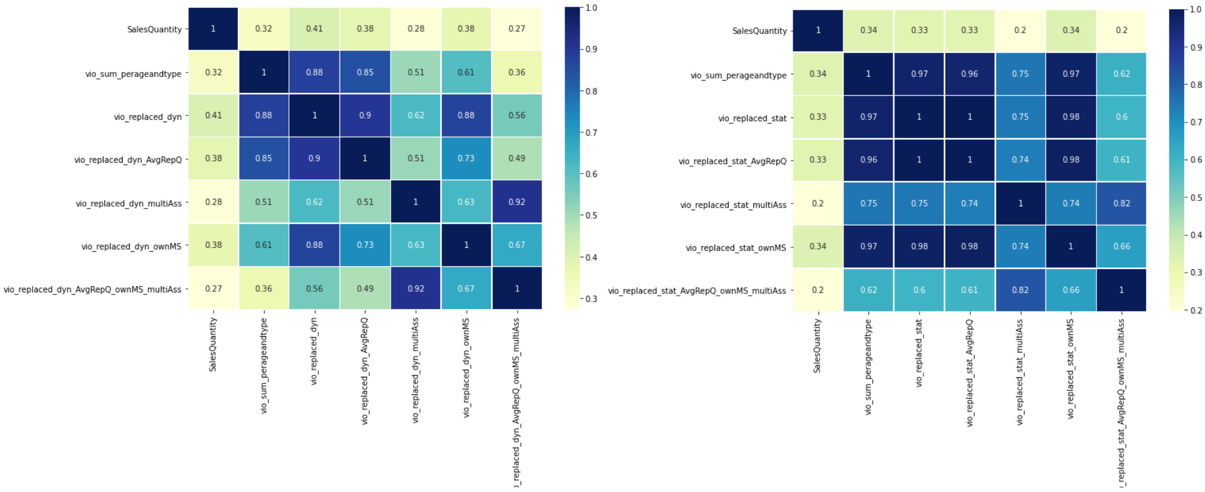


Figure 5.31 Correlation Testing for VIO-related Features

Importance and Feature Rank

For COM, the simulated dynamic failure probability is generally preferable to the static one. The importance score, which is determined by combining the VIO variable and the static failure probability deviates only slightly from that of MF1. (cf. Table 5.18 – COM) For TP, the static failure probability is a constant. Hence, importance scores are not affected. (cf. Table 5.18 – TP) However, similarly to MF3 (*AvgRepQ*) it allows better approximation of the absolute values. The quality of both constants can be considered positive since the information content of the composition of the vehicle population is not distorted. (cf. Figure 5.32)

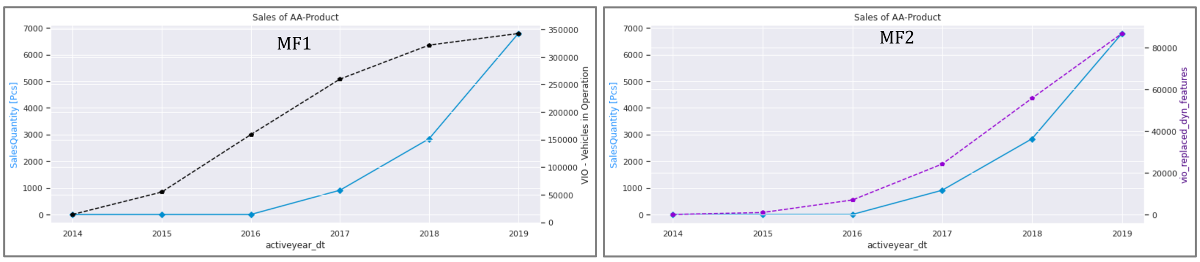


Figure 5.32 Approximation of VIO-related Features and PMC-Sales - COM (Example)

The factor value of the multiple assignment variable (MF4) only changes when the company's own product portfolio is enlarged or reduced. When including this variable, a distortion is caused, as the probability that a product is chosen over another one is not considered: The current assumption that part i from value contribution class A is purchased with the same probability as part $i + 1$ from class C is too simplified.

The use of the dynamic market share variable (MF5) also proves difficult. There are three reasons:

- Structural and organizational changes alter the composition of business units and product families.
- Transactions occur intracontinental and intercontinental.
- Values are based on human expertise, i.e. subjective and are hence not provable.
- Especially for COM, past and future evolutions are hard to estimate due to market consolidation and concentration processes (cf. Chapter 3.2 and Chapter 5.1.2).

Logically, there is a deterioration in information content for all three countries.

Table 5.18: Pre-Selection of VIO-related Features

RI^f				
COM		**MF1**		**MF2**
		'vio_sum_perage andtype'	**'vio_replaced_dyn'**	**'vio_replaced_stat'**
	overall	0.304	0.381	0.307
		MF1		**MF3**
COM		**'vio_sum_perage andtype'**	**'vio_replaced _dyn_AvgRepQ'**	**'vio_replaced _stat_AvgRepQ'**
	overall	0.300	0.386	0.307
COM		**MF1**		**MF4**
		'vio_sum_perage andtype'	**'vio_replaced _dyn_multiAss'**	**'vio_replaced _stat_multiAss'**
	overall	0.349	0.345	0.299
COM		**MF1**		**MF5**
		'vio_sum_perage andtype'	**'vio_replaced _dyn_ownMS'**	**'vio_replaced _stat_ownMS'**
	overall	0.351	0.345	0.296
TP		**MF1**	**MF2**	
		'vio_sum_perage andtype'	**'vio_replaced_stat'**	
	overall	0.495	0.493	
TP		**MF1**	**MF3**	
		'vio_sum_perage andtype'	**'vio_replaced _stat_AvgRepQ'**	
	overall	0.485	0.502	
TP		**MF1**	**MF4**	
		'vio_sum_perage andtype'	**'vio_replaced _dyn_multiAss'**	
	overall	0.504	0.483	
TP		**MF1**	**MF5**	
		'vio_sum_perage andtype'	**'vio_replaced _stat_ownMS'**	
	overall	0.505	0.483	

As a conclusion, one could assume that MFs, that represent VIOs and their corresponding age structure as well as the probability of product failure, add more value to the model than MFs that additionally consider internal and external market shares.

Thus, the feature space for the subsequent analysis is as follows:

... based on Frequency

For COM-A and COM-B, we fail to reject $H0$ of the Shapiro-Wilk test that the population is Gaussian distributed for all populations. Therefore, it is assumed that those populations are Gaussian distributed. Based on Bartlett’s test for homogeneity, it is assumed that the data related to these classes is heteroscedastic. The same applies for TP. For C-parts from COM and TP, we reject $H0$ of the Shapiro-Wilk test and hence assume that not all populations are Gaussian distributed.

Based on the results from the non-parametric Friedman/post-hoc Nemenyi test, we assume that there is no statistically significant difference between the mean values of the populations across product types and value contribution classes.[60] Hence, for all MFs, the prediction importance scores WRI^{f} are calculated. On behalf of those, the final prioritization and transfer of the MFs to the ML models are realized.

Table 5.19: Importance Testing for VIO-related Features - Frequency

min. observed p-values for $C = 8$		NON- / GAUSSIAN DISTRIBUTED	HOMOSKEDASTIC / HETEROSKEDASTIC	ROBUSTNESS of CENTRAL TENDENCIES
COM	A	0.034	0.000	0.693
	B	0.035	0.000	0.876
	C	0.005	-	0.174
TP	A	0.036	0.000	0.343
	B	0.039	0.000	0.430
	C	0.003	-	0.540

… based on Elasticity

For COM, MF1 is ranked highest with WRI^{f1} = 0.194. It supersedes the other MFs by a large margin. Age bin based MFs 3 and 2 are ranked second and third with WRI^{f3} = 0.139 and WRI^{f2} = 0.134. The least relevant ones are the 'older' vehicles and thus MF8 and MF7 with WRI^{f8} = 0.107 and WRI^{f7} = 0.076. The ranking for MFs related to TP is different: Here age-bin based MFs are preferred over MF1, i.e., that the information gain from the static failure probability is limited. Importance scores range from WRI^{f5} = 0.223 to WRI^{f7} = 0.039. Lowest values are again assigned to 'older' vehicles.

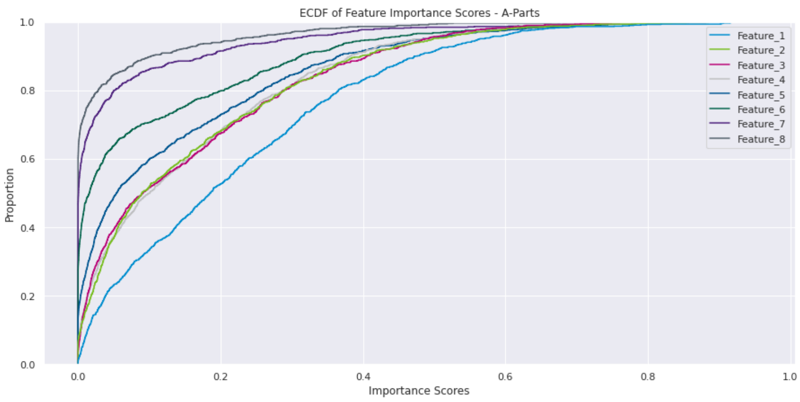


Figure 5.33 EDCF of Importance Scores for VIO-related Features- COM-A

[60] The ranks derived from the frequency analysis are represented in Table E 37 in the appendix.

Table 5.20: Importance Testing for VIO-related Features - Elasticity

		MF1	**MF2**	**MF3**	**MF4**
RI^f (WRI^f)		**'vio_replaced_dyn'**	**'vio_agebin_0-2'**	**'vio_agebin_3-5'**	**'vio_agebin_6-8'**
COM	A	0.215	0.160	0.155	0.157
	B	0.197	0.120	0.123	0.145
	C	0.172	0.039	0.060	0.094
	overall	0.180 (0.194)	0.068 (0.134)	0.103 (0.139)	0.106 (0.110)
		MF5	**MF6**	**MF7**	**MF8**
		'vio_agebin_9-11'	**'vio_agebin_12-14'**	**'vio_agebin_1516'**	**'vio_agebin_>16'**
COM	A	0.134	0.097	0.048	0.033
	B	0.159	0.122	0.076	0.059
	C	0.140	0.196	0.167	0.133
	overall	0.155 (0.118)	0.161 (0.122)	0.118 (0.076)	0.104 (0.107)
		MF1	**MF2**	**MF3**	**MF4**
		'vio_replaced_stat'	**'vio_agebin_0-2'**	**'vio_agebin_3-5'**	**'vio_agebin_6-8'**
TP	A	0.169	0.119	0.112	0.144
	B	0.184	0.099	0.093	0.159
	C	0.207	0.061	0.081	0.117
	overall	0.207 (0.151)	0.073 (0.135)	0.061 (0.096)	0.137 (0.122)
		MF5	**MF6**	**MF7**	**MF8**
		'vio_agebin_9-11'	**'vio_agebin_12-14'**	**'vio_agebin_1516'**	**'vio_agebin_>16'**
TP	A	0.190	0.128	0.083	0.057
	B	0.170	0.142	0.090	0.065
	C	0.159	0.156	0.122	0.098
	overall	0.166 (0.223)	0.155 (0.162)	0.093 (0.039)	0.094 (0.071)

All eight MFs are transferred to the ML models for determining their prediction importance. Additionally, the constant variable $AvRepQ$ will be considered.

5.2.2.4 Domain D_3: Pricing, Price Variablity and Target Agreements

This chapter examines price volatilities in the AA and tests the hypothesis if pricing and price adaptation related factors could serve as explaining variables for demand variability.

If so, the subsequent objective is twofold: Firstly, the ML models need to learn how and when prices change. This is directly related to the question 'How does the pricing variable evolve over time?' Speaking of pricing variables, it is not just the evolvement of the

- price for the product, which is in focus, but also the
- intra-categorical price, i.e., the price of related products from the portfolio (either substitutes or complementaries) as well as the
- inter-categorical price i.e., the price of non-related products from the portfolio

which need to be taken into consideration.

Secondly, the models need to learn the effect of price changes on demand. This is equivalent to the question 'How effective are price adaptations?'. Henry L. Moore answered this question in his early paper from 1922 by determining the *Coefficient of Flexibility* (Mills 1946: 439ff), a variable which is today commonly referred to as the *Elasticity of Demand*.

It is defined as the ratio of the relative change in the price per unit to the corresponding relative change in the quantity when the relative changes are infinitesimal. Own-price (point) elasticities (54) refer to the percentage change in demand of a specific product divided by the percentage change in its costs.

The second type of elasticities, cross-price (point) elasticities (55), refer to the percentage change in demand of a specific product divided by the percentage change in the costs of a second – be it complementary or substationary.

$$\eta_i = \frac{\frac{q_i - q_{i+1}}{q_i} \cdot 100}{\frac{pr_i - pr_{i+1}}{pr_i} \cdot 100} = \frac{\Delta q_i[\%]}{\Delta pr_i[\%]} \tag{54}$$

$$\eta_i = \frac{\frac{q_i - q_{i+1}}{q_i} \cdot 100}{\frac{pr_j - pr_{j+1}}{pr_j} \cdot 100} = \frac{\Delta q_i[\%]}{\Delta pr_j[\%]} \tag{55}$$

As pricing variables are not Gaussian distributed and hence do not meet certain assumptions of parametric statistical tests such as t-test, ANOVA, or linear regression, transformation is required to avoid misleading results. Logarithmization is a data transformation method for non-negative and non-zero values, in which each variable is replaced with a log(variable). Hereby, the log-transformation reduces or removes the skewness of the original data. Hence, rules for calculating elasticities of demand are as follows:

Table 5.21: Log-Transformation Rules to Determine Price Elasticities

Transformation	Function	Elasticity	
Level-Level	$pr_i = \beta_0 + \beta_1 q_i$	$\eta = \beta_1 \cdot \frac{q_i}{pr_i}$	(56)
Log-Level	$\log(pr_i) = \beta_0 + \beta_1 q_i$	$\eta = \beta_1 \cdot q_i$	(57)
Level-Log	$pr_i = \beta_0 + \beta_1 \cdot log(q_i)$	$\eta = \frac{\beta_1}{pr_i}$	(58)
Log-Log	$\log(pr_i) = \beta_0 + \beta_1 \cdot log(q_i)$	$\eta = \beta_1$	(59)

Own-price elasticities

For most goods, demand moves on the contrary direction of its costs, which means that $\Delta_{pr_i} > 0$ (resp. $\Delta_{pr_i} < 0$) results in $\Delta_{q_i} < 0$ (resp. $\Delta_{q_i} > 0$) and finally in $\eta_i < 0$. In case of $\eta_i > 0$, the product is referred to as a g*iffen good*, i.e., that customers buy larger volumes as the price rises and vice versa. (Wöhe and Döring 2010: 449ff)[61]

For price elasticities being greater than one i.e., $|\eta_i| > 1$ i.e., $\eta_i < -1$ and occasionally $\eta_i > +1$, price changes will lead to disproportionate changes in customer spendings and in corporate revenues. This is referred to as a *price-elastic demand response*, i.e., that the absolute percentage change in quantity demanded exceeds the absolute percentage change in price. If, in contrast, price elasticities are less than one ($|\eta_i| < 1$), the demand response is *price inelastic*. Price elasticities for most products cluster in-between [-1; 0].

Cross-price elasticities

For $\eta_i < 0$, the quantity demanded of one good decreases, when the price of another increases. This is most often valid for complements. A positive cross-price elasticity ($\eta_i > 0$), i.e., the quantity demanded of one good increases with the rising price of another, indicates substitution goods.

Subsequently, there is a brief presentation of the analytical results on own-price elasticities of demand in the aftermarket business according to Equation (59).[62] The two variables examined are *IPPinLc_perPc* and *IPPinGc_perPc*.[63]

Table 5.22: Share of Estimated Price Elasticities of Demand

		IPP in LC [%]	IPP in GC [%]
p<0.05	COM	10.33	19.55
	TP	14.84	23.24
thereof			
elasticity	COM	42.42	90.53
$\eta_i < 0$	TP	54.89	73.64
proportionality	COM	56.00	74.74
$\eta_i < \|1\|$	TP	23.43	42.22

[61] $\eta_i = 0$: perfectly inelastic demand i.e., the price has no effect on the quantity demanded; $\eta_i = 1$: unit-elastic demand i.e., changes in price have no effect on the quantity demanded.

[62] The market chosen for the analysis is a *low-promotion* market, i.e., B2B customers are offered lower standard prices throughout the year in preference to temporary price reductions and peak orders via selective special promotions. This means that the majority of price variability can be attributed to changes in exchange rates.

[63] To conclude whether changes in the predictor value are related to changes in the response variable, we either reject or confirm $H0$ with $H0$: the coefficient is equal to 0, i.e., the price has no effect and $H1$: the coefficient is smaller or larger than 0, i.e., the price is affecting demand.

In general, both variables are found to have a significant impact on PMC level when used as a single regressor (cf. Table 5.22). BU-wise, values are in the following ranges:

- IPP in LC: [5.98-14.84]
- IPP in GC: [19.55-23.24]

Prices in global currency (GC) are regarded more relevant than prices in local currency (LC). While IPPinGC has a significant influence as a single regressor for about 20 percent of COM and for more than 23 percent of TP, IPPinLC has about 10 percentage points less.

It is majorly negative own-price elasticities that are identified for *IPPinGC_perPc*. With about 90.53 percent, the share is more prominent for COM than for TP with an average of 73.64 percent. Consequently one could conclude that TP are also bought when the price increases and this more often than COM. BU-wise, values are in the following ranges:

- IPP in LC: [30.17-56.24]
- IPP in GC: [62.04-93.64]

For *IPPinLC_perPc*, the share of negative elasticities is significantly lower.

Changes in *IPPinGC_perPc* do not lead to a disproportionate reaction for the majority of COM with about 75 percent. This means that for the majority of products a price reduction leads to an increased demand, but not at the extent of the price change or above respectively vice versa. For approximately 25 percent, a price change leads to significantly increased or decreased demand. In general, the share of products that are price-elastic is higher for TP with about 58 percent. BU-wise, values are in the following ranges:

- IPP in LC: [20.45-57.56]
- IPP in GC: [40.04-85.45]

Based on the above results and conclusions, two types of influential factors need to be determined:

- the ones that cause price variability and
- the ones that affect the price elasticity of demand i.e., the covariates of prices

It is amongst others Wöhe and Döring (2010) who list five main factors, company-induced and external ones, causing price variability:

One of the first factors named, affecting the price of a product, is its cost and the related price definition strategies. It refers to the total of fixed, variable, and semi-variable costs incurred during the production, distribution and selling of the product. Pricing strategies refer to the methods companies use to list price their products or services. Here, researchers mostly agree that they can be categorized into three big groups:

- cost-based i.e., related to product costs especially resources,
- value-based i.e., related to brand, product quality and resources and
- competition-based i.e., related to country specifics, the nature of the market and market shares (Toni et. al. 2017). According to Hoch et. al. (1995), it is well known that sales of the focal product are subject to the negative impact of the prices and price reductions of competitor products, which also exacerbates the forecasting problem.

One more group, the target-based pricing strategy i.e., related to margin and profit maximization, is added by Seidenschwarz (2003: 437-453).

Following the definition of regional list prices, it is marketing objectives and measures, in here pricing and distributional instruments like on- and off-invoice discounts, packaging and transportation costs, which reduce list prices to invoice and ultimately net prices.

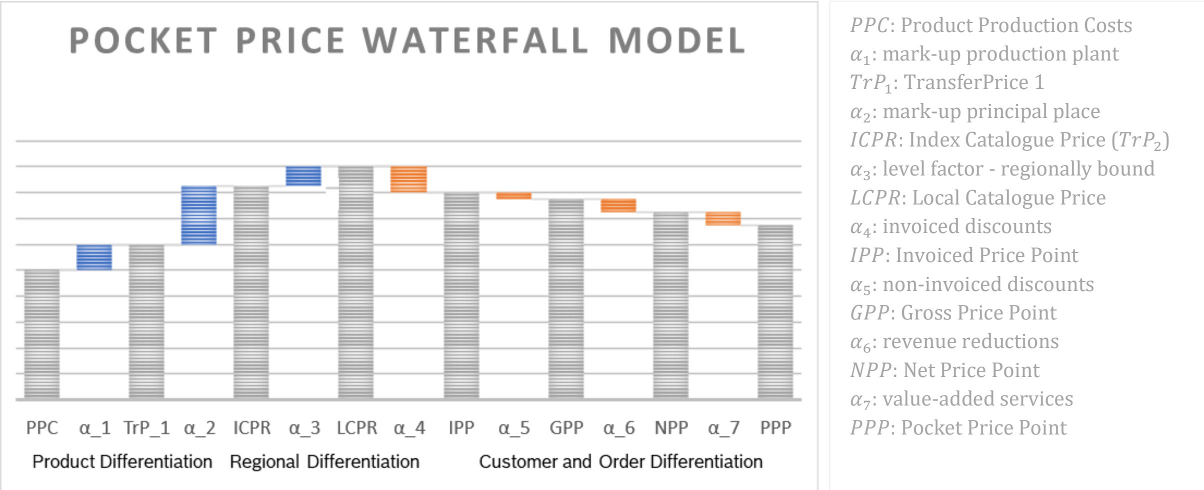


Figure 5.34 B2B Price Variability and Management based on Price Waterfall Model

The fourth, an external factor, that affects product costs, list prices and selling prices equally, is currency fluctuations. It basically determines at which value a shopping cart from one economy is exchanged for the identical cart from another economy. Hereby, the currency's strength is always dependent on the stability of the respective economy and vice versa.

Very specific knowledge with regards to country risks, market situations, etc. can lead to an increased understanding of exchange rate fluctuations and thus to a conditionally valid assessment of their trend. Their exact occurrence, extent, and persistency, however, remain mostly unknown. Hence, forecasted exchange rates as proposed by Lessard and Lorange (1976) are indeed difficult to determine and often do not correspond to reality.[64]

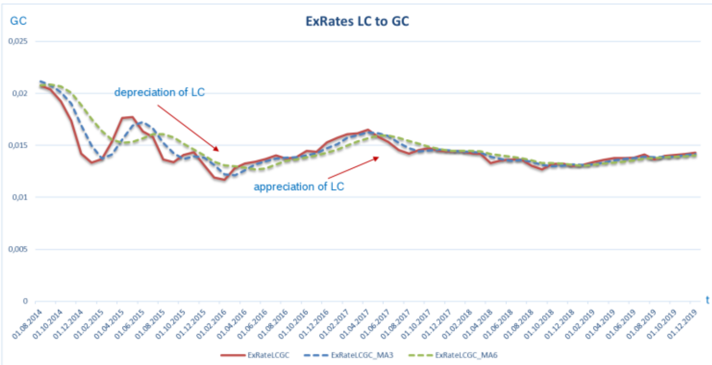


Figure 5.35 Exchange Rate Fluctuations and their Smoothed Versions

[64] Demirag (1992) instead proposes the calculation of an average rate. These are advantageous as short-terms fluctuations can be offset.

For this reason, it is considered more important to envision the prevailing scenarios, and to trace the reactions of the buying and selling parties triggered by currency fluctuations rather than the development of the currencies themselves.[65] Generally, there are two scenarios:

- Scenario 1: Export Business – Global Sourcing
  The selling party is not located in the sales market. Henceforth, sales are settled in the global currency (GC), making the currency fluctuations affect exporting and importing parties directly.

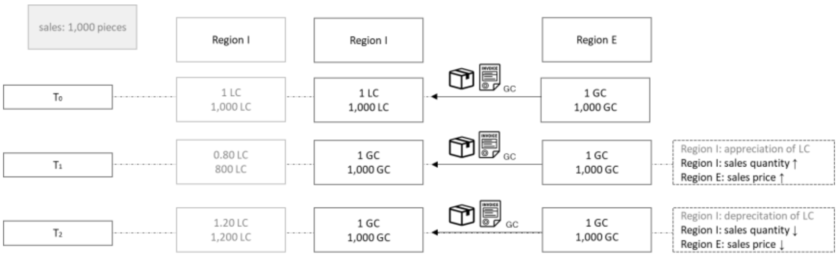


Figure 5.36 Pricing in Global Sourcing Transactions

T1: Appreciation of the Local Currency (LC)

T1.1: The importing party receives the same volume (cf. T0) for a reduced amount of money or more volume for the same amount of money (cf. T0).

T1.2: To seize currency effects, sales prices [in GC] are increased by the exporting party i.e., the same volume (cf. T0) is sold for the same amount of money (cf. T0) or more volume is sold for a higher amount of money.

T2: Depreciation of the Local Currency (LC)

T2.1: The importing party receives the same volume (cf. T0) for a higher amount of money or less volume for the same amount of money (cf. T0).

T2.2: To compensate currency effects, sales prices [in GC] are decreased by the exporting party i.e., the same volume (cf. T0) is sold for a reduced amount of money or more volume is sold for the same amount of money (cf. T0).

- Scenario 2: Local to Local – Local Sourcing
  Subsidiaries of the selling party are located in the sales market. Hence, sales are settled in the local currency, making the currency fluctuations affect both parties indirectly.

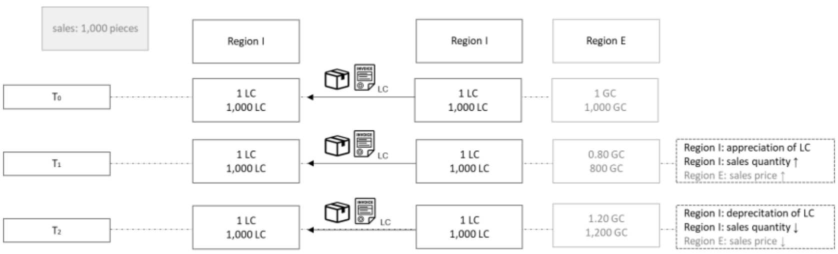


Figure 5.37 Pricing in Local Sourcing Transactions

[65] *Conditioning*: According to market-specific agreements, the exchange rates are frozen either at customer order date or at billing date. For the present dataset and related markets, option one is valid.

T1: Appreciation of the Local Currency (LC)

T1.1: The importing party receives the same volume (cf. T0) for the same amount of money (cf. T0) from the subsidiary.

T1.2: To compensate currency effects i.e., to prevent the appreciation of the local currency resulting in a lower amount in the global currency and to achieve fair performance appraisal, sales prices [in GC] are increased, or in case of short-term fluctuations held stable. This means the same volume (cf. T0) is sold for the same amount of money [in GC] or more volume is sold for a higher amount of money [in GC], whereas the buying party is consequently charged with increased prices [in LC].

T2: Depreciation of the Local Currency (LC)

T2.1: The importing party receives the same volume (cf. T0) for the same amount of money from the subsidiary.

T2.2: To compensate currency effects in terms of decreasing sales, sales prices [in GC] are reduced, or in case of short-term fluctuations held stable. This means the same volume (cf. T0) is sold for the same or a lower amount of money.

Thereof two hypotheses result:

$H1$: Short-term fluctuations have no or only little effect in forms of ad-hoc orders, which can neither be planned nor forecasted.

$H2$: Only in case of the exchange rate being subject to a continuously i.e., sustained rising or falling trend, pricing and hence numbers and volumes of orders change:

- Scenario 1: Export Business – Global Sourcing

T1: Longer-term Appreciation of the Local Currency (LC)

Goods become cheaper for the buying party and the anticipation of price increases [in GC] result in rising numbers of orders and volumes before the price adjustments.

T2: Longer-term Depreciation of the Local Currency (LC)

Goods become more expensive for the buying party and the anticipation of price reductions [in GC] result in decreasing numbers of orders and volumes before price adjustments.

- Scenario 2: Local to Local – Local Sourcing

T1: Longer-term Appreciation of the Local Currency (LC)

The anticipation of price increases [in GC] results in rising numbers of orders and volumes before price adjustments

T2: Longer-term Depreciation of the Local Currency (LC)

The anticipation of price reductions [in GC] results in decreasing numbers of orders and volumes before price adjustments.

The question which can be derived from these hypotheses is as follows: Are exchange rates suitable as predictors for prices (in [GC] and in [LC]) and their variabilities, and consequently also for corresponding changes in the buying behavior of the customers?

For this purpose, basic *Granger Causality* is applied. Originally proposed in 1969, it determines whether one times series is useful in forecasting another.[66]

Firstly, it is used to identify the best possible exchange rate feature affecting *IPPinGC_perPc* as well as the final dependent variable Y_t i.e. the *SalesQuantity*. According to the analysis, *ExRateLCGCMA2cent* (MF1) shows highest influence with about 70 percent (cf. Table 5.23 and Table 5.24). Secondly, from the best possible exchange rate feature, the percentage changes (MF2) are determined.

Finally, the buying behavior in C3 will be compared to the exchange rate evolution both with regards to the actual values and the percentage change of the best possible exchange rate feature.

Table 5.23: Granger Causality of Exchange Rates and IPPinGC_perPc

IPPinGC_perPc C_t	Granger_Causality _flag	Granger_share [%] lag_opt = Lag4[67]
'ExRateLCGC'	False	32.705
'ExRateLCGC'	True	67.295
'ExRateLCGCMA3'	False	33.716
'ExRateLCGCMA3'	True	66.284
'ExRateLCGCMA6'	False	32.284
'ExRateLCGCMA6'	True	67.717
'ExRateLCGCMA2cent'	False	29.787
'ExRateLCGCMA2cent'	**True**	**70.213**
'ExRateLCGCMA3cent'	False	31.873
'ExRateLCGCMA3cent'	True	68.128

Table 5.24: Granger Causality of Exchange Rates and SalesQuantity

Y_t = SalesQuantity C_t	Granger_Causality _flag	Granger_share [%] lag_opt = Lag4
'ExRateLCGC'	False	70.598
'ExRateLCGC'	True	29.402
'ExRateLCGCMA3'	False	69.381
'ExRateLCGCMA3'	True	30.620
'ExRateLCGCMA6'	False	70.203
'ExRateLCGCMA6'	True	29.797
'ExRateLCGCMA2cent'	False	69.335
'ExRateLCGCMA2cent'	**True**	**30.666**
'ExRateLCGCMA3cent'	False	70.598
'ExRateLCGCMA3cent'	True	29.402

[66] Basic *Granger Causality* is quite simple: Given the two terms Y_t, and C_t, Y_{t+1} is forecasted using past terms of Y_t and C_t. Also, Y_{t+1} is forecasted based on past terms of Y_t. If the first approach results in more accurate forecasts according to standard cost functions, it is assumed that C_t contains valuable information in forecasting Y_{t+1} which is not provided in the past values of Y_t. Hence, C_t is said to *Granger-cause* Y_{t+1} if the cause C_t happens prior to its effect seen in Y_{t+1} and if the cause C_t has unique information about the future values of its effect, that is not found in the past values of Y_t. Any lagged value of C_t is retained in the regression if first of all it is significant according to a t-test, and secondly it and the other lagged values of the variable jointly add explanatory power to the model according to an f-test. Hypotheses are as follows: $H0$: C_t does not Granger-cause time series Y_t. $H1$: C_t Granger-causes time series Y_t.

[67] Lag_opt is identified via time-lagged cross correlation.

Besides exchange rates, external factors also comprise the factor 'general demand and utilization' and related to this the customers' willingness to pay. The third group of external factors is related to government actions as well as social and economic concerns. (Wöhe and Döring 2010: 472)

For the covariates of pricing i.e., the factors that affect price elasticity of demand, there are eight general aspects, that could easily be transferred to the AA use case:

The availability of alternatives is the first to be named. With a higher number of substitute products or services being offered, the demand response becomes more price elastic; a small increase in the price levels of the focal goods may cause customers to buy its substitutes, be it

- from the company's portfolio,
- possibly from a different price segment with the segments available being LPP (low price point), MPP (mid price point) and HPP (high price point),
- from the company's competitor or
- in a neighboring country.[68]

An assessment of which respectively how many alternative products are offered in the company portfolio is determined based on the same logic as the multi-assignment variable in Chapter 5.2.2.3. It represents the depth of a specific product line and comprises all available products from the low, mid, and high price point segment.

Instead of estimating the number of competitors and competing products in the individual markets, the goods are subdivided according to the prevailing market form: Similar to the automotive industry, in which fourteen firms globally control more than 60 brands, the aftermarket business is also assigned to the oligopolistic market. Still, there is a clear divergence between the two product types COM and TP, with COM exhibiting polypolistic and TP monopolistic traits. The barrier for market entries is very high for TP; second and third sourcing is limited. Therefore, not only the customer's negotiating power but also the number of substitutes is severely limited. On the other hand, market dynamics in terms of competitors are much greater for COM than for TP. Here, bargaining power of customers and the choice for alternatives increases with the competition amongst suppliers. Also, the number of alternatives must always be considered in conjunction with the price offered by competitors and in neighboring regions. (cf. Chapter 3.2)

Besides the depth of the portfolio, it is also the breadth of the portfolio i.e., the product mix respectively the product lines a company has to offer, that matters. Cross-selling opportunities often enable sellers and buyers to take advantage of profitable economies of scale, which may then also offset price increases for certain products. It is hardly possible to represent the breadth of the portfolio using a single feature. For this reason, in addition to local models, global models are also tested, which are provided with the overall information collected from the portfolio. (cf. Chapter 3.3)

The third factor is related to the classification of the product as a necessity or luxury. When a good or service is a luxury or a comfort good, the demand response is highly price-sensitive i.e., price-elastic. Conversely, the demand for an essential good is generally price-

[68] The two MFs *RRP_neig* and *RRP_comp* will be dismissed for further analysis as historized data only comprises thirteen months.

inelastic because customers keep on buying it even if the price increases. In the following, the definition of a product as a luxury good or necessity is determined via the rotation rate. The rotation of products and services is the speed at which the same are sold. The higher the number of orderlines for a product per year, the higher the rotation of goods. Based on this procedure, the following categories result:

- Fast mover and slow mover label derivation based on the logic represented in Salinas et. al. (2020) and defining a quantile threshold at 0.75.
- Regular (R), irregular (I) and sporadic (S) items based on the coefficient of variance calculated for the sales quantity with threshold values 0.6 and 1.2

Fourth, the budget available respectively the customers' purchasing power affects elasticities. To consider firmographics and preferences of each customer segment, their purchasing power as well as the purchasing power of the top 3 customers per sales country are determined as a MF (cf. Chapter 5.2.2.2).

A fifth factor which is also directly connected to a product's type is brand and loyalty. The higher the brand awareness, the lower the elasticity. Brand awareness will be measured by the average number of unique customers per product and region over time (cf. Chapter 5.2.2.2).

The sixth one is the time elapsed since a change in price has happened. This factor however is strongly dependent on the product itself. Normally it holds true that the greater the time lapse since a price change, the more elastic the demand response. This factor can be equally represented by the products' single prices or the prices' percentage change.

The two final factors to be named are tariffs and communicational instruments i.e., non-monetary campaigns. Tariffs are charges imposed on imported goods. For the present use case, however, they are not at the expense of the customers in the affected sales market but compensated for by a corresponding Incoterm or an adaptation of the LCPR. Hence, they will remain unconsidered in the subsequent analysis. So will be the non-monetary campaigns.

Intra-Feature Correlation

Neither exchange rates nor price-related MFs are found to be significantly related to the dependent variable. However, a strong correlation can be found within the group of the latter. Coefficients for MF3, MF4 and MF5 exceed values of 0.9. Also highly correlated are the *IPPinGC_perPC* (MF5) and all downstream prices (MF6 and MF7) from the price waterfall. The values determined are as follows:

- $r_{f_3,f_4} = 0.96, r_{f_3,f_5} = 0.91, r_{f_4,f_5} = 0.99$
- $r_{f_4,f_7} = 0.84, r_{f_4,f_8} = 0.84, r_{f_4,f_9} = 0.84$
- $r_{f_7,f_8} = 0.90, r_{f_7,f_9} = 0.90$

Both MF1 and MF2 (exchange rate related) and MF10 and MF11 (marketing measure related) can said to be not correlated to any of the other MFs.

Based on the results, MF3, MF4, MF8 and MF9 are eliminated for subsequent evaluations.

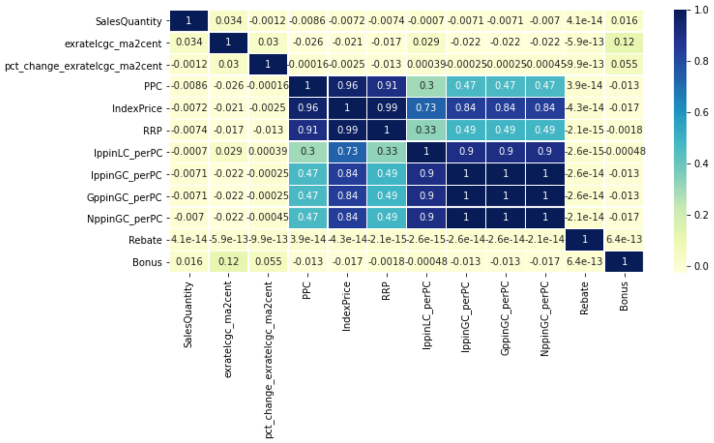


Figure 5.38 Correlation Testing for Pricing-related Features

Importance and Feature Rank

… based on Frequency

For COM, we fail to reject $H0$ that the population is Gaussian distributed for all populations. Therefore, we assume that all populations are Gaussian distributed. Based on Bartlett's test, we assume that our data is heteroscedastic. The same applies to TP-A. For TP-B and TP-C we reject $H0$ that the population is Gaussian distributed for the population of *IPPinGC_perPC*. Therefore, we assume that not all populations are Gaussian distributed. Based on the results from the non-parametric Friedman/post-hoc Nemenyi test, we assume that there is statistically significant difference between the mean values of the populations across product types and value contribution classes.[69] For COM as well as TP-C, MF5 and MF10 are eliminated. For TP-A and TP-B it is only MF5.

Table 5.25: Importance Testing for Price-related Features - Frequency

min. observed p-values for $C = 9$		NON- / GAUSSIAN DISTRIBUTED	HOMOSKEDASTIC / HETEROSKEDASTIC	ROBUSTNESS of CENTRAL TENDENCIES
COM	A	0.012	0.000	0.001
	B	0.009	0.000	0.000
	C	0.012	0.000	0.003
TP	A	0.006	0.000	0.001
	B	0.005	0.000	0.002
	C	0.004	0.000	0.002

… based on Elasticity

For COM, MF7 and MF1 are ranked highest with WRI^{f7} = 0.263 and WRI^{f1} = 0.251. The relevance of MF7 is expected as the invoice price is the price that symbolizes the agreement between the seller and the first-level customer in monetary terms. MF6 (IPPinLC_perPC) and MF3 (*PPC*) are ranked second highest with WRI^{f6} = 0.217 und WRI^{f3} = 0.142; MF2 and M11 with WRI^{f2} = 0.122 and WRI^{f11} = 0.030 are at rank 5 and

[69] The ranks derived from the frequency analysis are represented in Table E 39 in the appendix.

6. For TP, again MF1 and MF7 are most relevant. These are followed by MF6, MF3, MF11 and MF2. Hence, six distinct MFs representing currency fluctuations, production costs, invoice prices in [GC] and [LC] as well as bonus information will be transferred to the ML model for further analysis.

Table 5.26: Importance Testing for Price-related Features - Elasticity

		MF1	**MF2**	**MF3**	**MF4**
RI^f (WRI^f)		**'exratelcgc_ma2cent'**	**'pct_change_exratelcgc _ma2cent'**	**'PPC'**	**'IndexPrice inGC_perPc'**
COM	A	0.223	0.105	0.144	eliminated due to r_{f_3,f_4} >0.9
	B	0.252	0.105	0.128	
	C	0.278	0.121	0.111	
	overall	0.252 (0.251)	0.110 (0.122)	0.124 (0.142)	
		MF5	**MF6**	**MF7**	**MF8**
		'RRPinGC'	**'IPPinLC_perPc'**	**'IPPinGC_perPc'**	**'GPPinGC_perPc'**
COM	A	eliminated due to r_{f_3,f_5} >0.9	0.226	0.234	eliminated due to r_{f_7,f_8} >0.9
	B		0.208	0.240	
	C		0.191	0.230	
	overall		0.206 (0.217)	0.241 (0.263)	
		MF9	**MF10**	**MF11**	
		'NPPinGC_perPC'	**'Rebate [%]'**	**'Bonus [%]'**	
COM	A	eliminated due to r_{f_7,f_9} >0.9	eliminated due to Δ rank and mean rank > CD	0.066	
	B			0.066	
	C			0.065	
	overall			0.066 (0.030)	
		MF1	**MF2**	**MF3**	**MF4**
		'exratelcgc_ma2cent'	**'pct_change_exratelcgc _ma2cent'**	**'PPC'**	**'IndexPrice inGC_perPc'**
TP	A	0.235	0.112	0.125	eliminated due to r_{f_3,f_4} >0.9
	B	0.262	0.122	0.128	
	C	0.293	0.130	0.128	
	overall	0.279 (0.226)	0.126 (0.112)	0.127 (0.121)	
		MF5	**MF6**	**MF7**	**MF8**
		'RRPinGC'	**'IPPinLC_perPc'**	**'IPPinGC_perPc'**	**'GPPinGC_perPc'**
TP	A	eliminated due to r_{f_3,f_5} >0.9	0.199	0.223	eliminated due to r_{f_7,f_8} >0.9
	B		0.173	0.204	
	C		0.122	0.172	
	overall		0.146 (0.201)	0.182 (0.222)	
		MF9	**MF10**	**MF11**	
		'NPPinGC_perPC'	**'Rebate [%]'**	**'Bonus [%]'**	
TP	A	eliminated due to r_{f_7,f_9} >0.9	0.000	0.106	
	B		0.000	0.109	
	C		eliminated due to Δ rank and mean rank > CD	0.144	
	overall		0.000 (0.000)	0.132 (0.118)	

5.2.2.5 Domain D_4: Time-related Features

To integrate a potential temporal relationship between the purchase of spare parts and holiday seasons, the first time-related MF is the so-called *is_holiday* variable, marking

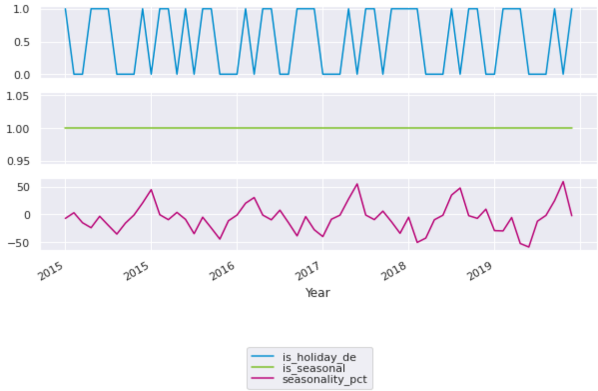


Figure 5.39 Evolution of Time-related Features

months in which meaningful national holidays and consequently vacation periods occur. The second category of time-related MFs points back to the time series decomposition in Chapter 4.2. and references natural as well as business-induced seasonality at product level. To adequately make use of them, indices are derived from the additive decomposition at PG2 level. The first MF from this category is a binary variable, which returns information whether strong seasonality is detectable for the corresponding item on PG2 level. If this is the case, the value is set at 1. The second MF is the percentage share of the total sales volume that is induced by the seasonality.

Intra-Feature Correlation

Correlation coefficients are neither high with regards to the dependent variable nor between the various MFs. The range of absolute values is between [0.0024; 0.078].

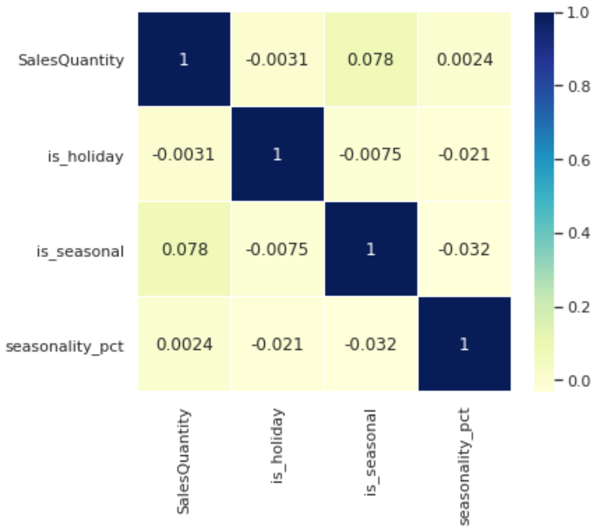


Figure 5.40 Correlation Testing for Time-related Features

Importance and Feature Rank

… based on Elastictiy

For COM and TP, MF3 is ranked highest with WRI^{f3} = 0.939 respectively 0.919. Hence, the discreate variable (MF3: *seasonality_pct*) is assigned a significantly higher relevance than the related binary variable (MF2: *is_seasonal*). MF1 (*is_holiday*) is ranked second with WRI^{f1} = 0.040 respectively 0.045.

Table 5.27: Importance Testing for Time-related Features - Elasticity

		MF1	**MF2**	**MF3**
RI^f (WRI^f)		**'is_holiday'**	**'is_seasonal'**	**'seasonality_pct'**
COM	A	0.050	0.079	0.868
	B	0.044	0.114	0.836
	C	0.039	0.409	0.534
	overall	0.040 (0.040)	0.327 (0.021)	0.622 (0.939)
		MF1	**MF2**	**MF3**
		'is_holiday'	**'is_seasonal'**	**'seasonality_pct'**
TP	A	0.055	0.130	0.803
	B	0.065	0.196	0.727
	C	0.060	0.351	0.566
	overall	0.059 (0.045)	0.313 (0.036)	0.607 (0.919)

5.2.3 Transfer of Leading Factors

Defining flexible threshold values for the critical distances as well as two fixed thresholds for intra-feature correlations (r_{c_o,c_u} ≤ 0.9) and WRI^k (WRI^k > 0.01), the original dataset comprising 356 SQL- and Python-engineered features is finally reduced to 83 MFs in the three categories *static knowns*, *time varying knowns* and *time varying unknowns*.

The latter do not only bring additional information, but also additional uncertainty into the model, since their usage depends on a prediction made for themselves. For this reason, a transfer into the final ML model depends on two conditions:

- The forecast quality of extrapolations for time varying unknowns needs to be high in terms of correctly capturing specific peaks, trends and other characteristic patterns (cf. Table E 48)
- The prediction importance under perfect condition needs to be significant. In this context, 'significant' means that extreme values of the error metrics in use (cf. Figure 6.3 and Table E 57 respectively Figure 6.4 and Table E 58) are reduced by adding or removing the corresponding MFs. The corresponding experiments and results can be found in Chapter 6.3 and in Appendix 18.

5.2.4 Conclusion

Overall, there are 83 MFs that are to be used in the forecasting experiments to determine the final prediction importance. As a conclusion we can say that

- 70 MFs are so-called dynamic respectively time varying features.
- 34 MFs have passed the pre-selection process and have hence been used very frequently and with corresponding influence by various regression models. They represent each of the previously identified potentially relevant domains

transactional and *portfolio information*, *customer* and *product* specifics, as well as *time* dependencies.

- 49 MFs are either regarded static as e.g. *MOQ* or *pallet_size*, rather static as the turn rate or RIS classification or related to lags.
- 18 MFs are considered to be so-called time-varying-unknown features that rely on a prediction for themselves to be used in the final ML model.
- 41 MFs are considered to be so-called *time varying unknown* features that can be derived from the extrapolations of related *time varying unknown* features. These include 38 MFs that represent the lagged versions of industry volumes or CSLs, two MFs that represent the relative versions of a CS's sales volume or exchange rates and one MF, that is a binary variable indicating the seasonality in a time series.

Table 5.28: Finalized Feature Space after Pre-Selection

Characterization	**Size**
Total Number of SF, HF and MFs	368
Total Number of SFs	144
Total Number of SFs after Pre-Selection	12 + 5
Total Number of HFs	152
Total Number of MFs	126
Total Number of MFs after Pre-Selection	83

The original number of SFs amounts to 144. Their dimensionality is reduced by relying on various traditional statistical and ML-based approaches as well as business knowledge. Finally, there are nine attributes in four respectively five categories, which are tested for their actual suitability in the final forecasting experiments. They contain information on the product type, the parts' value contribution and demand profile, obsolescence as well as the currency area.

6 Implementation of Behavior-based ML Models for Demand Planning

Given the complexity of ML models, they resist formal analysis methods. Therefore, one must learn about the algorithm's behavior empirically with respect to specific problems. The answers to questions such as which algorithm configuration works best on one 's data or which MFs to use, can only be found through the results of experimental trials. These trials are called *Controlled Experiments*. (Lebowitz 1987: 103ff)

Controlled experiments are experiments where all known independent variables are held constant and modified one at a time in order to determine their impact on the dependent variable. The results are compared to a baseline, or no-treatment, called a control, referring to a baseline method like the default-configuration of the model, an alternative but naïve algorithm, or a traditional forecasting algorithm.

Applied ML is special in that the user has complete control over the experiment and can run as few or as many trials as he or she wants to. Still, it is important that one is running the right types of experiments. According to Langley (1988) settings for experiments are based on

- ideas and references from scientific research
- search strategies and
- intuition

In the following, categories and sub-categories of experiments with their individual objectives will be explained. One will learn that it is not purposeful to run all experiments with all ML models. By rather relying on a specified algorithm, one could predefine a feature space (*Choose-Features Experiments*) and test cluster attributes (*Segmentation*) and then verify and adapt results using the other models.

The metrics in scope are $sMAPE^{200}$ and $MAAPE^{arctan2}$, the prioritized ML model for the subsequent analysis is GBR-XGB. The reasons are threefold (Hastie et. al. 2009: 351):

- It is less sensitive when running the model on a set of 'rather' static MFs.
- It is capable of handling NaNs.
- Compared with the other models in scope, it is computationally less expensive.

Tune-Model and *Compare-Model* Experiments will build on all three NN and boosting algorithms.

6.1 Setup of Experiments

The results of a single experiment are probabilistic, subjected to variance. The two main types of variances that are to be understood in controlled experiments are:

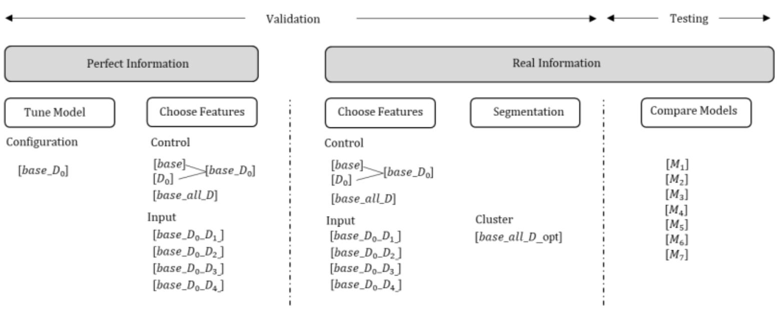
- **Variance in the model**, such as the use of randomness in the learning algorithm, such as random initial weights in NN, selection of cut points in bagging, shuffled order of data in stochastic gradient descent, and so on.
- **Variance in the data**, such as the data used to train the learning algorithm and the data used to evaluate its skill.

A result from a single run or trial of a controlled experiment would therefore be misleading given these sources of variance. Hence, the experiment must control for them. This is done by repeating the experimental trial multiple times in order to elicit the range of variance so that one can both report the expected result and the variance in the expected result, e.g., the mean and corresponding confidence intervals. For this reason, experimentation is a key part of applied machine learning.

Overall, there are three types of controlled experiments:

- **Tune-Model Experiments:** When finding the best-possible configuration of a ML model, the independent variables are the hyperparameters of the learning algorithm and the dependent variable is the estimated skill of the model on data within the validation set.
- **Choose-Features Experiments:** When determining which kind of input variables are most relevant to a model, the independent variables are the sales history and further related input features, and the dependent variable is the estimated skill of the model on unseen data. The subsequent approach is twofold and can be differentiated according to the information's status: perfect information and real information. Via these experiments the final number of factors is determined.
- **Compare-Models Experiments:** When comparing the performance of machine learning models, the independent variables may be the learning algorithms themselves with a specific configuration and the dependent variable is the estimated skill of the model on unseen data.

The category that is not officially listed in literature are so-called **Segmentation Experiments**: When defining the scale of a ML model i.e. making the decision whether it should run globally, semi-globally or locally, the overall data set builds the starting point. Cluster attributes serve to subdivide the scope for the ML model(s) und build logical groups. Running models per time series is the most granular level.



[$base$]: Predictions from semi-global model limited to static and time_varying_knowns.
[D_0]: Predictions from semi-global model with transactional and portfolio MFs.
[D_1]: Predictions from semi-global model with customer-related MFs.
[D_2]: Predictions from semi-global model with product-related MFs.
[D_3]: Predictions from semi-global model with price-related MFs;
[D_4]: Predictions from semi-global model with time-related MFs.
[$base\_all\_D$]: Predictions from semi-global model with all MFs.

Figure 6.1 Setup of Experiments

To objectively evaluate experiments, all the time series of the dataset are partitioned into three subsets – a training set for learning, a validation set for hyperparameter tuning and feature selection and a hold-out test set for the performance evaluation. For hyperparameter optimization *Bayesian Optimization* is chosen over *Random Search* with iterations = 200.

Based on Hyndman's and Athanasopoulos' ideas from 2012 to 2018, concluding to use time series cross-validation where possible rather than a simple training/test set split, *Time Series Cross Validation with Sliding Windows* is used. (Hyndman and Athanasopoulos 2018)

The roadmap für validation i.e., finding the best-possible models with regards to configuration and input features, is illustrated in the table below. The validation horizon comprises five time stamps. Forecasts are created in two overlapping samples. The evaluation during validation occurs via two statistical metrics sMAPE and MAAPE_ACR2.

Table 6.1: Roadmap for Validation of ML Models

STEP		RESULT
Configuration	Tune Model Parameters • model: semi-global[70] • scope: sample • feature space: $[base\_D_0]$	best possible model configuration
Control	Define ML-based Benchmark results • model: semi-global • scope: sample • feature space: transactional and portfolio related MFs vs. all MFs	baseline results
Input	Choose Feature Experiments • model: semi-global • scope: sample • feature space: cf. Appendix 17	final set of leading factors with feature and prediction importance
Cluster	Time Series Segmentation • scope: sample • attribute space: cf. Chapter 5.2.1	final set of cluster criteria and resulting clusters to train models cluster-wise updated model configuration per cluster.

For testing (i.e., Compare-Models experiments), sample n starts with the initial window length 48. In a rolling way, each additional sample (i.e., the training and validation dataset) gets two months of additional history by adding the forecasted values.

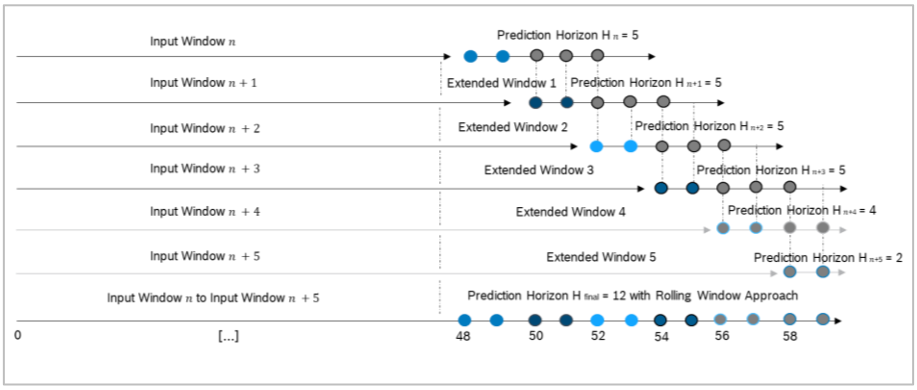


Figure 6.2 Sliding Window Approach for Forecasting

Figure 6.2 shows the series of training sets (in black) and test sets (in blue/grey colored bubbles). The evaluation occurs per sample for the five predicted months as well as across samples for the most current predictions with a total horizon of twelve months via using the selected statistical and business-based metrics from Chapter 2.3.3.

[70] Semi-global means that the model is trained separately for COM and TP.

6.2 Optimization of Parameters

Hyperparameters are the variables that govern the training process and the topology of an ML model. These variables remain constant over the training process and directly impact the performance of the ML program. They are of two types:

- Architecture Hyperparameters define the topology of a model e.g., the number and width of hidden layers.
- Training Hyperparameters influence the speed and quality of the learning algorithm e.g., the learning rate.

In Table 6.2, some of the most important model-specific configurations are illustrated. Values are based on the final runs of the semi-global models seizing the optimized feature space. In Table E 50 and Table E 51, search spaces and settings for further hyperparameters will be explained and referenced via literature.

Table 6.2: Optimized Model Configurations

MODEL	**DC-BI-LSTM**	**CNN-LSTM**	**LSTM-GRN**	**GBR-XGB**	
ARCHITECTURE					
NUM_LAYERS[71]	3	4	3	MAX_DEPTH	12
HIDDEN_SIZE	10, 10, 20	8, 12	256	TREE_METHOD	'auto'
OUTPUT_SIZE	1	1	3[72]	OUTPUT_SIZE	1
TRAINING					
ACTIVATION	relu, relu, elu	relu, sigmoid	softplus		
OPTIMIZER	adam	adam			
LOSS_FUNC	MSE	MSE	Quantile Loss	OBJECTIVE	reg: linear reg: tweedie
LEARNING_RATE	0.002	0.002	0.004	LEARNING_RATE	0.1

[71] Input and output as well as wrapper layers e.g., *Time Distributed* Layers are excluded in these numbers.
[72] Depending on the results during validation, quantiles are defined according to the products‘ PLC phase. Values are as follows: $quantile_{growth} = 0.6$; $quantile_{maturity} = 0.5$; $quantile_{decline\_phaseout} = 0.4$

6.3 Optimization of Features Spaces

To determine the optimal feature space, results are generated from two different scenarios: Scenario one is based on experimental forecasting with perfect information i.e., the model is assumed to have all *time varying unknowns* available in full transparency. Scenario two on the contrary returns results from experimental forecasting with realistic information. In the following, findings are paraphrased and illustrated. In addition, there is a tabular overview of all metrics with corresponding confidence intervals in Appendix 18.

Results from experimental forecasting with perfect information

Each of the five domains significantly improves the performance of the basic model when being added respectively worsens results of the all-feature-comprising model when being removed.

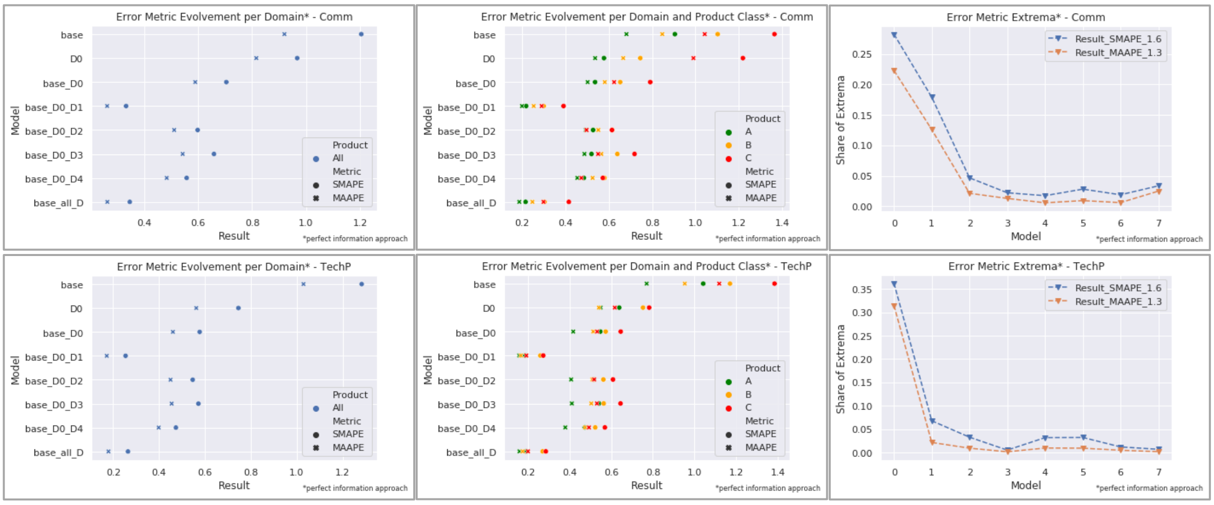


Figure 6.3 Error Metric Evolvement and Extremes - Perfect Information

Running experiments using individual MFs within the specific domains rather than the entire one (cf. Appendix 17) proves, that

- by using only static and time_varying_knowns[73] [*base*], the level of the forecasts deviates strongly from the level of the truth values. The relation to the *actual transaction* seems to be missing.
- By seizing [D_0], the predicted quantity moves at the correct level, with the majority represented as the average of past demands. This means that the model is not able to identify characteristic trajectories of demand based on portfolio features alone.
- Combining MFs from [*base*] and [D_0] improves the quality of control forecasts: product-specific patterns are more effectively learned by adding the other static and time-varying known MFs.
- High priority is given to the transactional MF *YearlySalesQuantity* ($r_{c,y} = 0.81$). It defines a realistic corridor for the model in which the demand quantity on PMC

[73] Time varying knowns comprise the following MFs

- from D_2 (product): *lifecycle_age_aa* and VIO related MFs
- form D_3 (price): *PPC-P AA_Mat10* and *bonus*
- form D_4 (time): *is_holiday*

level may move. This is similar for other portfolio-related MFs. Still, aggregation across time seems more important than aggregation across product families for model performance. Furthermore, it can be observed that the MF *item_category3_id_sum_item_cnt_month* is more important than the identical MF at a higher product family level.

- However, it is noteworthy here that the MFs that have the closest proximity to the dependent variable, i.e., the highest correlation, do not only add the highest value, but are also the most difficult to predict.
- The most important domain proofs to be customer-related with MFs *b*, *c* and *topcus_c*. c is only relevant for COM. Despite its importance, *topcus_c* is removed for both COM and TP as long as one keeps customer-related MFs that represent dominant customer segments: for the current use case the information content of *topcus_c* is represented by MF *b* and *e* for TP and by MFs *b* and *c* for COM.
- MFs with low variance (e.g. *c*) and hence a rather static nature do not seem to be useful for ML-based forecasting models. This holds true especially for NN. Instead of using it as a single MF or eliminating it completely from the feature space, though the quality of the forecasts improves as long as they are combined with another related MF. The advantage is that this preserves the information gain of the MF without dampening the forecasts for the unaffected products by the *zero content* of the MF.
- Amongst product-related input variables, MFs *lifecycle_age_aa* and *vio_replaced_dyn* are most important. Additional vio-related MFs (*avgrepq*, *multi_assignment_ph3* as well as *vio_sum_peragebin_andtype*-related MFs) do not add any specific value.
- The domain *pricing* shows higher influence for COM than for TP; for the latter only a very small effect is noticeable. Additionally, one can find that
    - if pricing is limited to global pricing information, metrics and forecasts for C3 worsen. Across regions, performance is best for *IPPinGC_Pc* in combination with *pct_change_exratelcgc_ma2cent* as well as *bonus*, C3-specifically this choice is second best. A possible reason for this phenomenon is the reduction of redundancy of *IPPinGC_perPc* and *IPPinLC_perPc* as well as the compensation of the information loss by the omission of *IPPinLC_perPc*. An inherent advantage is that the evolvement of *IPPinGC_perPc* is more transparent and can hence be forecasted with higher accuracy than *IPPinLC_perPc*.
    - Local pricing information (e.g. *IPPinLC_perPc, pct_change_exratelcgc_ma2cent*) has a minor impact on the metrics, but a major one when approaching more extreme characteristics of demand forecasts.
    - MF *bonus* majorly influences forecasts and metrics.
- MFs *is_holiday* and *is_seasonal* is removed from the domain *time* as they do not add any specific value.

Results from experimental forecasting with realistic information

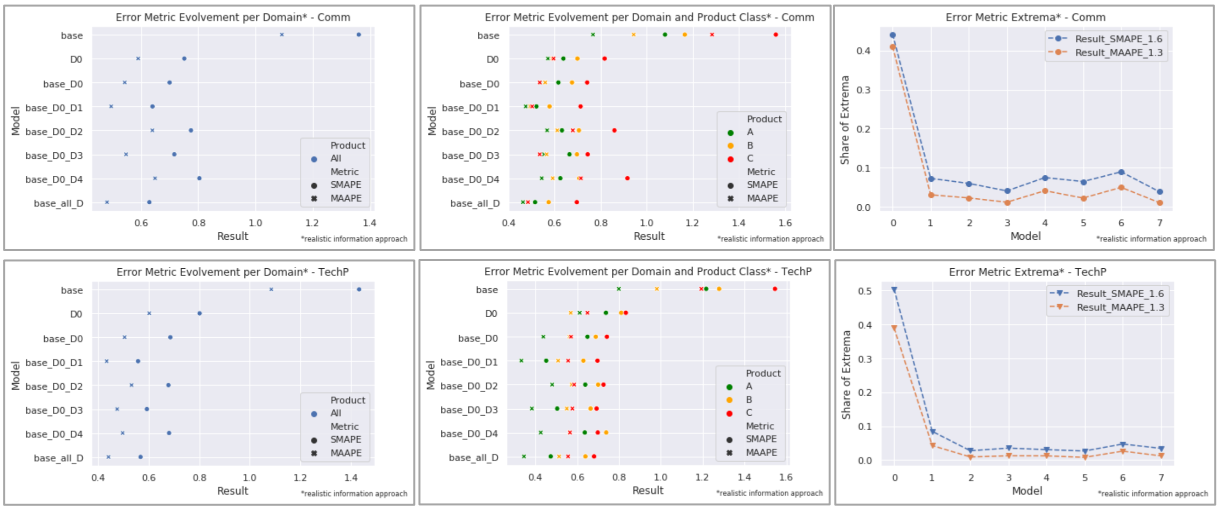


Figure 6.4 Error Metric Evolvement and Extremes - Realistic Information

The number of MFs for scenario two is reduced again by eliminating customer-related MF *e* and the lagged version of the CSL. This results in the following final feature space for GBR-XGB.

Table 6.3: Finalized Feature Space for GBR-XGB

DOMAIN	COM	TP
D_0	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'
D_1	'b_c' 'b_pct' 'unique_cus_month'	'b' 'b_pct' unique_cus_month'
D_2	'lifecycle_age_aa' 'vio_replaced_dyn' 'csl1'	'lifecycle_age_aa' 'vio_replaced_stat' 'csl1'
D_3	'ippingc' 'bonus' 'pct_change_exratelcgc_ma2cent'	'ippingc' 'bonus' 'pct_change_exratelcgc_ma2cent'
D_4	'seasonality_pct'	'seasonality_pct'

Table 6.4: Finalized Feature Space for NN

	BI-LSTM		**CNN-LSTM**		**LSTM-GRN**	
DOMAIN	COM	TP	COM	TP	COM	TP
D_0	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'	'YearlySalesQuantity' 'item_category3_id_sum_item_cnt_month' 'item_category1_id_sum_item_cnt_month'
D_1	'b_c' 'b_pct' 'unique_cus_month'	'b' 'b_pct' 'e'[74] 'unique_cus_month'	'b_c' 'b_pct' 'unique_cus_month'	'b' 'b_pct' 'e'[74] 'unique_cus_month'	'b_c' 'b_pct' 'unique_cus_month'	'b' 'b_pct' 'unique_cus_month'
D_2	'vio_replaced_dyn' 'csl1'	'vio_replaced_stat' 'csl1'	'vio_replaced_dyn' 'csl1'	'vio_replaced_stat' 'csl1'	'vio_replaced_dyn' 'csl1'	'lifecycle_age_aa' 'vio_replaced_stat' 'csl1'
D_3	'ippingc' 'bonus' 'pct_change_exratelcgc_ma2cent'[74]	'pct_change_exratelcgc_ma2cent'[74]	'ippingc' 'bonus' 'pct_change_exratelcgc_ma2cent'[74]	'pct_change_exratelcgc_ma2cent'[74]	'ippingc' 'bonus' 'pct_change_exratelcgc_ma2cent'[74]	'ippingc' 'bonus' 'pct_change_exratelcgc_ma2cent'[74]
D_4	'seasonality_pct'	'seasonality_pct'	'seasonality_pct'	'seasonality_pct'	'seasonality_pct'	'seasonality_pct'

[74] only if the currency region in scope is equal to *C3*.

6.4 Optimization of Constraints for Post-Processing

One phenomenon that becomes evident during validation is that outlier forecasts i.e. forecasts that significantly exceed the general level of actual demands are generated. It is model- though not product type or product family related and only occurs occasionally. Similar to the constraint for feature correction ($YearlySalesQuantiy_{t\epsilon H}$ < $vio\_replaced\_dyn_{t\epsilon H}$), a constraint for forecast correction is formulated:

if: $f_{i,t,m==ML} > p \cdot YearlySalesQuantity_{t\epsilon H}$ then: $f_{i,t,m==ML} = f_{i,t,m==MA12}$

else: $f_{i,t,m==ML} = f_{i,t,m==ML}$

The initialization of thresholds p occurs via the ADI_i; an iterative improvement is based on the minimization of extreme values of sMAPE and MAAPE_arc2 (cf. Table E 58) resulting from forecasts of the validation set. The adjustment of threshold values p also depends on the demand profile of PMC i:

Table 6.5: Threshold Determination for Constraint Building

	THRESHOLDS p	
MODEL	intermittent	non-intermittent
XGB	0.60	0.17
LSTM-GRN	-	-
BI-LSTM	0.55	0.15
CNN-LSTM	0.45	0.135

The forecast that is too high is replaced by the forecast of the best conventional statistical method.

6.5 Conclusion

Based on the validation set and the two statistical metrics sMAPE and MAAPE_arctan2, three different types of experiments (*Choose-Feature* experiments, *Tune-Model* experiments, *Segmentation* experiments) are conducted. Based on the outcome, the number of MFs can finally be reduced to fourteen, with the selection and composition varying slightly per cluster.

Table 6.6: Finalized Feature Space after Choose-Feature Experiments

CHARACTERIZATION	**SIZE**
Total Number of SF, HF and MFs	368
Total Number of SFs	144
Total Number of SFs after Pre-Selection	12 + 5
Total Number of SFs after Experimentation	4
Total Number of HFs	152
Total Number of MFs	126
Total Number of MFs after Pre-Selection	83
Total Number of MFs after Experimentation	14

Besides the MFs, it is also the time series identifier (*PMC*), the transaction time (*CustomerDueDate*), the dependent variable (*SalesQuantity*) and four classifiers (*PartType*, *ValueContribution* and *DemandProfile*, *Obsolescence – PLC*, *Currency Area*) that are to be used for the *Compare-Models* experiments. Their outcome is based on the test set and verified by two statistical and three business-related KPIs.

7 Verification of the Results of the Behavior-based ML Model for Demand Planning

Results, i.e., forecasts and metrics will be evaluated on different levels: They are viewed at

- on the individual time series,
- aggregated across the portfolio,
- clustered by part type, value contribution class, as well as according to their structure in non-intermittent and intermittent-lumpy time series
- and finally, across the different snapshot dates.

Via these levels the hypothesis, that the forecasting accuracy for the total demand of a SKU in an oligopolistic market can be significantly improved, also for longer horizons, by systematically integrating and combining a wide variety of region, market-/customer- and product-specific factors into an ML-based model is to be acknowledged or rejected (cf. Chapter 1.2).

To simplify the comparison from model to model, the results are shown as a heat map row by row. The weakest accuracy tends to be colored red, while the best result is colored green.

7.1 Verification on SKU-, Product Family- and Business Unit- Level

If one focuses only on forecasts, it may be seen that all the ML models have learned a lot of local dynamics as their values and evolvements look different for each individual time series of the different product families and BUs. Regardless of the scale of sales volumes, i.e., whether the SKU sells in tens, hundreds, or thousands, the models can generate reasonable forecasts. The situation is similar for the AA ensemble.

For the other conventional statistical methods, levels are often adhered to as well. However, it is especially MA12 and RWq3 that show their smoothing effect and thus miss the approximation of the specific upward and downward deflections.

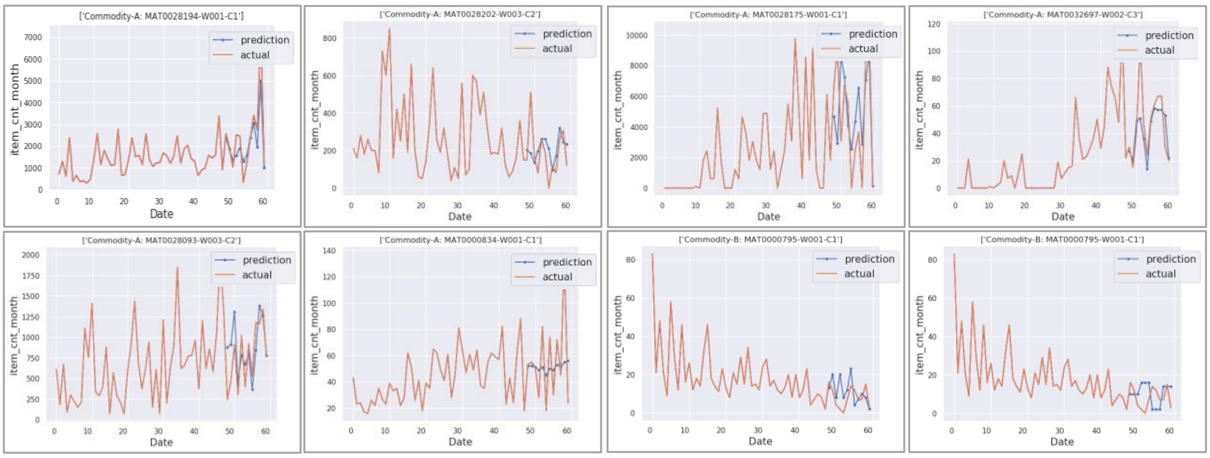


Figure 7.1 Forecasts for COM-A and COM-B

Figure 7.1 exemplifies *good* forecasts from the different models: The upper row references from left to right GBR-XGB, LSTM-GRN, DC-BI-LSTM and CNN-LSTM, the lower row denotes also from left to right the AA-Ensemble, MA12, RQ12 and RWq3.

The plots below (Figure 7.2) show a comparison of the actual and forecasted demands based on the best performing model overall, the GBR-XGB. A comparison of the bins confirms that the values are on a similar level with a clear dominance of zero demands for both real and forecasted values. Figure 7.3 represents the same for the best conventional statistical model (AA-Ensemble) with the difference being that here we see a stronger tendency to underforecasting.

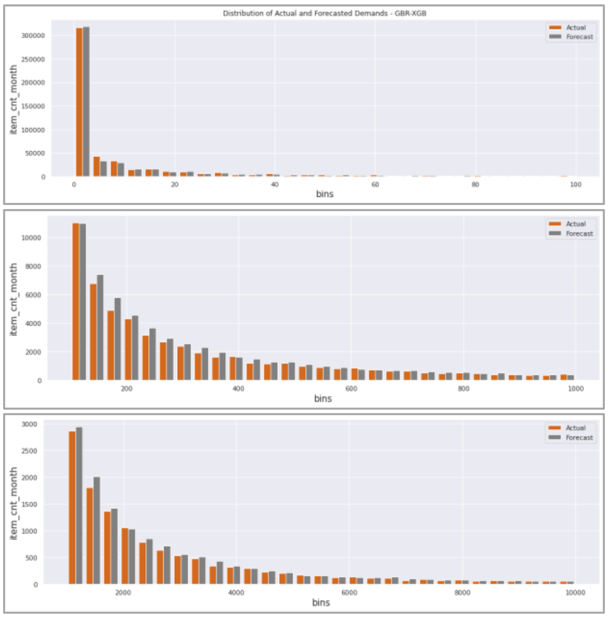


Figure 7.3 Distribution of Actual Demand and Forecasts- Best Model Overall

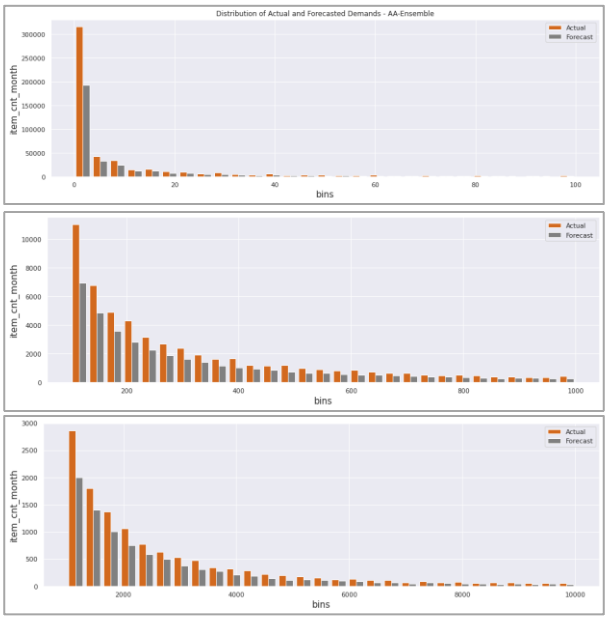


Figure 7.2 Distribution of Actual Demand and Forecasts- Best Conventional Model

When one observes forecasts which have been declared as *good* or *bad* in terms of the statistical evaluation metrics and business KPIs, one may see that also GBR-XGB is the best performing algorithm for both COM and TP. LSTM-GRN is the weakest learner amongst all ML-models and when compared to the AA-Ensemble – also with regards to its behavior to continuous underforecasting. Amongst the conventional statistical methods, it is MA12 that is hard to outperform with both ML models and the AA ensemble from a purely metric-based perspective.

Table 7.1: Verification on Part Type Level [%-points]

difference in %points		**sMAPE**	**MAAPE_ACR2**	**wFAI**	**MBA**
best model overall – worst ML model	COM	4.1	6.2	11.0	20.9
	TP	12.1	6.6	8.0	21.1
best model overall – best conventional statistical model	COM	2.7	2.2	5.4	5.0
	TP	3.9	1.0	3.5	9.1

Table 7.2: Verification on Part Type Level [%][39]

sMAPE	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
COM	**0.576**	0.617	0.630	0.614	*0.603*	0.667	0.759	0.751
TP	**0.594**	0.715	0.674	0.654	*0.633*	0.778	0.788	0.753

MAAPE_ARC2	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
COM	**0.384**	0.446	0.504	0.490	*0.406*	0.556	0.563	0.553
TP	*0.418*	0.484	0.515	0.499	**0.410**	0.611	0.548	0.522

wFAI	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
COM	**0.627**	0.517	0.543	0.555	*0.573*	0.563	0.504	0.381
TP	**0.592**	0.512	0.568	*0.578*	0.557	0.476	0.320	0.335

MBA	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
COM	**0.633**	0.842	0.762	0.693	0.691	*0.681*	0.988	0.946
TP	**0.657**	0.868	0.815	0.734	0.749	*0.727*	1.025	1.093

TS	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
COM	0.049	-0.129	0.079	0.046	-0.032	0.045	0.085	0.019
TP	-0.067	-0.013	0.070	0.044	-0.048	0.060	0.082	0.036

In general, one can also conclude that evaluation metrics for TP are worse than for COM. Here, two reasons are assumed:

- The share of intermittent-lumpy time series is higher for TP.
- The influence of MFs seems smaller as e.g., from D_2 – failure probabilities (*vio_replaced_stat*) and from D_3 – *IPPinGC_perPc.*

7.2 Verification per Value Contribution Class and Demand Profile

Across the majority of models, it can be confirmed that A/B Parts or products whose demand is represented by non-intermittent time series achieve the best results: sMAPE is on average at 0.591 (A) respectively 0.674 (B) and at 0.636 (non-intermittent); MAAPE_arc2 at 0.467 (A) respectively at 0.532 (B) and at 0.482 (non-intermittent). Values for *wFAI* are at 0.553 (A) respectively at 0.499 (B) and at 0.546 (non-intermittent); for MBA at 0.782 (A) respectively at 0.885 (B) and at 0.746 (non-intermittent).

This contrasts with the results for intermittent time series and thus C-parts. Here, the results for the statistical error metrics are higher across all clusters with mean values of 0.706 (C) and 0.715 (intermittent) for sMAPE and 0.517 (C) and 0.523 (intermittent) for MAAPE_ARC2. Business KPIs give a similar indication. (cf. Table 7.3) RW12 and RWq3 are exceptional – according to the error metrics, predictions are better for C-parts than for B-parts and occasionally also A-parts.

Across the various clusters, GBR-XGB, the AA ensemble and occasionally CNN-LSTM are the models with best results. Negative performers are LSTM-GRN and the two naïve models. Consequently, the assumption that an estimation of the quarterly demand (e.g.

via RWq3) can be used to better estimate the actual demand of intermittent sequences cannot be confirmed.

Table 7.3: Verification on Value Contribution and Demand Profile Level [%][39]

sMAPE	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
A	*0.531*	0.603	0.548	**0.525**	0.553	0.572	0.689	0.703
B	**0.587**	0.631	0.643	0.619	*0.608*	0.704	0.809	0.792
C	**0.617**	0.733	0.688	0.674	*0.638*	0.779	0.782	0.748
intermittent	**0.616**	0.718	0.706	0.690	*0.676*	0.818	0.773	0.724
non-intermittent	**0.562**	0.628	0.585	*0.564*	0.583	0.606	0.776	0.788

MAAPE_ARC2	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
A	**0.365**	0.393	0.494	0.470	*0.390*	0.515	0.555	0.558
B	**0.434**	0.530	0.535	0.512	*0.455*	0.597	0.606	0.592
C	**0.457**	0.583	0.502	0.494	*0.461*	0.599	0.532	0.505
intermittent	*0.493*	0.575	0.522	0.500	0.517	0.607	0.500	**0.467**
non-intermittent	**0.328**	0.384	0.500	0.491	*0.344*	0.557	0.627	0.626

wFAI	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
A	**0.648**	0.563	0.576	0.581	*0.614*	0.584	0.458	0.397
B	**0.581**	0.494	0.535	0.557	*0.561*	0.544	0.385	0.337
C	*0.524*	0.474	0.523	**0.543**	0.516	0.510	0,419	0.365
intermittent	0.474	*0.481*	0.412	**0.525**	0.471	0.434	0.349	0.291
non-intermittent	**0.620**	0.507	0.594	*0.601*	0.586	0.578	0.493	0.386

MBA	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
A	**0.623**	0.768	0.736	0.703	*0.640*	0.735	1.059	0.988
B	**0.657**	0.949	0.933	0.743	*0.731*	0.814	1.088	1.163
C	*0.679*	0.989	0.815	0.863	0.781	*0.660*	0.955	0.968
intermittent	**0.677**	0.921	0.904	0.854	0.758	*0.736*	1.037	1.026
non-intermittent	**0.632**	0.804	*0.643*	0.594	*0.643*	0.665	0.969	1.018

TS	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
A	0.040	-0.038	0.027	0.013	0.022	0.047	0.096	0.009
B	0.051	-0.015	0.075	0.077	-0.039	0.075	0.099	0.046
C	-0.072	-0.102	0.116	0.105	-0.055	0.044	0.072	0.025
intermittent	-0.073	-0.136	0.098	0.046	-0.055	0.086	0.040	0.012
non-intermittent	0.051	0.016	0.056	0.045	-0.018	0.027	0.138	0.049

The analysis, which is shown subsequently as an example for the sMAPE, follows the analogous principle for all other metrics and models:

The positive skewness of the distribution is to be assessed encouraging. This can be observed independently of the cluster and for both the best model overall and the best conventional model. However, it is especially for intermittent time series and for C-parts, that there is a high concentration around the zero. One could hence conclude that the dominance of the zero periods in these time series and the associated zero forecasts positively affect the mean value of the error terms. This in turn can be seen as evidence, that the sole averaged value of error metrics is not very representative: Predicting zero periods in a correct manner is important. However, it is more complex and essential to correctly determine periods with a demand larger than zero amongst all the zero periods. Therefore, the proportions of the other expressions of the metrics are also of particular interest.

The medians for non-intermittent parts are at 0.485 (GBR-XGB) and at 0.526 (AA-Ensemble). A/B parts show slightly better results with 0.480 (GBR-XGB) and 0.500 (AA-Ensemble). For intermittent parts, 50 percent of the time series exceed a value of 0.552 (GBR-XGB) respectively 0.673 (AA-Ensemble); for C-parts it is 0.565 (GBR-XGB) and 0.633 (AA-Ensemble).

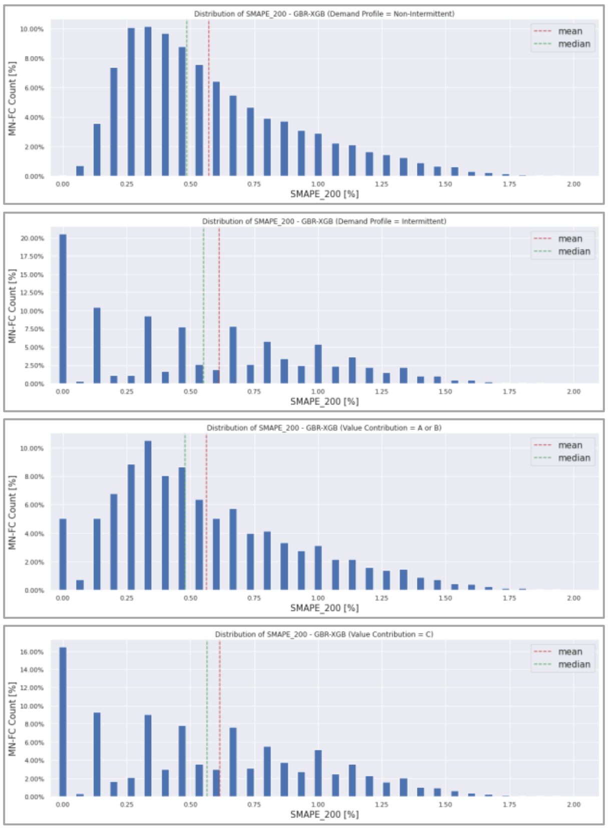


Figure 7.5 Distribution of $sMAPE^{200}$ values - Best Model Overall

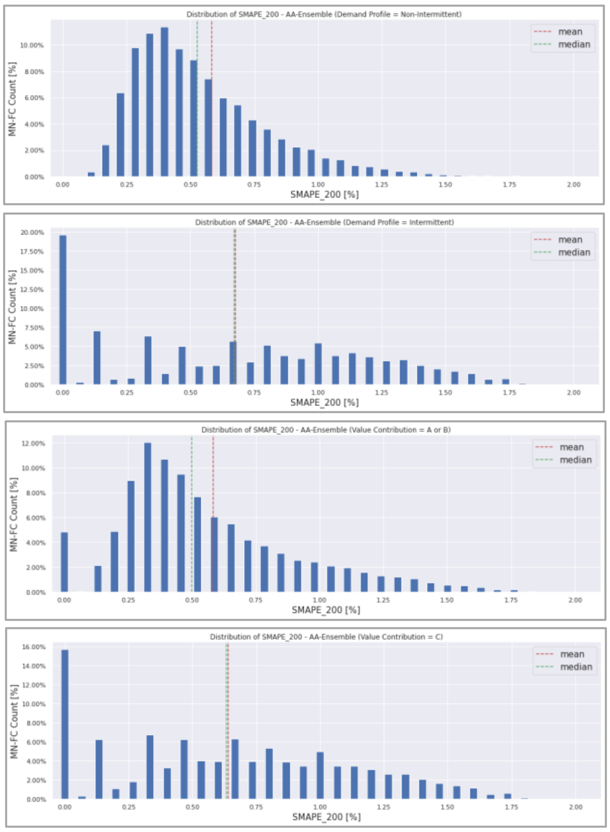


Figure 7.4 Distribution of $sMAPE^{200}$ values - Best Conventional Model

7.3 Verification over Time

Product Life Cycle Phase

Best are results for SKUs in the decline und maturity phase. sMAPE is on average at 0.587 (Decline) respectively 0.668 (Maturity) and at 0.587 (Growth); MAAPE_arctan2 is at 0.424 (Decline) respectively at 0.522 (Maturity) and at 0.522 (Growth). Values for $wFAI$ are at 0.0.567 (Decline) respectively at 0.521 (Maturity) and at 0.504 (Growth); for MBA at 0.705 (Decline) respectively at 0.782 (Maturity) and at 0.837 (Growth). The TS for time series from cluster *PLC1* (Growth) is strictly negative and – when considering the absolute values – also more pronounced.

Potential reasons for this phenomenon are twofold:

- Historical time series data is shorter for SKUs in the growth phase than for mature and declining products. The chance to recognize patterns and learn from them is hence limited.

- Models have difficulties to capture the increasing trend in the corresponding strength.

Adapting the quantiles-parameter already improved the results for LSTM-GRN. Another potential solution is to pass a statistical MF for growth products to each of the multi-variate models to emphasize the trend separately – alike the one for seasonality.

Model wise, one can again highlight GBR-XGB as the best model: Across PLCs it assumes best values for sMAPE (0.575), $wFAI$ (0.608) and MBA (0.608). The AA-Ensemble is best when MAAPE_arctan2 (0.390) is chosen, GBR-XGB is the follow-up.

Table 7.4: Verification per PLC-Phase [%][39]

sMAPE		GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
	PLC0	-	-	-	-	-	-	-	-
	PLC1	**0.621**	0.677	0.710	0.692	*0.651*	0.806	0.810	0.800
	PLC2	**0.568**	0.719	0.648	0.604	*0.602*	0.621	0.795	0.790
	PLC3	*0.535*	0.578	0.547	0.547	**0.518**	0.640	0.693	0.638
	PLC4	-	-	-	-	-	-	-	-

MAAPE_ARC2		GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
	PLC0	-	-	-	-	-	-	-	-
	PLC1	*0.464*	0.557	0.547	0.522	**0.377**	0.634	0.573	0.598
	PLC2	**0.410**	0.483	0.541	0.524	*0.463*	0.536	0.567	0.555
	PLC3	0.339	0.365	0.424	0.421	**0.330**	0.523	0.526	0.462
	PLC4	-	-	-	-	-	-	-	-

wFAI		GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
	PLC0	-	-	-	-	-	-	-	-
	PLC1	**0.580**	0.486	0.488	0.506	0.532	*0.556*	0.475	0.411
	PLC2	**0.608**	0.539	0.539	0.554	*0.583*	0.546	0.464	0.332
	PLC3	**0.635**	0.536	*0.633*	*0.633*	0.606	0.593	0.495	0.401
	PLC4	-	-	-	-	-	-	-	-

MBA		GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
	PLC0	-	-	-	-	-	-	-	-
	PLC1	**0.664**	0.770	0.863	0.712	*0.697*	0.777	1.093	1.119
	PLC2	**0.624**	0.923	0.790	0.813	0.765	*0.679*	0.723	0.936
	PLC3	**0.536**	0.773	0.646	0.618	*0.549*	0.555	1.059	0.906
	PLC4	-	-	-	-	-	-	-	-

TS		GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
	PLC0	-	-	-	-	-	-	-	-
	PLC1	-0.114	-0.265	-0.094	-0.082	-0.131	-0.116	-0.141	-0.053
	PLC2	0.052	0.017	0.075	0.031	-0.068	0.061	0.097	0.033
	PLC3	0.087	-0.042	0.060	0.047	0.112	0.107	0.225	0.081
	PLC4	-	-	-	-	-	-	-	-

Snapshots

The snapshots lead to one of the most important findings: the models are robust in the sense that they deliver stable forecast quality regardless of the forecast origin i.e., one can say that the forecast is *good* as it is reliable over time.

Figure 7.6 represents the deviations per ML-based model that are calculated between the metrics sMAPE and MAAPE_arctan2 in-between the different snapshots. According to it, it is the CNN-LSTM and DC-BI-LSTM that are most stable. LSTM-GRN reveals some albeit outliers but still remains in an acceptable range.

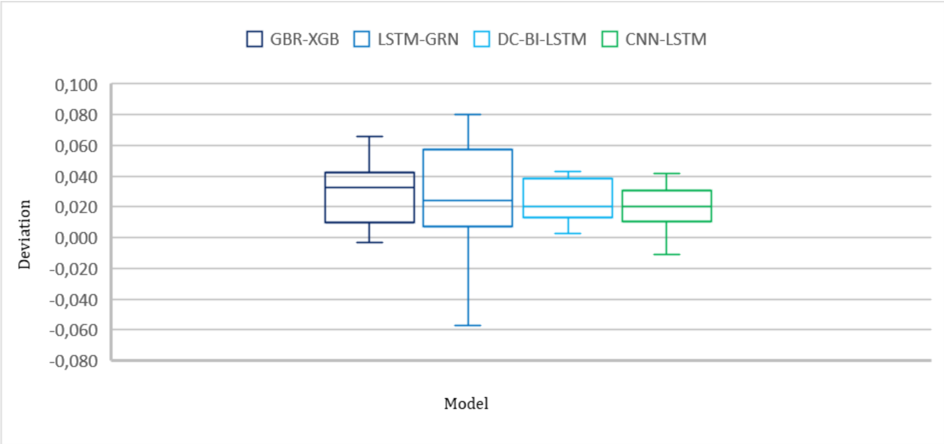


Figure 7.6 Boxplots of Deviations between Error Metrics across Snapshots

It is intuitively clear that a forecast referring to a more distant future is generally more uncertain than a forecast referring to the near future. Here too, empirical evidence proves the point: Forecasts for bins that are closer to the forecast origin are more accurate than those that are further away.

- Metrics on snapshot level (H=5 via one-shot) are worse than for overall horizon (H=12 with window-shift = 2). Hence one can conclude that a rolling planning and more frequent training respectively seizing updated historical data improves the performance.
- Metrics from snapshot 4 show slight improvement compared to those from snapshot 1. One can thus assume a dependency on the number of historical time bins which means more historical data adds value with respect to forecast quality.

Table 7.5: Verification across Snapshots [%][39]

sMAPE	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM
Snap1	**0.631**	0.696	0.734	0.705
Snap2	**0.597**	0.676	0.708	0.684
Snap3	**0.588**	0.616	0.695	0.663
Snap4	**0.591**	0.673	0.691	0.674

MAAPE_ARC2	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM
Snap1	**0.445**	0.485	0.574	0.552
Snap2	**0.426**	0.460	0.551	0.533
Snap3	**0.414**	0.413	0.538	0.513
Snap4	**0.379**	0.437	0.535	0.522

TS	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM
Snap1	-0.045	-0.015	0.115	0.103
Snap2	-0.045	-0.045	0.079	0.065
Snap3	-0.021	-0.069	0.066	0.052
Snap4	-0.055	-0.080	0.061	0.051

7.4 Conclusion

Based on the results represented in Chapters 7.1, 7.2 and 7.3, the hypothesis, that the forecasting accuracy for the total demand of a SKU in an oligopolistic market can be significantly improved, also for longer horizons, by systematically integrating and combining a wide variety of region, market-/customer- and product-specific factors into an ML-based model is to be acknowledged.

Table 7.6: Verification – Full Horizon [%][39]

sMAPE	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
Full Horizon	**0.586**	0.664	0.653	0.635	0.615	0.725	0.774	0.752

MAAPE_ARC2	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
Full Horizon	**0.400**	0.448	0.510	0.495	0.411	0.585	0.555	0.537

wFAI	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
Full Horizon	**0.602**	0.540	0.556	*0.567*	0.560	0.529	0.474	0.375

MBA	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
Full Horizon	**0.640**	0.856	0.789	*0.705*	0.738	*0.705*	1.007	1.023

TS	GBR-XGB	LSTM-GRN	DC-BI-LSTM	CNN-LSTM	AA-Ensemble	MA12	RW12	RWq3
Full Horizon	-0.049	-0.069	0.075	0.045	-0.039	0.053	0.083	0.028

As shown in all previous tables as well as in Table 7.6, the two univariate, naïve models RW12 and RWq3 are the most inaccurate ones. The approach of switching from the required time bucket to the quarterly level and then splitting the forecasts top-down amongst the monthly bins also proves not to be useful.

The results of MA12 related error metrics prove to be a good benchmark. However, if one compares the actual forecasts, ML based models and the AA ensemble prove to be much more trainable and accurate.

Amongst ML based models, GBR-XGB proves to be best for most clusters. In case of rating its performance for the full horizon it achieves values that are on average 6.47 (sMAPE), 8.43 (MAAPE_arctan2), 4.77 ($wFAI$) and 14.33 (MBA) percentage points better than other tested ML models and 13.05 (sMAPE), 12.20 (MAAPE_arctan2), 11.75 ($wFAI$) and 22.83 (MBA) percentage points better than the conventional statistical models.

8 Final Consideration

The final chapter summarizes the main findings of the thesis and verifies to what extent the objectives have been achieved. As every dissertation deals with and answers the research questions only up to a certain degree, Chapter 8.2 will elaborate on future potentials and prospects. These are related to the historical data of the dependent variables as well as the MFs, the use of a hybrid model and the concept of probabilistic forecasting.

8.1 Summary and Critical Evaluation

Up to now ML superiority claims are characterized by the following two major limitations (Leitner and Leopold-Wildburger 2011, Benidis et. al. 2020):

- Their conclusions are based on a few, or even a single time series, raising questions about the statistical significance of the results and their generalization and
- ML inferiority claims were never tackled by a profound factor respectively feature selection and integration.

This second condition is the focus of this dissertation. RQ1 and RQ2 can also be derived from it:

- **RQ1**: Which factors are most likely to affect individual demand behavior and its forecasting in the Automotive Aftermarket and how can they be categorized and prioritized?
- **RQ2:** How to represent respectively model qualitatively prioritized factors via data and how to determine their relevance from a quantitative perspective?

To profoundly answer these questions, a comprehensive literature research on factors influencing demand in general, influencing demand in the automotive industry and influencing demand for spare parts has been conducted. This approach also formed the foundation for a subsequent qualitative analysis in form of expert interviews. More than 40 participants from different large sized, international companies and domains (sales, logistics and business units) shared their knowledge with regards to demand behavior and demand triggers in the B2B area.

Based on this, more than 88 factors of variance are identified and prioritized, and as a second step translated into so-called main features (MFs) via various external and internal data sources and engineering steps. In order to ensure the highest possible degree of comprehensiveness and meaningfulness, a reasonable categorization of the factors as well as the features plays an important role. The associated criteria are as follows:

- Context
- Level of Influence
- Type of Influence
- Texture of Feature

For a subsequent transfer of those MFs into the ML-model, one should rely on a profound quantitative pre-selection process. In this context, an approach based on Demšar 2006 is chosen and incorporated into a wrapper method. The result is a ranking of all identified MFs based on weighted importance (FI) scores and different regression models.

In relation to the other two limitations addressed above, as well as to the final determination of prediction importance (PI), RQ3 is defined.

- **RQ3:** In conjunction with portfolio- and behavioral-based forecasting, which ML-based algorithms promise to lead to sustainable improvements in demand planning in the Automotive Aftermarket?

It is shown that semi-global behavior-based ML models lead to better results than their univariate conventional statistical counterparts - both in terms of evaluation metrics and forecasts. This though does not hold true for every type of ML-based model.

The integration of factors from a wide range of domains is indispensable to transmit the cause-effect relationships to the behavior-based models in a meaningful way. The fact that all domains are relevant has already been shown in the pre-selection and then again when testing different perfect information and realistic information scenarios. The term 'behavior-based' refers on the one hand to factors of variance and on the other hand to the clustering approach. Herein, tiering by intermittence and value contribution, and also by obsolescence, currency area and product type prove beneficial.

Likewise, it must be made clear that a forecast is not a perfect reflection of reality: The lack of accuracy can proportionally be explained by a lack of information due to either total (e.g. promotional information) or partial non-availability of MFs (e.g. *vio_replaced_dyn*) as well as quality issues of already integrated MFs. Further potential for improving the results is seen in the models' architecture and hyperparameter optimization.

8.2 Prospects

The results of the dissertation show that it is worth the effort to follow a data-centric approach instead of a model-centric approach. Furthermore, the upstream analysis generates great transparency as to how often and how much the *Factors of Variance* and the associated MFs influence the dependent variable, i.e. sales quantities in the AA. The question that is not answered yet, but still interesting, is whether the variance in the MFs correlates with the rank respectively the importance scores. This could confirm or reject the hypothesis that more volatile MFs add more value. Answers to this question may be found by determining the coefficient of variance per MF (CV^C) and a test for each model, cluster and scenario.

A topic that has already been touched upon is the transformation or adaptation of the dependent variable, either by outlier correction or so-called label transformation (cf. Chapter 4.2). Upcoming experiments may comprise the

- benchmarking of individual outlier corrections with a concomitant qualitative root-cause analysis for identified outliers as well as a
- benchmarking of single label transformations and back-transformations.

Here, one must also consider the computational effort of the individual calculations.

As best models per cluster are of varying types, one can additionally think of switching from a one-type-of-model-fits-all to a build-an-ensemble strategy. The best (conventional statistical and ML) models can be identified and combined not only per cluster but also per horizon, a process that is commonly known as *demand sensing*.

So far, the proposed methods rely on point forecasts representing the mean value of the future demand. This approach has a significant drawback: It fully ignores the uncertainty of the future demand. Depending on the probability distribution of the demand, the mean value of the forecast can differ significantly from the other potential scenarios, as demonstrated exemplarily in the below figure: While both distributions have the same mean value of 6,000 pieces, their 90 percent quantile, e.g. corresponding to a service level target of 90%, differs significantly (7,000 pieces versus 8,000 pieces).

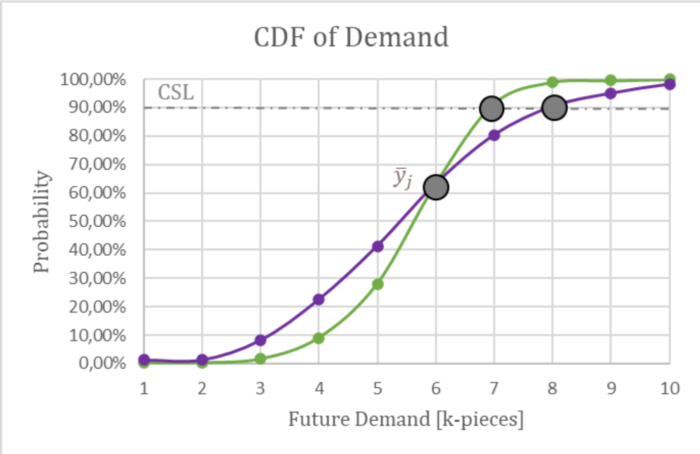


Figure 8.1 CDFs of Future Demand for Two Different PMCs

This bears several risks (cf. Figure 1.2). Hence, an improved understanding of the uncertainty of the future demand is crucial. Based on probabilistic forecasting, the fourth prospect, multiple use cases and scenarios can be implemented, varying in complexity as successfully demonstrated by participants in the B2C area. To the best of my knowledge, it has not been realized in the B2B environment up to now.

Bibliography

Aburto L. and R. Weber (2007). Improved supply chain management based on hybrid demand forecasts. Applied Soft Computing. Vol. 7. pp. 136-144.

Adya, M. P. and F. L. Collopy (1998). How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation.

Aghabozorgi S., T. Y. Wah, T. Herawan, H. A. Jalab., M. A. Shaygan, and A. Jalali (2014). A hybrid algorithm for clustering of time series data based on affinity search technique. *Scientific World Journal*. 2014:562194. DOI: 10.1155/2014/562194.

Aghabozorgi, S, A. S. Shirkhorshidi and T. Y. Wah (2015). Time-series clustering – A decade review. *Information Systems*. Vol. 53. pp. 16-38, ISSN 0306-4379, DOI: https://doi.org/10.1016/j.is.2015.04.007.

Agrawal R., C. Faloutsos and A. Swami. (1993). Efficient Similarity Search in Sequence Databases. *Lecture Notes in Computer Science*. Vol. 730. pp. 69-84.

Ailawadi, K., K. Gedenk, C. Lutzky and S. Neslin (2006). Decomposition of the Sales Impact of Promotion-Induced Stockpiling. *Journal of Marketing Research*. DOI: 44. 10.1509/jmkr.44.3.450.

Alpaydin, E. (2019) *Maschinelles Lernen*. Berlin, Boston: De Gruyter Oldenbourg.

Amin-Naseri, M. R. and B. Rostami-Tabar. (2008). Neural network approach to lumpy demand forecasting for spare parts in process industries. *ICCCE 2008*. pp. 1378-1382. 10.1109/ICCCE.2008.4580831.

Angkiriwang R., P. Nyoman P and B. Santosa (2014). Managing uncertainty through supply chain flexibility: reactive vs. proactive approaches. *Production & Manufacturing Research*, 2:1, 50-70, DOI: 10.1080/21693277.2014.882804.

Armstrong J. S. (1985). Measures of Accuracy. *Long-Range Forecasting: From Crystal Ball to Computer*. A Wiley-Interscience Publication. Wiley. pp. 346–354.

Armstrong J. S. and F. Collopy (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting.* Vol. 8, no. 1. pp. 69–80, 1992.

Arslankaya, S. and V. Öz (2018). Time Series Analysis on Sales Quantity in an Automotive Company and Estimation by Artificial Neural Networks. Vol. 22, Issue 5, pp. 1482-1492.

Athanasopoulos G., R. A. Ahmed, R. J. Hyndman (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*. Vol. 25. pp. 146–166.

Athanasopoulos, G, R. Hyndman, N. Kourentzes and F. Petropoulos (2017). Forecasting with Temporal Hierarchies. *European Journal of Operational Research*. Vol. 262. DOI: 10.1016/j.ejor.2017.02.046.

Autostat (1) (2021). Average car mileage in Russia. https://eng.autostat.ru/news/17121/ (access: 28.06.2021)

Autostat (2) (2021). Vehicle registration in Russia. https://eng.autostat.ru/ (Access: 28.06.2021)

Axsater, S. (2000). Exact Analysis of Continuous Review (R, Q) Policies in Two-Echelon Inventory Systems with Compound Poisson Demand. *Operations Research*. Vol. 48, pp. 686-696.

B. P. Shapiro and T. V. Bonoma (1984). How to Segment Industrial Markets. *Segmenting the Industrial Market*. San Francisco: Lexington Books. 8th edition.

Bacchetti, A. and N. Saccani (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega.* Vol. 40. pp. 722-737. 10.1016/j.omega.2011.06.008.

Backhaus, K., B. Erichson, S. Gensler, R. Weiber and T. Weiber, Thomas (2021): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Vol. 16. Wiesbaden: Springer Fachmedien.

Bahdanau, D, C. Kyunghyun and Y. Bengio (2014). Neural Machine Translation by Jointly Learning to Align and Translate. DOI: CoRR abs/1409.0473.

Bala, P. K. (2012): Improving inventory performance with clustering based demand forecasts. *Journal of modelling in management*. Vol. 7.pp. 23-37.

Bartezzaghi E, R Verganti, and G Zotteri (1999). A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*. Vol. 59. pp. 499-510.

Batista, G., X. Wang, and E. Keogh (2011). A Complexity-Invariant Distance Measure for Time Series. *Proceedings of the 11th SIAM International Conference on Data Mining*. pp. 699-710. DOI: https://doi.org/10.1137/1.9781611972818.60

Benidis, K., S. S. Rangapuram, V. Flunkert, B. Wang, D. C. Maddix, A. C. Türkmen, J. Gasthaus, M. Bohlke-Schneider, D. Salinas, L. Stella, L. Callot and T. Januschowski (2020). Neural forecasting: Introduction and literature overview. DOI: *arXiv* abs/2004.10240

Bianchi, F. M., E. Maiorino, M. Kampffmeyer, A. Rizzi and R. Jenssen (2017). An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting. *Springer Briefs in Computer Science.* DOI: https://doi.org/10.48550/arXiv.1705.04378.

Bode C. and S. M. Wagner (2015). Structural drivers of upstream supply chain complexity and the frequency of supply chain disruptions. *Journal of Operations Management*. Vol. 36, pp. 215-228, ISSN 0272-6963, DOI: https://doi.org/10.1016/j.jom.2014.12.004.

Bogner A., B. Littig and W Menz (2014). Interviews mit Experten: Eine praxisorientierte Einführung (Qualitative Sozialforschung). Springer VS.

Böttcher, M., M. Spott, D. Nauck and R. Kruse (2009). Mining changing customer segments in dynamic markets. *Expert Systems with Applications*. Vol. 36. pp. 155-164. DOI: 10.1016/j.eswa.2007.09.006.

Bousonville, T. (2016). *Die digitale Transformation der Wertschöpfungskette*. Heidelberg: Springer.

Boylan J. (1997) The centralisation of inventory and the modelling of demand. PhD thesis. University of Warwick, UK.

Boylan J. and A. Syntetos (2021). Intermi*ttent Demand Forecasting: Context, Methods and Applications*. Hoboken, NJ: Wiley.

Boylan J. E. and A. A. Syntetos (2010).Spare parts management: a review of forecasting research and extensions. *IMA Journal of Management Mathematics*, Vol. 21(3):227-237.

Boylan, J., A. Syntetos and G. Karakostas (2008). Classification for forecasting and stock control: A case study. *Journal of the Operational Research Society*. Vol. 59. DOI: 10.1057/palgrave.jors.2602312.

Breiman, L. (1998). Arcing classifiers. The Annals of Statistics. Vol. 26. pp. 801-849. DOI: https://doi.org/10.1214/aos/1024691079

Brown, G. (2011). Ensemble Learning. *Encyclopedia of Machine Learning*. Boston, MA: Springer.

Burgin, T. A. and J. M. Norman (1976). Table for Determining Probability of a Stock out and Potential Lost Sales for a Gamma-Distributed Demand. *Operational Research Quarterly*. Vol. 27. pp. 621-631.

Buxmann, P. and H. Schmidt (2019): *Künstliche Intelligenz – Mit Algorithmen zum wirtschaftlichen Erfolg.* Wiesbaden: Springer.

Caliński T and J. Harabasz (1974) A dendrite method for cluster analysis. *Communications in Statistics*, Vol. 3. pp. 1-27. DOI: 10.1080/03610927408827101

Caniato F., M. Kalchschmidt, S. Ronchi, R. Verganti and G. Zotteri (2005). Clustering customers to forecast demand, *Production Planning and Control*, Vol. 16:1. pp. 32-43, DOI: 10.1080/09537280512331325155

Caniato, F., M. Kalchschmidt and S. Ronchi (2011). Integrating quantitative and qualitative forecasting approaches: Organizational learning in an action research case. *JORS*. Vol. 62. pp. 413-424. DOI: 10.1057/jors.2010.142.

Carbonneau, R., R. Vahidov and K. Laframboise. (2007). Machine Learning-Based Demand Forecasting in Supply Chains. *IJIIT*. Vol. 3. pp. 40-57. DOI: 10.4018/jiit.2007100103.

Carbonneau, R., R. Vahidov and K. Laframboise. (2012). Forecasting Supply Chain Demand Using Machine Learning Algorithms. *IJIIT*. DOI: 10.4018/978-1-60960-818-7.ch609.

Chase C. W. (2013). *Demand-Driven Forecasting: A Structured Approach to Forecasting*. Hoboken, NJ: John Wiley & Sons Inc. 2nd edition.

Chen T and C. Guestrin (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. DOI. https://doi.org/10.1145/2939672.2939785.

Chen, Z. and Y. Yang. Assessing Forecast Accuracy Measures. (2004).

Christ M., N. Braun, J. Neuffer and A. W. Kempa-Lieh (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*. Vol. 307 pp. 72-77. DOI: 10.1016/j.neucom.2018.03.067.

Chung J., C. Gulcehre, K. Cho and Yoshua Bengio. (*2014*). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *Deep Learning and Representation Learning Workshop.* DOI: https://doi.org/10.48550/arXiv.1412.3555

Cochran, W. G. (1963). Sampling Techniques. New York: John Wiley and Sons, Inc. 2nd Ed.

Cohen, M., A. Narendra Agrawal, and V. Agrawal (2006). Winning in the aftermarket. *Harvard Business Review.* Vol. 84.5. pp. 129-138.

Crone S. F. (2010): Neuronale Netze zur Prognose und Disposition im Handel. PhD thesis.

Croston, J.D. (1972) Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly*. Vol. 23. pp. 289-303. DOI: http://dx.doi.org/10.2307/3007885

Danese P. and M. Kalchschmidt (2011). The role of the forecasting process in improving forecast accuracy and operational performance. *International Journal of Production Economics*. Vol. 131, Issue 1. pp. 204-214.

Dangerfield, B. J. and J. S. Morris (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations, *International Journal of Forecasting*. Vol. 8, Issue 2. pp. 233-241, https://EconPapers.repec.org/RePEc:eee:intfor:v:8:y:1992:i:2:p:233-241.

Demšar J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research.* Vol. 7. pp. 1–30. DOI: https://www.jmlr.org/papers /volume7/demsar 06a/demsar06a.pdf

Dolnicar, S., B. Grün, and F. Leisch (2018). Market Segmentation Analysis: Understanding It, Doing It, and Making It Useful. DOI: 10.1007/978-981-10-8818-6_2.

Donoho, D.L. and P. J. Huber (1983). The notion of breakdown point. *A Festschrift for Erich Lehmann*. Belmont, CA: Wadsworth. pp. 157–184.

Eaves, A and B. Kingsman (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of The Operational Research Society*. Issue 55. pp. 431-437. DOI: 10.1057/palgrave.jors.2601697.

Elman J. L. (1990). Finding structure in time. *Cognitive Science*. Vol. 14. Issue 2. pp. 179-211. DOI: ISSN 0364-0213.

Enderdata (2021). CHANGE IN DISTANCE TRAVELLED BY CAR. https://www.odyssee-mure.eu/publications/efficiency-by-sector/transport/distance-travelled-by-car.html. (Access: 25.06.2021)

European Commission (2021). DIRECTIVE 2007/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?Uri= CELEX:32007L0046&from=DE (Access: 21.06.2021)

European Parliament (2035). *EU ban on the sale of new petrol and diesel cars from 2035 explained*. https://www.europarl.europa.eu/news/en/headlines/economy/20221019STO 44572/ eu-ban-on-sale-of-new-petrol-and-diesel-cars-from-2035-explained (Access: 21.06.2021)

Evans J. W., D. E. Kretschmann and D. W. Green (2019). Procedures for Estimation of Weibull Parameters. General Technical Report. *US Department of Agriculture, Forest Service, Forest Products Laboratory*.

Ferreira A. J. and M. A.T. Figueiredo (2012). Efficient feature selection filters for high-dimensional data, Pattern Recognition Letters. Vol. 33. Issue 13. pp. 1794-1804. ISSN 0167-8655.

Fildes, R. and S. Makridakis (1988). Forecasting and loss functions. *International Journal of Forecasting.* Elsevier. Vol. pp. 545-550.

Fildes, R., K. Nikolopoulos, S. F. Crone, and A.. A. Syntetos (2008). Forecasting and Operational Research: A Review. *The Journal of the Operational Research Societ*y. Vol. 59, Issue 9. pp. 1150–72.

Forge, S. (2009). Forecasting quantitatively using micro/meso/macro-economics with scenarios for qualitative balance. foresight. DOI: 11. 43-60. 10.1108/14636680910936431.

Freund Y. and R. E Schapire (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*. Vol. 55. Issue 1, pp. 119-139. ISSN 0022-0000.

Friedrichs, J. (1990). Methoden empirischer Sozialforschung. Wiesbaden: Westdeutscher Verlag.

Frowein, B., N. Lang, F. Schmieg and G. Sticher (2014), Returning to Growth: A Look at the European Automotive Aftermarket. *Marktstudie, The Boston Consulting Group*. Berlin and München.

Galicia A., R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez (2019). Multi-step forecasting for big data time series based on ensemble learning. *Knowledge-Based Systems*. Vol. 163. pp. 830-841. ISSN 0950-7051,

Garcia Asuero, A., A. Sayago and G. González (2006). The Correlation Coefficient: An Overview. *Critical Reviews in Analytical Chemistry*. Vol. 36. pp. 41-59. DOI: 10.1080/10408340500526766.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*. Vol. 4. Issue 1. pp. 1–28.

Gaskin, S. P., A. Griffin, J. R. Hauser, G. M. Katz and R. L. Klein (2010). Voice of the Customer. *Wiley International Encyclopedia of Marketing*. Vol. 5.

Gelders, L. F. and P. M. Van Looy (1978). An Inventory Policy for Slow and Fast Movers in a Petrol-Chemical Plant: A Case Study. *Journal of the Operation Research Society*. Vol. 29. Issue. 9. pp. 867-874.

Ghobbar, A. A. and C. H. Friend (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Comput. Oper. Res.* Vol. 30. pp. 2097-2114.

Gilbert, R. O. (1987). Statistical Methods for Environmental Pollution Monitoring, Wiley, NY.

Glatzel C. (2021): Die Neuausrichtung der Supply Chain. *Das Magazin der Bundesvereinigung Logistik 2021*. Vol. 1.

Glorot X., A. Bordes, and Y. Bengio (2011). Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Vol. 15. pp. 315-323.

Goodfellow I., Y. Bengio, and A. Courville (2018). Deep Learning. Das umfassende Handbuch. Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze. *MITP*. 2018.

Grand View Research (2020). Automotive Aftermarket Industry Size, Share & Trends Analysis Report By Replacement Part (Tire, Battery, Brake Parts, Filters, And Others), By Distribution Channel, By Service Channel, By Certification, By Region, And Segment Forecasts, 2020 – 2030. https://www.grandviewresearch.com/industry-analysis /aftermarket-automotive-parts-market. (Access: 28.06.2021)

Guest, G., E. Namey, and K. Mckenna (2016). How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. Field Methods. Vol. 29.

Günay, M. E. (2016). Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. Energy Policy, Elsevier. Vol. 90. pp. 92-101.

Gutierrez R. S., A. O. Solis and S. Mukhopadhyay (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics.* Vol. 111. Issue 2., pp. 409-420, 2008.

Hair, J. F. Jr., R. P. Bush and D. J. Ortinau (2003). Marketing Research: Within a Changing Information Environment. 2nd edition. McGraw Hill: Boston.

Han, J., M. Kamber, and J. Pei (2012). Data mining. Concepts and techniques. 3rd edition. Waltham, MA: Morgan Kaufmann/Elsevier.

Hansmann, K-W. (1983) Kurzlehrbuch Prognoseverfahren: mit Aufgaben u. Lösungen, Wiesbaden: Gabler.

Hastie, T., R. Tibshirani and J. H. Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Vol. 2. pp. 1-758. New York: Springer.

Hecker F., J. Hurth and H.-G. Seeba (2012): Aftersales in der Automobilwirtschaft. München: Springer Automotive.

Hecker F., J. Hurth and H.-G. Seeba (2017): Aftersales in der Automobilwirtschaft. München: Springer Automotive.

Heinrich, B. and M. Klier (2011). Assessing data currency – A probabilistic approach. *Journal of Information Science.* Vol. 37. pp. 86-100.

Hennig, C. and M. Meila. (2015) Cluster Analysis: An Overview. *Handbook of Cluster Analysis*. New York: Chapman and Hall/CRC.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, TUM, (advisor J. Schmidhuber).

Hochreiter, S. and J. Schmidhuber (1997). Long Short-term Memory. *Neural computation*. Vol. 9. pp. 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA.

Hsu, D. (2017). Time Series Compression Based on Adaptive Piecewise Recurrent Autoencoder. DOI: *ArXiv* abs/1707.07961.

Hu, Q., J.E. Boylan, H. Chen and L. Ashraf (2018). OR in spare parts management: A review. *European Journal of Operational Research*. Vol. 266. Issue 2. pp. 395-414.

Huang, T., R. Fildes, and D. Soopramanien (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European journal of operational research*. Vol. 237. Issue 2. pp. 738-748. DOI: ISSN 0377-2217, ZDB-ID 243003-4

Huber, P. J. (1981) Robust Statistics. New York: John Wiley and Sons.

Huhtala, Y., J. Kärkkäinen and J. Toivonen (1999). Mining for similarities in aligned time series using wavelets. *Defense, Security, and Sensing*.

Hyndman R. J. (2006), Another look at forecast accuracy Metrics for Intermittent Demand, *Foresight: The International Journal of Applied Forecasting*. Vol. 4. pp. 43-46.

Hyndman R. J., and A. B. Koehler (2006). Another look at measures of forecast accuracy, *International Journal of Forecasting.* Vol. 22. Issue 4. pp. 679-688, DOI: ISSN 0169-2070.

Hyndman R. J., E. Wang and N. Laptev (2015). Large-Scale Unusual Time Series Detection. *International Conference on Data Mining Workshop* (ICDMW). pp. 1616-1619, DOI: 10.1109/ICDMW.2015.104.

Hyndman, R. J. and G. Athanasopoulos (2018). *Forecasting: Principles and Practice*. 2nd edition, OTexts: Melbourne, Australia.

Hyndman, R. J., A. B. Koehler, K. J. Ord and R. D. Snyder (2008). Forecasting with exponential smoothing: The state space approach. Berlin: Springer.

Indyk, P., N. Koudas, and S. Muthukrishnan (2000). Identifying representative trends in massive time series data sets using sketches. *Proceedings of the 26th International Conference on Very Large Data Bases*. pp. 363-372.

Jack, H. (2022). Chapter 8 - Reliability and system design. *Engineering Design, Planning, and Management.* 2nd Edition. pp. 293-329. Academic Press: Cambridge.

Januschowski, T., J. Gasthaus, Jan, Y. Wang, D. Salinas, V. Flunkert, M. Bohlke-Schneider, and L. Callot (2019). Criteria for classifying forecasting methods. *International Journal of Forecasting*. DOI: 36. 10.1016/j.ijforecast.2019.05.008.

John, G.H., R. Kohavi and K. Pfleger (1994). Irrelevant Feature and the Subset Selection Problem. *Proceedings of 11th International Conference on Machine Learning*. New Brunswick. pp. 121-129.

Jouni, P., J. Huiskonen, and T. Pirttila, (2011). Improving global spare parts distribution chain performance through part categorization: A case study. *International Journal of Production Economics*. DOI: 133. 164-171. 10.1016/j.ijpe.2010.12.025.

Kalpakis K., D. Gada and V. Puttagunta (2001). Distance measures for effective clustering of ARIMA time-series, *Proceedings 2001 IEEE International Conference on Data Mining*. pp. 273-280, DOI: 10.1109/ICDM.2001.989529.

Kareem B. and A.S. Lawal (2015): Spare parts failure prediction of an automobile under criticality condition. *Engineering Failure Analysis*. pp. 67-79.

Kendall, M. G. (1975). *Rank Correlation Methods.* 4th edition, Charles Griffin: London.

Keogh E. and J. Lin (2005). Clustering of time-series subsequences is meaningless: Implications for previous and future research. *Knowledge and Information Systems*. Vol. 8. Issue: 2. pp. 154–177.

Keogh E. and S. Kasetty (2002). On the need for time series data mining benchmarks: a survey and empirical demonstration. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '02). Association for Computing Machinery, New York, NY, USA. pp. 102–111.

Keogh E., K. Chakrabarti, M. Pazzani, and S. Mehrotra (2001). Locally adaptive dimensionality reduction for indexing large time series databases. *Proceedings of the 2001 ACM SIGMOD international conference on Management of data* (SIGMOD '01). Association for Computing Machinery, New York, NY, USA, pp. 151–162.

Keogh E., S. Chu, D. Hart David and M. Pazzani (2004). Segmenting Time Series: A Survey and Novel Approach. *Data Mining in Time Series Databases*. pp. 1-21.

Keogh, E. (2006). A decade of progress in indexing and mining large time series databases. *VLDB*. pp. 1268–1268.

Keogh, E. and C. Ratanamahatana (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*. Vol. 7. pp. 358-386. DOI: 10.1007/s10115-004-0154-9.

Keogh, E. and P. Smyth (1997). A Probabilistic Approach to Fast Pattern Matching in Time Series Databases. Association for Computing Machinery, New York, NY, USA. pp. 24-30.

Keogh, E., C. Ratanamahatana (2005). Exact indexing of dynamic time warping. Knowledge Information Systems. Vol. 7. pp. 358–386.

Keras API Reference (2022). Losses. https://keras.io/api/. (Access: 22.01.2022)

Kilger, C., and M. Wagner (2008). Demand Planning. *Supply Chain Management and Advanced Planning*. Springer, Berlin, Heidelberg.

Kilimci, Z. H., A. O. Akyuz, M. Uysal, S. Akyokus, M. Uysal, B. A. Bulbul, and M. A. Ekmis, (2019). An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain Complexity. *Machine Learning and Sentiment Analysis Lab (MacSent).* DOI: 10.1155/2019/9067367.

Kingma, D. and J. Ba (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*. DOI: arXiv:1412.6980.

Klug, F. (2010): *Logistikmanagement in der Automobilindustrie: Grundlagen der Logistik im Automobilbau*. Berlin/Heidelberg. Springer.

Kosky, P., R. Balmer, W. Keat, and G. Wise (2020). An Introduction to Engineering and Design. *Exploring Engineering.* 5th edition. pp. 229-257.

Kotsakos, D., G. Trajcevski, D. Gunopulos, and C. C. Aggarwal (2014). Time-Series Data Clustering. *Data Clustering*. 1st edition. pp.24-33.

Kotsakos, D., G. Trajcevski, D. Gunopulos, Dimitrios and C. Aggarwal (2018). Chapter 15: Time-Series Data Clustering. *Data Clustering*. DOI: 10.1201/9781315373515-15.

Kourentzes N. (2014). On intermittent demand model optimization and selection. *International Journal of Production Economics*. Vol. 156, pp. 180-190, ISSN 0925-5273.

Kraftfahrt-Bundesamt (1) (2021). Segmente 2021. https://www.kba.de/DE/Statistik/Fahrzeuge/Neuzulassungen/Segmente/n_segmente_node.html (Access: 21.06.2021)

Kraftfahrt-Bundesamt (2) (2021). Inländerfahrleistung 2021. https://www.kba.de/DE/Statistik/Kraftverkehr /VerkehrKilometer/vk_inlaenderfahrleistung/vk_inlaenderfahrleistung_node.html (Access: 28.06.2021)

Kraftfahrt-Bundesamt (3) 2021. Neuzulassungen 2021. https://www.kba.de/DE/Statistik/Fahrzeuge/Neuzulassungen/Jahresbilanz_Neuzulassungen/jahresbilanz_node.html?yearFilter=2019 (Access: 30.06.2021)

Kromrey H. (2002). *Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung*. 10th edition. VS Verlag für Sozialwissenschaften: Wiesbaden.

Kruskal J. B. and M. Liberman (1983). The Symmetric Time Warping Problem: From Continous to Discrete. *Vol. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, pp. 125–161.

Kuhn M. and K. Johnson (2013): *Applied Predictive Modeling*. Springer: New York.

Kumar, D., J. Bezdek, M. Palaniswami, S. Rajasegarar, C. Leckie and T. Havens (2015). A Hybrid Approach to Clustering in Big Data. *IEEE Transactions on Cybernetics*. Vol. 46.

Kwon, O., N. Lee and B. Shin (2014). Data Quality Management, Data Usage Experience and Acquisition Intention of Big Data Analytics. *International Journal of Information Management*. DOI: 34. 10.1016/j.ijinfomgt.2014.02.002.

Lang S., W. J. Steiner, A. Weber, P. Wechselberger (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *European Journal of Operational Research*. Vol. 246. Issue 1. pp. 232-241, ISSN 0377-2217.

Längkvist M., L. Karlsson, and A. Loutfi (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters.* Vol. 42. pp. 11-24.

Lee, J. and Y. Cho (2009). Demand forecasting of diesel passenger car considering consumer preference and government regulation in South Korea. *Transportation Research Part A: Policy and Practice*. Vol. 43. Issue 4. pp. 420-429.

Leitner J., and U. Leopold-Wildburger (2011): Experiments on forecasting behavior with several sources of information – A review of the literature. *European Journal of Operational Research*. Vol. 213. Issue 3. pp. 459-469.

Lengu, D., A. A. Syntetos and M. Z. Babai (2014). Spare parts management: Linking distributional assumptions to demand classification. *European Journal of Operational Research*. Vol. 235. pp. 624–635. DOI: 10.1016/j.ejor.2013.12.043.

Lessard, D. R. and P. Lorange (1977). Currency Changes and Management Control: Resolving the Centralization/Decentralization Dilemma. *The Accounting Review.* Vol. 52. Issue 3. pp. 628–637.

Liao, T. W. (2005). Clustering of Time-Series Data—A Survey. *Pattern Recognition.* Vol. 38, pp. 1857-1874.

Lim B., S. Ö. Arık, N. Loeff, and T. Pfister (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*. Vol. 37. Issue 4. pp. 1748-1764, ISSN 0169-2070.

Liu, H. and H. Motoda (2000). Feature Selection for Knowledge Discovery and Data Mining. *The Springer International Series in Engineering and Computer Science*. Kluwer Academic, USA. DOI: 10.1007/978-1-4615-5689-3.

Liu, Y, Z. Li, X. Hui, G. Xuedong and J. Wu (2010). Understanding of Internal Clustering Validation Measures. *International Conference on Data Mining*. pp. 911-916. DOI: 10.1109/ICDM.2010.35.

Locklin, S. (2020). Neglected machine learning ideas. *Deep Learning Daily*. https://scottlocklin.wordpress.com/2014/07/22/neglected-machine-learning-ideas/ (Access: 30.04.2020)

Lolli, F., R. Gamberini, A. Regattieri, E. Balugani, T. Gatos, and S. Gucci (2017). Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics.* Vol. 183. pp. 116-128.

Lossio-Ventura J. A., S. Gonzales, J. Morzan, H. Alatrista-Salas, T. Hernandez-Boussard, and J. Bian (2021). Evaluation of clustering and topic modeling methods over health-related tweets and emails. *Artificial Intelligence Medicine*. DOI. 10.1016/j.artmed.2021.102096.

Loukmidis, G. and H. Luczak (2006): Lebenszyklusorientierte Planungsstrategien für den Ersatzteilbedarf. *Erfolgreich mit After Sales Services: Geschäftsstrategien für Servicemanagement und Ersatzteillogistik*. Springer. Berlin u. a. pp. 251-270.

Lynham, J. (2018). *Principles of Microeconomics*. OpenStax Economics – Pressbooks.

Ma, S., R. Fildes and T. Huang (2015). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*. Vol. 249. DOI: 10.1016/j.ejor.2015.08.029.

Ma, X, T. Zhimin, W. Yinhai, Y. Haiyang, and W. Yunpeng (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*. Vol. 54. pp. 187-197. ISSN 0968-090X.

Mack, N, C. Woodsong, K. M. MacQueen, G. Guest and E. Namey (2005). Qualitative Research Methods: A Data Collector's Field Guide. *Family Health International* (FHI), USA.

Mack, N., C. Woodsong, K. Macqueen, G. Guest and E. Namey (2005). Qualitative Research Methods: A Data Collector's Field Guide. *Open Access Library Journal*. Vol 2. Issue 1.

Makridakis S (1993). Accuracy measures: theoretical and practical concerns, *International Journal of Forecasting*. Volume 9. Issue 4. DOI: 527-529, ISSN 0169-2070.

Makridakis S., E. Spiliotis, V. Assimakopoulos (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting.* Vol. 36. Issue 1. pp. 54-74.

Makridakis, S., E. Spiliotis and V. Assimakopoulos (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*. Vol. 13. DOI: 10.1371/journal.pone.0194889.

Malhotra, N. and D. Birks (2000). Marketing-Research – An Applied Approach. *European Edition*. Essex.

Mann N. R. (1968). Point and Interval Estimation Procedures for the Two-Parameter Weibull and Extreme-Value Distributions. *Technometrics*. Vol. 10. Issue: 2. pp. 231-256, DOI: 10.1080/00401706.1968.10490558.

Mann, H. B. (1945). Non-parametric tests against trend, *Econometrica.* Vol.13. pp. 163-171.

Martin, D., P. Spitzer and N. Kühl (2020). A New Metric for Lumpy and Intermittent Demand Forecasts: Stock-keeping-oriented Prediction Error Costs. *Proceedings of the 53rd Annual Hawaii International Conference on System Sciences*. DOI: arXiv:2004.10537.

Masini, R., M. Medeiros and E. Mendes (2020). Machine Learning Advances for Time Series Forecasting. *Econometrics*.

Meireles Pereira M., Enzo Morosini Frazzon (2021). A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains. *International Journal of Information Management*. Volume 57, DOI. 102165.

Mills, F. C. (1946). Elasticity of Physical Quantities and Flexibility of Unit Prices in the Dimension of Time. *Journal of the American Statistical Association*. Vol. 41. Issue 236. pp. 439-467. DOI: 10.1080/01621459.1946.10501890.

Minner, S. (2000): *Strategic Safety Stocks in Supply Chains*. Springer: Berlin.

Minto B. (2021): The Pyramid Principle: Logic in Writing and Thinking. *Financial Times Prent*. 3rd edition.

Mitsa, T. (2010). *Temporal Data Mining*. Chapman & Hall/CRC: Boca Raton, FL. DOI: 10.1201/9781420089776.

Mukhopadhyay, S., A. O. Solis and R. S. Gutierrez (2012). The Accuracy of Non-traditional versus Traditional Methods of Forecasting Lumpy Demand. *Journal of Forecasting*. Vol. 31. Issue 8. pp. 721-735.

Murray, P. W., B. Agard and M (2018). Market segmentation through data mining: A method to extract behaviors from a noisy data set. . *Computational Industrial Engineering.* Vol. 118 pp. 33-43.

Murray, P. W., B. Agard and M. Barajas (2018). Forecast of individual customer's demand from a large and noisy dataset. *Computational Industrial Engineering.* Vol. 109. pp. 233-252.

Murtagh, F. (2015). Cluster Analysis: An Overview. *Handbook of Cluster Analysis*. CRC Press LLC: New York. pp. 21-29.

Musalem, E. P. and R. Dekker (2005). Controlling inventories in a supply chain: A case study. *International Journal of Production Economics*. Issue 1, pp. 179-188.

Navneet V., L. Doina, L. Florin, and M. Fitore (2019). Demand Forecasting Using Random Forest and Artificial Neural Network for Supply Chain Management. *Computational Collective Intelligence: 11th International Conference*. Springer-Verlag, Berlin, Heidelberg, pp. 328–339.

Nemati A. K., A. Baboli, M. K. Shahzad, R. Tonadre (2017): Demand Forecasting for Irregular Demands in Business Aircraft Spare Parts Supply Chains by using Artificial Intelligence (AI). *IFAC-Papers OnLine*. Vol. 50. Issue 1. pp. 15221-15226.

Ng, A. (2023): *Data-centric AI*. https://datacentricai.org/. (Access: 06.05.2023).

Ng, A. and R. Sagar (2021). Big Data to Good Data: Andrew Ng Urges ML Community To Be More Data-Centric And Less Model-Centric. *The Belamy: People and Technology*. https://analyticsindiamag.com/big-data-to-good-data-andrew-ng-urges-ml-community-to-be-more-data-centric-and-less-model-centric/.

Nijs, V. R., G. Marnik, J.B. Dekimpe, E. M. Steenkamp, and D. M. Hanssens (2001). The Category-Demand Effects of Price Promotions. *Marketing Science.* Vol. 20, Issue 1. pp. 1–22.

Nikolopoulos, K. I., M. Z. Babai, K. Bozos (2016). Forecasting supply chain sporadic demand with nearest neighbor approaches. *International Journal of Production Economics*. Vol. 177, pp. 139-148, ISSN 0925-5273.

Niminet, L. A. (2013). Automotive Market – From A General To A Market Segmentation Approach Studies and Scientific Researchers. Economics Edition. Faculty of Economic Sciences, Issue 18.

Odette (2012): *Forecast Accuracy Measurement*. White Paper.

Ord K. (2001). Commentaries on the M3-Competition An introduction, some comments and a scorecard. *International Journal of Forecasting*. Vol. 17. Issue 4. pp.537–541.

Osborne, J. and A. Overbay (2004). The Power of Outliers (and Why Researchers Should Always Check for Them). *Practical Assessment, Research, and Evaluation*. Vol. 9. Article 6. DOI: https://doi.org/10.7275/qf69-7k43.

Özden G. A., S. Sayın, T. v. Woensel, J. Fransoo (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*. Vol. 36. Issue 10. pp. 12340-12348, ISSN 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2009.04.052.

Palacio Niño J. and F. Berzal (2019). Evaluation Metrics for Unsupervised Learning Algorithms. *CoRR*. https://doi.org/10.48550/arXiv.1905.05667.

Paparrizos J. and L. Gravano (2015). k-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. pp. 1855–1870.

Pegels, C. C. (1969). Exponential forecasting: Some new variations. *Management Science*, Vol. 15. Issue 5. pp. 311–315. DOI: http://dx.doi.org/10.1287/mnsc.15.5.311

Pfohl H.-C. (2018) Ersatzteillogistik. *Logistiksysteme*. Springer: Berlin. 9th edition. pp. 235-246.

Pfohl H.-C. (2018): Lagerhaltung (Lagerbestände). *Logistiksysteme*. Springer: Berlin. 9th edition. pp. 99-126.

Pinçe C., L. Turrini and J. Meissner (2021): Intermittent demand forecasting for spare parts: A Critical review. *Omega*. Vol. 105, 102513, DOI: https://doi.org/10.1016/j.omega.2021.102513.

Porst, R. (2014): Fragebogen. Ein Arbeitsbuch. 4th. edition. Springer VS Verlag: Berlin.

Prechelt, L. (1998). Early Stopping - But When? *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science.* Vol. 1524. Springer: Berlin, Heidelberg. DOI: https://doi.org/10.1007/3-540-49430-8_3

de Rego, J. and M. de Mesquita (2015). Demand forecasting and inventory control: A simulation study on automotive spare parts. *International Journal of Production Economics*. DOI: 161. 10.1016/j.ijpe.2014.11.009.

de Souto, M., A. Coelho, K. Faceli, T. C. Sakata, V. Bonadia and I. Costa (2012). A Comparison of External Clustering Evaluation Indices in the Context of Imbalanced Data Sets. *Conference: Neural Networks (SBRN).* pp. 49-54. DOI: 10.1109/SBRN.2012.25.

Regattieri, A., M. Gamberi, R. Gamberini and R. Manzini (2005). Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*. Vol. 11. Issue 6. pp. 426-431.

Rizand-Mure, S., T. Grenier, C. Guttmann and H. Benoit-Cattin (2016). Unsupervised time-series clustering of distorted and asynchronous temporal patterns. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* DOI: 10.1109/ICASSP.2016.7471879.

Roland Berger GmbH and HSH Nordbank AG (2018). *Konsolidierung im Europäischen KFZ-Markt: Transformation einer Branche: Übernahmen und Digitalisierung verändern die Wettbewerbslandschaft*. White Paper.

Romeijnders W., R. Teunter, W. v. Jaarsveld (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research.* Vol. 220. Issue 2. pp. 386–393.

Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*. Vol. 20, pp. 53-65. DOI: 10.1016/0377-0427(87)90125-7.

Russel S. and P. Norvig (2012). *Künstliche Intelligenz – Ein moderner Ansatz*. 3rd edition. Pearson: Munich.

Sagaert Y. R., A. E.-H. Aghezzafa, N. Kourentzes, B. Desmetc (2018): Tactical sales forecasting using a very large set of macroeconomic indicators. *European journal of operational research*. Vol. 264. Issue 2. pp. 558-569.

Salais, T., J. Saucedo, R. Rodriguez Aguilar and J. Vela-Haro. (2020). Demand Prediction Using a Soft-Computing Approach: A Case Study of Automotive Industry. *Applied Sciences*. Vol. 10. DOI: 829. 10.3390/app10030829.

Salinas D., V. Flunkert, J. Gasthaus, T. Januschowski (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*. Vol. 36. Issue 3. pp. 1181-1191. DOI: https://doi.org/10.1016/j.ijforecast.2019.07.001.

Samuelson P. A. (1998). *Economics: The Original 1948 Edition*. McGraw-Hill Professional: New York.

Schuster, M. and K. Paliwal (1997). Bidirectional recurrent neural networks. *Signal Processing IEEE*. Vol. 45. pp. 2673 - 2681. DOI: 10.1109/78.650093.

Schütz A. and T. Luckmann (2017). *Strukturen der Lebenswelt*. 2nd edition. UTB GmbH: Stuttgart.

Seidenschwarz W (2003). Target Costing. *Handbuch Preispolitik: Strategien – Planung – Organisation – Umsetzung*. Gabler: Wiesbaden. pp 437–453.

Sharkey, A.J.C. (1996). On combining artificial neural nets. *Connection Science: Special Issue on Combining Artificial Neural Nets. Ensemble Approaches*. Vol. 8. pp. 299–314.

Shmueli G., P. C. Bruce, I. Yahav, N. R. Patel, K. C. Lichtendahl Jr. (2017). *Data Mining for Business Analytics: Concepts, Techniques, and Applications*. Wiley:

Siami-Namini S., N. Tavakoli and S. Namin Akbar (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA,. pp. 3285-3292. DOI: 10.1109/BigData47090.2019.9005997.

Silver, E. A. (1980) Comments on the lognormal approximation of lead time demand. *Omega.* Vol. 8, p. 701.

Simangunsong, E., L. Hendry, and M. Stevenson (2012). Supply Chain Uncertainty: A Review and Theoretical Foundation for Future Research. *International Journal of Production Research*. Vol. 50. pp. 4493-4523. DOI: 10.1080/00207543.2011.613864.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*. Vol. 15. pp. 201-293.

Srivastava N., G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*. Vol. 15. Issue 56. pp. 1929–1958.

Stathakis D. (2009). How Many Hidden Layers and Nodes. International Journal of Remote Sensing. Vol. 30. Issue 8. pp. 2133-2147.

Statista (2021): https://www.statista.com/statistics/692853/average-annual-distance-covered-by-drivers-in-italy/ (Access: 28.06.2021)

Strijbosch, L.W.G. and J. J. A. Moors (2006). Modified normal demand distributions in (R, S)-inventory control. *European Journal of Operational Research*. Vol. 172. Issue 1. pp. 201-212.

Sungil K. and Heeyoung K. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting.* Vol. 32. pp. 669-679.

Sutskever I., J. Martens, G. Dahl, and G. Hinton (2013). On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning*. Vol. 28. Issue 3. pp. 1139-1147.

Sutskever I., O. Vinyals, and Q. V. Le (2014). Sequence to Sequence Learning with Neural Networks. *Conference on Neural Information Processing Systems (NIPS)*. Vol. 2. pp. 3104–3112.

Swamidass P. M. (2000): Seven Rights Of Logistics. *Encyclopedia of Production and Manufacturing Managemen*t. Springer: Boston, MA. DOI: https://doi.org/10.1007/1-4020-0612-8_871.

Syntetos A. A., J.E Boylan (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics.* Vol. 71, Issues 1–3. DOI: https://doi.org/10.1016/S0925-5273(00)00143-2.

Syntetos A., J. E. Boylan and R. Teunter (2011). Classification for Forecasting and Inventory. *Foresight: The International Journal of Applied Forecasting*. *International Institute of Forecasters.* Issue 20. pp. 12-17.

Syntetos Aris A., M. Z. Babai, J. E. Boylan, S. Kolassa, and K. Nikolopoulos (2016). Supply chain forecasting:theory, practice, their gap and the future. *European Journal of Operational*

Research. Vol. 252. Issue 1. pp. 1-26. ISSN 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2015.11.010.

Syntetos, A. and J. Boylan (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*. Vol. 21. pp. 303-314. DOI: 10.1016/j.ijforecast.2004.10.001.

Syntetos, A., M. Z. Babai, and E. Gardner (2015). Forecasting intermittent inventory demands: Simple parametric methods vs. bootstrapping. *Journal of Business Research*. Vol. 68. DOI: 10.1016/j.jbusres.2015.03.034.

Syntetos, Aris A. and J. E. Boylan (2006). Comments on the attribution of an intermittent demand estimator. *International Journal of Forecasting*. Vol. 22. Issue 1. pp. 201-201.

Syntetos, M., J. Boylan and J. D. Croston (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*. Vol. 56. DOI: 10.1057/palgrave.jors.2601841.

Tadikamalla, P. (1979). An Inspection Policy for the Gamma Failure Distributions. *Journal of Operations Research*. Vol. 30. pp. 77–80. DOI: https://doi.org/10.1057/jors.1979.9

Tadikamalla, P. R. (1978). Eoq Inventory Model for Items with Gamma Distributed Deterioration. *Aiie Transactions.* Vol. 10. pp. 100-103.

Taulli T. (2019): Artificial Intelligence Basics – A Non-Technical Introduction. Springer/APress: New York.

Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*. Vol. 19. Issue 4. pp. 715–725.

Tealab A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*. Vol. 3. Issue 2. pp. 334-340, ISSN 2314-7288. DOI: https://doi.org/10.1016/j.fcij.2018.10.003.

Tempelmeier, H. (2003). *Material-Logistik: Modelle und Algorithmen für die Produktionsplanung und -steuerung und das Supply Chain Management*. 5th edition. Springer: Berlin et. al.

Teunter, R., and S. Babangida (2009). On the bias of Croston's forecasting method. *European Journal of Operational Research*. Vol. 194. Issue 1. DOI: https://doi.org/10.1016/j.ejor.2007.12.001.

Trapero, J., N. Kourentzes and R. Fildes (2012). Impact of information exchange on supplier forecasting performance. *OMEGA the International Journal of Management Science*. Vol. 40. DOI: 738-747. 10.1016/j.omega.2011.08.009.

Treiblmaier, H. (2014): A classification framework for supply chain forecasting literature: *Conference: International Conference on Industrial Logistics.* pp. 11-13

Tripathi, S., A. Bhardwaj and E. Poovammal (2018). Approaches to Clustering in Customer Segmentation. *International Journal of Engineering & Technology*. Vol. 7. pp. 802-807.

Valiant L. (2014) *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books: New York.

Verband der Automobilindustrie e.V. – VDA 9000 (2010): *Measuring Aftermarket Forecast Accuracy*. White Paper.

Vermorel, J. (2018). *The Quantitative Supply Chain*. Lokad: Paris.

Vlachos M., D. Gunopulos and G. Das (2004). Indexing time-series under conditions of noise. *Data Mining in Time Series Databases*. World Scientific: Singapore.

Voss, H. (2006). Life Cycle Logistics: Der Weg zur Produktlebenszyklus-orientierten Ersatzteillogistik. *Schriftenreihe Logistik der Kühne-Stiftung*. Haupt Verlag: Bern.

Wang, X., K. Smith-Miles and R. Hyndman (2006). Characteristic-Based Clustering for Time Series Data. Data Min. *Knowledge Discovery*. Vol. 13. pp. 335-364. DOI: 10.1007/s10618-005-0039-x.

Weinstein, A. (2014). Segmenting B2B technology markets via psychographics: An exploratory study. *Journal of Strategic Marketing*. Vol. 22. DOI: 10.1080/0965254X.2013.876072.

Werbos P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. Thesis. Harvard University, Cambridge USA.

Wilker, H. (2010). *Weibull-Statistik in der Praxis: Leitfaden zur Zuverlässigkeitsermittlung technischer Komponenten.* Band 3. BoD: Noderstedt.

Willemain, T., C. Smart and H. Schwarz (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*. Vol. 20. pp. 375-387. DOI: 10.1016/S0169-2070(03)00013-X.

Williams T. M. (1984). Stock Control with Sporadic and Slow-Moving Demand. *The Journal of the Operational Research Society*. Vol. 35. Issue 10. pp. 939-948.

Williams, B. D., M. A. Waller, S. Ahire and G. D. Ferrier (2014). Predicting retailer orders with POS and order data: The inventory balance effect. *European Journal of Operational Research*. Vol. 232. Issue 3. pp. 593-600.

Wind, Y. and R. N. Cardozo (1974). Industrial market segmentation. *Industrial Marketing Management*. Vol. 3. Issue 3. pp. DOI: https://doi.org/10.1016/0019-8501(74)90025-X.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*. Vol. *6*. Issue 3. pp. 324–342.

Wöhe, G. and U. Döring (2010). *Einführung in die Allgemeine Betriebswirtschaftslehre*. 24th ed. Vahlen: München.

Wunsch, D. and R. Xu (2008): *Clustering*. John Wiley & Sons: New York.

XGBoost (2022). *XGBoost Documentation*. https://xgboost.readthedocs.io/en/stable/parameter.html (Access: 02.01.2022).

Xueheng, Q., L. Zhang, Y. Ren, P. N. Suganthan and G. Amaratunga (2014), Ensemble deep learning for regression and time series forecasting, *IEEE Symposium on Computational*

Intelligence in Ensemble Learning (CIEL), Orlando, FL, USA, pp. 1-6, DOI: 10.1109/CIEL.2014.7015739.

Yamane, T. (1967). *Statistics, An Introductory Analysis*. 2nd edition. New York: Harper and Row.

Yang Z., A. Zhang and A. Sudjianto (2021). Enhancing Explainability of Neural Networks Through Architecture Constraints. *IEEE Trans Neural Network Learning Systems*. Vol. 32. Issue 6. pp. 2610-2621. DOI: 10.1109/TNNLS.2020.3007259.

Yeung P. (1972). A Note on the Rules of Derived Demand. *The Quarterly Journal of Economics.* Vol. 86. Issue 3. pp. 511–517. DOI: https://doi.org/10.2307/1880811.

Zhang P. (2012). Neural Networks for Time-Series Forecasting. *Handbook of Natural Computing*. G. Berlin-Heidelberg. Springer-Verlag.

Zhang P., B. E. Patuwo and M. Hu (1998). Forecasting with artificial neural networks: The state of the art, *International Journal of Forecasting.* Vol. 14. Issue 1. pp. 35-62.

Zhang T., R. Ramakrishnan, and M. Livny. (1996). BIRCH: an efficient data clustering method for very large databases. *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*. Association for Computing Machinery, New York, NY, USA, pp. 103–114.

Zhang W., C. K. Chang, H.-I. Yang and H.-Y. Jiang (2010). A Hybrid Approach to Data Clustering Analysis with K-Means and Enhanced Ant-Based Template Mechanism. *International Conference on Web Intelligence and Intelligent Agent Technology*. pp. 390-397, DOI: 10.1109/WI-IAT.2010.197.

Zhao, Y. and M. Katehakis. (2006). On the structure of optimal ordering policies for stochastic inventory systems with minimum order quantity. *Probability in the Engineering and Informational Sciences*. Vol. 20. pp. 257 - 270. DOI: 10.1017/S0269964806060165.

Appendix

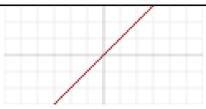
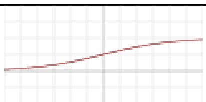
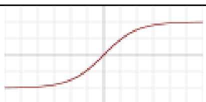
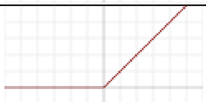
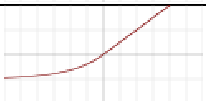
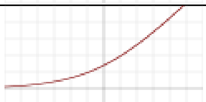
Appendix 1: Activation Functions in Neural Networks

Activation Functions, also known as *Transfer Functions,* are used to determine the output from a node respectively nodes in a layer of NNs. They can basically be divided into two types:

- Linear activation function: The range is defined as $]-\infty;\ \infty[$ i.e., the output of the functions is not confined in any range.
- Non-linear activation functions: When the activation function is non-linear, then a three-layer NN can be proven to be a universal function approximator, solving certain limitations of linear activation functions. This is known as the *Universal Approximation Theorem*.[75]

The latter group allows the stacking of multiple layers of neurons as the output would now be a non-linear combination of input passed through multiple layers.

Table E 1: Activation Functions and their Characteristics

NAME	PLOT	EQUATION	DERIVATIVE	RANGE	COMMENT
Identity (Linear)		$f(y) = y$	$f'(y) = 1$	]-∞; ∞[	used in output layer for regression problems
Sigmoid (Logistic; Softstep)		$f(y) = \frac{1}{1+e^{-y}}$	$f'(y) = f(y) \cdot (1 - f(y))$	[0;1]	equivalent to 2-element Softmax w/ second element = 2, often used in RNNs
hyperbolic tangens (tanh)		$f(y) = \tanh(x) = \frac{e^{y} - e^{-y}}{e^{y} + e^{-y}}$	$f'(y) = 1 - tanh(y)^2$	[-1;1]	zero-centered output
Rectified Linear Unit (ReLU)		$f(y) = \begin{cases} 0 \ if\ y \leq 0 \\ y \ if\ y > 0 \end{cases}$	$f'(y) = \begin{cases} 0 \ if\ y < 0 \\ 1 \ if\ y > 0 \end{cases}$	[0; ∞[	not to be used within hidden layers
Eponential Linear Unit (ELU)		$f(\alpha, y) = \begin{cases} \alpha(e^{y} - 1) \ if\ y \leq 0 \\ y \quad if\ y > 0 \end{cases}$	$f'(\alpha, y) = \begin{cases} \alpha e^{y} \ if\ y < 0 \\ 1 \ if\ y > 0 \end{cases}$	[-1; ∞[	same range as ReLU for non-negative inputs
Softplus		$f(y) = \ln(1 + e^{y})$	$f'(y) = 1/(1 + e^{-y})$	[0; ∞[	smooth version of ReLU

[75] The theorem is often cited to indicate that one hidden layer is theoretically sufficient and that there is no need for deep NNs. However, it does not give any statement about the expected number of neurons in the hidden layer. Depending on the function to be approximated, an enormously large number of neurons may be necessary. Here, adding more hidden layers often allows to achieve the same approximation quality with fewer neurons.

Appendix 2: Error Metrics and KPIs

The forecast error e_t is decomposed of the three elements *bias*, *variance*, and an irreducible error σ_e^2, making (12) correspond to (60). The bias is the difference between the model's average forecast and the corresponding truth y_t. Models with high bias pay very little attention to the training data and oversimplify the model, a phenomenon called *underfitting*, resulting in high errors on training and test data. Variance denotes the variability of f_t. A model with high variance pays high attention to within-sample data and does not generalize on out-of-sample data (*overfitting*). As a result, such models perform very well on training data but have high error rates on validation respectively test data.

$$e_t = \left(E(f_t - y_t)\right)^2 + E\left(\left(f_t - E(f_t)\right)^2\right) + \sigma_e^2 \tag{60}$$

Table E 2: Properties of Statistical Error Metrics and Business KPIs

		STATISTICAL PERSPECTIVE					BUSINESS PERSPECTIVE	
		Scale-independent	$y_t = 0$	Outlier Insensitivity	Symmetry	Variance-Bias-Trade-off	Interpreta-bility	Economic Considera-tions
Absolute Errors	MAE	×	✓	×	✓	×	✓	×
	MdAE	×	✓	×	✓	×	✓	×
Squared Errors	MSE	×	✓	×	✓	×	✓	×
	RMSE	×	✓	×	✓	×	✓	×
	GMSE	×	✓	×	✓	×	×	×
	GRMSE	×	✓	×	✓	×	×	×
Percentage Errors	MAPE	✓	×	×	×	×	✓	×
	MdAPE	✓	×	×	×	×	×	×
	sMAPE	✓	✓	×	×_✓	×	×	×
	MAAPE	✓	×_✓	×	×	×	×	×
	R^2	✓	✓	✓	✓	×	×	×
Relative Errors	MRAE	×	×	×	✓	×	✓	×
	MdRAE	×	×	×	✓	×	×	×
	GMRAE	×	×	×	✓	×	×	×
	MASE	✓	✓	✓	✓	×	✓	×
	FAI	✓	✓	×_✓	×	×	×_✓	×
	wFAI	✓	✓	×_✓	×	×	×_✓	✓
	MBA	✓	×_✓	✓	✓	×	✓	×
	TS	✓	✓	×	✓	✓	✓	×

Appendix 3: Customer and Product Hierarchies

Beside relating customers to sales channels and formats, a further categorization is made according to the organizational and legal structures of the customers. This is represented in Figure E 34.

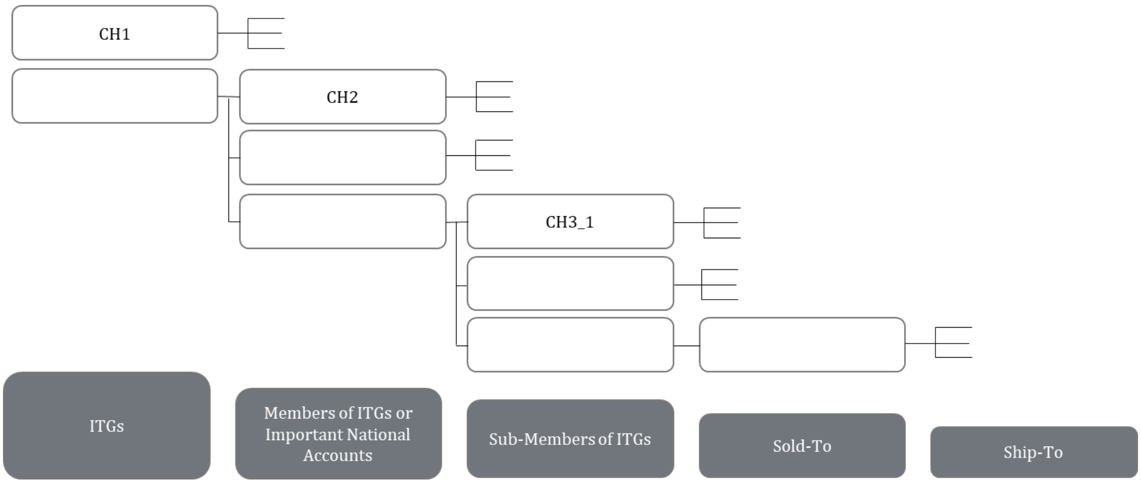


Figure E 1 Customer Hierarchies in the IAM

As with the customers, there is also an attribute-based categorization of parts into so called product families (cf. Figure E 2). In case that those families are to be defined from a technical point of view, they are called *product hierarchies* (PH). For the commercial point of view, they are referenced as *product groups* (PG). The inherent goal of the latter is to have a standard structure for price and condition management.

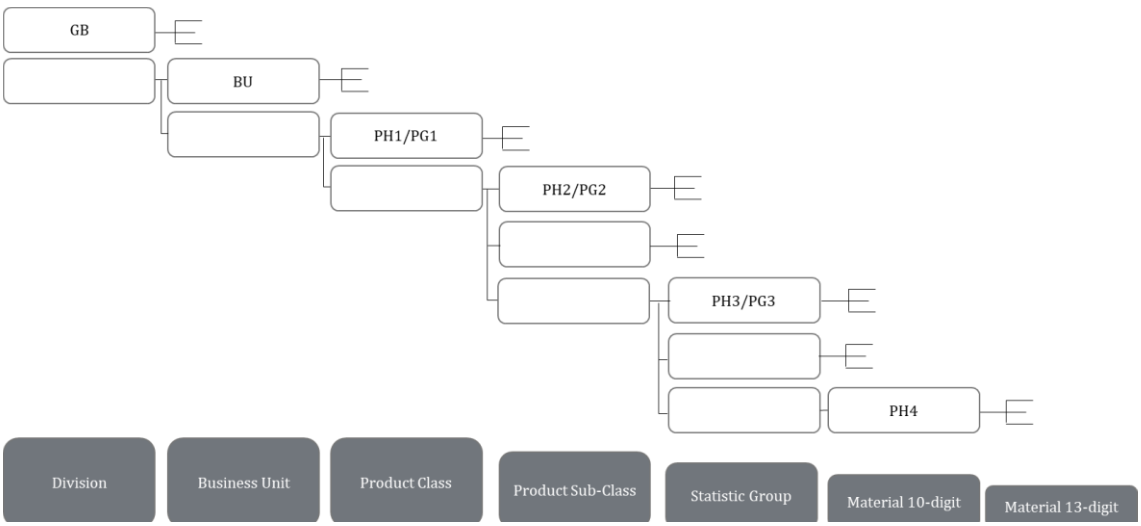


Figure E 2 Product Hierarchies in the IAM

Categories on the same level are disjunct, i.e., a SKU is exclusively assigned to a category.

Appendix 4: Data Management and Analytics: Distributions

Distributions have two functions in the field of forecasting: Firstly, it is *probabilistic* demand forecasting, meaning demand estimation by distributional assumptions, that requires that the actual probability function $P_y(y_t)$ is precisely specified, so that $P_y(Y = y_t) = p_t)$ is known for all values of Y. Instead of an empirical distributional function of the actual demand, a simplified approximation by a theoretical demand distribution is commonly used for this purpose. Furthermore, they give an indication how to define the objective i.e., the minimization function of the ML based algorithm.

Syntetos and Boylan (2021: 65) list four criteria for a good demand distribution, which were originally recommended by Boylan (1997) and Lengu et. al. (2014):

- empirical evidence in support of its goodness of fit,
- a priori grounds for its choice,
- flexibility to represent different types of demand patterns and
- computational ease

Accordingly, the valid selection of an appropriate distribution function requires the calculation of established statistical distribution tests of the chi-square goodness-of-fit test or the Kolmogorov-Smirnov goodness-of-fit test on the sample of demand. The hypotheses that are to be tested are as follows:

- $H0$: The sample data follow the hypothesized distribution.
- $H1$: The fitted distribution is statistically different to the observed data distribution.

For the present use case, the infrequent demand occurrences respectively the irregular demand sizes, when demand occurs, do not allow lead time demand to be represented by the Gaussian distribution. A different distribution may be more appropriate. Hence, apart from the commonly known and applied continuous distributions *norm* (Strijbosch und Moors 2006), *lognorm* (Tadikamalla 1979; Silver 1980) *t,* and *gamma* (Burgin und Norman 1976; Tadikamalla 1978), also the discrete distributions *poisson* (Axsater 2000; Tempelmeier 2003: 94) and *negative-binomial* (Minner 2000: 13) as well as the *poisson-gamma* compound distribution *Tweedie* were fitted to the frequency data of each series and then sorted according to their goodness of fit. The Tweedie distribution is a family of distributions that are a subset of the so-called *Exponential Dispersion Models* (EDMs). EDMs are two-parameter distributions from the linear exponential family that have a dispersion parameter φ and an additional shape parameter p^* with $p^* \in \mathbb{R}$.

This family of distributions has the following characteristics:

- a mean of $E(y_t)$,
- a variance of $var(y_t) = \varphi \cdot E(y_t)^{p^*}$ with p^* determining the sub-class of distributions within the family and
- as illustrated in Figure E 3 by a high point mass being concentrated at zero.

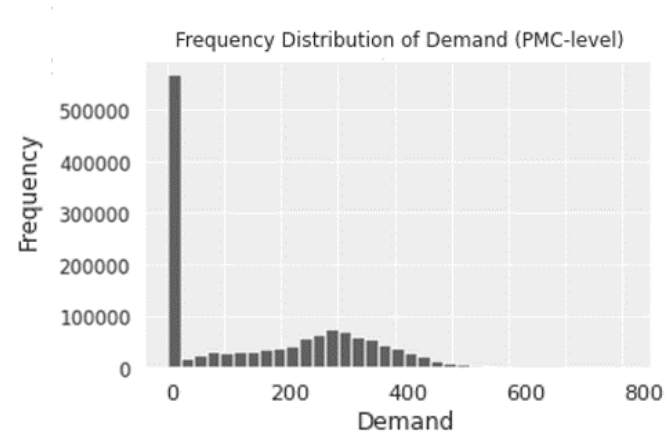


Figure E 3 Tweedie-like Distribution of AA Demand

Some familiar distributions are special cases of the Tweedie distribution. For

- p^* = 0: Gaussian distribution,
- p^* = 1: Poisson distribution,
- 1 < p^* < 2: Compound Poisson/gamma distribution,
- p^* = 2: gamma distribution.

To estimate the Tweedie parameter p^* for AA-specific demand data, a GLM using the Tweedie distribution can be fitted and used to derive the max log-likelihood at a certain parameter p^*. Table E 3 list the individual values derived from the analysis performed on monthly level:

Table E 3: Power Parameter p^*

LEVEL	VAR_POWER (p^*)	LogL
AS-IS	1.817	2.25261e+06
NORMALIZED	1.687	344581.927

With a proportion of 40.97%, Tweedie is the distribution, which generally approximates AA demand best. Negative Binomial amounts to 17.37 % whilst shares of other fittings do not exceed values of 10%.

One interesting insight, which is additionally gained, is that, in case of a min-max normalization of the dependent variable, the data is inflated to a Tweedie-like distribution even more. As the scaling of the data – especially when using a portfolio of time series – is regarded a mandatory precondition whilst preparing input variables for ML algorithms, close attention is to be set on the subsequently listed proportions for normalized data:

Table E 4: Distributional Fitting of AA Demand

	DISTRIBUTION	PROPORTION [%]
CONTINUOUS	norm	2.59
	lognorm	9.25
	t	7.63
	gamma	2.47
DISCRETE	poisson	5.75
	negative binomial	17.37
COMPOUND	tweedie_1.7	40.97

In a subsequent analysis, it is shown that Tweedie-distributed demands are majorly found in intermittent-lumpy time series, a minor share is also related to non-intermittent erratic ones:

Table E 5: Tweedie-distributed Demands per Demand Profile and Part Type

PART TYPE	INTERMITTENT-LUMPY	NON-INTERMITTENT-ERRATIC
COM	91%	9%
TP	98 %	2%

Appendix 5: Questionnaire for Focus Group Interviews

Fragebogen zur Diskussion der Nachfrageentwicklung im IAM und deren Einflussfaktoren

Sehr geehrter Teilnehmer, sehr geehrte Teilnehmerin,

herzlichen Dank für Ihre Bereitschaft, diesen Fragebogen auszufüllen.
Ich bin Doktorandin bei der Robert Bosch GmbH und beschäftige mich mit der Frage, unter Berücksichtigung welcher Daten und mithilfe welcher Algorithmen die Absatz- und Bestandsplanung im Oligopolmarkt verbessert werden kann.

Eine erste Forschungsfrage ist hierbei, welche Leitindikatoren die Individualnachfrage, d.h. die Kundennachfrage nach einem bestimmten Produkt, und damit auch die Gesamtnachfrage im IAM (Independent Aftermarket) beeinflussen.

Jede Interaktion mit Kunden führt zu einer Vielzahl von Datenpunkten – im System oder bei Ihnen im Gedächtnis, in denen Wissen über das Kundenverhalten verankert ist. Hiervon wertvolle Erkenntnisse für die Zukunft abzuleiten, ist eine große Herausforderung. Gleichzeitig bietet es aber auch neue Möglichkeiten, das Verhalten der Kunden genau(er) zu verstehen.

Die Ziele des Fragebogens gestalten sich demnach wie folgt:

- Abfrage der kollektiven Erfahrungswerte in Bezug auf das Kundenverhalten und die Nachfrage.
- Identifikation und Priorisierung von Indikatoren, welche das Kundenverhalten beeinflussen.
- Kategorisierung der Indikatoren gemäß der Attribute:

  - AA-aktiv z.B. Rabattaktionen
  - AA-reaktiv z.B. Zölle sowie

  - in bestimmten Abständen wiederkehrend, d.h. saisonal oder zyklisch
  - und längerfristig wirkend

- Hypothesengewinnung für eine nachfolgende quantitative Analyse.

Der Fragebogen umfasst die Sektoren A mit offenen Fragen sowie B und C mit skalierten Multiple-Choice-Fragen. Die Bearbeitungsdauer beträgt etwa 30 – 40 Minuten.

Alle Daten werden selbstverständlich anonym und streng vertraulich behandelt. Sie werden ausschließlich zur Analyse des oben dargestellten Sachverhalts in der Dissertation verwendet, ohne etwaige Rückschlüsse auf ein einzelnes Unternehmen oder eine Person zu ermöglichen.

Persönliche Angaben

Fokusgruppe I / II / III:

Abteilung: ☐ Logistik
☐ Vertrieb
☐ andere

Position:
Seit:

Verantwortungsbereich:

Verantworten Sie die Betreuung eines / mehrerer Märkte (~Regionen)?

Wenn ja, welche(n)?

Betreuen / Verantworten Sie ein bestimmtes Produkt / eine bestimmte Produktgruppe / eine Business Unit?

☐ ja ☐ nein

Wenn ja, welche?

Schlüsselfragen

(AA) Welche Indikatoren beeinflussen nach Ihrem Empfinden die Gesamtnachfrage im IAM?

(AB) Welche Indikatoren beeinflussen nach Ihrem Empfinden das individuelle Kundenbestellverhalten und damit die Gesamtnachfrage im IAM?

(B) + (C) Im Nachfolgenden werden verschiedene Kategorien sowie Einzelindikatoren genannt. Bitte wählen Sie nach Ihrem Empfinden, wie signifikant die Kategorie bzw. der Einzelindikator die Kundennachfrage nach Produkten aus den Bereichen 'Technical Parts' bzw. 'Commodities' beeinflusst.

Wählen Sie die Farbe 'blau' für 'Technical Parts'[76] sowie 'rot' für 'Commodities'[77].

Skala: 0 bis 5

0: Es gibt keinen Zusammenhang zwischen dem Kriterium und der Kundennachfrage.
1: Das Kriterium beeinflusst den Kunden in keinster Weise.
3: Das Kriterium beeinflusst den Kunden.
5: Das Kriterium beeinflusst den Kunden nachhaltig in seinem Bestellverhalten.

- o AA-aktiv
- o AA-reaktiv
- o wiederkehrend; zyklisch
- o längerfristig wirkend

Zusätzlich zu der Einschätzung bitte ich Sie, wenn möglich, eine Begründung für die Auswahl zu geben.

Ordnen Sie den Einzelindikatoren ebenso die Attribute 'AA-aktiv', 'AA-reaktiv', 'wiederkehrend/zyklisch' und/oder 'längerfristig wirkend' zu. Eine Mehrfachauswahl ist möglich.

[76] Hohe Produktkomplexität, spezifisches Know-How, zumeist Abdeckung aller, aber insbesondere der oberen Preissegmente (UHPP und HPP), wenige Anbieter
[77] Standardisierte Güter gekennzeichnet durch ein hohes Maß an Konformität und Fungibilität, wenig Differenzierungspotenzial

(B) Indikatorkategorien

(BA) Marktkomponente (~ Markt- und Umweltbedingungen)

- Makroökonomische Faktoren

ökonomisch	politisch	rechtlich	technisch	kulturell	ökologisch

- Mikroökonomische Faktoren

Marktbeschaffenheit und Marktentwicklung	Marktposition	Branchenstruktur und Marktteilnehmer

(BB) Absatzpolitische Instrumente

Preispolitische Instrumente	Produktpolitische Instrumente	Kommunkationspolitische Instrumente	Distributionspolitische Instrumente

(BC) Kundenkomponente

Firmographics	Psychographics

(BD) Produktkomponente

Produkt Status	Produktkategorie	Produktlebenszyklus	Produktbezogene Eigenschaften
(Fast-Mover vs. Slow-Mover)	(Service Verschleiß Unfall)	(wirtschaftlicher PLZ und technischer PLZ)	(Marke und Qualität; Teilekomplexität)

(BE) Zeitkomponente

Klimabedingte Saisonalität	Unternehmensinitiierte Saisonalität

(C) Einzelfaktoren

(CA) Marktkomponenten (~ Markt- und Umweltbedingungen)

- Makroökonomische Faktoren

ökonomisch	politisch	rechtlich
- Volkswirtschaftliche Leistungsfähigkeit - Währungseffekte - Konjunkturimpulse - Lohnkosten	- Zölle und Embargos	- kartellrechtliche Bestimmungen - Typgenehmigungs-Rahmenrichtlinien - Umweltgesetzgebung - Straßenverkehrsordnungen
technisch	kulturell	ökologisch
- Antriebsmix	- Mobility Mix - Trends (Sharing Economy, Mobility Shift) - Soziokulturelle Bedingungen (Feiertage, Urlaub)	- Klimamaßnahmen - atmosphärisches Bedingungen

- ökonomisch / politisch / rechtlich

- o AA-aktiv
- o AA-reaktiv
- o wiederkehrend; zyklisch
- o längerfristig wirkend

- technisch

- o AA-aktiv
- o AA-reaktiv
- o wiederkehrend; zyklisch
- o längerfristig wirkend

- kulturell

- o AA-aktiv
- o AA-reaktiv
- o wiederkehrend; zyklisch
- o längerfristig wirkend

- ökologisch

- o AA-aktiv
- o AA-reaktiv
- o wiederkehrend; zyklisch
- o längerfristig wirkend

(CA) Marktkomponenten (~ Markt- und Umweltbedingungen)

- ### Mikroökonomische Faktoren

Marktbeschaffenheit und Marktentwicklung	Marktposition	Branchenstruktur und Marktteilnehmer
- Marktform und Markteintritte/ -austritte - Marktvolatilität	- Markt- und Absatzvolumen; Marktanteil - Markt- und Absatzpotential	- Vertriebskanal (IAM – OES) - Distributions- und Vertriebsstruktur (Stufenmodell) - Kundenstruktur - Wettbewerbsstruktur

- Marktbeschaffenheit und Marktentwicklung

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Marktposition

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Branchenstruktur und Marktteilnehmer

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

(CB) Absatzpolitische Instrumente

Preispolitische Instrumente			Produktpolitische Instrumente	Kommunikationspolitische Instrumente	Distributionspolitische Instrumente	
Preissegment und Preise	Concessions	Revenue Reductions	Value Added Services	Kampagnen und Aktionen	Service Level	Vertragsvereinbarungen - MOQ - Umsatzziele

- Preissegment und Preise

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

- Concessions (verrechnete, ausgewiesene Rabatte)

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

- Revenue Reductions (nachgelagerte Aktionsgutschriften und Boni)

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

- Value Added Services

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

- Kampagnen und Aktionen

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

- Service Level

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

- Vertragsvereinbarungen (Zielumsatz)

0 1 2 3 4 5

  - o AA-aktiv
  - o AA-reaktiv
  - o wiederkehrend; zyklisch
  - o längerfristig wirkend

(CC) Kundenkomponente

Firmographics		Psychographics	
- Größe und Hierarchie - Distributions- und Vertriebsstruktur	ERP-System	- Bestellpolitik der Einkaufskooperationen und Großändler - Kaufverhalten Endkonsumenten	Verhandlungsstärke und Kommunikation

- Firmographics: Größe und Vertriebsnetzwerk

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Firmographics: ERP-Sytem

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Psychographics: Bestellpolitik und Einkaufsverhalten der Endkonsumenten

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Psychographics: Verhandlungsstärke und Kommunikation

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

(BD) Produktkomponente

Produktstatus	Produktkategorie	Produktlebenszyklus		Produktbezogene Attribute	
Fast-Mover vs. Slow-Mover	Service, Unfall, Verschleiß	Wirtschaftlich: phase-in, maturity phase-out	Technisch: Verschleißlebenszyklus / Laufleistung	Marke Qualität	Teilekomplexität

- Produktstatus

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Produktkategorie

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Wirtschaftlicher PLZ

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Technischer PLZ

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Produktbezogene Attribute: Marke und Qualität

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Produktbezogene Attribute: Teilekomplexität

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

(BE) Zeitkomponente

Klima-bedingte Saisonalität	Unternehmensinitiierte Saisonalität
	Zielvereinbarungen bzw. Zielperioden

- Klimabedingte Saisonalität

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

- Unternehmensinitiierte Saisonalität

0 1 2 3 4 5

  - AA-aktiv
  - AA-reaktiv
  - wiederkehrend; zyklisch
  - längerfristig wirkend

Abschließende Bemerkungen

Gibt es noch zusätzliche Aspekte, Ideen oder Einwände oder gibt es etwas, das Ihrem Empfinden nach vergessen bzw. vernachlässigt wurde?

Fazit und Empfehlung

Questionnaire to Discuss Demand Developments and their Indicators in the IAM

Dear Participant,

Thank you very much for your willingness to fill out this questionnaire.

I am a doctoral candidate at Robert Bosch GmbH and am working on the question of which data and algorithms can be used to improve sales and inventory planning in an oligopoly market.

A first research question here is which leading indicators influence individual demand, i.e., customer demand for a specific product, and thus also total demand in the IAM (Independent Aftermarket).

Every interaction with customers leads to a large number of data points - in the system or in your memory - in which knowledge about customer behavior is anchored. Deriving valuable insights for the future from this is a major challenge. At the same time, however, it also offers new opportunities to understand customer behavior more (more) precisely.

The objectives of the questionnaire are therefore as follows:

- Inquiry of the collective empirical values regarding customer behavior and demand.
- Identification and prioritization of indicators that influence customer behavior.
- Categorization of indicators according to attributes:

    - AA-active e.g., discounts
    - AA-reactive e.g., duties as well as

    - recurring at certain intervals, i.e., seasonal or cyclical
    - and effective in the long term

- Hypothesis generation for subsequent quantitative analysis.

The questionnaire comprises sectors A with open questions and B and C with scaled multiple-choice questions. The completion time is approximately 30 - 40 minutes.

All data will of course be treated anonymously and in strict confidence. They will be used exclusively for the analysis of the facts presented above in the dissertation, without allowing any conclusions to be drawn about an individual company or person.

Personal Details

Focus Group I / II / III: Participants: Duration:	
Department: Position: since: Main Responsibilities: Are you responsible for the support of one / several market(s) or regions? If yes, for which market(s)? Are you responsible for a specific product, product group or business unit? ☐ yes ☐ no If yes, for which products, product groups or business units?	

Key Questions

(AA) Which indicators do you think influence the overall demand in the IAM?

(AB) Which indicators, do you think, influence individual customer order behavior and thus overall demand in the IAM?

(B) + (C) In the following, various categories as well as single factors are listed. Please rate according to your own perception how significantly these categories and factors influence customer demand for products from the two sectors 'Technical Parts' and 'Commodities'.

Please choose color ‘blue‘ for ‘Technical Parts’ and ‘red‘ for ‘Commodities’.

Scale: 0 bis 5

0: There is no correlation between the criterion and customer demand.
1: This criterion does not influence the customer in any way.
3: This criterion influences the customer in his/her demand behavior.
5: This criterion has a lasting effect on the customer's ordering behavior.

- o managerial
- o random
- o structural
- o effective in the long-term

In addition to the assessment, I would ask you to give a reason for the selection in each case.

Please also map each factor with the named attributes ‘managerial’ (AA-active), ‘random’ (AA-reactive), ‘structural’ (seasonal or cyclical) and/or ‘effective in the long-term’.

(B) Indicator Categories

(BA) Market Component (~ Market- and Environmental Conditions)

- Macro-economic Factors

Economics	Politics	Legislature	Technology	Culture	Ecology

0 1 2 3 4 5

- Micro-economic Factors

Market Structures and Dynamics	Market Position	Industry Structure and Market Participants

0 1 2 3 4 5

(BB) Marketing Instruments

Price Instruments	Product-related Instruments	Communicational Instruments	Distributional Instruments

0 1 2 3 4 5

(BC) Customer Component

Firmographics	Psychographics

0 1 2 3 4 5

(BD) Product Component

Product Status (Fast-Mover vs. Slow-Mover)	Product Category (service wear-and-tear accident)	Product Life Cycle (economic and technical)	Product Attributes (brand and quality; complexity)

(BE) Time Component

Climate-related Seasonality	Business-initiated Seasonlity (target agreements)

(C) Single Indicators

(CA) Market Component (~ Market- and Environmental Conditions)

- Macro-economic Factors

Economics	Politics	Legislature
- economic performance - currency effects - economic stimulus - labor costs	- tariffs and embargos	- antitrust regulations - type-approval frameworks - environmental legislation - traffic regulations
Technology	**Culture**	**Ecology**
- powertrain mix	- mobility mix - trends (Sharing Economy, Mobility Shift) - socio-cultural conditions (holidays, vacation)	- climate measures - atmospheric conditions

- economic / political / legal

0 1 2 3 4 5

- o managerial
- o random
- o structural
- o effective in the long-term

- technical

0 1 2 3 4 5

- o managerial
- o random
- o structural
- o effective in the long-term

- cultural

0 1 2 3 4 5

- o managerial
- o random
- o structural
- o effective in the long-term

- ecological

0 1 2 3 4 5

- o managerial
- o random
- o structural
- o effective in the long-term

(CA) Market Component (~ Market- and Environmental Conditions)

- ### Microeconomic Factors

Market Structures and Dynamics	Market Position	Industry Structure and Market Participants
- market form and market entries / exits - market volatility	market- and sales volume; market share market- and sales potential	- sales channel (IAM – OES) - distribution and sales structure (tiered model) - customer structure - competitive structure

- market structures and dynamics

  - managerial
  - random
  - structural
  - effective in the long-term

- market position

  - managerial
  - random
  - structural
  - effective in the long-term

- industry structure and market participants

  - managerial
  - random
  - structural
  - effective in the long-term

(CB) Marketing Instruments

Price Instruments			Product-related Instruments	Communicational Instruments	Distributional Instruments	
price segment and prices	concessions	revenue reductions	value added services	campaigns	service level	contractual agreements - MOQ - sales agreements

- price segments and prices

  - managerial
  - random
  - structural
  - effective in the long-term

- concessions

  - managerial
  - random
  - structural
  - effective in the long-term

- revenue reductions

  - managerial
  - random
  - structural
  - effective in the long-term

- value added services

  - managerial
  - random
  - structural
  - effective in the long-term

- campaigns

  - managerial
  - random
  - structural
  - effective in the long-term

- service level

  - managerial
  - random
  - structural
  - effective in the long-term

- contractual agreements (sales targets)

- managerial
- random
- structural
- effective in the long-term

(CC) Customer Component

Firmographics		Psychographics	
size ownership and hierarchy distribution and sales structure	ERP-System	- order policy of purchasing cooperations and wholesalers - buying behavior of workshops and final customers	negotiating power and communication

- firmographics: size, ownerhisp and sales structure

- managerial
- random
- structural
- effective in the long-term

- firmographics: ERP-System

- managerial
- random
- structural
- effective in the long-term

- psychographics: order policy and Buying behavior

- managerial
- random
- structural
- effective in the long-term

- psychographics: negotiating power and communication

- managerial
- random
- structural
- effective in the long-term

(BD) Product Component

Product Status	Product Category	Product Life Cycle		Product Attributes	
Fast Mover Slow Mover	service wear-and-tear accident	economic: phase-in, maturity phase-out	technical (wear-and-tear): mileage	brand quality	degree of complexity

- product status

0 1 2 3 4 5

  - managerial
  - random
  - structural
  - effective in the long-term

- product category

0 1 2 3 4 5

  - managerial
  - random
  - structural
  - effective in the long-term

- economic PLC

0 1 2 3 4 5

  - managerial
  - random
  - structural
  - effective in the long-term

- technical PLC

0 1 2 3 4 5

  - managerial
  - random
  - structural
  - effective in the long-term

- product attributes: brand and quality

0 1 2 3 4 5

  - managerial
  - random
  - structural
  - effective in the long-term

- product attributes: degree of complexity

0 1 2 3 4 5

  - managerial
  - random
  - structural
  - effective in the long-term

(BE) Time Component

Climate-related Seasonality	Business-initated Seasonality
	target periods

- climate-related seasonality

  - managerial
  - random
  - structural
  - effective in the long-term

- business-initiated seasonality

  - managerial
  - random
  - structural
  - effective in the long-term

Final Remarks

Are there any additional aspects, ideas or objections? Is there anything that has been forgotten according to your opinion?

Do you have any conclusions or recommendations?

Appendix 6: Item Batteries and Items

Subsequently, all answers provided to questions 1 and 2 from the questionnaire are displayed. In order to identify the item categories, the deductive approach was chosen i.e., clusters are derived from the individual responses. The respective number per item category is given in parentheses. As interviewees were to answer the questions in German or English, responses are provided in either of the two.

Table E 6: Item-Set related to Influences on Individual Demand

Influencing Factors on Individual Demand (55)

Mikroökonomische Faktoren und Entwicklungen:

Marktteilnehmer - Anbieter (13)	**Marktteilnehmer - Nachfrager (20)**

• Branding: Marke und Qualität
 	
• Marktposition / -anteil
 	
• Interne Ziele / Business Plan
Basis für Konditionssysteme und Incentives
 	
• Geschäftsbeziehung
  • Langfristige Zusammenarbeit
  • Kundenbindung:
    o membership in 'Business Clubs'
    o 'exclusive supplier' status
  • Zuverlässigkeit
 	
• Breite und Tiefe des Produktportfolios – Abdeckung (Coverage):
Kunden legen Wert auf das Vorhandensein von C-Teilen (~Verfügbarkeit)
Möglichkeit des "Cherry-Pickings" bei großem Lieferantenportfolio
 	
• Verfügbarkeit
• Lieferperformance (log. Service-Level v.a. bei Angebotspolypol/Multi-Sourcing)
 	
• Preisstrategie und Preissegmente (UHPP – HPP – MPP – LPP)
  • Trade-Off zw. Wettbewerbsfähigkeit und Margenpotenzial
 	
• Distributionspolitik bzw. Vertriebskanäle: Channel (e.g., E-Commerce)
	• Marktposition / -anteil
	
	• Unternehmensgröße
	• Unternehmensstruktur und Hierarchie (Organisationsstruktur – Zugehörigkeit zu Int. Handels- und Einkaufskooperationen)
	• Distributionspolitik - Vertriebsstruktur
	Handelsstruktur: Preise in benachbarten Ländern / Märkten
	Vertriebskanäle
	Filialeröffnung
	
	• Verhandlungsebene
	• Verhandlungsstärke
	• Verhandlungskomplexitäten
	• Kommunikation, Offenheit
	
	• ERP- / IT-System
	• Datenmanagement - Information-Sharing
	
	• Bestell- und Bestandspolitik
	• Umsatzstärke Buy-in und Lagerkapazitäten
	• Umsatzstärke – Sell-out
	
	• Kaufkraft der WS-Kunden
	• Reparaturverhalten der WS-Kunden (Liebhaber/Statussymbol vs. Gebrauchsgegenstand)
	• Markenbewusstsein der WS-Kunden
	
	• Produktportfolio - Abdeckung – Größe des Portfolios (Breite und Tiefe)
	• Beschaffungselemente

Marktteilnehmer - Wettbewerber (6)	**Produkt (16)**

• Performance der Wettbewerber
  • Produktportfolio – Abdeckung (Coverage)
  • Verfügbarkeit
  • Lieferperformance (log. Service-Level)
	• Produktart: Fast Mover – Slow Mover
	  • Fahrzeugpopulation und Altersstruktur
	  • Laufleistung

• Lieferanten-(Wettbewerber)ausfälle: Second-Source-Nachfragen
 	
• Preissegmente der Wettbewerber
• Preise der Wettbewerber
	• Marke und Qualität
	• Komplexität ~ Wertigkeit
	• Produktlebenszyklus: Phase-in / Maturity / Phase-out
	
	• Verträge bzw. Vertragsvereinbarungen
	• Preise und Preisanpassungen
	• Konditionssysteme bzw. kommerzielle Vereinbarungen
	  • Concessions /
	  • Revenue Reductions (Cashbacks)
	  • Anfragen über Sonderpreise inkl. Mengenangabe
	  • Aktionen (Frequenz und Laufzeit) im GH und in WS
	
	• Gestaltung der Zielvereinbarungen (Quartalsziele / Halbjahresziele / Jahresziele)
	
	• Leistungspaket
	  • After-Sales-Service
	  • Vertragsstrafen – CSL
	  • Service-Angebot (technische Unterstützung, Sell-out Unterstützung)
	• Marketingmaßnahmen – Werbung
	
	• Minimum Order Quantity (MOQ)

Table E 7: Item-Set related to Influences on Total Demand

Influencing Factors on Total Demand (33)	
Makroökonomische Faktoren und Entwicklungen (14)	**Mikroökonomische Faktoren und Entwicklungen (14)**

• Politische Entscheidungen z.B. Brexit (= Bevorratungseffekt)
 	
• Konjunktur bzw. wirschaftliche Lage allgemein
 	
• Ausbildungnivau:
Ausbildungsgrad der Workshop Mitarbeiter: Teilekomplexität ↑; Wertigkeit und Anzahl der Reparaturaufträge ↑; Sortimentenvielfalt ↑
 	
• Lohnkosten der WS Mitarbeiter
Lohnniveau:
Wertigkeit und Anzahl der Reparaturen ↓; Sortimentenvielfalt ↓
 	
• Währungseffekte
Wechselkursschwankungen
(= Bevorratungseffekte)
 	
• Staatsökonomische bzw. wirtschaftspolitische Entscheidungen
  • Konjunkturimpulse (Subventionen etc. *stimuli cheque*; Abwrackprämie)
  • Zölle
  • Embargos
	• Marktform: Angebotspolypol vs. Angebotsoligopol/-monopol
	
	• Marktvolumen
	
	• Marktstrukturen – länderspezifische sowie regionale Cluster
	
	• Wettbewerbssituation IAM – OES
	
	• Channelvielfalt /-variabilität
	Direktvertrieb vs. indirekter Vertrieb
	Multichannel-Konzept: Wholesale, Retail, e-tail, Workshop
	
	• Marktteilnehmer:
	  • Kundenstrukturen und Kundenmix
	    • Konzentrationsprozesse:
	Entwicklung vom traditionellen Groß- und Einzelhändler zu Einkaufskooperationen
	Lieferantenportfolio ↑: Verhandlungsposition erstärkt.
	Cherry-Picking
	    • M&A, Markteintritte und –austritte

    - Anbieter und Wettbwerber
        - M&A, Markteintritte und –austritte
        - Wettbewerb durch Copy-Anbieter

- Gesetzgebung
    - Kartellrechtliche Bestimmungen
    - “Right to Repair”-Act
    - Klimaschutzmaßnahmen

- Wetter
- Klima (Winter, Überflutungen …)

- Soziokulturelle Gegebenheiten:
    - Überregionale Ferien- und Urlaubszeiten (Schließtage)

- Marktverzerrungen:
    - Streik
    - Transportmittelengpässe
    - Ressourcenknappheit
    - Schließtage
- Technische / industrielle Neuerungen: Veränderung des Portfolios (~ u.a. Korrelation zu Klimaschutz- und Umweltmaßnahmen)

Produkt (3)

- Produktart: Fast Mover – Slow Mover definiert durch
    - Fahrzeugpopulation: Typ, Anzahl und Altersstruktur) je Markt
    - Laufleistung

- Ausfallraten der Komponenten

Zeit (2)

- Saisonalität – klimabedingt
- Saisonalität – unternehmensinitiiert durch Zielvereinbarungen

The list of item batteries and items was concluded both from the literature research and iteratively by means of the expert interviews. The presentation and structure correspond to the structure of the questionnaires.

Table E 8: General Factors - Macroeconomics

Economics	Politics	Legislature
• economic performance • economic stimulus • currency effects • labor costs	• tariffs and embargos	• antitrust regulations • type-approval frameworks • environmental legislation • traffic regulations
Technology	**Culture**	**Ecology**
• powertrain mix • high-grade fittings	• mobility mix • trends e.g., sharing economy and mobility shift • socio-cultural conditions (holidays, vacation)	• climate measures • atmospheric conditions

Table E 9: General Factors - Microeconomics

Market Structures and Dynamics	Industry Structure and Market Participants	Market Position
• market form and market entries respectively exits • market volatility	• sales channel (IAM – OES) • distribution and sales structure (tiered model) • customer structure • competitive structure	• market- and sales volume; market shares • market- and sales potential

Table E 10: Specific Factors - Microeconomics (Customer)

Firmographics		Psychographics	
• size • ownership and hierarchy • distribution and sales structure	• ERP-System	• order policy of purchasing cooperations and wholesalers • buying behavior of workshops and final customers	• negotiating power and communication

Table E 11: Specific Factors - Microeconomics (Product Attributes)

Product Status	Product Life Cycle	
• Fast-Mover Slow-Mover	• economic phase-in, maturity phase-out	• technical (wear-and-tear): mileage
Product Category	**Product Attributes**	
• service wear-and-tear accident	• brand • quality	• degree of complexity

Table E 12: Specific Factors - Microeconomics (Sales Promoting Measures)

Price Instruments			Product-related Instruments
• price segment and prices	• concessions	• revenue reductions	• value added services
Distributional Instruments			**Communicational Instruments**
• service level	• contractual agreements incentivized sales targets MOQ	• portfolio (coverage)	• campaigns

Table E 13: General Factors - Time

Natural	Business-initiated
• climate-related seasonality	• target periods

Appendix 7: Overview of Factors

The following two remarks help to increase the understanding of the tables:

- Features which are printed in grey, can be attributed to more than one domain.
- The two columns *FI* and *PI* are exclusively related to the features' relevance when used in forecasting models.

Table E 14: Business-induced Statistical Features and Demand Profiles

Nr.	MECE	Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Statistical Features based on Trans-actional Data	ZeroPeriods	Number of Months with zero-demands per year and PMC	4	metric - int	transactional - dynamic	HF	-	-
2.		NonZeroPeriods	Number of Month with no zero-demands per year and PMC	8	metric - int	transactional - dynamic	HF	-	-
5.		GrowthRate_SA_PMC … level = [PMC, SKU, SalesCountry]	Percentage change of turnover from year to year on specified levels	-0.06	metric - float	transactional - dynamic	SF	-	-
8.		AvGrowthRate_SA_PMC … level = [PMC, SKU, SalesCountry]	Averaged Percentage change of turnover from year to year on specified levels	0.12	metric - float	transactional - dynamic	SF	-	-
9.		AvOrderCount	Average order count for specified SKU on yearly level	104	metric - float	transactional - dynamic	SF		
10.		AvCustCount	Average customer count per SKU on yearly level	12	metric - float	transactional - dynamic	SF		
13.		APFR_Year_PMC … level = [PMC, SKU, SalesCountry]	Average Purchase Frequency Rate; Yearly Order Count/Average Customer Count	8.666667	metric - float	transactional - dynamic	SF	-	-
16.		CusVal_PMC … level = [PMC, SKU, SalesCountry]	Customer Value (~Market Value); APV_Year/APFR_Year	4.953696e+05	metric - float	transactional - dynamic	SF	-	-
19.		Mean_SA_Year_PMC		75.83			HF/SF	-	-

	… level = [PMC, SKU, SalesCountry]	Mean Sales Quantity per Year and PMC		metric - float	transactional - dynamic			
22.	Mean_TO_Year_PMC (APV) … level = [PMC, SKU, SalesCountry]	Mean Turnover per Year and specified level (Average Purchase Value = total revenue/numer of orders)	824.65	metric - float	transactional - dynamic	HF/SF	-	-
25.	Mean_SA_PMC … level = [PMC, SKU, SalesCountry]	Mean Sales Quantity on PMC level	68.50	metric - float	transactional - dynamic	HF/SF	-	-
28.	Mean_TO_PMC (AvAPV) … level = [PMC, SKU, SalesCountry]	Mean Turnover on PMC level	732.16	metric - float	transactional - dynamic	HF/SF	-	-
31.	Std_SA_Year_PMC … level = [PMC, SKU, SalesCountry]	Standard Deviation of Sales Quantity per Year and PMC	12.5	metric - float	transactional - dynamic	HF/SF	-	-
34.	Std_SA_PMC … level = [PMC, SKU, SalesCountry]	Standard Deviation of Sales Quantity on PMC level	20.83	metric - float	transactional - dynamic	HF/SF	-	-
37.	CV_SA_Year_PMC … level = [PMC, SKU, SalesCountry]	Coefficient of Variance of Sales Quantity per Year and PMC	2.12	metric - float	transactional - dynamic	HF/SF	-	-
40.	CV_SA_PMC … level = [PMC, SKU, SalesCountry]	Coefficient of Variance of Sales Quantity on PMC level	3.54	metric - float	transactional - dynamic	HF/SF	-	-
43.	CV_Squared_Year_PMC … level = [PMC, SKU, SalesCountry]	Squared Coefficient of Variance per Year and PMC	4.46	metric - float	transactional - dynamic	HF/SF	-	-
47.	CV_Squared_PMC … level = [PMC, SKU, SalesCountry]	Squared Coefficient of Variance on PMC level	12.54	metric - float	transactional - dynamic	HF/SF	-	-
48.	ADI	Average Demand Interval	1.2	metric - float	transactional - dynamic	HF/SF	-	-
51.	AvOrderCount_PMC	Averaged number of	8	metric - float	transactional - dynamic	SF	-	-

			… level = [PMC, SKU, SalesCountry]	orders on PMC level per year						
52.	Product Status and Demand Profiles based on Trans-actional Data		ProductStatus1	Classification of PMC into R, I and S Parts (regular, irregular and sporadic) with $CV \in [0.6, 1.2;\ \infty]$	R	string	transactional - dynamic	MF/SF	-	NR
53.			ProductStatus2	Classification of PMC into Fast Mover and Slow Mover acc. to number of orderlines (turn rate)	FM	string	transactional - dynamic	MF/SF	-	NR
54.			ProductStatus3	Classification of PMC into A, B and C Parts based on turnover with thresholds [80-95-100]	A	string	transactional - dynamic	SF	-	R
55.			ProductStatus4	Classification of PMC into X, Y and Z Parts based on sales with thresholds [80-95-100]	X	string	transactional - dynamic	SF	-	-
56.			ProductLifeCycle	Categorization of PMC into the *Phase-in*, *Growth*, *Maturity, Decline* and *Phase-out*	Growth	string	transactional - dynamic	SF	-	R
57.			DemandProfile1	Classification of PMC into intermittent-non-intermittent	intermittent	string	transactional - dynamic	SF	-	-
58.			DemandProfile2	Classification of PMC into clumped-lumpy-smooth-erratic	lumpy	string	transactional - dynamic	SF	-	-

Table E 15: Transactional Information

Nr.	MECE		Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Micro-economics	Date	CustomerDueDate	Delivery Date required by Customer (CDD)	2021-01-01	string	transactional - dynamic	MF	R	R
2.			activeyear_dt	Year of transaction	2021	string	transactional - dynamic	HF	-	-
3.			activemonth_dt	Month of transaction	01	string	transactional - dynamic	HF	-	-

4.			PMC_Transactiontime	Time Index for Delivery Date required by Customer (CDD)	123	metric-int	transactional - dynamic	HF	-	-
5.		Market	SalesRegion	Sales Region of Destination	SEC	string	transactional - static	SF	NR	NR
6.			SalesCountry	Sales Country of Destination	DE	string	transactional - static	SF	R	R
7.			Warehouse	Warehouse-ID related to Country of Departure	W123	string	transactional - static	MF	R	R
8.		Product	Mat10_ID	SKU Identification No (10-digit)	1234567890	string	transactional - static	MF	R	R
9.			rg_sub_ID	ID for successors of specific SKU in a certain region	234567891	string	transactional - static	HF	NR	NR
10.			PMC	Mat_Warehouse_ SalesCountry	1234567890_ W123_DE	string	transactional - static	HF	R	R
11.		Historical Sales Quantity	ActualSales	Invoiced Volume (Sales Quantity) for a SKU in a specific SalesCountry at CDD	4300	metric-int	transactional - dynamic	HF	NR	NR
12.			LostSales	Volume (Sales Quantity) of specific SKU the customer would have ordered if availabilty in a specific SalesCountry at CDD had been given	500	metric-int	transactional - dynamic	HF	NR	NR
13.			SalesQuantity	Sum of ActualSales and LostSales of a specific SKU in a specific SalesCountry at CDD	4800	metric-int	transactional - dynamic	MF	R	R

Table E 16: Transactional Information - Portfolio

Nr.	MECE	Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Transactional Features – Portfolio Information	item_category3ID_avg_ite m_cnt_month	Company's Industry Volume on Average (acc. to ActualSales) for specific product family (Level 3) in a specific market at CDD	150	metric-int	transactional - dynamic	MF	R	NR
2.		item_category2ID_avg_ite m_cnt_month	Company's Industry Volume on Average (acc. to ActualSales) for specific product family (Level 2) in a	300	metric-int	transactional - dynamic	MF	R	NR

			specific market at CDD						
3.		item_category1ID_avg_item_cnt_month	Company's Industry Volume on Average (acc. to ActualSales) for specific product family (Level 1) in a specific market at CDD	450	metric-int	transactional - dynamic	MF	R	NR
4.		BU_avg_item_cnt_month	Company's Industry Volume on Average (acc. to ActualSales) for specific product family (Level BU) in a specific market at CDD	550	metric-int	transactional - dynamic	MF	NR	NR
5.		item_category3ID_sum_item_cnt_month	Company's Industry Volume (acc. to ActualSales) for specific product family (Level 3) in a specific market at CDD	2000	metric-int	transactional - dynamic	MF	R	R
6.		item_category2ID_sum_item_cnt_month	Company's Industry Volume (acc. to ActualSales) for specific product family (Level 2) in a specific market at CDD	4500	metric-int	transactional - dynamic	MF	R	NR
7.		item_category1ID_sum_item_cnt_month	Company's Industry Volume (acc. to ActualSales) for specific product family (Level 1) in a specific market at CDD	6200	metric-int	transactional - dynamic	MF	R	R
8.		BU_sum_item_cnt_month	Company's Industry Volume (acc. to ActualSales) for specific product family (Level BU) in a specific market at CDD	8500	metric-int	transactional - dynamic	MF	NR	NR
9.		YearlySalesQuantity	Aggregated SalesQuantity of a specific SKU in a specific SalesCountry on yearly level	16000	metric-int	transactional - dynamic	MF	R	R
15.	Lagged Transactional Features – Portfolio Information	item_category3ID_avg_item_cnt_month_lag_1 … lags = [1 ,2 ,3 ,4, 5, 6]	Lag Version of Company's Industry Volume on Average for specific product family level 3	1200	metric-int	transactional - dynamic	MF	R	NR
21.		item_category2ID_avg_item_cnt_month_lag_1	Lag Version of Company's Industry Volume	1200	metric-int	transactional - dynamic	MF	R	NR

	... lags = [1 ,2 ,3 ,4, 5, 6]	on Average for specific product family level 3						
27.	item_category1ID_avg_item_cnt_month_lag_1 ... lags = [1 ,2 ,3 ,4, 5, 6]	Lag Version of Company's Industry Volume on Average for specific product family level 3	1200	metric-int	transactional - dynamic	MF	R	NR
33.	item_category3ID_sum_item_cnt_month_lag_1 ... lags = [1 ,2 ,3 ,4, 5, 6]	Lag Version of Company's Industry Volume for specific product family level 3	5500	metric-int	transactional - dynamic	MF	R	NR
39.	item_category2_ID_sum_item_cnt_month_lag_1 ... lags = [1 ,2 ,3 ,4, 5, 6]	Lag Version of Company's Industry Volume for specific product family level 2	7000	metric-int	transactional - dynamic	MF	R	NR
45.	item_category1_ID_sum_item_cnt_month_lag_1 ... lags = [1 ,2 ,3 ,4, 5, 6]	Lag Version of Company's Industry Volume for specific product family level 1	8500	metric-int	transactional - dynamic	MF	R	NR

Table E 17: Customer-related Features

Nr.	MECE			Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Micro-economics	Market Participants – First Level Customers	Firmographics	ShipTo	Shipping Location of Customer – Contractual Partner	A12345678	string	transactional – static	HF	NR	NR
2.				SoldTo	Customer – Contractual Partner (reg. transaction and invoice)	A1234567	string	transactional – static	HF	NR	NR
3.				CH3	Customer Hierarchy 3	A123456	string	meta – static	HF	NR	NR
4.				CH2	Customer Hierarchy 2	A12345	string	meta – static	HF	NR	NR
5.				CH1	Customer Hierarchy 1	A1234	string	meta – static	HF	NR	NR
6.				CS3	Customer Segment 3	Special Wholesale rel. to specific product group	string	meta – static	HF	NR	NR
7.				CS2	Customer Segment 2	Special Wholesale	string	meta – static	HF	NR	NR
8.				CS1	Customer Segment 1	Wholesaler	string	meta – static	HF	NR	NR
9.				Dom_CS	Dominating Customer	b	string	meta – static	SF	-	-

		Segment in a SalesRegion						
12.	Top1_C11 … countries C = [1, 2, 3]	Sales of Top 1 Customer in Country 1	10	metric – int	transactional – dynamic	HF/MF	NR	NR
15.	Top2_C11 … countries C = [1, 2, 3]	Sales of Top 2 Customer in Country 1	7	metric – int	transactional – dynamic	HF/MF	NR	NR
18.	Top3_C11 … countries C = [1, 2, 3]	Sales of Top 3 Customer in Country 1	3	metric – int	transactional – dynamic	HF/MF	NR	NR
21.	Top1_C11_pct … countries C = [1, 2, 3]	Percentage Share of Sales of Top 1 Customer in Country 1	0.35	metric – float	transactional – dynamic	HF/MF	NR	NR
24.	Top2_C11_pct … countries C = [1, 2, 3]	Percentage Share of Sales of Top 2 Customer in Country 1	0.2	metric – float	transactional – dynamic	HF/MF	NR	NR
27.	Top3_C11_pct … countries C = [1, 2, 3]	Percentage Share of Sales of Top 3 Customer in Country 1	0.1	metric – float	transactional – dynamic	HF/MF	NR	NR
30.	TopCus_C11 … countries C = [1, 2, 3]	Cumulated Sales of Top Customers (amounting to 80% of turnover and sales) in Country 1	21	metric – int	transactional – dynamic	MF	R	NR
33.	TopCus_C11_pct … countries C = [1, 2, 3]	Cumulated Percentage Share of Sales of Top Customers (amounting to 80% of turnover and sales) in Country 1	21	metric – float	transactional – dynamic	MF	NR	NR
36.	B_C11 … countries C = [1, 2, 3]	Sales of Customer Segment B in Country 1	20	metric – int	transactional – dynamic	HF/MF	R	R
39.	C_C11 … countries C = [1, 2, 3]	Sales of Customer Segment C in Country 1	10	metric – int	transactional – dynamic	HF/MF	R	R
42.	E_C11 … countries C = [1, 2, 3]	Sales of Customer Segment E in Country 1	5	metric – int	transactional – dynamic	MF	R	R*
45.	F_C11	Sales of Customer	0			MF	NR	NR

	... countries C = [1, 2, 3]	Segment F in Country 1		metric – int	transactional – dynamic			
48.	B_C11_pct ... countries C = [1, 2, 3]	Percentage Share of Sales of Customer Segment B in Country 1	0.72	metric – float	transactional – dynamic	MF	R	R
51.	C_C11_pct ... countries C = [1, 2, 3]	Percentage Share of Sales of Customer Segment C in Country 1	0.13	metric – float	transactional – dynamic	MF	NR	NR
54.	E_C11_pct ... countries C = [1, 2, 3]	Percentage Share of Sales of Customer Segment E in Country 1	0.08	metric – float	transactional – dynamic	MF	NR	NR
57.	F_C11_pct ... countries C = [1, 2, 3]	Percentage Share of Sales of Customer Segment F in Country 1	0.07	metric – float	transactional – dynamic	MF	NR	NR
58.	unique_cus_month	Net Reach: Number of unique customers (CH2-Level) per Mat10 and Month	5	metric – int	transactional – dynamic	MF	R	R
59.	unique_cus_year	Net Reach: Number of unique customers (CH2-Level) per Mat10 and Year	12	metric – int	transactional – dynamic	MF	R	NR

Table E 18: Product-related Features

Nr.	MECE			Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Micro-econo-mics	Product	Product Category	PH3	Technical item category ID level 3	A1234 A1234000000 01	string	meta - static	HF/SF	-	-
2.				PH2	Technical item category ID level 2	A1234 A123400	string	meta - static	HF/SF	-	-
3.				PH1	Technical item category ID level 1	A1234	string	meta - static	HF/SF	-	-
4.				PG3	Commerical item category ID level 3	AA111222233 333333	string	meta - static	HF/SF	-	-
5.				PG2	Commerical item category ID level 2	AA1112222	string	meta - static	HF/SF	-	-
6.				PG1	Commerical item category ID level 1	AA111	string	meta - static	HF/SF	-	-

7.		BU	item category ID level Business Unit	AA-BBB	string	meta - static	HF/SF	-	-
8.		PartSector	Classification of SKU into AutoParts, WTE and Battery	AutoParts	string	meta - static	HF/SF	-	-
9.	Product Value and Comple-xity	PartType	Differentiation into Technical Parts and Commodities	COM	string	meta - static	MF/SF	R	R
10.		PPC-P_ AA_Mat10	Material and Production Costs on SKU level per piece or set	3.65	metric-float	meta - dynamic	MF/SF	R	NR
11.	Economic PLC	SOP	Start of Production for AA ~ first date of selling in the Automotive Aftermarket	01.01.2011	string	meta - static	HF	NR	NR
12.		LifeCycleAge_ AA	CustomerDueDate - SOP [No of years resp. months]	11	metric - int	transactional - dynamic	MF	R	NR
13.		EODOBL	End of Delivery Obligation for OES (ususally 10 to 15 years)	01.01.2015	string	meta - static	HF	NR	NR
14.		LifeCycleAge_ AAonly	CustomerDueDate [No of years resp. months]-EndOfDeliveryObligation	6	metric - int	transactional - dynamic	MF	NR	NR
15.	Marketing Instru-ments	MOQ	Minimum Order Quantity for SKU to be ordered by Customer	60	metric - int	meta - static	MF	R	NR
16.		pallet_size	Quantity of SKU fitting on one palett	450	metric - int	meta - static	MF	NR	NR
17.		LeadTime	Time taken from releasing an order for a specific SKU to receiving the ordered SKU [days]	45	metric - int	meta - static	MF	NR	NR
18.		CSL	Customer Service Level Quota per SKU and Month in a specific SalesCountry; ~Out-of-Stock Indicator	75	metric - int	meta - dynamic	MF	R	R
19.		CSL_Lag1	Out-of-Stock Indicator one month before ordering	65	metric - int	meta - dynamic	MF	R	NR
20.		CSL_Lag2	Out-of-Stock Indicator two months before ordering	50	metric - int	meta - dynamic	MF	R	NR

Brand and Quality	unique_cus_month	Net Reach: Number of unique customers (CH2-Level) per Mat10 and Month	8	metric - int	meta - dynamic	MF	R	R
	unique_cus_year	Net Reach: Number of unique customers (CH2-Level) per Mat10 and Year	12	metric - int	meta - dynamic	MF	NR	NR

Table E 19: Product - Vehicles in Population

Nr.	MECE			Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Micro-econo-mics	Product Status	VIO	rbk	ID for for a specific VIO (brand and model, powertrain system, equipment)	ABC0001234	string	meta - static	HF	NR	NR
2.				vio_key	ID for a specific VIO sector-segment-fueltype combination	Passenger Car_Micro Car_Gasoline	string	meta - static	HF	NR	NR
3.				fueltype	fuel type of a specific VIO	Electric	string	meta - static	HF	NR	NR
4.				vio_market_ sector	VIO type	LCV	string	meta - static	HF	NR	NR
5.				vio_market_ segment	VIO class (acc. to KBA)	Middle Class	string	meta - static	HF	NR	NR
6.				vio_age	age of rbk in a specific country and year (~ years active)	8	metric - int	meta-dynamic	HF	NR	NR
7.				vio_sum_ perageandtype	sum of VIO per SKU-vio_key combination of a certain age in a specific country and year	45000	metric-int	meta-dynamic	HF	NR	NR
26.				vio_sum_ perageandtype_0	sum of VIO per SKU-vio_key combination of age 0 in a specific country and year	45000	metric-int	meta-dynamic	HF	NR	NR
				... age = [0, ... ,18]							
33.				vio_sum_peragebinandtype_012	sum of VIO per SKU-vio_key combination of age 0, 1 and 2 in a specific country and year	80000	metric-int	meta-dynamic	HF	R	R
				... age bin = [0_1_2, ... , 17_18]							
34.				vio_sum_perage	sum of VIO per SKU of a certain age in a specific country and year	100000	metric-int	meta-dynamic	HF	NR	NR

			regardless of vio_key						
35.		vio_sum_pertype	sum of VIO per SKU-vio_key combination in a specific country and year regardless of age	60000	metric-int	meta-dynamic	HF	NR	NR
36.		vio_sum_perageand sector	sum of VIO per SKU-vio_sector combination of a certain age in a specific country and year	350000	metric-int	meta-dynamic	HF	NR	NR
37.		vio_sum_persector	sum of VIO per SKU-vio_sector combination in a specific country and year regardless of age	910000	metric-int	meta-dynamic	HF	NR	NR
40.		vio_sum_persector_ HCV(>6t) … sector = [LCV, … , PassengerCar]	sum of VIO per SKU in a specific country and year - sector-specific "HCV"	55000	metric-int	meta-dynamic	HF	NR	NR
49.		vio_sum_persegment_ MiddleClass … segment = [MicroCar, … , UpperClass]	sum of VIO per SKU in a specific country and year - segment-specific "MiddleClass"	80000	metric-int	meta-dynamic	HF	NR	NR
53.		vio_sum_perfueltype_ Gasoline … fueltype = [Diesel, … , Electric]	sum of VIO per SKU in a specific country and year - fueltype-specific "Gasoline"	110000	metric-int	meta-dynamic	HF	NR	NR
54.		vio_sum	sum of VIO per SKU in a specific country and year	45000000	metric-int	meta-dynamic	HF	NR	NR
55.	Techni-cal PLC	start_registration	start of registration of mileage status for a specific vio_key in a certain country [age ~ years]	1995	metric-int	meta-dynamic	HF	NR	NR
56.		end_registration	end of registration of mileage status for a specific vio_key in a certain country [age ~ years]	2021	metric-int	meta-dynamic	HF	NR	NR
57.		start_mileage	mileage status of specific vio_key in a certain country at start of registration	0	metric-int	meta-dynamic	HF	NR	NR
58.		end_mileage	mileage status of specific vio_key in a certain	175000	metric-int	meta-dynamic	HF	NR	NR

			country at end of registration (start of registration + x-years)						
59.		end_mileage_per_activeyear_dt	average mileage of a specific vio_key of a certain age in a certain country within one year	30000	metric-int	meta-dynamic	HF	NR	NR
60.		cum_sum_mileage	accumulated mileage of a specific vio_key of a certain age in a certain country	120000	metric-int	meta-dynamic	HF	NR	NR
61.	Failure	FailureProb_static	Failure Probability of a product on PH1 Level - static	0.002	metric-float	meta-dynamic	HF	NR	NR
62.		SurvivalProb_static	Survival Probability of a product on PH1 Level - static	0.998	metric-float	meta-dynamic	HF	NR	NR
63.		LifeSpan_exp_km	Expected Life Span in kilometres of a product on PH1 Level	45000	metric-float	meta-dynamic	HF	NR	NR
64.		LifeSpan_stdev_km	Standard Deviation reg. Life Span in kilometres of a product on PH1 Level	5000	metric-float	meta-dynamic	HF	NR	NR
65.		LifeSpan_exp_years	Expected Life Span in years of a product on PH1 Level in a specific country	3	metric-float	meta-dynamic	HF	NR	NR
66.		prod_age	Age of product on PH1 level at a specific time	9	metric-float	meta-dynamic	HF	NR	NR
67.		FailureProb_dyn_1stFailure_b_opt	Failure Probability (for 1st failure) of a product on PH1 Level - dynamic (i.e. age and mileage-dependent)	0.012	metric-float	meta-dynamic	HF	NR	NR
68.		FailureProb_dyn_nthFailure_b_opt	Failure Probability (for ith failure) of a product on PH1 Level - dynamic (i.e. age and mileage-dependent)	0.015	metric-float	meta-dynamic	HF	NR	NR
69.		combFailureProb_dynamic_b_opt	General Failure Probability of a product on PH1 Level - dynamic (i.e. age and mileage-dependent)	0.062	metric-float	meta-dynamic	HF	NR	NR

70.			FailureRate_b_opt	hazard rate of a product on PH1 level based on Weibullian distributed failure probalities for yearly bins	0.00200400801603206	metric-float	meta-dynamic	HF	NR	NR
71.			AvgRepQ	Average Replacement Quantity of Mat10 if one Product breaks	4	metric-int	meta-static	HF	NR	NR
72.		Portfo-lio	multi_assignment_ph2	No of unique SKU per PH2 which fit into unique RBK; representation of internal competition	6	metric-int	meta-dynamic	HF	NR	NR
73.			multi_assignment_factor_ph2	Resulting Probability of multi_assignment_ph2	0.167	metric-float	meta-dynamic	HF	NR	NR
74.			multi_assignment_ph3	No of unique SKU per PH3 which fit into unique RBK; representation of internal competition	3	metric-int	meta-dynamic	HF	NR	NR
75.			multi_assignment_factor_ph3	Resulting Probability of multi_assignment_ph3	0.333	metric-float	meta-dynamic	HF	NR	NR
76.	Resulting Features		vio_replaced_stat	vio_sum_perageandtype * static failure probability	8743	metric-float	meta-dynamic	MF	R	R
77.			vio_replaced_dyn	vio_sum_perageandtype * dynamic failure probability	9425	metric-float	meta-dynamic	MF	R	R
78.			vio_replaced_stat_AvgRepQ	vio_replaced_stat * AvgRepQ	17486	metric-float	meta-dynamic	MF	R	NR
79.			vio_replaced_dyn_AvgRepQ	vio_replaced_dyn * AvgRepQ	18850	metric-float	meta-dynamic	MF	R	NR
80.			vio_replaced_stat_ownMS	vio_replaced_stat * market share	6120	metric-float	meta-dynamic	MF	NR	NR
81.			vio_replaced_dyn_ownMS	vio_replaced_dyn * market share	6598	metric-float	meta-dynamic	MF	NR	NR
82.			vio_replaced_stat_multiAss	vio_replaced_stat * multi_assignment_ factor_ph3	2886	metric-float	meta-dynamic	MF	NR	NR
83.			vio_replaced_dyn_multiAss	vio_replaced_dyn * multi_assignment_ factor_ph3	3111	metric-float	meta-dynamic	MF	NR	NR
84.			vio_replaced_stat_AvRepQ_ownMS_multiAss	vio_replaced_stat * AvgRepQ * market share * multi_assignment_factor_ph3	4040	metric-float	meta-dynamic	MF	NR	NR

85.				vio_replaced_dyn_AvRepQ_ownMS_multiAss	vio_replaced_dyn * AvgRepQ * market share * multi_assignment_factor_ph3	4355	metric-float	meta-dynamic	MF	NR	NR

Table E 20: Price, Price Variability and Target Agreements

Nr.	MECE			Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
1.	Macro-economics	Economics	Currency Fluctuations	ExRateLCGC	Exchange Rates	75.05	metric-float	meta - dynamic	HF	NR	NR
2.				ExRateLCGC_MA3	Averaged Exchange Rates $m_{MA}^{(3)}$	73.5	metric-float	meta - dynamic	HF/MF	NR	NR
3.				ExRateLCGC_MA6	Averaged Exchange Rates $m_{MA}^{(6)}$	70.5	metric-float	meta - dynamic	HF/MF	NR	NR
4.				ExRateLCGC_MA2cent	Averaged Exchange Rates *centered* $m_{MA}^{(2)}$	73.8	metric-float	meta - dynamic	HF/MF	R	NR
5.				ExRateLCGC_MA3cent	Averaged Exchange Rates *centered* $m_{MA}^{(3)}$	73.0	metric-float	meta - dynamic	HF/MF	NR	NR
6.				pct_change_ExRateLCGC	Percentage Change ExRates from Month to Month	0.05	metric-float	meta - dynamic	MF	NR	NR
7.				pct_change_ExRateLCGC_MA3	Percentage Change ExRates from Month to Month	0.03	metric-float	meta - dynamic	MF	NR	NR
8.				pct_change_ExRateLCGC_MA2cent	Percentage Change ExRates from Month to Month	0.04	metric-float	meta - dynamic	MF	R	R
	Micro-economics	Product	Portfolio Information	ProductStatus_1	Classification of Mat10 into Fast Mover and Slow Mover	FM	string	meta - dynamic	MF	-	NR
				ProductStatus_2	Classification of SKU into R, I and S Parts (regular, irregular and sporadic) with $CV \in [0.6, 1.2;\ \infty]$	I	string	meta - dynamic	MF	-	NR
				multi_assignment_ph3	Number of Substitutes per SKU	3	metric-int	meta- dynamic	MF	-	NR
			Product Value and Complexity	PartType	Classification of Mat10 into Technical Parts and Commodities	TP	string	meta - static	MF/SF	-	R
				PPC-P AA_Mat10	Material and Production Costs on Mat10 level per piece or set	3.65	metric-float	meta - dynamic	MF	R	NR

9.	Market Partici-pants – First Level Custo-mers	Marketing Instruments	ICPR_own	Index Catalogue Price of Mat10 per piece or set	100	metric-float	transactional - dynamic	MF	NR	NR
10.			RRP	Catalogue Price of Mat10 per piece or set, which is bound to a certain region (regional retail price)	95	metric-float	transactional - dynamic	MF	NR	NR
11.			RRP_neig	Catalogue Price of Mat10 per piece or set, which is bound to a neighbouring region (regional retail price)	92	metric-float	transactional - dynamic	MF	NR	NR
12.			RRP_comp	Catalogue Price of Mat10 per piece or set, which is defined by a competitor (regional retail price)	94	metric-float	transactional - dynamic	MF	NR	NR
13.		Marketing Instruments and Psycho-graphics (~ cf. negotia-tion power and communi-cation)	IPP_inLC_perPc	Invoiced Price Point [€/Pcs] in the buyers' currency w* exchange rate status at order placement date	72	metric-float	transactional - dynamic	MF	R	NR
14.			IPP_inGC_perPc	Invoiced Price Point [€/Pcs] in the sellers' currency	72	metric-float	transactional - dynamic	MF	R	R
15.			GPP_inGC_perPc	Gross Price Point [€/Pcs] in the sellers' currency	71	metric-float	transactional - dynamic	MF	NR	NR
16.			NPP_inGC_perPc	Net Price Point [€/Pcs] in the sellers' currency (IPP - revenue reductions)	65	metric-float	transactional - dynamic	MF	NR	NR
17.			Rebate	Non-Invoiced Discounts [%] - PMC specific	0	metric-float	transactional - dynamic	MF	NR	NR
18.			Bonus	Revenue Reductions [%] - PMC specific	3	metric-float	transactional - dynamic	MF	R	R
		Purchasing Power	TopCus_C11 … countries C = [1, 2, 3]	Cummulated Sales of Top Customers (amounting to 80% of turnover and sales) per SKU in Country 1	21	metric – int	transactional – dynamic	MF	R	R*

Table E 21: Time and Calendar-related Features

Nr.	MECE	Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI

3.	Time	Calender	is_holiday_C11 … countries C = [1 ,2 ,3]	National Holidays and Vacational Periods	1	metric - int	transactional - dynamic	MF	R	NR
4.		Seasonality and Trends	is_seasonal	Binary variable indicating if Mat10 reveals seasonality on PG2 level	0	metric - int	transactional - dynamic	MF	R	R
5.			seasonal_ind	Seasonal index derived on PG3 level per Mat10 (additive)	-271.974507	metric - float	transactional - dynamic	HF	NR	NR
6.			seasonal_pct	Percentage Share incurred due to Seasonality derived on PG3 level per Mat10	-24.14	metric - float	transactional - dynamic	MF	R	R
7.			trend_ind	trend index derived on PG3 level per Mat10 (additive)	515.860917	metric - float	transactional - dynamic	HF	NR	NR
8.			resid_ind	residual index derived on PG3 level per Mat10 (additive)	629.113590	metric - float	transactional - dynamic	HF	NR	NR

Table E 22: Statistical Cluster Attributes – TS-Fresh

Nr.	MECE		Feature	Description	Example	Data Type	Level of Influence and Texture	Status	FI	PI
9.	General time series characteristics	Quantiles	quantile__q_0.1 … number Q = [0.1 , 0.2 , 0.9]	Calculates the q quantile of the time series	123	metric - float	transactional - dynamic	SF	-	-
12.		Entropy	approximate_entropy__m_2__r_0.1 …. number R = [0.1, 0.2, …, 0.9]	Implements a vectorized approximate entropy algorithm	2.684657	metric - float	transactional - dynamic	SF	-	-
17.		Trend	linear_trend__attr_slope …. attributes = [intercept, slope, rvalue, pvalue, stderr]	Calculates a linear least-squares regression for the values of the time series versus the sequence from 0 to length of the time series minus one and returns attributes	-0.158831	metric - float	transactional - dynamic	SF	-	-
			agg_linear_trend__attr_slope	Calculates a linear least-squares	-0.020838	metric - float	transactional - dynamic	SF	-	-

27.		…. attributes = [intercept, slope, rvalue, pvalue, stderr] …. chunks = [6; 12]	regression for values of the time series that were aggregated over chunks versus the sequence from 0 up to the number of chunks minus one.						
36.	Autocorrelation	partial_autocorrelation__lag_2 …. lags = [2, 3, …, 12]	Calculates the value of the partial autocorrelation function at the given lag.	-0.0334	metric - float	transactional - dynamic	SF	-	-
45.		autocorrelation__lag_2 …. lags = [2, 3, …, 12]	Calculates the autocorrelation of the specified lag	-0.023845	metric - float	transactional - dynamic	SF	-	-
46.	benford_correlation	benford_correlation	Returns the correlation from first digit distribution when compared to the Newcomb-Benford's Law distribution. Useful for anomaly detection applications	0.864123	metric - float	transactional - dynamic	SF	-	-
52.	Structural charactericstics	number_peaks__n_1 … number N = [1 ,2 , 6]	Calculates the number of peaks of at least support n in the time series	5	metric - int	transactional - dynamic	SF	-	-
53.		length	Returns the length of the time series	60	metric - float	transactional - dynamic	SF	-	-
54.		maximum	Returns the maximum value of the time series	1256	metric - float	transactional - dynamic	SF	-	-
55.		minimum	Returns the minimum value of the time series	0	metric - float	transactional - dynamic	SF	-	-
56.		first_location_of_maximum	Returns the first location of the maximum value of the time series	5.0	metric - float	transactional - dynamic	SF	-	-
57.		first_location_of_minimum	Returns the first location of the minimum value of the time series	6.0	metric - float	transactional - dynamic	SF	-	-
58.		last_location_of_maximum	Returns the last location of the maximum value of the time series	32.0	metric - float	transactional - dynamic	SF	-	-
59.		last_location_of_minimum	Returns the last location of the minimum value of the time series	49.0	metric - float	transactional - dynamic	SF	-	-
60.		has_duplicate_max	Checks if the maximum value of the time series	1	metric - float	transactional - dynamic	SF	-	-

		is observed more than once						
61.	has_duplicate_min	Checks if the minimum value of the time series is observed more than once	1	metric - float	transactional - dynamic	SF	-	-
62.	variance_larger_than_standard_deviation	Binary variable: Is variance higher than the standard deviation?	0	metric - float	transactional - dynamic	SF	-	-

Appendix 8: Assessment of Factors

The feature selection process described in Chapter 5.2.2 relies on twelve models (cf. Table E 23) and the number of features provided in the respective tables. Dependent on this, specific values for the critical distance are obtained.

Amongst regression models, ridge regression was chosen over multiple linear regression for two reasons: Approximately two thirds of multiple linear regression models failed to show significant results respectively failed to pass the f-test. For this reason, the number of PMCs being assigned a score via linear regression would differ greatly from the original sample size. Also, the fact that some of the MFs are correlated leads to ridge regression replacing multiple linear regression.

With regards to MFs, three constraints apply:

- The approach is not reliable if $T_i^k < 10$.
- The approach is not reliable if $r_{k_1,k_2} \geq 0.9$
- All regression models applied are local models i.e., they are not capable of learning from static MFs e.g. *minimumlotsize*

Table E 23: Regression Models used for Feature Selection

MODELS m	CONVENTIONAL	ML-BASED
FSD		
• forward • backward	• Ridge Regression • Lasso Regression	• Gradient Boosting Regressor • Decision Tree Regressor • K-nearest Neighbors Regressor • Support Vector Regression

The family-wise significance level of the tests is α = 0.050.

Table E 24: Critical Distances

α	NO OF MODELS m	NO OF MFs c	CRITICAL DISTANCE cd
0.05	12	3	0.956
0.05	12	4	1.353
0.05	12	5	
0.05	12	6	2.176
0.05	12	7	
0.05	12	8	3.030
0.05	12	9	3.468
0.05	12	18	7.604

Appendix 9: Stationarity of Main Features

Table E 25: Stationarity of Industry Volume Features

	MF1	**MF2**	**MF3**	**MF4**
COM	**item_category3_ID_ sum_item_cnt_month**	**item_category2_ID_ sum_item_cnt_month**	**item_category1_ID_ sum_item_cnt_month**	**BU_ID_ sum_item_cnt_month**
Stationarity [Yes in %]	95%: 66.9 90%: 71.5	95%: 71.9 90%: 72.4	95%: 78.9 90%: 78.9	95%: 83.1 90%: 83.1
	MF5	**MF6**	**MF7**	**MF8**
COM	**item_category3_ID_ avg_item_cnt_month**	**item_category2_ID_ avg_item_cnt_month**	**item_category1_ID_ avg_item_cnt_month**	**BU_ID_ avg_item_cnt_month**
Stationarity [Yes in %]	95%: 66.9 90%: 71.5	95%: 71.9 90%: 72.4	95%: 78.9 90%: 78.9	95%: 83.1 90%: 83.1
	MF1	**MF2**	**MF3**	**MF4**
TP	**item_category3_ID_ sum_item_cnt_month**	**item_category2_ID_ sum_item_cnt_month**	**item_category1_ID_ sum_item_cnt_month**	**BU_ID_ sum_item_cnt_month**
Stationarity [Yes in %]	95%: 44.1 90%: 47.8	95%: 35.2 90%: 36.1	95%: 24.4 90%: 24.4	95%: 18.6 90%: 18.6
	MF5	**MF6**	**MF7**	**MF8**
TP	**item_category3_ID_ avg_item_cnt_month**	**item_category2_ID_ avg_item_cnt_month**	**item_category1_ID_ avg_item_cnt_month**	**BU_ID_ avg_item_cnt_month**
Stationarity [Yes in %]	95%: 44.1 90%: 47.8	95%: 35.2 90%: 36.1	95%: 24.4 90%: 24.4	95%: 18.6 90%: 18.6

Table E 26: Stationarity of Customer-related Features

	MF1	**MF2**	**MF3**	**MF4**
COM	**topcus_c**	**top1_c**	**top2_c**	**top3_c**
Stationarity [Yes in %]	95%: 71.0 90%: 73.1	95%: 73.0 90%: 75.3	95%: 74.0 90%: 75.9	95%: 74.0 90%: 75.8
	MF5	**MF6**	**MF7**	**MF8**
COM	**topcus_c_pct**	**top1_c_pct**	**top2_c_pct**	**top3_c_pct**
Stationarity [Yes in %]	95%: 63.9 90%: 66.4	95%: 77.1 90%: 78.9	95%: 73.3 90%: 77.0	95%: 79.4 90%: 80.6

	MF9	MF10	MF11	MF12
COM	**b** (wholesale)	**c** (retail)	**e** (others)	**f** (workshops)
Stationarity [Yes in %]	95%: 66.2 90%: 70.1	95%: 16.7 90%: 17.1	95% 32.8 90%: 32.9	95%: 56.2 90%: 62.1
	MF13	**MF14**	**MF15**	**MF16**
COM	**b_pct**	**c_pct**	**e_pct**	**f_pct**
Stationarity [Yes in %]	95%: 63.0 90%: 64.8	95%: 17.1 90%: 17.9	95%: 32.3 90%: 32.5	95%: 58.4 90%: 63.0
	MF17	**MF18**		
COM	**unique_cus(_month)**	**unique_cus_year**		
Stationarity [Yes in %]	95%: 66.7 90%: 69.5	95%: 5.2 90%: 5.2		
	MF1	**MF2**	**MF3**	**MF4**
TP	**topcus_c**	**top1_c**	**top2_c**	**top3_c**
Stationarity [Yes in %]	95%: 75.8 90%: 77.5	95%: 54.7 90%: 55.8	95%: 65.6 90%: 66.5	95%: 66.4 90%: 67.8
	MF5	**MF6**	**MF7**	**MF8**
TP	**topcus_c_pct**	**top1_c_pct**	**top2_c_pct**	**top3_c_pct**
Stationarity [Yes in %]	95%: 74.7 90%: 76.4	95%: 54.5 90%: 56.0	95%: 67.3 90%: 68.1	95%: 67.5 90%: 68.5
	MF9	**MF10**	**MF11**	**MF12**
TP	**b** (wholesale)	**c** (retail)	**e** (others)	**f** (workshops)
Stationarity [Yes in %]	95%: 75.7 90%: 78.0	95%: 2.5 90%: 2.5	95% 22.5 90%: 23.2	95% 10.4 90%: 10.8
	MF13	**MF14**	**MF15**	**MF16**
TP	**b_pct**	**c_pct**	**e_pct**	**f_pct**
Stationarity [Yes in %]	95%: 73.8 90%: 76.5	95%: 2.6 90%: 2.7	95%: 24.4 90%: 24.6	95%: 10.7 90%: 11.0
	MF17	**MF18**		
TP	**unique_cus(_month)**	**unique_cus_year**		
Stationarity [Yes in %]	95%: 73.4 90%: 75.9	95%: 7.3 90%: 7.4		

Table E 27: Stationarity of Product-related Features

	MF1	MF2	MF3	MF4
COM	**ppc**	**pallet_qty**	**minimumlotsize**	**leadtime**
Stationarity [Yes in %]	95%: 3.4 90%: 4.2	95%: - 90%: -	95%: - 90%: -	95%: 97.3 90%: 97.9
	MF5	**MF6**	**MF7**	**MF8**
COM	**lifecycle_age_aa**	**csl1**	**csl1_lag1**	**csl1_lag2**
Stationarity [Yes in %]	95%: 36.3 90%: 36.4	95%: 72.7 90%: 75.6	95%: 72.7 90%: 75.6	95%: 72.7 90%: 75.6
	MF1	**MF2**	**MF3**	**MF4**
TP	**ppc**	**pallet_qty**	**minimumlotsize**	**leadtime**
Stationarity [Yes in %]	95%: 2.4 90%: 3.2	95%: - 90%: -	95%: - 90%: -	95%: 98.2 90%: 98.3
	MF5	**MF6**	**MF7**	**MF8**
TP	**lifecycle_age_aa**	**csl1**	**csl1_lag1**	**csl1_lag2**
Stationarity [Yes in %]	95%: 37.4 90%:37.6	95%: 75.8 90%: 78.1	95%: 72.7 90%: 75.6	95%: 72.7 90%: 75.6

Table E 28: Stationarity of VIO-related Features

	MF1	MF2	MF3	MF4
COM	**vio_sum_perage andtype**	**vio_replaced_dyn**	**vio_replaced _dyn_AvgRepQ**	**vio_replaced _dyn_multiAss**[78]

[78] MF4 (*vio_replaced_dyn_multiAss*) is the product of MF2 (*vio_replaced_dyn*) and a so-called *multi-assignment probability*, defined as $\frac{1}{\text{n\_unique SKUs per rbk (key) within a specific PH3}}$.

Stationarity [Yes in %]	95%: 26.44 90%: 28.15	95%: 32.56 90%: 33.99	95%: 32.56 90%: 33.99	95%: 32.56 90%: 34.0
	MF5			
COM	**vio_replaced_dyn _ownMS**			
Stationarity [Yes in %]	95%: 12.11 90%: 18.02			
	MF1	**MF2**	**MF3**	**MF4**
TP	**vio_sum_perage andtype**	**vio_replaced_stat**	**vio_replaced _stat_AvgRepQ**	**vio_replaced _stat_multiAss**
Stationarity [Yes in %]	95%: 26.44 90%: 28.15	95%: 26.44 90%: 28.15	95%: 26.44 90%: 28.15	95%: 26.44 90%: 28.15
	MF5			
TP	**vio_replaced_stat _ownMS**			
Stationarity [Yes in %]	95%: 30.10 90%: 31.86			

Table E 29: Stationarity of Price-related Features

	MF1	**MF2**	**MF3**	**MF4**
COM	**exratelcgc_ma2cent**	**pct_change_exratelcgc _ma2cent**	**PPCinGC**	**IndexPriceinGC**
Stationarity [Yes in %]	stationary	non-stationary	95%: 47.9 90%: 49.7	95%: 74.0 90%: 74.0
	MF5	**MF6**	**MF7**	**MF8**
COM	**RRPinGC**	**IPPinLC_perPc**	**IPPinGC_perPc**	**GPPinGC_perPc**
Stationarity [Yes in %]	95%: 74.0 90%: 74.0	95%: 23.4 90%: 24.8	95%: 28.7 90%: 31.9	95%: 28.7 90%: 31.9
	MF9	**MF10**	**MF11**	
COM	**NPPinGC_perPc**	**Rebate [%]**	**Bonus [%]**	
Stationarity [Yes in %]	95%: 29.0 90%: 32.2	95%: - 90%: -	95%: - 90%: -	
	MF1	**MF2**	**MF3**	**MF4**
TP	**exratelcgc_ma2cent**	**pct_change_exratelcgc _ma2cent**	**PPCinGC**	**IndexPriceinGC**
Stationarity [Yes in %]	stationary	non-stationary	95%: 34.9 90%: 36.3	95%: 87.0 90%: 89.2
	MF5	**MF6**	**MF7**	**MF8**
TP	**RRPinGC**	**IPPinLC_perPc**	**IPPinGC_perPc**	**GPPinGC_perPc**
Stationarity [Yes in %]	95%: 87.0 90%: 89.2	95%: 27.11 90%: 29.47	95%: 55.5 90%: 57.9	95%: 55.5 90%: 57.9
	MF9	**MF10**	**MF11**	
TP	**NPPinGC_perPc**	**Rebate [%]**	**Bonus [%]**	
Stationarity [Yes in %]	95%: 57.34 90%: 60.12	95%: - 90%: -	95%: - 90%: -	

Table E 30: Stationarity of Time-related Features

	MF1	**MF2**	**MF3**
COM	**is_holiday**	**is_seasonal**	**seasonality_pct**
Stationarity [Yes in %]	95%: - 90%: -	95%: - 90%: -	95%: 55.2 90%: 57.1
	MF1	**MF2**	**MF3**
TP	**is_holiday**	**is_seasonal**	**seasonality_pct**
Stationarity [Yes in %]	95%: - 90%: -	95%: - 90%: -	95%: - 90%: -

Appendix 10: Intra-Feature Correlation Testing of Main Features

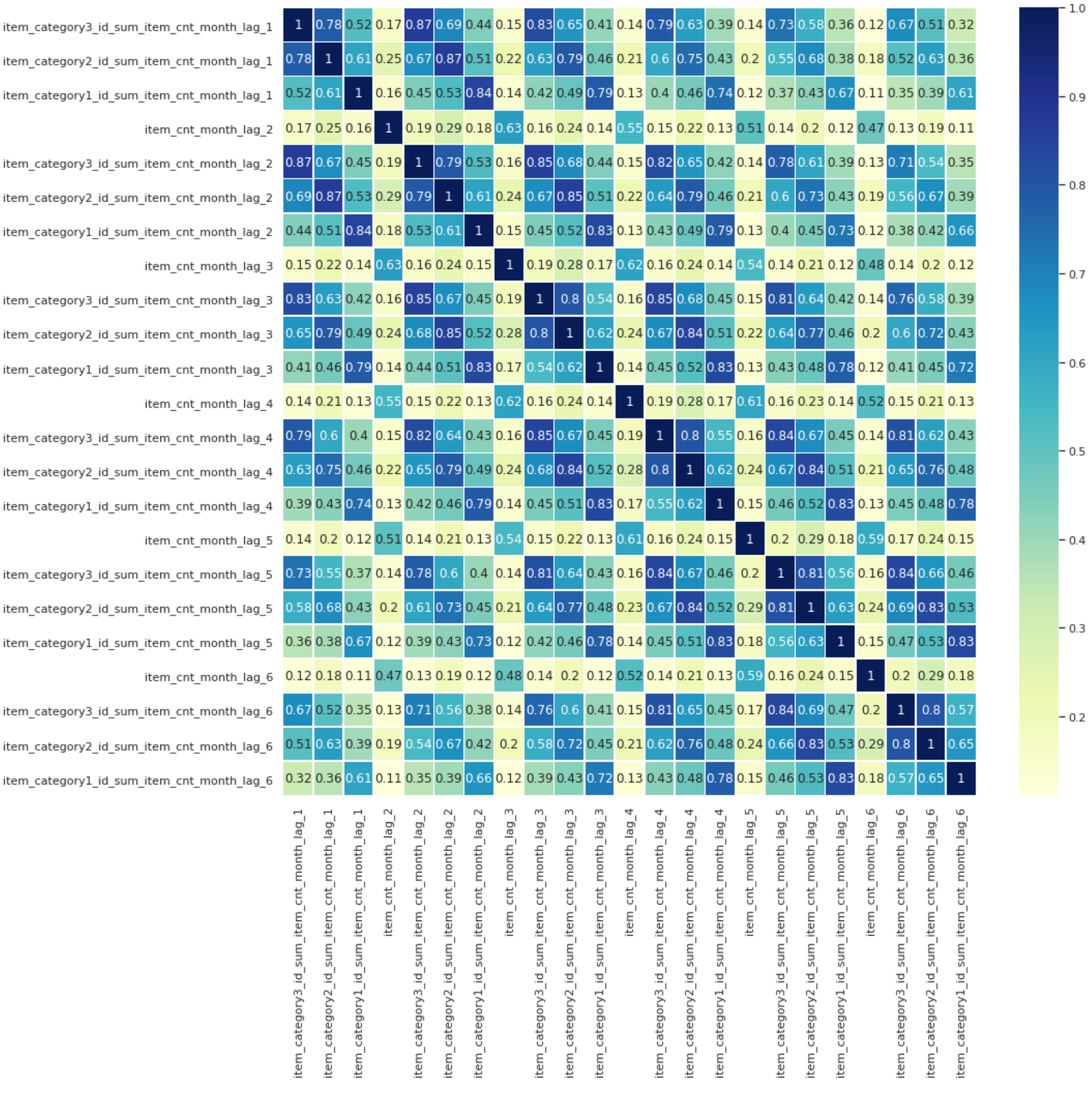


Figure E 4 Correlation Testing for Lagged Industry Volume Features

Appendix 11: Importance Testing of Main Features

Table E 31: Ranks of Industry Volume Features - Frequency

	Rank	A		B		C	
COM	1	item_category3_id_ avg_item_cnt_month	mean: 0.558 +-0.183 σ: 0.198	item_category3_id_ avg_item_cnt_month	mean: 0.544 +-0.174 σ: 0.188	item_category2_id_ avg_item_cnt_month	mean: 0.561 +-0.093 σ: 0.101
	2	item_category3_id_ sum_item_cnt_month	mean: 0.576 +-0.121 σ: 0.131	item_category3_id_ sum_item_cnt_month	mean: 0.531 +-0.109 σ: 0.117	item_category1_id_ avg_item_cnt_month	mean: 0.516 +-0.178 σ: 0.192
	3	item_category1_id_ avg_item_cnt_month	mean: 0.480 +-0.183 σ: 0.197	item_category1_id_ avg_item_cnt_month	mean: 0.518 +-0.175 σ: 0.189	item_category3_id_ avg_item_cnt_month	mean: 0.553 +-0.169 σ: 0.183
	4	item_category2_id_ avg_item_cnt_month	mean: 0.489 +-0.089 σ: 0.097	item_category2_id_ avg_item_cnt_month	mean: 0.503 +-0.075 σ: 0.081	item_category3_id_ sum_item_cnt_month	mean: 0.485 +-0.090 σ: 0.097
	5	item_category2_id_ sum_item_cnt_month	mean: 0.492 +-0.124 σ: 0.133	item_category2_id_ sum_item_cnt_month	mean: 0.466 +-0.091 σ: 0.099	item_category2_id_ sum_item_cnt_month	mean: 0.442 +-0.109 σ: 0.117
	6	item_category1_id_ sum_item_cnt_month	mean: 0.405 +-0.066 σ: 0.071	item_category1_id_ sum_item_cnt_month	mean: 0.438 +-0.074 σ: 0.080	item_category1_id_ sum_item_cnt_month	mean: 0.443 +-0.079 σ: 0.086
TP	1	item_category2_id_ avg_item_cnt_month	mean: 0.548 +-0.080 σ: 0.100	item_category3_id_ avg_item_cnt_month	mean: 0.554 +-0.083 σ: 0.161	item_category3_id_ avg_item_cnt_month	mean: 0.576 +-0.094 σ: 0.180
	2	item_category3_id_ avg_item_cnt_month	mean: 0.558 +-0.080 σ: 0.182	item_category1_id_ avg_item_cnt_month	mean: 0.536 +-0.083 σ: 0.184	item_category1_id_ avg_item_cnt_month	mean: 0.545 +-0.094 σ: 0.210
	3	item_category3_id_ sum_item_cnt_month	mean: 0.501 +-0.080 σ: .112	item_category2_id_ avg_item_cnt_month	mean: 0.554 +-0.083 σ: 0.116	item_category2_id_ avg_item_cnt_month	mean: 0.551 +-0.094 σ: 0.142
	4	item_category1_id_ avg_item_cnt_month	mean: 0.538 +-0.080 σ: 0.164	item_category3_id_ sum_item_cnt_month	mean: 0.478 +-0.083 σ: 0.132	item_category3_id_ sum_item_cnt_month	mean: 0.479 +-0.094 σ: 0.153
	5	item_category2_id_ sum_item_cnt_month	mean: 0.457 +-0.080 σ: 0.108	item_category2_id_ sum_item_cnt_month	mean: 0.456 +-0.083 σ: 0.110	item_category2_id_ sum_item_cnt_month	mean: 0.420 +-0.094 σ: 0.118
	6	item_category1_id_ sum_item_cnt_month	mean: 0.398 +-0.080 σ: 0.107	item_category1_id_ sum_item_cnt_month	mean: 0.421 +-0.083 σ: 0.107	item_category1_id_ sum_item_cnt_month	mean: 0.429 +-0.094 σ: 0.117

Table E 32: ECDFs derived from Importance Scores for Industry Volume Features

		MF1	**MF2**	**MF3**	**MF4**
$ECDF$		**item_category3_ID_ sum_item_cnt_month**	**item_category2_ID_ sum_item_cnt_month**	**item_category1_ID_ sum_item_cnt_month**	**BU_ID_ sum_item_cnt_month**
COM	A	P(x<0.1): 0.169 P(x<0.2): 0.556 P(x<0.5): 0.978 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.345 P(x<0.2): 0.808 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.343 P(x<0.2): 0.748 P(x<0.5): 0.994 P(x<0.7): 1.000 P(x<0.9): 1.000	eliminated due to r_{f_3,f_4} >0.9
	B	P(x<0.1): 0.163 P(x<0.2): 0.674 P(x<0.5): 0.986 P(x<0.7): 0.997 P(x<0.9): 1.000	P(x<0.1): 0.273 P(x<0.2): 0.835 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.205 P(x<0.2): 0.641 P(x<0.5): 0.997 P(x<0.7): 1.000 P(x<0.9): 1.000	
	C	P(x<0.1): 0.187 P(x<0.2): 0.695 P(x<0.5): 0.993 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.255 P(x<0.2): 0.798 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.181 P(x<0.2): 0.652 P(x<0.5): 0.996 P(x<0.7): 0.999 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.198 P(x<0.2): 0.646 P(x<0.5): 0.991 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.267 P(x<0.2): 0.800 P(x<0.5): 0.996 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.218 P(x<0.2): 0.682 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	

		MF5 item_category3_ID_ avg_item_cnt_month	MF6 item_category2_ID_ avg_item_cnt_month	MF7 item_category1_ID_ avg_item_cnt_month	MF8 BU_ID_ avg_item_cnt_month
COM	A	P(x<0.1): 0.169 P(x<0.2): 0.544 P(x<0.5): 0.984 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.368 P(x<0.2): 0.813 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.350 P(x<0.2): 0.745 P(x<0.5): 0.995 P(x<0.7): 1.000 P(x<0.9): 1.000	eliminated due to r_{f_3,f_4} >0.9
	B	P(x<0.1): 0.164 P(x<0.2): 0.680 P(x<0.5): 0.983 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.276 P(x<0.2): 0.829 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.204 P(x<0.2): 0.644 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	
	C	P(x<0.1): 0.162 P(x<0.2): 0.705 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.231 P(x<0.2): 0.836 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.170 P(x<0.2): 0.647 P(x<0.5): 0.995 P(x<0.7): 1.000 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.166 P(x<0.2): 0.664 P(x<0.5): 0.993 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.250 P(x<0.2): 0.820 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.229 P(x<0.2): 0.677 P(x<0.5): 0.993 P(x<0.7): 0.999 P(x<0.9): 1.000	

		MF1 item_category3_ID_ sum_item_cnt_month	MF2 item_category2_ID_ sum_item_cnt_month	MF3 item_category1_ID_ sum_item_cnt_month	MF4 BU_ID_ sum_item_cnt_month
TP	A	P(x<0.1): 0.189 P(x<0.2): 0.613 P(x<0.5): 0.985 P(x<0.7): 0.998 P(x<0.9): 1.000	P(x<0.1): 0.306 P(x<0.2): 0.789 P(x<0.5): 0.997 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.258 P(x<0.2): 0.726 P(x<0.5): 0.995 P(x<0.7): 1.000 P(x<0.9): 1.000	eliminated due to r_{f_7,f_8} >0.9
	B	P(x<0.1): 0.218 P(x<0.2): 0.685 P(x<0.5): 0.994 P(x<0.7): 0.998 P(x<0.9): 1.000	P(x<0.1): 0.276 P(x<0.2): 0.778 P(x<0.5): 0.995 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.212 P(x<0.2): 0.663 P(x<0.5): 0.995 P(x<0.7): 0.999 P(x<0.9): 1.000	
	C	P(x<0.1): 0.288 P(x<0.2): 0.722 P(x<0.5): 0.984 P(x<0.7): 0.998 P(x<0.9): 1.000	P(x<0.1): 0.306 P(x<0.2): 0.750 P(x<0.5): 0.989 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.246 P(x<0.2): 0.637 P(x<0.5): 0.977 P(x<0.7): 0.996 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.267 P(x<0.2): 0.696 P(x<0.5): 0.991 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.296 P(x<0.2): 0.756 P(x<0.5): 0.993 P(x<0.7): 0.998 P(x<0.9): 1.000	P(x<0.1): 0.248 P(x<0.2): 0.658 P(x<0.5): 0.986 P(x<0.7): 0.998 P(x<0.9): 1.000	

		MF5 item_category3_ID_ avg_item_cnt_month	MF6 item_category2_ID_ avg_item_cnt_month	MF7 item_category1_ID_ avg_item_cnt_month	MF8 BU_ID_ avg_item_cnt_month
TP	A	P(x<0.1): 0.187 P(x<0.2): 0.616 P(x<0.5): 0.984 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.291 P(x<0.2): 0.784 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.251 P(x<0.2): 0.731 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000	eliminated due to r_{f_7,f_8} >0.9
	B	P(x<0.1): 0.224 P(x<0.2): 0.697 P(x<0.5): 0.993 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.271 P(x<0.2): 0.777 P(x<0.5): 0.995 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.217 P(x<0.2): 0.670 P(x<0.5): 0.991 P(x<0.7): 0.998 P(x<0.9): 0.999	
	C	P(x<0.1): 0.282 P(x<0.2): 0.721 P(x<0.5): 0.991 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.297 P(x<0.2): 0.756 P(x<0.5): 0.992 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.237 P(x<0.2): 0.639 P(x<0.5): 0.983 P(x<0.7): 0.998 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.260 P(x<0.2): 0.695 P(x<0.5): 0.993 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.290 P(x<0.2): 0.760 P(x<0.5): 0.994 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.234 P(x<0.2): 0.668 P(x<0.5): 0.984 P(x<0.7): 0.997 P(x<0.9): 1.000	

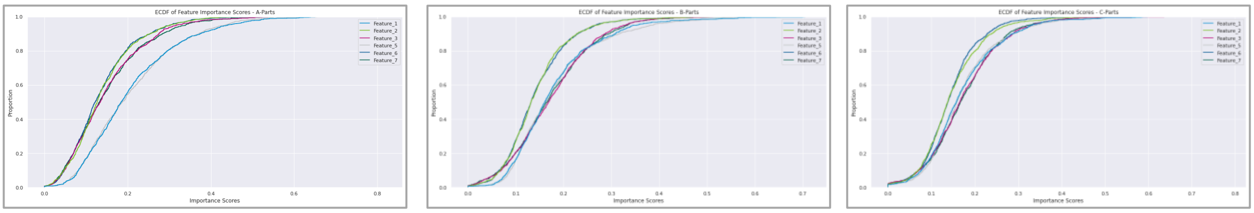


Figure E 6 ECDF of Importance Scores for Industry Volume Features - COM

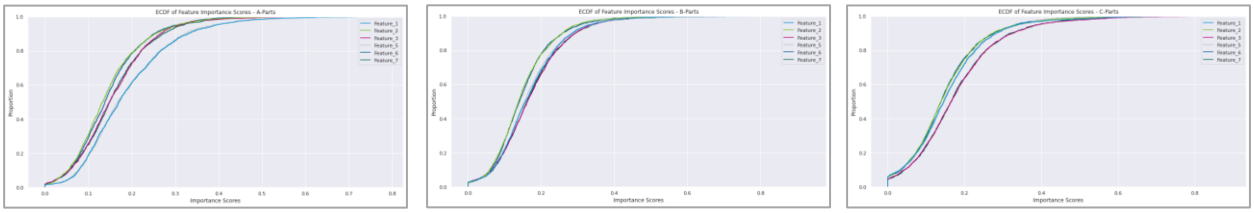


Figure E 5 ECDF of Importance Scores for Industry Volume Features - TP

Table E 33: Ranks of Customer-related Features - Frequency

	Rank	A		B		C	
COM	1	**b**	median: 0.723 +-0.270 MAD: 0.035	**b**	median: 0.743 +-0.255 MAD: 0.057	**b**	median: 0.684 +-0.150 MAD: 0.024
	2	**topcus_c1**	median: 0.383 +-0.195 MAD: 0.083	**topcus_c1**	median: 0.293 +-0.148 MAD: 0.040	**topcus_c1**	median: 0.199 +-0.095 MAD: 0.034
	3	**c**	median: 0.163 +-0.087 MAD: 0.032	**unique_cus_month**	median: 0.183 +-0.153 MAD: 0.055	**unique_cus_month**	median: 0.194 +-0.148 MAD: 0.089
	4	**unique_cus_month**	median: 0.123 +-0.130 MAD: 0.052	**e**	median: 0.127 +-0.080 MAD: 0.040	**b_pct**	median: 0.115 +-0.139 MAD: 0.054
	5	**e**	median: 0.148 +-0.065 MAD: 0.033	**b_pct**	median: 0.095 +-0.035 MAD: 0.013	**unique_cus_year**	median: 0.101 +-0.100 MAD: 0.037
	6	**unique_cus_year**	median: 0.077 +-0.177 MAD: 0.040	**unique_cus_year**	median: 0.090 +-0.138 MAD: 0.023	**e**	median: 0.092 +-0.058 MAD: 0.031
	7	**b_pct**	median: 0.073 +-0.043 MAD: 0.013	**e_pct**	median: 0.085 +-0.117 MAD: 0.038	**e_pct**	median: 0.082 +-0.076 MAD: 0.045
	8	**e_pct**	median: 0.058 +-0.090 MAD: 0.022	**topcus_c1_pct**	median: 0.045 +-0.080 MAD: 0.010	**top1_c1_pct**	median: 0.062 +-0.047 MAD: 0.015
	9	**f_pct**	median: 0.073 +-0.143 MAD: 0.062	**c**	median: 0.057 +-0.043 MAD: 0.010	**f_pct**	median: 0.075 +-0.138 MAD: 0.075
	10	**topcus_c1_pct**	median: 0.037 +-0.187 MAD: 0.010	**f_pct**	median: 0.068 +-0.198 MAD: 0.065	top1_c1	median: 0.053 +-0.031 MAD: 0.017
	11	top1_c1	median: 0.033 +-0.025 MAD: 0.008	top1_c1	median: 0.033 +-0.020 MAD: 0.007	top1_c1_pct	median: 0.044 +-0.062 MAD: 0.012
	12	f	median: 0.023 +-0.013 MAD: 0.007	top2_c1_pct	median: 0.032 +-0.018 MAD:	top3_c1	median: 0.058 +-0.033 MAD: 0.014
	13	top2_c1	median: 0.023 +-0.010 MAD: 0.007	top3_c1	median: 0.025 +-0.015 MAD: 0.005	top2_c1	median: 0.040 +-0.027 MAD: 0.020
	14	top3_c1	median: 0.020 +-0.010 MAD: 0.007	top2_c1	median: 0.023 +-0.013 MAD: 0.005	top3_c1_pct	median: 0.043 +-0.043 MAD: 0.016
	15	c_pct	median: 0.008 +-0.097 MAD: 0.007	c_pct	median: 0.013 +-0.078 MAD: 0.012	top2_c1_pct	median: 0.045 +-0.033 MAD: 0.022

	16	top3_c1_pct	median: 0.015 +-0.032 MAD: 0.007	f	median: 0.015 +-0.022 MAD: 0.008	c	median: 0.023 +-0.100 MAD: 0.021
	17	top1_c1_pct	median: 0.013 +-0.010 MAD: 0.003	top1_c1_pct	median: 0.022 +-0.022 MAD: 0.012	c_pct	median: 0.012 +-0.025 MAD: 0.007
	18	top2_c1_pct	median: 0.010 +-0.008 MAD: 0.003	top3_c1_pct	median: 0.017 +-0.042 MAD: 0.005	f	median: 0.000 +-0.013 MAD: 0.000
TP	1	**b**	median: 0.784 +-0.279 MAD: 0.038	**b**	median: 0.757 +-0.270 MAD: 0.027	**b**	median: 0.535 +-0.170 MAD: 0.066
	2	**topcus_c1**	median: 0.217 +-0.112 MAD: 0.052	**topcus_c1**	median: 0.185 +-0.100 MAD: 0.045	**unique_cus_month**	median: 0.207 +-0.257 MAD: 0.138
	3	**unique_cus_month**	median: 0.193 +-0.207 MAD: 0.072	**unique_cus_month**	median: 0.213 +-0.193 MAD: 0.098	**b_pct**	median: 0.180 +-0.179 MAD: 0.104
	4	**e**	median: 0.152 +-0.114 MAD: 0.055	**b_pct**	median: 0.125 +-0.110 MAD: 0.047	**topcus_c1**	median: 0.114 +-0.109 MAD: 0.059
	5	**e_pct**	median: 0.105 +-0.136 MAD: 0.047	**e_pct**	median: 0.108 +-0.105 MAD: 0.060	**unique_cus_year**	median: 0.122 +-0.181 MAD: 0.067
	6	**b_pct**	median: 0.102 +-0.068 MAD: 0.018	**e**	median: 0.092 +-0.092 MAD: 0.037	**e_pct**	median: 0.127 +-0.121 MAD: 0.109
	7	**unique_cus_year**	median: 0.098 +-0.175 MAD: 0.043	**unique_cus_year**	median: 0.092 +-0.157 MAD: 0.033	**e**	median: 0.064 +-0.034 MAD: 0.020
	8	**topcus_c1_pct**	median: 0.049 +-0.059 MAD: 0.015	**topcus_c1_pct**	median: 0.050 +-0.033 MAD: 0.020	**topcus_c1_pct**	median: 0.053 +-0.045 MAD: 0.017
	9	**f_pct**	median: 0.073 +-0.272 MAD: 0.069	**f_pct**	median: 0.082 +-0.247 MAD: 0.082	**f_pct**	median: 0.159 +-0.264 MAD: 0.159
	10	top1_c1	median: 0.025 +-0.022 MAD: 0.012	top1_c1	median: 0.028 +-0.028 MAD: 0.012	**top1_c1**	median: 0.037 +-0.081 MAD: 0.022
	11	top3_c1	median: 0.018 +-0.012 MAD: 0.004	top1_c1_pct	median: 0.020 +-0.018 MAD: 0.007	top3_c1_pct	median: 0.033 +-0.050 MAD: 0.011
	12	top3_c1_pct	median: 0.014 +-0.017 MAD: 0.004	top3_c1	median: 0.017 +-0.015 MAD: 0.007	top2_c1	median: 0.037 +-0.039 MAD: 0.015
	13	top2_c1	median: 0.019 +-0.011 MAD: 0.005	top3_c1_pct	median: 0.017 +-0.017 MAD: 0.008	top2_c1_pct	median: 0.026 +-0.042 MAD: 0.012
	14	top1_c1_pct	median: 0.015 +-0.029 MAD: 0.004	top2_c1_pct	median: 0.013 +-0.037 MAD: 0.003	top3_c1	median: 0.033 +-0.032 MAD: 0.021
	15	top2_c1_pct	median: 0.016 +-0.010 MAD: 0.004	top2_c1	median: 0.017 +-0.017 MAD: 0.008	top1_c1_pct	median: 0.022 +-0.041 MAD: 0.012
	16	c	median: 0.010 +-0.045 MAD: 0.010	c	median: 0.003 +-0.043 MAD: 0.003	c	median: 0.007 +-0.045 MAD: 0.007
	17	c_pct	median: 0.004 +-0.082 MAD: 0.004	c_pct	median: 0.000 +-0.072 MAD: 0.000	c_pct	median: 0.002 +-0.011 MAD: 0.002
	18	f	median: 0.006 +-0.010 MAD: 0.004	f	median: 0.000 +-0.017 MAD: 0.000	f	median: 0.000 +-0.005 MAD: 0.000

Table E 34: ECDFs derived from Importance Scores for Customer-related Features

		MF1	MF2	MF3	MF4
	$ECDF$	**topcus_c**	**top1_c**	**top2_c**	**top3_c**
COM	A	P(x<0.1): 0.577 P(x<0.2): 0.632 P(x<0.5): 0.757 P(x<0.7): 0.846 P(x<0.9): 0.981	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.596 P(x<0.2): 0.666 P(x<0.5): 0.784 P(x<0.7): 0.865 P(x<0.9): 0.985			
	C	P(x<0.1): 0.590			

		P(x<0.2): 0.697 P(x<0.5): 0.862 P(x<0.7): 0.934 P(x<0.9): 0.991			
	overall	P(x<0.1): 0.509 P(x<0.2): 0.619 P(x<0.5): 0.773 P(x<0.7): 0.878 P(x<0.9): 0.992			
		MF5	**MF6**	**MF7**	**MF8**
		topcus_c_pct	**top1_c_pct**	**top2_c_pct**	**top3_c_pct**
COM	A	P(x<0.1): 0.991 P(x<0.2): 0.996 P(x<0.5): 0.999 P(x<0.7): 0.999 P(x<0.9): 1.000	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.991 P(x<0.2): 0.995 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000			
	C	P(x<0.1): 0.961 P(x<0.2): 0.978 P(x<0.5): 0.995 P(x<0.7): 0.999 P(x<0.9): 1.000			
	overall	P(x<0.1): 0.984 P(x<0.2): 0.993 P(x<0.5): 0.998 P(x<0.7): 0.999 P(x<0.9): 1.000			
		MF9	**MF10**	**MF11**	**MF12**
		b (wholesale)	**c** (retail)	**e** (others)	**f** (workshops)
COM	A	P(x<0.1): 0.106 P(x<0.2): 0.204 P(x<0.5): 0.344 P(x<0.7): 0.440 P(x<0.9): 0.628	P(x<0.1): 0.910 P(x<0.2): 0.935 P(x<0.5): 0.963 P(x<0.7): 0.970 P(x<0.9): 0.980	P(x<0.1): 0.941 P(x<0.2): 0.972 P(x<0.5): 0.993 P(x<0.7): 0.997 P(x<0.9): 0.999	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.074 P(x<0.2): 0.161 P(x<0.5): 0.322 P(x<0.7): 0.429 P(x<0.9): 0.671	P(x<0.1): 0.959 P(x<0.2): 0.962 P(x<0.5): 0.968 P(x<0.7): 0.972 P(x<0.9): 0.979	P(x<0.1): 0.948 P(x<0.2): 0.979 P(x<0.5): 0.995 P(x<0.7): 0.998 P(x<0.9): 1.000	
	C	P(x<0.1): 0.076 P(x<0.2): 0.133 P(x<0.5): 0.321 P(x<0.7): 0.450 P(x<0.9): 0.751	eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.957 P(x<0.2): 0.979 P(x<0.5): 0.991 P(x<0.7): 0.997 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.079 P(x<0.2): 0.166 P(x<0.5): 0.360 P(x<0.7): 0.490 P(x<0.9): 0.754	P(x<0.1): 0.936 P(x<0.2): 0.953 P(x<0.5): 0.963 P(x<0.7): 0.969 P(x<0.9): 0.983	P(x<0.1): 0.968 P(x<0.2): 0.989 P(x<0.5): 0.998 P(x<0.7): 0.999 P(x<0.9): 1.000	
		MF13	**MF14**	**MF15**	**MF16**
		b_pct (wholesale)	**c_pct** (retail)	**e** (others)	**f_pct** (workshops)
COM	A	P(x<0.1): 0.983 P(x<0.2): 0.991 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.00	eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.995 P(x<0.2): 0.998 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	B	P(x<0.1): 0.979 P(x<0.2): 0.990 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000		P(x<0.1): 0.992 P(x<0.2): 0.995 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.997 P(x<0.2): 0.998 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000
	C	P(x<0.1): 0.913 P(x<0.2): 0.951 P(x<0.5): 0.991 P(x<0.7): 0.999 P(x<0.9): 0.999		P(x<0.1): 0.991 P(x<0.2): 0.993 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	overall	P(x<0.1): 0.966 P(x<0.2): 0.981 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000		P(x<0.1): 0.996 P(x<0.2): 0.998 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.999 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
		MF17	**MF18**		
		unique_cus(_month)	**unique_cus_year**		
COM	A	P(x<0.1): 0.957 P(x<0.2): 0.981 P(x<0.5): 0.994 P(x<0.7): 0.999	P(x<0.1): 0.982 P(x<0.2): 0.993 P(x<0.5): 0.998 P(x<0.7): 1.000		

		P(x<0.9): 1.000	P(x<0.9): 1.000		
	B	P(x<0.1): 0.920 P(x<0.2): 0.958 P(x<0.5): 0.980 P(x<0.7): 0.990 P(x<0.9): 0.998	P(x<0.1): 0.977 P(x<0.2): 0.992 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000		
	C	P(x<0.1): 0.818 P(x<0.2): 0.887 P(x<0.5): 0.962 P(x<0.7): 0.980 P(x<0.9): 0.994	P(x<0.1): 0.973 P(x<0.2): 0.989 P(x<0.5): 0.997 P(x<0.7): 0.998 P(x<0.9): 0.999		
	overall	P(x<0.1): 0.914 P(x<0.2): 0.957 P(x<0.5): 0.982 P(x<0.7): 0.990 P(x<0.9): 0.997	P(x<0.1): 0.967 P(x<0.2): 0.986 P(x<0.5): 0.998 P(x<0.7): 0.999 P(x<0.9): 1.000		
		MF1	**MF2**	**MF3**	**MF4**
		topcus_c	**top1_c**	**top2_c**	**top3_c**
TP	A	P(x<0.1): 0.722 P(x<0.2): 0.779 P(x<0.5): 0.876 P(x<0.7): 0.919 P(x<0.9): 0.983	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.757 P(x<0.2): 0.814 P(x<0.5): 0.895 P(x<0.7): 0.947 P(x<0.9): 0.996			
	C	P(x<0.1): 0.691 P(x<0.2): 0.764 P(x<0.5): 0.912 P(x<0.7): 0.974 P(x<0.9): 0.994			
	overall	P(x<0.1): 0.680 P(x<0.2): 0.770 P(x<0.5): 0.902 P(x<0.7): 0.957 P(x<0.9): 0.995			
		MF5	**MF6**	**MF7**	**MF8**
		topcus_c_pct	**top1_c_pct**	**top2_c_pct**	**top3_c_pct**
TP	A	P(x<0.1): 0.980 P(x<0.2): 0.990 P(x<0.5): 0.998 P(x<0.7): 1.000 P(x<0.9): 1.000	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.966 P(x<0.2): 0.984 P(x<0.5): 0.997 P(x<0.7): 0.999 P(x<0.9): 1.000			
	C	P(x<0.1): 0.899 P(x<0.2): 0.934 P(x<0.5): 0.987 P(x<0.7): 1.000 P(x<0.9): 1.000			
	overall	P(x<0.1): 0.939 P(x<0.2): 0.968 P(x<0.5): 0.993 P(x<0.7): 0.999 P(x<0.9): 1.000			
		MF9	**MF10**	**MF11**	**MF12**
		b (wholesale)	**c** (retail)	**e** (others)	**f** (workshops)
TP	A	P(x<0.1): 0.045 P(x<0.2): 0.093 P(x<0.5): 0.214 P(x<0.7): 0.314 P(x<0.9): 0.581	eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.936 P(x<0.2): 0.975 P(x<0.5): 0.993 P(x<0.7): 0.997 P(x<0.9): 0.999	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.058 P(x<0.2): 0.119 P(x<0.5): 0.285 P(x<0.7): 0.424 P(x<0.9): 0.698		P(x<0.1): 0.920 P(x<0.2): 0.961 P(x<0.5): 0.985 P(x<0.7): 0.989 P(x<0.9): 0.991	
	C	P(x<0.1): 0.093 P(x<0.2): 0.150 P(x<0.5): 0.393 P(x<0.7): 0.537 P(x<0.9): 0.739		P(x<0.1): 0.978 P(x<0.2): 0.994 P(x<0.5): 0.998 P(x<0.7): 0.999 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.060 P(x<0.2): 0.111 P(x<0.5): 0.306 P(x<0.7): 0.452 P(x<0.9): 0.730		P(x<0.1): 0.961 P(x<0.2): 0.986 P(x<0.5): 0.993 P(x<0.7): 0.994 P(x<0.9): 0.995	

		MF13	MF14	MF15	MF16
		b_pct (wholesale)	**c_pct** (retail)	**e** (others)	**f_pct** (workshops)
TP	A	P(x<0.1): 0.948 P(x<0.2): 0.969 P(x<0.5): 0.996 P(x<0.7): 0.998 P(x<0.9): 1.000	eliminated due to Δ rank and mean rank > *CD*	P(x<0.1): 0.996 P(x<0.2): 0.999 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	B	P(x<0.1): 0.902 P(x<0.2): 0.935 P(x<0.5): 0.988 P(x<0.7): 0.999 P(x<0.9): 1.000		P(x<0.1): 0.993 P(x<0.2): 0.996 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.999 P(x<0.2): 0.999 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	C	P(x<0.1): 0.750 P(x<0.2): 0.825 P(x<0.5): 0.961 P(x<0.7): 0.994 P(x<0.9): 0.999		P(x<0.1): 0.999 P(x<0.2): 0.999 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	overall	P(x<0.1): 0.864 P(x<0.2): 0.912 P(x<0.5): 0.980 P(x<0.7): 0.998 P(x<0.9): 1.000		P(x<0.1): 0.998 P(x<0.2): 0.999 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.999 P(x<0.2): 0.999 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
		MF17	**MF18**		
		unique_cus(_month)	**unique_cus_year**		
TP	A	P(x<0.1): 0.885 P(x<0.2): 0.942 P(x<0.5): 0.981 P(x<0.7): 0.988 P(x<0.9): 0.995	P(x<0.1): 0.975 P(x<0.2): 0.992 P(x<0.5): 0.999 P(x<0.7): 0.999 P(x<0.9): 1.000		
	B	P(x<0.1): 0.766 P(x<0.2): 0.862 P(x<0.5): 0.950 P(x<0.7): 0.971 P(x<0.9): 0.988	P(x<0.1): 0.946 P(x<0.2): 0.988 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000		
	C	P(x<0.1): 0.761 P(x<0.2): 0.832 P(x<0.5): 0.945 P(x<0.7): 0.965 P(x<0.9): 0.983	P(x<0.1): 0.958 P(x<0.2): 0.984 P(x<0.5): 0.997 P(x<0.7): 0.998 P(x<0.9): 0.999		
	overall	P(x<0.1): 0.818 P(x<0.2): 0.889 P(x<0.5): 0.958 P(x<0.7): 0.977 P(x<0.9): 0.990	P(x<0.1): 0.926 P(x<0.2): 0.980 P(x<0.5): 0.999 P(x<0.7): 1.000 P(x<0.9): 1.000		

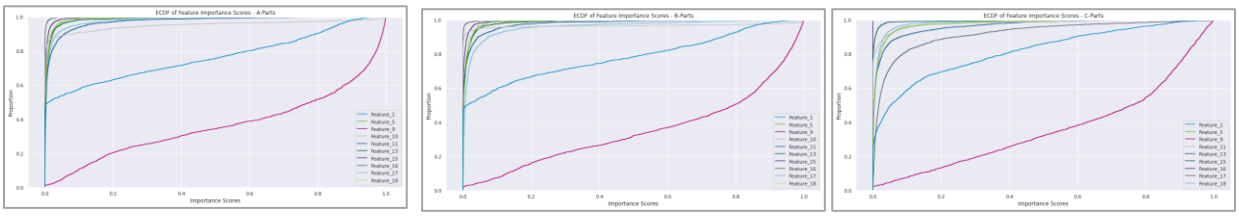


Figure E 7 ECDF of Importance Scores for Customer-related Features - COM

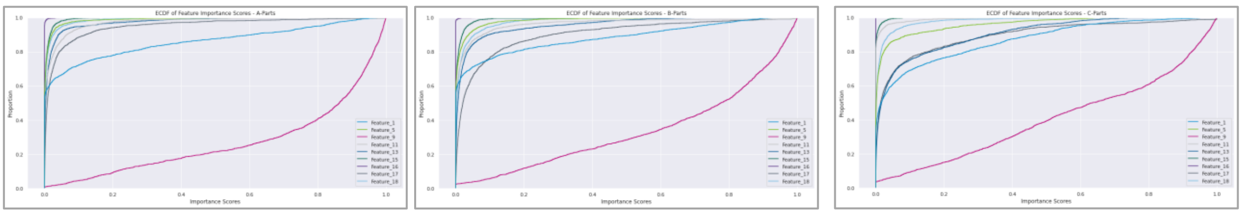


Figure E 8 ECDF of Importance Scores for Customer-related Features - TP

Table E 35: Ranks of Product-related Features - Frequency

	Rank	A		B		C	
COM	1	**csl1**	median: 0.307 +-0.148 MAD: 0.037	**csl1**	median: 0.355 +-0.098 MAD: 0.043	**csl1**	median: 0.603 +-0.120 MAD: 0.028
	2	**lifecycle_age_aa**	median: 0.330 +-0.207 MAD: 0.120	**lifecycle_age_aa**	median: 0.280 +-0.200 MAD: 0.137	**lifecycle_age_aa**	median: 0.237 +-0.147 MAD: 0.102
	3	**ppc**	median: 0.253 +-0.127 MAD: 0.065	**csl1_lag2**	median: 0.270 +-0.095 MAD: 0.030	**csl1_lag2**	median: 0.223 +-0.040 MAD: 0.012
	4	**csl1_lag2**	median: 0.218 +-0.103 MAD: 0.030	**ppc**	median: 0.250 +-0.152 MAD: 0.037	**ppc**	median: 0.232 +-0.142 MAD: 0.043
	5	**csl1_lag1**	median: 0.207 +-0.097 MAD: 0.017	**csl1_lag1**	median: 0.250 +-0.073 MAD: 0.040	**csl1_lag1**	median: 0.210 +-0.045 MAD: 0.018
	6	pallet_qty	median: 0.115 +-0.355 MAD: 0.115	leadtime	median: 0.235 +-0.235 MAD: 0.220	**leadtime**	median: 0.255 +-0.200 MAD: 0.135
	7	leadtime	median: 0.222 +-0.242 MAD: 0.222	pallet_qty	median: 0.092 +-0.295 MAD: 0.092	pallet_qty	median: 0.083 +-0.245 MAD: 0.083
	8	minimumlotsize	median: 0.188 +-0.117 MAD: 0.033	minimumlotsize	median: 0.107 +-0.075 MAD: 0.012	minimumlotsize	median: 0.080 +-0.080 MAD: 0.022
TP	1	**csl1**	median: 0.522 +-0.119 MAD: 0.031	**csl1**	median: 0.522 +-0.112 MAD: 0.033	**csl1**	mean: 0.659 +-0.090 σ: 0.093
	2	**lifecycle_age_aa**	median: 0.217 +-0.154 MAD: 0.078	**csl1_lag1**	median: 0.225 +-0.062 MAD: 0.032	**csl1_lag1**	mean: 0.212 +-0.051 σ: 0.053
	3	**csl1_lag1**	median: 0.225 +-0.059 MAD: 0.029	**lifecycle_age_aa**	median: 0.213 +-0.148 MAD: 0.077	**lifecycle_age_aa**	mean: 0.271 +-0.154 σ: 0.159
	4	**ppc**	median: 0.258 +-0.105 MAD: 0.047	**ppc**	median: 0.255 +-0.107 MAD: 0.043	**leadtime**	mean: 0.216 +-0.200 σ: 0.205
	5	leadtime	median: 0.236 +-0.237 MAD: 0.222	leadtime	median: 0.235 +-0.237 MAD: 0.217	ppc	mean: 0.178 +-0.049 σ: 0.050
	6	pallet_qty	median: 0.105 +-0.297 MAD: 0.105	pallet_qty	median: 0.103 +-0.302 MAD: 0.103	csl1_lag2	mean: 0.169 +-0.025 σ: 0.025
	7	csl1_lag2	median: 0.186 +-0.063 MAD: 0.008	csl1_lag2	median: 0.190 +-0.063 MAD: 0.010	pallet_qty	mean: 0.195 +-0.205 σ:
	8	minimumlotsize	median: 0.132 +-0.081 MAD: 0.034	minimumlotsize	median: 0.130 +-0.080 MAD: 0.033	minimumlotsize	mean: 0.100 +-0.045 σ: 0.046

Table E 36: ECDFs derived from Importance Scores for Product-related Features

		MF1	**MF2**	**MF3**	**MF4**
	ECDF	**ppc**	**pallet_qty**	**minimumlotsize**	**leadtime**
COM	A	P(x<0.1): 0.239 P(x<0.2): 0.501 P(x<0.5): 0.892 P(x<0.7): 0.983 P(x<0.9): 0.999	eliminated due to Δ rank and mean rank > *CD*	eliminated due to Δ rank and mean rank > *CD*	eliminated due to Δ rank and mean rank > *CD*
	B	P(x<0.1): 0.238 P(x<0.2): 0.492 P(x<0.5): 0.882 P(x<0.7): 0.974 P(x<0.9): 0.995			
	C	P(x<0.1): 0.501 P(x<0.2): 0.721 P(x<0.5): 0.966 P(x<0.7): 0.993 P(x<0.9): 0.998			P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	overall	P(x<0.1): 0.423 P(x<0.2): 0.657 P(x<0.5): 0.940 P(x<0.7): 0.988 P(x<0.9): 0.999			P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000

		MF5	MF6	MF7	MF8
		lifecycle_age_aa	**csl1**	**csl1_lag1**	**csl1_lag2**
COM	A	P(x<0.1): 0.317 P(x<0.2): 0.588 P(x<0.5): 0.920 P(x<0.7): 0.987 P(x<0.9): 1.000	P(x<0.1): 0.192 P(x<0.2): 0.527 P(x<0.5): 0.938 P(x<0.7): 0.985 P(x<0.9): 0.999	P(x<0.1): 0.312 P(x<0.2): 0.698 P(x<0.5): 0.990 P(x<0.7): 0.998 P(x<0.9): 1.000	P(x<0.1): 0.344 P(x<0.2): 0.743 P(x<0.5): 0.992 P(x<0.7): 0.999 P(x<0.9): 0.999
	B	P(x<0.1): 0.263 P(x<0.2): 0.539 P(x<0.5): 0.874 P(x<0.7): 0.971 P(x<0.9): 0.999	P(x<0.1): 0.274 P(x<0.2): 0.570 P(x<0.5): 0.931 P(x<0.7): 0.986 P(x<0.9): 0.999	P(x<0.1): 0.343 P(x<0.2): 0.701 P(x<0.5): 0.990 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.431 P(x<0.2): 0.773 P(x<0.5): 0.992 P(x<0.7): 0.999 P(x<0.9): 1.000
	C	P(x<0.1): 0.389 P(x<0.2): 0.697 P(x<0.5): 0.959 P(x<0.7): 0.986 P(x<0.9): 1.000	P(x<0.1): 0.089 P(x<0.2): 0.211 P(x<0.5): 0.602 P(x<0.7): 0.786 P(x<0.9): 0.939	P(x<0.1): 0.468 P(x<0.2): 0.780 P(x<0.5): 0.984 P(x<0.7): 0.995 P(x<0.9): 0.999	P(x<0.1): 0.583 P(x<0.2): 0.862 P(x<0.5): 0.993 P(x<0.7): 0.999 P(x<0.9): 1.000
	overall	P(x<0.1): 0.355 P(x<0.2): 0.646 P(x<0.5): 0.942 P(x<0.7): 0.982 P(x<0.9): 1.000	P(x<0.1): 0.144 P(x<0.2): 0.316 P(x<0.5): 0.703 P(x<0.7): 0.847 P(x<0.9): 0.960	P(x<0.1): 0.423 P(x<0.2): 0.754 P(x<0.5): 0.986 P(x<0.7): 0.995 P(x<0.9): 0.999	P(x<0.1): 0.524 P(x<0.2): 0.835 P(x<0.5): 0.994 P(x<0.7): 0.999 P(x<0.9): 1.000
		MF1	**MF2**	**MF3**	**MF4**
		ppc	**pallet_qty**	**minimumlotsize**	**leadtime**
TP	A	P(x<0.1): 0.265 P(x<0.2): 0.504 P(x<0.5): 0.882 P(x<0.7): 0.973 P(x<0.9): 0.998	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.226 P(x<0.2): 0.438 P(x<0.5): 0.819 P(x<0.7): 0.941 P(x<0.9): 0.990			
	C	eliminated due to Δ rank and mean rank > CD			P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
	overall	P(x<0.1): 0.372 P(x<0.2): 0.538 P(x<0.5): 0.838 P(x<0.7): 0.937 P(x<0.9): 0.983			P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000
		MF5	**MF6**	**MF7**	**MF8**
		lifecycle_age_aa	**csl1**	**csl1_lag1**	**csl1_lag2**
TP	A	P(x<0.1): 0.272 P(x<0.2): 0.509 P(x<0.5): 0.883 P(x<0.7): 0.976 P(x<0.9): 0.999	P(x<0.1): 0.155 P(x<0.2): 0.350 P(x<0.5): 0.835 P(x<0.7): 0.948 P(x<0.9): 0.990	P(x<0.1): 0.233 P(x<0.2): 0.558 P(x<0.5): 0.971 P(x<0.7): 0.996 P(x<0.9): 0.999	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.357 P(x<0.2): 0.629 P(x<0.5): 0.931 P(x<0.7): 0.979 P(x<0.9): 0.999	P(x<0.1): 0.140 P(x<0.2): 0.340 P(x<0.5): 0.753 P(x<0.7): 0.927 P(x<0.9): 0.988	P(x<0.1): 0.324 P(x<0.2): 0.642 P(x<0.5): 0.970 P(x<0.7): 0.997 P(x<0.9): 0.999	
	C	P(x<0.1): 0.088 P(x<0.2): 0.257 P(x<0.5): 0.731 P(x<0.7): 0.898 P(x<0.9): 0.993	P(x<0.1): 0.112 P(x<0.2): 0.238 P(x<0.5): 0.656 P(x<0.7): 0.864 P(x<0.9): 0.979	P(x<0.1): 0.235 P(x<0.2): 0.521 P(x<0.5): 0.944 P(x<0.7): 0.991 P(x<0.9): 0.997	
	overall	P(x<0.1): 0.527 P(x<0.2): 0.732 P(x<0.5): 0.944 P(x<0.7): 0.983 P(x<0.9): 1.000	P(x<0.1): 0.121 P(x<0.2): 0.235 P(x<0.5): 0.545 P(x<0.7): 0.719 P(x<0.9): 0.880	P(x<0.1): 0.552 P(x<0.2): 0.777 P(x<0.5): 0.982 P(x<0.7): 0.996 P(x<0.9): 0.999	

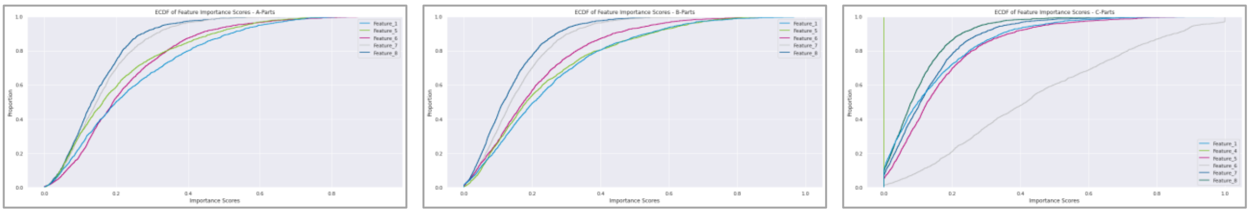


Figure E 10 ECDF of Importance Scores for Product-related Features - COM

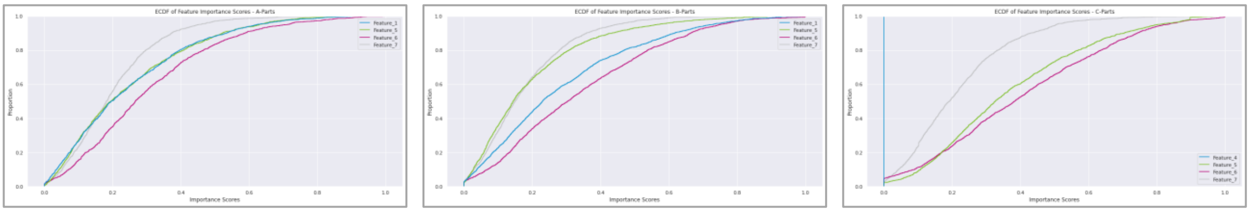


Figure E 9 ECDF of Importance Scores for Product-related Features - TP

Table E 37: Ranks of Vio-related Features - Frequency

	Rank	A		B		C	
COM	1	vio_replaced_dyn	mean: 0.325 +-0.241 σ: 0.248	vio_replaced_dyn	mean: 0.348 +-0.249 σ: 0.256	vio_agebin_1516	median: 0.192 +-0.083 MAD: 0.032
	2	vio_agebin_3-5	mean: 0.275 +-0.177 σ: 0.182	vio_agebin_>16	mean: 0.333 +-0.256 σ: 0.263	vio_agebin_>16	median: 0.295 +-0.393 MAD: 0.225
	3	vio_agebin_1516	mean: 0.214 +-0.046 σ: 0.047	vio_agebin_0-2	mean: 0.357 +-0.285 σ: 0.293	vio_replaced_dyn	median: 0.333 +-0.433 MAD: 0.230
	4	vio_agebin_0-2	mean: 0.321 +-0.256 σ: 0.264	vio_agebin_1516	mean: 0.182 +-0.064 σ: 0.065	vio_agebin_3-5	median: 0.290 +-0.333 MAD: 0.185
	5	vio_agebin_>16	mean: 0.339 +-0.256 σ: 0.263	vio_agebin_6-8	mean: 0.183 +-0.077 σ: 0.080	vio_agebin_0-2	median: 0.368 +-0.435 MAD: 0.337
	6	vio_agebin_6-8	mean: 0.176 +-0.076 σ: 0.079	vio_agebin_3-5	mean: 0.263 +-0.196 σ: 0.201	vio_agebin_12-14	median: 0.138 +-0.088 MAD: 0.017
	7	vio_agebin_9-11	mean: 0.186 +-0.069 σ: 0.071	vio_agebin_12-14	mean: 0.164 +-0.056 σ: 0.057	vio_agebin_6-8	median: 0.123 +-0.087 MAD: 0.028
	8	vio_agebin_12-14	mean: 0.163 +-0.066 σ: 0.068	vio_agebin_9-11	mean: 0.170 +-0.058 σ: 0.060	vio_agebin_9-11	median: 0.123 +-0.087 MAD: 0.040
TP	1	vio_agebin_>16	mean: 0.357 +-0.256 σ: 0.263	vio_agebin_1516	mean: 0.201 +-0.054 σ: 0.055	vio_agebin_12-14	median: 0.140 +-0.082 MAD: 0.040
	2	vio_agebin_1516	mean: 0.200 +-0.052 σ: 0.054	vio_agebin_>16	mean: 0.345 +-0.255 σ: 0.262	vio_replaced_stat	median: 0.257 +-0.435 MAD: 0.232
	3	vio_replaced_stat	mean: 0.317 +-0.270 σ: 0.278	vio_agebin_12-14	mean: 0.182 +-0.064 σ: 0.066	vio_agebin_1516	median: 0.153+- 0.103 MAD: 0.033
	4	vio_agebin_3-5	mean: 0.272 +-0.204 σ: 0.209	vio_agebin_0-2	mean: 0.388 +-0.312 σ: 0.321	vio_agebin_>16	median: 0.315 +-0.388 MAD: 0.243
	5	vio_agebin_0-2	mean: 0.382 +-0.337 σ: 0.347	vio_agebin_3-5	mean: 0.273 +-0.201 σ: 0.207	vio_agebin_0-2	median: 0.380 +-0.440 MAD: 0.355
	6	vio_agebin_9-11	mean: 0.168 +-0.071 σ: 0.073	vio_replaced_stat	mean: 0.306 +-0.273 σ: 0.280	vio_agebin_3-5	median: 0.280 +-0.323 MAD: 0.242
	7	vio_agebin_12-14	mean: 0.165 +-0.072 σ: 0.074	vio_agebin_9-11	mean: 0.164 +-0.069 σ: 0.071	vio_agebin_9-11	median: 0.120 +-0.080 MAD: 0.050

8	vio_agebin_6-8	mean: 0.139 +-0.066 σ: 0.068	vio_agebin_6-8	mean: 0.141 +-0.070 σ: 0.072	vio_agebin_6-8	median: 0.097 +-0.058 MAD: 0.025

Table E 38: ECDFs derived from Importance Scores for VIO-related Features

		MF1	**MF2**	**MF3**	**MF4**
	ECDF	**vio_replaced_dyn**	**vio_agebin_0-2**	**vio_agebin_3-5**	**vio_agebin_6-8**
COM	A	P(x<0.1): 0.343 P(x<0.2): 0.521 P(x<0.5): 0.933 P(x<0.7): 0.987 P(x<0.9): 0.996	P(x<0.1): 0.508 P(x<0.2): 0.663 P(x<0.5): 0.966 P(x<0.7): 0.992 P(x<0.9): 0.999	P(x<0.1): 0.515 P(x<0.2): 0.693 P(x<0.5): 0.950 P(x<0.7): 0.995 P(x<0.9): 1.000	P(x<0.1): 0.505 P(x<0.2): 0.655 P(x<0.5)**: 0.963** **P(x<0.7): 0.**996 P(x<0.9): 0.999
	B	P(x<0.1): 0.434 P(x<0.2): 0.587 P(x<0.5): 0.920 P(x<0.7): 0.986 P(x<0.9): 0.996	P(x<0.1): 0.657 P(x<0.2): 0.764 P(x<0.5): 0.964 P(x<0.7): 0.988 P(x<0.9): 0.996	P(x<0.1): 0.635 P(x<0.2): 0.756 P(x<0.5): 0.965 P(x<0.7): 0.989 P(x<0.9): 0.998	P(x<0.1): 0.581 P(x<0.2): 0.710 P(x<0.5): 0.950 P(x<0.7): 0.989 P(x<0.9): 0.999
	C	P(x<0.1): 0.491 P(x<0.2): 0.691 P(x<0.5): 0.936 P(x<0.7): 0.972 P(x<0.9): 0.989	P(x<0.1): 0.896 P(x<0.2): 0.939 P(x<0.5): 0.986 P(x<0.7): 0.995 P(x<0.9): 0.997	P(x<0.1): 0.844 P(x<0.2): 0.908 P(x<0.5): 0.967 P(x<0.7): 0.984 P(x<0.9): 0.995	P(x<0.1): 0.756 P(x<0.2): 0.847 P(x<0.5): 0.951 P(x<0.7): 0.974 P(x<0.9): 0.998
	overall	P(x<0.1): 0.494 P(x<0.2): 0.654 P(x<0.5): 0.921 P(x<0.7): 0.963 P(x<0.9): 0.984	P(x<0.1): 0.810 P(x<0.2): 0.879 P(x<0.5): 0.972 P(x<0.7): 0.984 P(x<0.9): 0.991	P(x<0.1): 0.749 P(x<0.2): 0.834 P(x<0.5): 0.934 P(x<0.7): 0.961 P(x<0.9): 0.980	P(x<0.1): 0.714 P(x<0.2): 0.812 P(x<0.5): 0.950 P(x<0.7): 0.977 P(x<0.9): 0.992
		MF5	**MF6**	**MF7**	**MF8**
		vio_agebin_9-11	**vio_agebin_12-14**	**vio_agebin_1516**	**vio_agebin_>16**
COM	A	P(x<0.1): 0.587 P(x<0.2): 0.719 P(x<0.5): 0.962 P(x<0.7): 0.991 P(x<0.9): 0.999	P(x<0.1): 0.707 P(x<0.2): 0.805 P(x<0.5): 0.963 P(x<0.7): 0.989 P(x<0.9): 0.999	P(x<0.1): 0.866 P(x<0.2): 0.922 P(x<0.5): 0.982 P(x<0.7): 0.994 P(x<0.9): 1.000	P(x<0.1): 0.898 P(x<0.2): 0.937 P(x<0.5): 0.995 P(x<0.7): 0.998 P(x<0.9): 1.000
	B	P(x<0.1): 0.561 P(x<0.2): 0.718 P(x<0.5): 0.917 P(x<0.7): 0.963 P(x<0.9): 0.995	P(x<0.1): 0.670 P(x<0.2): 0.776 P(x<0.5): 0.939 P(x<0.7): 0.980 P(x<0.9): 1.000	P(x<0.1): 0.799 P(x<0.2): 0.877 P(x<0.5): 0.963 P(x<0.7): 0.987 P(x<0.9): 0.999	P(x<0.1): 0.853 P(x<0.2): 0.899 P(x<0.5): 0.967 P(x<0.7): 0.982 P(x<0.9): 0.999
	C	P(x<0.1): 0.672 P(x<0.2): 0.783 P(x<0.5): 0.899 P(x<0.7): 0.951 P(x<0.9): 0.992	P(x<0.1): 0.545 P(x<0.2): 0.691 P(x<0.5): 0.843 P(x<0.7): 0.922 P(x<0.9): 0.984	P(x<0.1): 0.578 P(x<0.2): 0.720 P(x<0.5): 0.885 P(x<0.7): 0.938 P(x<0.9): 0.988	P(x<0.1): 0.658 P(x<0.2): 0.771 P(x<0.5): 0.921 P(x<0.7): 0.956 P(x<0.9): 0.994
	overall	P(x<0.1): 0.634 P(x<0.2): 0.735 P(x<0.5): 0.902 P(x<0.7): 0.938 P(x<0.9): 0.977	P(x<0.1): 0.616 P(x<0.2): 0.741 P(x<0.5): 0.885 P(x<0.7): 0.927 P(x<0.9): 0.978	P(x<0.1): 0.709 P(x<0.2): 0.804 P(x<0.5): 0.924 P(x<0.7): 0.956 P(x<0.9): 0.988	P(x<0.1): 0.765 P(x<0.2): 0.838 P(x<0.5): 0.925 P(x<0.7): 0.954 P(x<0.9): 0.981
		MF1	**MF2**	**MF3**	**MF4**
		vio_replaced_stat	**vio_agebin_0-2**	**vio_agebin_3-5**	**vio_agebin_6-8**
TP	A	P(x<0.1): 0.441 P(x<0.2): 0.650 P(x<0.5): 0.966 P(x<0.7): 0.990 P(x<0.9): 1.000	P(x<0.1): 0.626 P(x<0.2): 0.761 P(x<0.5): 0.973 P(x<0.7): 0.997 P(x<0.9): 0.997	P(x<0.1): 0.636 P(x<0.2): 0.781 P(x<0.5): 0.990 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.572 P(x<0.2): 0.741 P(x<0.5): 0.943 P(x<0.7): 0.983 P(x<0.9): 0.997
	B	P(x<0.1): 0.409 P(x<0.2): 0.592 P(x<0.5): 0.923 P(x<0.7): 0.977 P(x<0.9): 0.991	P(x<0.1): 0.704 P(x<0.2): 0.814 P(x<0.5): 0.961 P(x<0.7): 0.987 P(x<0.9): 0.997	P(x<0.1): 0.696 P(x<0.2): 0.836 P(x<0.5): 0.986 P(x<0.7): 0.996 P(x<0.9): 0.998	P(x<0.1): 0.593 P(x<0.2): 0.712 P(x<0.5): 0.934 P(x<0.7): 0.966 P(x<0.9): 0.994
	C	P(x<0.1): 0.447 P(x<0.2): 0.605 P(x<0.5): 0.902 P(x<0.7): 0.958 P(x<0.9): 0.984	P(x<0.1): 0.839 P(x<0.2): 0.885 P(x<0.5): 0.973 P(x<0.7): 0.988 P(x<0.9): 0.994	P(x<0.1): 0.784 P(x<0.2): 0.864 P(x<0.5): 0.966 P(x<0.7): 0.980 P(x<0.9): 0.989	P(x<0.1): 0.708 P(x<0.2): 0.811 P(x<0.5): 0.928 P(x<0.7): 0.967 P(x<0.9): 0.988
	overall	P(x<0.1): 0.444 P(x<0.2): 0.630 P(x<0.5): 0.897 P(x<0.7): 0.943 P(x<0.9): 0.967	P(x<0.1): 0.811 P(x<0.2): 0.891 P(x<0.5): 0.959 P(x<0.7): 0.973 P(x<0.9): 0.982	P(x<0.1): 0.797 P(x<0.2): 0.906 P(x<0.5): 0.984 P(x<0.7): 0.990 P(x<0.9): 0.996	P(x<0.1): 0.667 P(x<0.2): 0.784 P(x<0.5): 0.897 P(x<0.7): 0.957 P(x<0.9): 0.982
		MF5	**MF6**	**MF7**	**MF8**
		vio_agebin_9-11	**vio_agebin_12-14**	**vio_agebin_1516**	**vio_agebin_>16**
TP	A	P(x<0.1): 0.461 P(x<0.2): 0.626	P(x<0.1): 0.593 P(x<0.2): 0.764	P(x<0.1): 0.761 P(x<0.2): 0.852	P(x<0.1): 0.822 P(x<0.2): 0.906

	P(x<0.5): 0.919 P(x<0.7): 0.973 P(x<0.9): 0.983	P(x<0.5): 0.946 P(x<0.7): 0.976 P(x<0.9): 0.997	P(x<0.5): 0.970 P(x<0.7): 0.983 P(x<0.9): 0.997	P(x<0.5): 0.987 P(x<0.7): 0.993 P(x<0.9): 0.993
B	P(x<0.1): 0.540 P(x<0.2): 0.696 P(x<0.5): 0.930 P(x<0.7): 0.967 P(x<0.9): 0.994	P(x<0.1): 0.599 P(x<0.2): 0.745 P(x<0.5): 0.928 P(x<0.7): 0.966 P(x<0.9): 0.994	P(x<0.1): 0.735 P(x<0.2): 0.840 P(x<0.5): 0.962 P(x<0.7): 0.984 P(x<0.9): 0.998	P(x<0.1): 0.814 P(x<0.2): 0.889 P(x<0.5): 0.972 P(x<0.7): 0.988 P(x<0.9): 0.998
C	P(x<0.1): 0.594 P(x<0.2): 0.714 P(x<0.5): 0.911 P(x<0.7): 0.951 P(x<0.9): 0.985	P(x<0.1): 0.607 P(x<0.2): 0.728 P(x<0.5): 0.898 P(x<0.7): 0.951 P(x<0.9): 0.987	P(x<0.1): 0.689 P(x<0.2): 0.785 P(x<0.5): 0.927 P(x<0.7): 0.962 P(x<0.9): 0.994	P(x<0.1): 0.748 P(x<0.2): 0.828 P(x<0.5): 0.949 P(x<0.7): 0.972 P(x<0.9): 0.989
overall	P(x<0.1): 0.600 P(x<0.2): 0.708 P(x<0.5): 0.901 P(x<0.7): 0.940 P(x<0.9): 0.975	P(x<0.1): 0.606 P(x<0.2): 0.741 P(x<0.5): 0.912 P(x<0.7): 0.949 P(x<0.9): 0.963	P(x<0.1): 0.752 P(x<0.2): 0.832 P(x<0.5): 0.951 P(x<0.7): 0.977 P(x<0.9): 0.979	P(x<0.1): 0.805 P(x<0.2): 0.842 P(x<0.5): 0.934 P(x<0.7): 0.963 P(x<0.9): 0.977

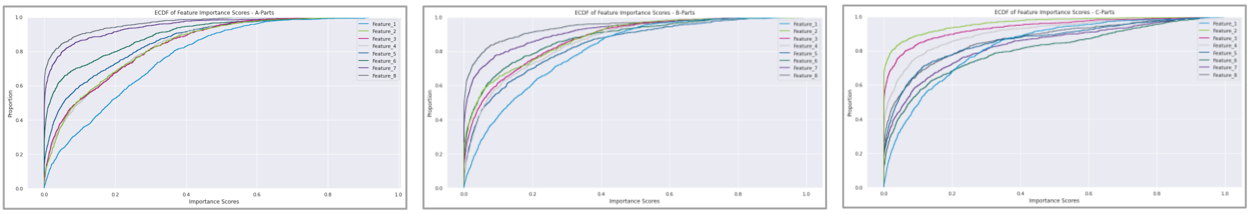


Figure E 11 ECDF of Importance Scores for VIO-related Features - COM

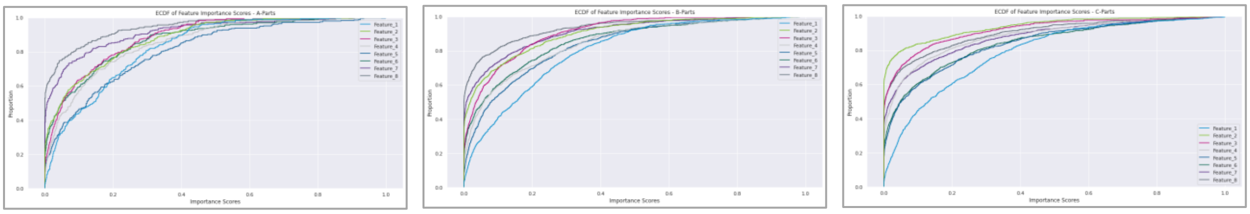


Figure E 12 ECDF of Importance Scores for VIO-related Features - TP

Table E 39: Ranks of Price-related Features - Frequency

	Rank	A		B		C	
COM	1	**IppinGC_perPC**	mean: 0.364 +-0.110 σ: 0.111	**IppinGC_perPC**	mean: 0.305 +-0.104 σ: 0.105	**pct_change_exratelcgc_ma2cent**	mean: 0.298 +-0.158 σ: 0.159
	2	**IppinLC_perPC**	mean: 0.262 +-0.126 σ: 0.127	**pct_change_exratelcgc_ma2cent**	mean: 0.268 +-0.135 σ: 0.136	**IppinGC_perPC**	mean: 0.269 +-0.092 σ: 0.093
	3	**pct_change_exratelcgc_ma2cent**	mean: 0.245 +-0.115 σ: 0.115	**IppinLC_perPC**	mean: 0.250 +-0.113 σ: 0.114	**IppinLC_perPC**	mean: 0.323 +-0.255 σ: 0.257
	4	**PPC**	mean: 0.239 +-0.190 σ: 0.191	**exratelcgc_ma2cent**	mean: 0.304 +-0.257 σ: 0.259	**exratelcgc_ma2cent**	mean: 0.233 +-0.103 σ: 0.104
	5	**bonus**	mean: 0.263 +-0.303 σ: 0.267	**bonus**	mean: 0.283 +-0.306 σ: 0.309	**bonus**	mean: 0.292 +-0.317 σ: 0.319
	6	**exratelcgc_ma2cent**	mean: 0.297 +-0.265 σ: 0.305	**PPC**	mean: 0.232 +-0.209 σ: 0.211	**PPC**	mean: 0.225 +-0.185 σ: 0.186
	7	rebate	mean: 0.151 +-0.144 σ: 0.145	rebate	mean: 0.162 +-0.164 σ: 0.165	rebate	mean: 0.169 +-0.170 σ: 0.172
TP	1	**bonus**	mean: 0.371 +-0.269 σ: 0.272	**bonus**	median: 0.263 +-0.435 MAD: 0.155	**pct_change_exratelcgc_ma2cent**	median: 0.300 +-0.340 MAD: 0.100

2	**pct_change_exratelcgc_ma2cent**	mean: 0.304 +-0.175 σ: 0.177	**pct_change_exratelcgc_ma2cent**	median: 0.307 +-0.343 MAD: 0.089	**IppinGC_perPC**	median: 0.230 +-0.140 MAD: 0.080
3	**exratelcgc_ma2cent**	mean: 0.313 +-0.267 σ: 0.269	**exratelcgc_ma2cent**	median: 0.263 +-0.481 MAD: 0.130	**bonus**	median: 0.220 +-0.440 MAD: 0.160
4	**IppinGC_perPC**	mean: 0.199 +-0.069 σ: 0.069	**IppinGC_perPC**	median: 0.196 +-0.118 MAD: 0.046	**IppinLC_perPC**	median: 0.240 +-0.480 MAD: 0.170
5	**IppinLC_perPC**	mean: 0.192 +-0.077 σ: 0.078	**IppinLC_perPC**	median: 0.174 +-0.138 MAD: 0.048	**exratelcgc_ma2cent**	median: 0.140+-0.280 MAD: 0.100
6	**PPC**	mean: 0.209 +-0.198 σ: 0.200	**PPC**	median: 0.106 +-0.266 MAD: 0.051	**PPC**	median: 0.190 +-0.150 MAD: 0.050
7	**rebate**	mean: 0.225 +-0.220 σ: 0.222	**rebate**	median: 0.220 +-0.321 MAD: 0.220	rebate	median: 0.220 +-0.310 MAD: 0.200

Table E 40: ECDFs derived from Importance Scores for Price-related Features

$ECDF$		**MF1** **exratelcgc_ma2cent**	**MF2** **pct_change_exratelcgc_ma2cent**	**MF3** **PPC**	**MF4** **IndexPrice**
COM	A	P(x<0.1): 0.126 P(x<0.2): 0.498 P(x<0.5): 0.969 P(x<0.7): 0.993 P(x<0.9): 0.999	P(x<0.1): 0.563 P(x<0.2): 0.889 P(x<0.5): 0.995 P(x<0.7): 0.998 P(x<0.9): 0.999	P(x<0.1): 0.564 P(x<0.2): 0.764 P(x<0.5): 0.947 P(x<0.7): 0.996 P(x<0.9): 1.000	Eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.115 P(x<0.2): 0.421 P(x<0.5): 0.936 P(x<0.7): 0.985 P(x<0.9): 0.995	P(x<0.1): 0.558 P(x<0.2): 0.888 P(x<0.5): 0.998 P(x<0.7): 0.999 P(x<0.9): 1.000	P(x<0.1): 0.615 P(x<0.2): 0.791 P(x<0.5): 0.963 P(x<0.7): 0.996 P(x<0.9): 1.000	
	C	P(x<0.1): 0.105 P(x<0.2): 0.387 P(x<0.5): 0.891 P(x<0.7): 0.971 P(x<0.9): 0.991	P(x<0.1): 0.496 P(x<0.2): 0.846 P(x<0.5): 0.988 P(x<0.7): 0.996 P(x<0.9): 0.998	P(x<0.1): 0.642 P(x<0.2): 0.820 P(x<0.5): 0.977 P(x<0.7): 0.997 P(x<0.9): 1.000	
	overall	P(x<0.1): 0.125 P(x<0.2): 0.423 P(x<0.5): 0.927 P(x<0.7): 0.986 P(x<0.9): 0.996	P(x<0.1): 0.540 P(x<0.2): 0.881 P(x<0.5): 0.995 P(x<0.7): 0.998 P(x<0.9): 0.999	P(x<0.1): 0.621 P(x<0.2): 0.796 P(x<0.5): 0.966 P(x<0.7): 0.997 P(x<0.9): 1.000	
		MF5 **RRP**	**MF6** **IPPinLC_perPC**	**MF7** **IPPinGC_perPC**	**MF8** **GPPinGC_perPC**
COM	A	Eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.192 P(x<0.2): 0.516 P(x<0.5): 0.938 P(x<0.7): 0.986 P(x<0.9): 1.000	P(x<0.1): 0.158 P(x<0.2): 0.496 P(x<0.5): 0.935 P(x<0.7): 0.992 P(x<0.9): 0.999	eliminated due to r_{f_7,f_8} >0.9
	B		P(x<0.1): 0.247 P(x<0.2): 0.585 P(x<0.5): 0.943 P(x<0.7): 0.987 P(x<0.9): 1.000	P(x<0.1): 0.175 P(x<0.2): 0.464 P(x<0.5): 0.933 P(x<0.7): 0.991 P(x<0.9): 0.999	
	C		P(x<0.1): 0.302 P(x<0.2): 0.613 P(x<0.5): 0.960 P(x<0.7): 0.993 P(x<0.9): 0.998	P(x<0.1): 0.204 P(x<0.2): 0.501 P(x<0.5): 0.933 P(x<0.7): 0.990 P(x<0.9): 0.999	
	overall		P(x<0.1): 0.251 P(x<0.2): 0.577 P(x<0.5): 0.952 P(x<0.7): 0.988 P(x<0.9): 0.999	P(x<0.1): 0.171 P(x<0.2): 0.466 P(x<0.5): 0.934 P(x<0.7): 0.987 P(x<0.9): 0.999	
		MF9 **NPPinGC_perPC**	**MF10** **Rebate [%]**	**MF11** **Bonus [%]**	
COM	A	eliminated due to r_{f_7,f_9} >0.9	eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.795 P(x<0.2): 0.883 P(x<0.5): 0.981 P(x<0.7): 0.997 P(x<0.9): 0.999	
	B			P(x<0.1): 0.794 P(x<0.2): 0.891 P(x<0.5): 0.980	

				P(x<0.7): 0.997 P(x<0.9): 1.000	
	C			P(x<0.1): 0.791 P(x<0.2): 0.878 P(x<0.5): 0.983 P(x<0.7): 0.997 P(x<0.9): 0.999	
	overall			P(x<0.1): 0.795 P(x<0.2): 0.883 P(x<0.5): 0.981 P(x<0.7): 0.997 P(x<0.9): 0.999	
		MF1	**MF2**	**MF3**	**MF4**
		exratelcgc_ma2cent	**pct_change_exratelcgc _ma2cent**	**PPC**	**IndexPrice**
TP	A	P(x<0.1): 0.115 P(x<0.2): 0.463 P(x<0.5): 0.949 P(x<0.7): 0.994 P(x<0.9): 0.999	P(x<0.1): 0.538 P(x<0.2): 0.857 P(x<0.5): 0.996 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.595 P(x<0.2): 0.809 P(x<0.5): 0.963 P(x<0.7): 0.995 P(x<0.9): 1.000	eliminated due to Δ rank and mean rank > CD
	B	P(x<0.1): 0.103 P(x<0.2): 0.398 P(x<0.5): 0.923 P(x<0.7): 0.986 P(x<0.9): 0.996	P(x<0.1): 0.513 P(x<0.2): 0.841 P(x<0.5): 0.992 P(x<0.7): 0.996 P(x<0.9): 1.000	P(x<0.1): 0.551 P(x<0.2): 0.782 P(x<0.5): 0.982 P(x<0.7): 0.998 P(x<0.9): 0.999	
	C	P(x<0.1): 0.107 P(x<0.2): 0.333 P(x<0.5): 0.878 P(x<0.7): 0.973 P(x<0.9): 0.992	P(x<0.1): 0.470 P(x<0.2): 0.812 P(x<0.5): 0.987 P(x<0.7): 0.995 P(x<0.9): 0.999	P(x<0.1): 0.540 P(x<0.2): 0.777 P(x<0.5): 0.978 P(x<0.7): 0.996 P(x<0.9): 0.999	
	overall	P(x<0.1): 0.108 P(x<0.2): 0.371 P(x<0.5): 0.892 P(x<0.7): 0.977 P(x<0.9): 0.994	P(x<0.1): 0.491 P(x<0.2): 0.815 P(x<0.5): 0.991 P(x<0.7): 0.996 P(x<0.9): 0.999	P(x<0.1): 0.556 P(x<0.2): 0.775 P(x<0.5): 0.981 P(x<0.7): 0.997 P(x<0.9): 1.000	
		MF5	**MF6**	**MF7**	**MF8**
		RRP	**IPPinLC_perPC**	**IPPinGC_perPC**	**GPPinGC_perPC**
TP	A	eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.265 P(x<0.2): 0.619 P(x<0.5): 0.948 P(x<0.7): 0.988 P(x<0.9): 1.000	P(x<0.1): 0.184 P(x<0.2): 0.536 P(x<0.5): 0.944 P(x<0.7): 0.982 P(x<0.9): 1.000	eliminated due to r_{f_7,f_8} >0.9
	B		P(x<0.1): 0.362 P(x<0.2): 0.683 P(x<0.5): 0.962 P(x<0.7): 0.989 P(x<0.9): 1.000	P(x<0.1): 0.233 P(x<0.2): 0.572 P(x<0.5): 0.956 P(x<0.7): 0.991 P(x<0.9): 1.000	
	C		P(x<0.1): 0.539 P(x<0.2): 0.803 P(x<0.5): 0.983 P(x<0.7): 0.997 P(x<0.9): 1.000	P(x<0.1): 0.346 P(x<0.2): 0.634 P(x<0.5): 0.972 P(x<0.7): 0.991 P(x<0.9): 0.999	
	overall		P(x<0.1): 0.466 P(x<0.2): 0.733 P(x<0.5): 0.975 P(x<0.7): 0.993 P(x<0.9): 1.000	P(x<0.1): 0.319 P(x<0.2): 0.626 P(x<0.5): 0.964 P(x<0.7): 0.989 P(x<0.9): 0.999	
		MF9	**MF10**	**MF11**	
		NPPinGC_perPC	**Rebate [%]**	**Bonus [%]**	
TP	A	eliminated due to r_{f_7,f_9} >0.9	P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.669 P(x<0.2): 0.831 P(x<0.5): 0.968 P(x<0.7): 0.999 P(x<0.9): 1.000	
	B		P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.602 P(x<0.2): 0.812 P(x<0.5): 0.982 P(x<0.7): 0.999 P(x<0.9): 1.000	
	C		eliminated due to Δ rank and mean rank > CD	P(x<0.1): 0.550 P(x<0.2): 0.749 P(x<0.5): 0.949 P(x<0.7): 0.975 P(x<0.9): 0.990	
	overall		P(x<0.1): 1.000 P(x<0.2): 1.000 P(x<0.5): 1.000 P(x<0.7): 1.000 P(x<0.9): 1.000	P(x<0.1): 0.584 P(x<0.2): 0.769 P(x<0.5): 0.954 P(x<0.7): 0.985 P(x<0.9): 0.995	

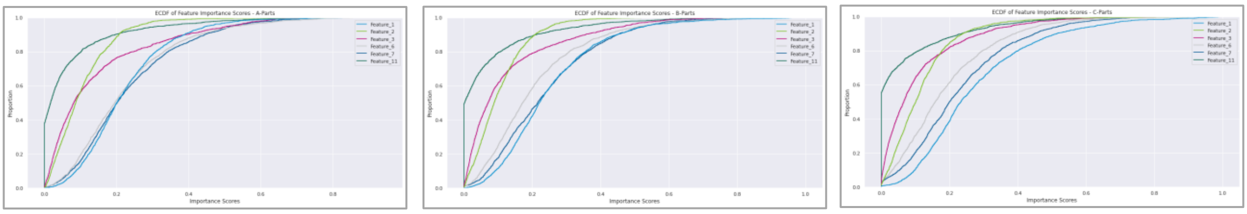


Figure E 14 ECDF of Importance Scores for Price-related Features - COM

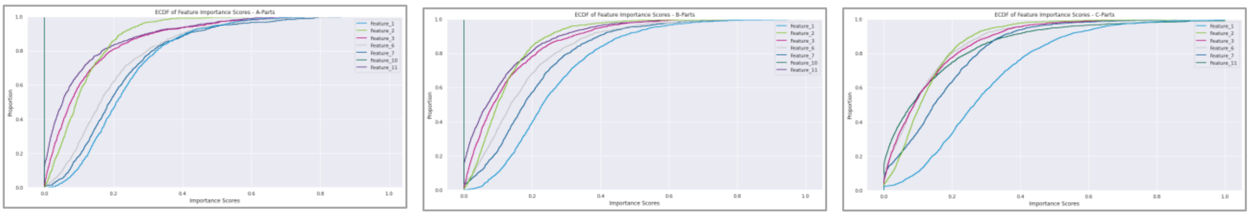


Figure E 13 ECDF of Importance Scores for Price-related Features - TP

Table E 41: ECDFs derived from Importance Scores for Time-related Features

		MF1	**MF2**	**MF3**
	ECDF	**is_holdiay**	**is_seasonal**	**seasonality_pct**
COM	A	P(x<0.1): 0.929	P(x<0.1): 0.805	P(x<0.1): 0.022
		P(x<0.2): 0.981	P(x<0.2): 0.859	P(x<0.2): 0.031
		P(x<0.5): 0.986	P(x<0.5): 0.955	P(x<0.5): 0.070
		P(x<0.7): 0.986	P(x<0.7): 0.980	P(x<0.7): 0.134
		P(x<0.9): 0.986	P(x<0.9): 0.997	P(x<0.9): 0.301
	B	P(x<0.1): 0.935	P(x<0.1): 0.675	P(x<0.1): 0.017
		P(x<0.2): 0.989	P(x<0.2): 0.784	P(x<0.2): 0.021
		P(x<0.5): 0.991	P(x<0.5): 0.950	P(x<0.5): 0.068
		P(x<0.7): 0.992	P(x<0.7): 0.987	P(x<0.7): 0.172
		P(x<0.9): 0.992	P(x<0.9): 0.998	P(x<0.9): 0.464
	C	P(x<0.1): 0.933	P(x<0.1): 0.221	P(x<0.1): 0.103
		P(x<0.2): 0.978	P(x<0.2): 0.313	P(x<0.2): 0.175
		P(x<0.5): 0.988	P(x<0.5): 0.605	P(x<0.5): 0.457
		P(x<0.7): 0.989	P(x<0.7): 0.795	P(x<0.7): 0.654
		P(x<0.9): 0.989	P(x<0.9): 0.946	P(x<0.9): 0.847
	overall	P(x<0.1): 0.940	P(x<0.1): 0.360	P(x<0.1): 0.079
		P(x<0.2): 0.980	P(x<0.2): 0.452	P(x<0.2): 0.131
		P(x<0.5): 0.989	P(x<0.5): 0.700	P(x<0.5): 0.346
		P(x<0.7): 0.990	P(x<0.7): 0.844	P(x<0.7): 0.520
		P(x<0.9): 0.990	P(x<0.9): 0.959	P(x<0.9): 0.724
		MF1	**MF2**	**MF3**
		is_holdiay	**is_seasonal**	**seasonality_pct**
TP	A	P(x<0.1): 0.918	P(x<0.1): 0.672	P(x<0.1): 0.039
		P(x<0.2): 0.972	P(x<0.2): 0.759	P(x<0.2): 0.051
		P(x<0.5): 0.982	P(x<0.5): 0.922	P(x<0.5): 0.116
		P(x<0.7): 0.984	P(x<0.7): 0.966	P(x<0.7): 0.224
		P(x<0.9): 0.984	P(x<0.9): 0.991	P(x<0.9): 0.473
	B	P(x<0.1): 0.889	P(x<0.1): 0.492	P(x<0.1): 0.052
		P(x<0.2): 0.959	P(x<0.2): 0.618	P(x<0.2): 0.062
		P(x<0.5): 0.976	P(x<0.5): 0.885	P(x<0.5): 0.173
		P(x<0.7): 0.976	P(x<0.7): 0.961	P(x<0.7): 0.360
		P(x<0.9): 0.977	P(x<0.9): 0.990	P(x<0.9): 0.660
	C	P(x<0.1): 0.887	P(x<0.1): 0.362	P(x<0.1): 0.132
		P(x<0.2): 0.942	P(x<0.2): 0.434	P(x<0.2): 0.192
		P(x<0.5): 0.970	P(x<0.5): 0.649	P(x<0.5): 0.427
		P(x<0.7): 0.971	P(x<0.7): 0.814	P(x<0.7): 0.589
		P(x<0.9): 0.971	P(x<0.9): 0.935	P(x<0.9): 0.750
	overall	P(x<0.1): 0.895	P(x<0.1): 0.413	P(x<0.1): 0.117
		P(x<0.2): 0.954	P(x<0.2): 0.489	P(x<0.2): 0.169
		P(x<0.5): 0.972	P(x<0.5): 0.698	P(x<0.5): 0.373
		P(x<0.7): 0.972	P(x<0.7): 0.839	P(x<0.7): 0.525
		P(x<0.9): 0.972	P(x<0.9): 0.945	P(x<0.9): 0.702

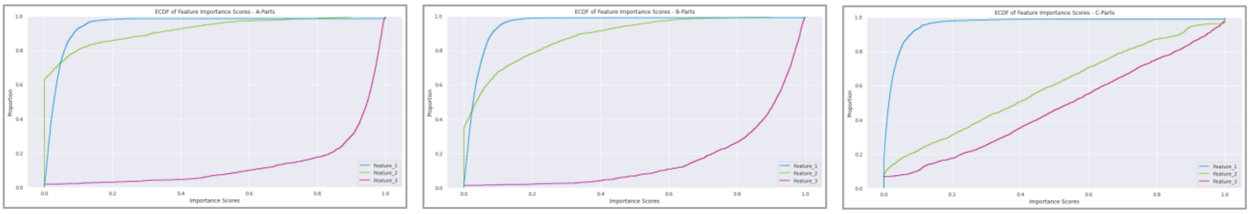


Figure E 16 ECDF of Importance Scores for Time-related Features - COM

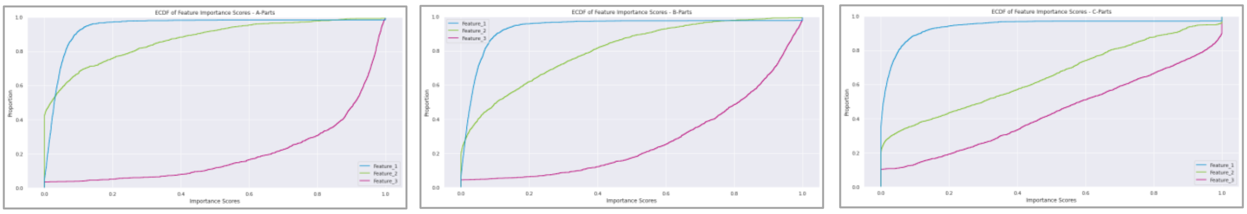


Figure E 15 ECDF of Importance Scores for Time-related Features - TP

Appendix 12: Characteristics of Vehicles in Operation (VIO)

In order to verify to what extent the simulated data correspond to the real conditions, the countries vehicle populations are characterized with regard to age, sector, segment and fuel type.

Age

For C1, passenger cars amount to approximately 48.2 million vehicles in 2020 and are thus the vehicle class with the highest share. It is a rather young vehicle fleet – in 2010, the average age was 8.1 years; in 2021, it was 9.1 years. The trend is upward. A cross-check between figures from KBA (Kraftfahrt-Bundesamt (1): 2021) and the simulated data set reveal high congruency and the following figures:

- 18.7% are aged zero to two years.
- 25% are aged five to nine years.
- 18.9% are aged 15 to 29 years.

For C2, passenger cars amount to about 39.5 million in 2020. VIO in total amount to 43 million. The average age is 11.4 years.

- According to the simulated data 25% are aged zero to 5 years.
- Another 25% are between six and ten years.
- Another 30% are aged eleven to 15 years.
- 10% are older than 20 years.

For C3, passenger cars amount to 49.1 million vehicles in 2020. The overall population amounts to 55.8 million. The average age is 12.5 years.

- According to the simulated data 21% are zero to five years old; according to Autostat (Autostat (2): 2021) it is a share of 28%.
- One quarter of the VIO are aged five to ten years.
- More than 10% are older than 20 years.

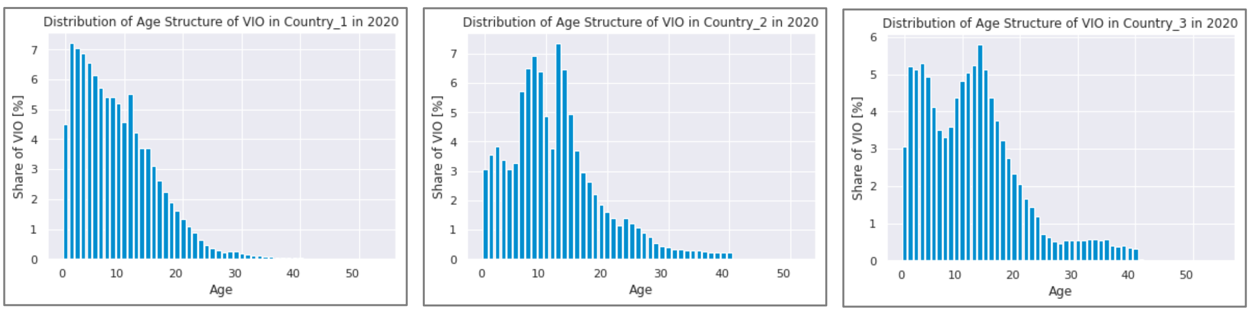


Figure E 17 Distribution of the Age Structure according to Simulated Data

To analyze if the age structure of VIOs is a relevant factor, so-called age bin features, representing the proportional strength of each age class, are built. In total eight different MFs are tested for their effect: vio, vio_agebin_0-2, vio_agebin_3-5, vio_agebin_6-8, vio_agebin_9-11, vio_agebin_12-14, vio_agebin_1516, and vio_agebin_>16.

As shown in Table E 42, Table E 43 and Table E 44, the age structure of the vehicle fleet is a relevant indicator in three respects:

- The older the vehicle, the higher the mileage and the likelihood of age-related wear and tear, hence the higher the absolute replacement frequency of the installed parts. As demonstrated for C1, high importance is assigned to
  - age bin 0-2: they are assumed to be related to OES business and random failures.
  - age bins 3 – 14: they are assumed to be related to IAM business, random and wear-and-tear failures.

  Significantly lower significance is attributed to bins starting at age fifteen.
- The proportional strength of the respective age group defines the product status i.e., the relative replacement frequency of the installed parts: Products that can be assigned to the age group that is most strongly represented fall into category A. If the products are installed in age groups with decreasing proportional strength, the probability of an assignment to value contribution classes B and C increases.
- Moreover, it is noticeable that products from class A are majorly being installed in 'young' vehicles.

Hence, once could conclude, that a products status with regard to value contribution shifts with age and proportional strength of a country's vehicle population.

These phenomena are very evident for COM. A possible explanation is the market position of the company considered: While people are turning to COM for 'younger' vehicles from the company considered and to price leaders/competitors for 'older' vehicles, demand for TP across the various age bins is more levelled due to the oligopolistic respectively market-leading position of the company considered.

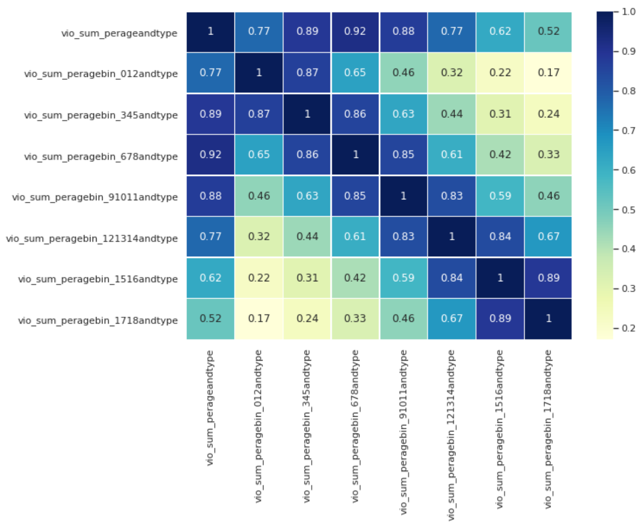


Figure E 18 Intra-Feature Correlation Testing for Age Structure of VIO

Table E 42: Importance Testing for Age Structure of VIO in C1 - Elasticity

	RI^f ($RI^f w/ MF0$)	MF0	MF1	MF2	MF3
		vio	vio_agebin_0-2	vio_agebin_3-5	vio_agebin_6-8
COM	A	-	0.193	0.211	0.222
	B	-	0.151	0.158	0.204
	C	-	0.047	0.078	0.122
	overall	- (0.169)	0.082 (0.070)	0.122 (0.095)	0.147 (0.116)
	accumulated VIO [%]	-	18.729	38.312	54.873
		MF4	**MF5**	**MF6**	**MF7**
		vio_agebin_9-11	vio_agebin_12-14	vio_agebin_1516	vio_agebin_>16
COM	A	0.167	0.114	0.056	0.037
	B	0.189	0.140	0.091	0.067
	C	0.169	0.226	0.199	**0.160**
	overall	0.189 (0.160)	0.190 (0.164)	0.143 (0.117)	0.119 (0.102)
	accumulated VIO [%]	70.154	81.799	87.546	100
		MF0	**MF1**	**MF2**	**MF3**
		vio	vio_agebin_0-2	vio_agebin_3-5	vio_agebin_6-8
TP	A	-	0.141	0.147	0.186
	B	-	0.121	0.124	0.185
	C	-	0.071	0.106	0.150
	overall	- (0.201)	0.068 (0.058)	0.086 (0.060)	0.147 (0.114)
	accumulated VIO [%]	-	18.729	38.312	54.873
		MF4	**MF5**	**MF6**	**MF7**
		vio_agebin_9-11	vio_agebin_12-14	vio_agebin_1516	vio_agebin_>16
TP	A	0.208	0.144	0.102	0.071
	B	0.202	0.174	0.111	0.084
	C	0.191	0.199	0.153	0.128
	overall	0.173 (0.145)	0.187 (0.155)	0.160 (0.127)	0.168 (0.129)
	accumulated VIO [%]	70.154	81.799	87.546	100

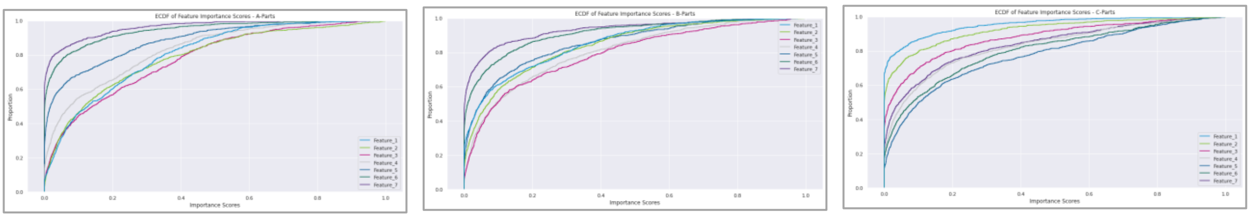


Figure E 19 ECDF of Importance Scores for Age Structure of VIO in C1 - COM

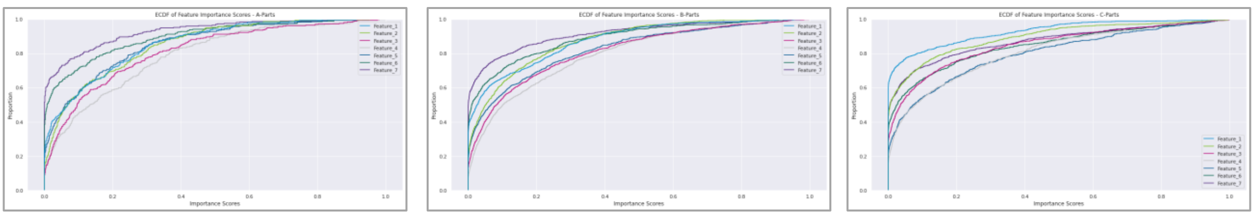


Figure E 20 ECDF of Importance Scores for Age Structure of VIO in C1 - TP

Table E 43: Importance Testing for Age Structure of VIO in C2 - Elasticity

	RI^f ($RI^f w/ MF0$)	**MF0**	**MF1**	**MF2**	**MF3**
		vio	vio_agebin_0-2	vio_agebin_3-5	vio_agebin_6-8
COM	A	-	0.143	0.163	0.221
	B	-	0.119	0.147	0.221
	C	-	0.060	0.094	0.134
	overall	- (0.160822)	0.073 (0.063)	0.112 (0.099)	0.160 (0.137)
	accumulated VIO [%]	-	10.433	20.102	39.237
		MF4	**MF5**	**MF6**	**MF7**
		vio_agebin_9-11	vio_agebin_12-14	vio_agebin_1516	vio_agebin_>16
COM	A	0.160	0.141	0.102	0.071
	B	0.182	0.151	0.106	0.075
	C	0.177	0.178	0.184	0.173
	overall	0.173 (0.139)	0.152 (0.116)	0.139 (0.109)	0.110 (0.094)
	accumulated VIO [%]	54.258	72.998	79.644	100
		MF0	**MF1**	**MF2**	**MF3**
		vio	vio_agebin_0-2	vio_agebin_3-5	vio_agebin_6-8
TP	A	-	0.187	0.186	0.179
	B	-	0.157	0.191	0.192
	C	-	0.075	0.100	0.167
	overall	- (0.179)	0.056 (0.045)	0.091 (0.072)	0.138 (0.110)
	accumulated VIO [%]	-	10.433	20.102	39.237
		MF4	**MF5**	**MF6**	**MF7**
		vio_agebin_9-11	vio_agebin_12-14	vio_agebin_1516	vio_agebin_>16
TP	A	0.154	0.168	0.080	0.047
	B	0.177	0.141	0.087	0.055
	C	0.189	0.191	0.149	0.128
	overall	0.151 (0.127)	0.193 (0.161)	0.192 (0.159)	0.173 (0.142)
	accumulated VIO [%]	54.258	72.998	79.644	100

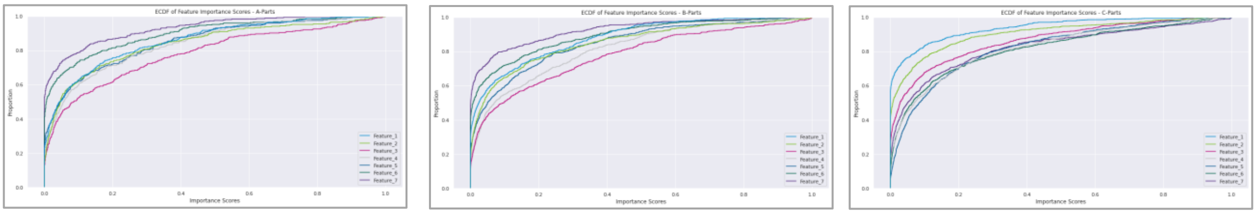


Figure E 21 ECDF of Importance Scores for Age Structure of VIO in C2 - COM

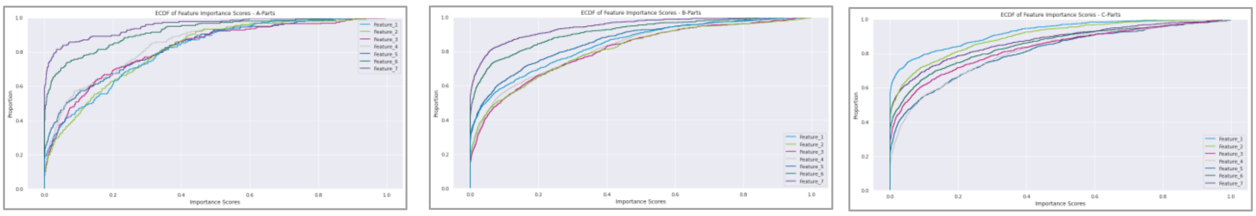


Figure E 22 ECDF of Importance Scores for Age Structure of VIO in C2 - TP

Table E 44: Importance Testing for Age Structure of VIO in C3 - Elasticity

RI^f ($RI^f w/ MF0$)		**MF0**	**MF1**	**MF2**	**MF3**
		vio	vio_agebin_0-2	vio_agebin_3-5	vio_agebin_6-8
COM	A	-	0.214	0.254	0.231
	B	-	0.164	0.190	0.204
	C	-	0.054	0.078	0.122
	overall	- (0.173)	0.077 (0.068)	0.099 (0.080)	0.146 (0.115)
	accumulated VIO [%]	-	13.421	27.771	38.189
		MF4	**MF5**	**MF6**	**MF7**
		vio_agebin_9-11	vio_agebin_12-14	vio_agebin_1516	vio_agebin_>16
COM	A	0.130	0.083	0.051	0.036
	B	0.186	0.139	0.074	0.043
	C	0.160	0.215	0.198	0.172
	overall	0.166 (0.137)	0.186 (0.158)	0.167 (0.138)	0.144 (0.116)
	accumulated VIO [%]	52.426	68.617	76.758	100
		MF0	**MF1**	**MF2**	**MF3**
		vio	vio_agebin_0-2	vio_agebin_3-5	vio_agebin_6-8
TP	A	-	0.121	0.127	0.190
	B	-	0.106	0.108	0.201
	C	-	0.089	0.109	0.196
	overall	- (0.179)	0.066 (0.053)	0.099 (0.085)	0.171 (0.143)
	accumulated VIO [%]	-	13.421	27.771	38.189
		MF4	**MF5**	**MF6**	**MF7**
		vio_agebin_9-11	vio_agebin_12-14	vio_agebin_1516	vio_agebin_>16
TP	A	0.183	0.1745	0.100	0.105
	B	0.188	0.176	0.120	0.103
	C	0.224	0.182	0.108	0.092
	overall	0.211 (0.174)	0.183 (0.150)	0.142 (0.107)	0.125 (0.106)
	accumulated VIO [%]	52.426	68.617	76.758	100

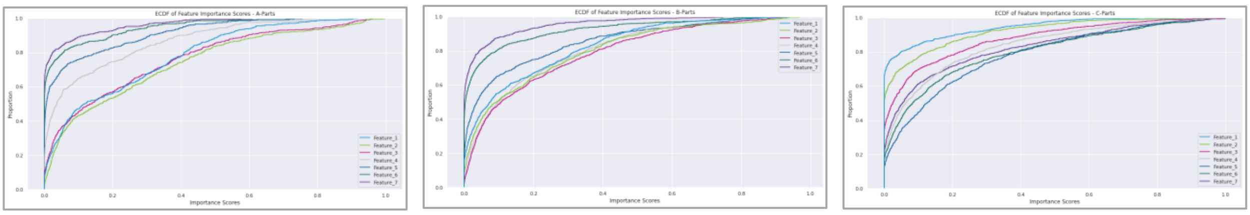


Figure E 23 ECDF of Importance Scores for Age Structure of VIO in C3 - COM

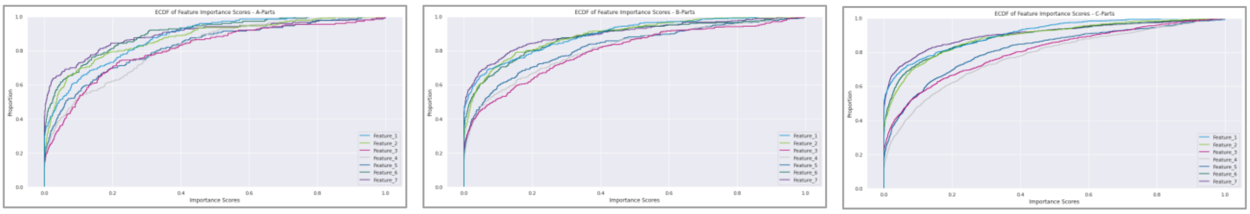


Figure E 24 ECDF of Importance Scores for Age Structure of VIO in C3 - TP

Sector

The market is dominated by the sector PC –therefore it forms the predominant target group. This dominance is also reflected in the importance of the MFs.

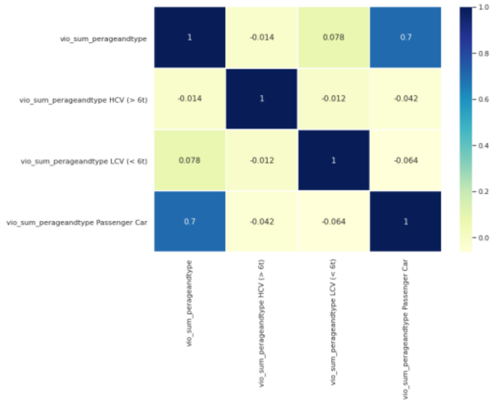


Figure E 25 Intra-Feature Correlation Testing for Sectors of VIO

Table E 45: Importance Testing for Sectors of VIO - Elasticity

	RI^f ($RI^f w/ MF0$)	MF0	MF1	MF2	MF3
		vio	vio_HCV(>6t)	vio_LCV(<6t)	vio_Passenger Car
COM	A	-	0.021	0.160	0.818
	B	-	0.012	0.116	0.871
	C	-	0.009	0120	0.871
	overall	- (0.497)	0.021 (0.008)	0.181 (0.098)	0.790 (0.390)
	accumulated VIO [%]	-	6.78		100
		MF0	**MF1**	**MF2**	**MF3**
		vio	vio_HCV(>6t)	vio_LCV(<6t)	vio_Passenger Car
TP	A	-	0.060	0.171	0.769
	B	-	0.083	0.124	0.793
	C	-	0.033	0.061	0.902
	overall	- (0.476)	0.080 (0.039)	0.092 (0.058)	0.815 (0.414)
	accumulated VIO [%]	-	6.78		100

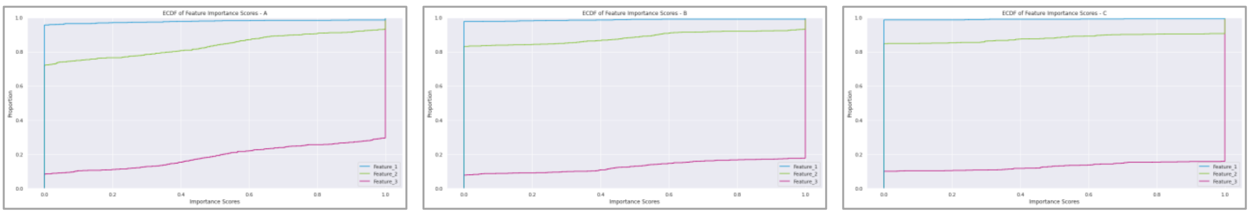


Figure E 26 ECDF of Importance Scores for Sector of VIO - COM

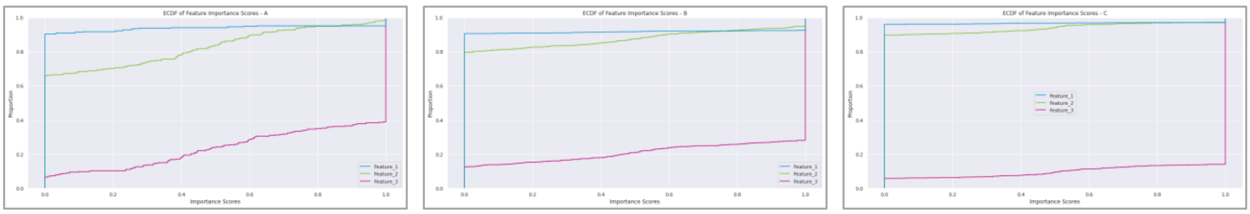


Figure E 27 ECDF of Importance Scores for Sector of VIO - TP

Segment

Not only the segment's proportional strength seems to be decisive – it is probably also brand and quality. This becomes evident by the example of vehicles from the lower middle class and the middle-class segment. The first amounts to a share of 36.11 percent whereas the latter amounts to only 23.99 percent. Despite the 13-percentage points difference, it is attributed more importance.

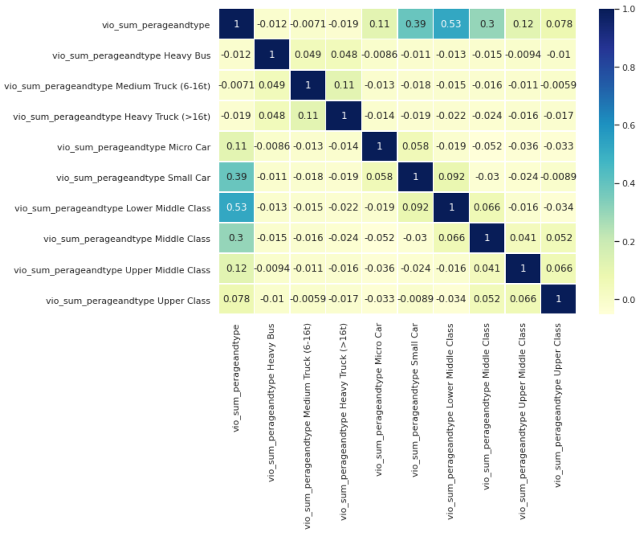


Figure E 28 Intra-Feature Correlation Testing for Segments of VIO

Table E 46: Importance Testing for Segments of VIO - Elasticity

RI^f ($RI^f w/ MF0$)		MF0	MF1	MF2	MF3
		vio	Vio_HeavyBus	Vio_MediumTruck	Vio_HeavyTruck
COM	A	-	0.005	0.012	0.005
	B	-	0.003	0.008	0.002
	C	-	0.000	0.004	0.004
	overall	- (0.423)	0.003 (0.002)	0.014 (0.009)	0.003 (0.002)
	accumulated VIO [%]	-	0.088	1.260	1.607
		MF4	**MF5**	**MF6**	**MF7**
		vio_MicroCar	vio_SmallCar	vio_LowerMiddleClass	vio_MiddleClass
COM	A	0.068	0.151	0.239	0.292
	B	0.062	0.151	0.210	0.318
	C	0.067240	0.186	0.236	0.333
	overall	0.100 (0.060)	0.214 (0.127)	0.264 (0.148)	0.310 (0.166)
	accumulated VIO [%]	8.562	32.548	68.660	92.646
		MF8	**MF9**		
		vio_UpperMiddleClass	vio_UpperClass		
COM	A	0.146	0.082		
	B	0.150	0.096		
	C	0.113	0.058		
	overall	0.051 (0.033)	0.032 (0.024)		
	accumulated VIO [%]	99.72508	100		

		MF0	**MF1**	**MF2**	**MF3**
		vio	Vio_HeavyBus	Vio_MediumTruck	Vio_HeavyTruck
TP	A	-	0.024	0.011	0.024
	B	-	0.029	0.020	0.032
	C	-	0.009	0.011	0.013
	overall	- (0.358)	0.022 (0.013)	0.023 (0.013)	0.034 (0.021)
	accumulated VIO [%]	-	0.088	1.260	1.607
		MF4	**MF5**	**MF6**	**MF7**
		vio_MicroCar	vio_SmallCar	vio_LowerMiddleClass	vio_MiddleClass
TP	A	0.043	0.131	0.191	0.237
	B	0.053	0.109	0.179	0.257
	C	0.035	0.101	0.197	0.281
	overall	0.037 (0.026)	0.102 (0.072)	0.172 (0.112)	0.285 (0.179)
	accumulated VIO [%]	8.562	32.548	68.660	92.646
		MF8	**MF9**		
		vio_UpperMiddleClass	vio_UpperClass		
TP	A	0.183	0.157		
	B	0.180	0.141		
	C	0.174	0.174		
	overall	0.172 (0.109)	0.139 (0.086)		
	accumulated VIO [%]	99.725	100		

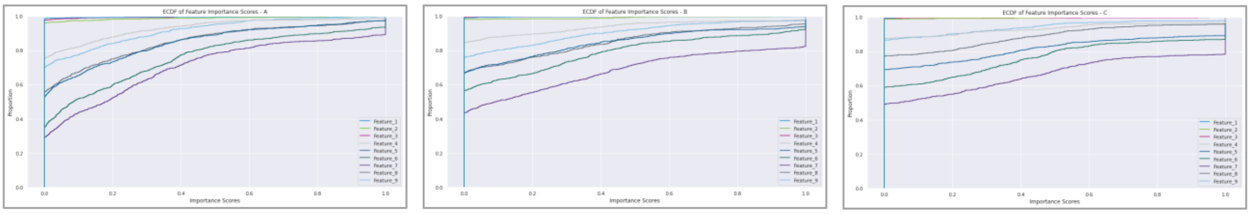


Figure E 29 ECDF of Importance Scores for Segment of VIO - COM

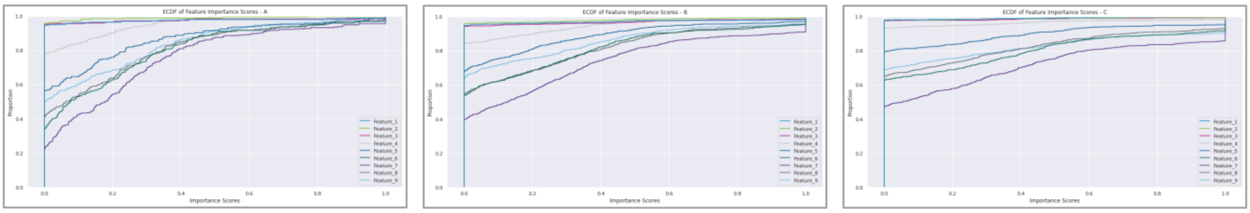


Figure E 30 ECDF of Importance Scores for Segment of VIO - TP

Fuel Type

In C1 in 2019, about 60 percent of newly registered PC were gasoline-powered. Around 30 percent of newly registered passenger cars were powered by diesel. Vehicles with alternative drive systems accounted for the remaining registrations. (Kraftfahrt-Bundesamt 2021(3)) With regards to latest ecologically-motivated respectively political regulations tough a disruptive cut is to be expected from 2035, significantly influencing the shares of the respective fuel types. (European Parliament 2023)

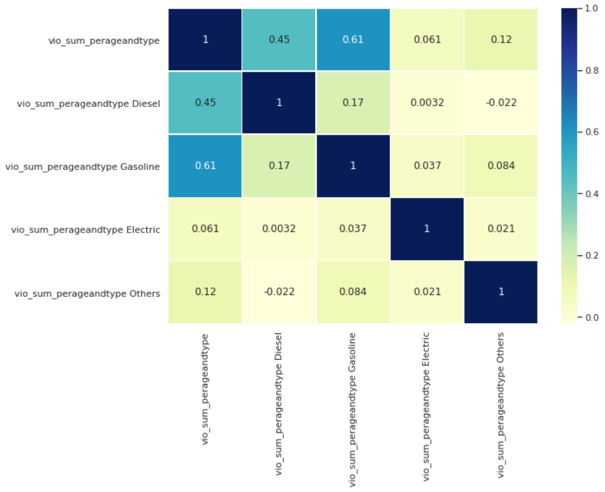


Figure E 31 Intra-Feature Correlation Testing for Fuel Types of VIO

Table E 47: Importance Testing for Fuel Type of VIO - Elasticity

RI^f ($RI^f w/ MF0$)		**MF0**	**MF1**	**MF2**	**MF3**
		vio	vio_Diesel	vio_Gasoline	vio_Electric
COM	A	-	0.385	0.436	0.039
	B	-	0.409	0.522	0.004
	C	-	0.409	0.565	0.003
	overall	- (0.363)	0.444 (0.282)	0.500 (0.311)	0.007 (0.005)
	accumulated VIO [%]	-	34.581	95.966	96.474

		MF4
		vio_Others[79]
COM	A	0.141
	B	0.064
	C	0.022
	overall	0.042 (0.031)
	accumulated VIO [%]	100

		MF0	**MF1**	**MF2**	**MF3**
		vio	vio_Diesel	vio_Gasoline	vio_Electric
TP	A	-	0.349	0.543	0.010
	B	-	0.350	0.586	0.006
	C	-	0.192	0.774	0.000
	overall	- (0.465)	0.404 (0.213)	0.552 (0.287)	0.000 (0.000)
	accumulated VIO [%]	-	34.581	95.966	96.474

		MF4
		vio_Others
TP	A	0.099
	B	0.058
	C	0.030
	overall	0.031 (0.022)
	accumulated VIO [%]	100

[79] Powertrains that are not diesel, gasoline, electric are labelled "Others".

Figure E 33 ECDF of Importance Scores for Fuel Type of VIO - COM

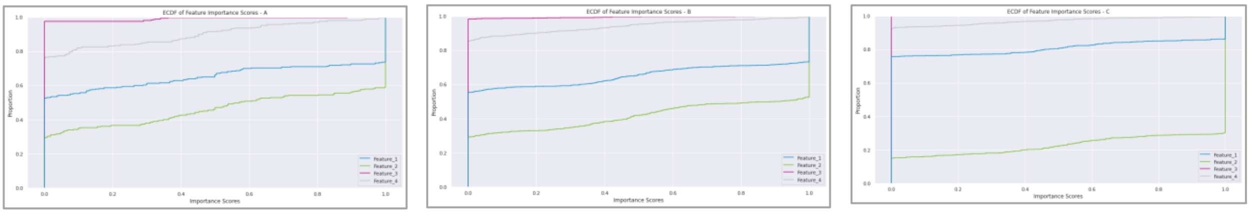


Figure E 32 ECDF of Importance Scores for Fuel Type of VIO - TP

Appendix 13: Mileage Distribution

The objective of the procedure described in this chapter is to obtain a complete progression of the VIOs mileage per sector-segment-fuel type class and country over the years. For this purpose, the empirically derived data points for the start mileage (d_0^v) and end mileage (d_{max}^v) status are complemented by means of a piecewise linear interpolation.[80]

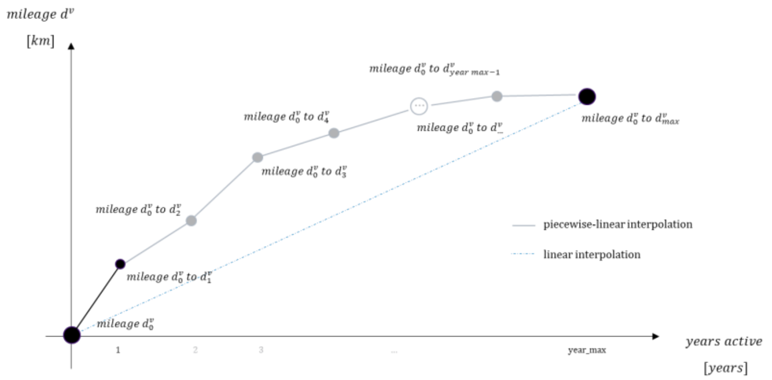


Figure E 34 Piecewise-linear Interpolation of VIO Mileage

To do so, the following procedure is applied:

- Define the start mileage d_0^v by setting d_0^v = 0.
- Determine the expected first year mileage d_1^v by defining $d_1^v = \Sigma_0^1 d_{years_{active}}^v = \sum_1^{rec} \frac{end\_mileage^{rec}}{vio\_age^{rec}} \cdot \frac{1}{\sum rec}$ per sector-segment-fuel type class and country, only considering recordings rec where $0 < vio\_age^{rec} \leq 3$
- Determine the expected end mileage $\Sigma_0^{\max(vio\_age^{rec})} d_{years\_active}^v$ for $\max(vio\_age^{rec})$ = 19[81] by defining $\Sigma_0^{19} d_{years\_active}^v = d_0^v + d_1^v + \cdots + d_{19}^v = \sum \frac{end\_mileage^{rec}}{vio\_age^{rec}} \cdot \frac{1}{\sum rec} \cdot \max(vio\_age^{rec})$ per sector-segment-fuel type class and country, only considering recordings rec where $vio\_age^{rec} \geq 17$
- Determine the expected end year's mileage $d_{\max(vio\_age^{rec})}^v$ by defining $d_{\max(vio\_age^{rec})}^v = d_{19}^v = \sum \frac{end\_mileage^{rec}}{vio\_age^{rec}} \cdot \frac{1}{\sum rec}$, only considering recordings rec where $vio\_age^{rec} \geq 17$

[80] The data is obtained from a variety of external databases (Kraftfahrt-Bundesamt 2021 (2), Autostat 2021 (1), Autostat 2021 (2) Statista 2021 and Enerdata 2021) and needs to be understood as a snapshot at the time of retrieval. The information gained comprises HF3 (fuel_type), HF4 (vio_market_sector), and HF5 (vio_market_segment) as well as HF55 (start_registration), HF56 (end_registration), HF57 (start_mileage), and HF58 (end_mileage) as displayed in Table E 14 in Appendix 6. HF6 (vio_age) is calculated from HF55 and HF56 and displays the VIOs active years at the end registration date. Fuel types which are neither *Diesel*, *Gasoline*, nor *Electric* are grouped under the concept *Others*. The level chosen (sector-segment-fuel type-country) results in a total of 188 combinations for the present use case. For each of the combinations, different recordings *rec* are available, displaying the mileage in kilometers for different aged VIOs.
[81] Using the available data, mileage values for a maximum span of 19 years can be determined.

The three data points determined serve as base varialbes for the piecewise-linear interpolation to derive all mileage values travelled by a specific sector-segment-fuel type class in a specific country and year active.

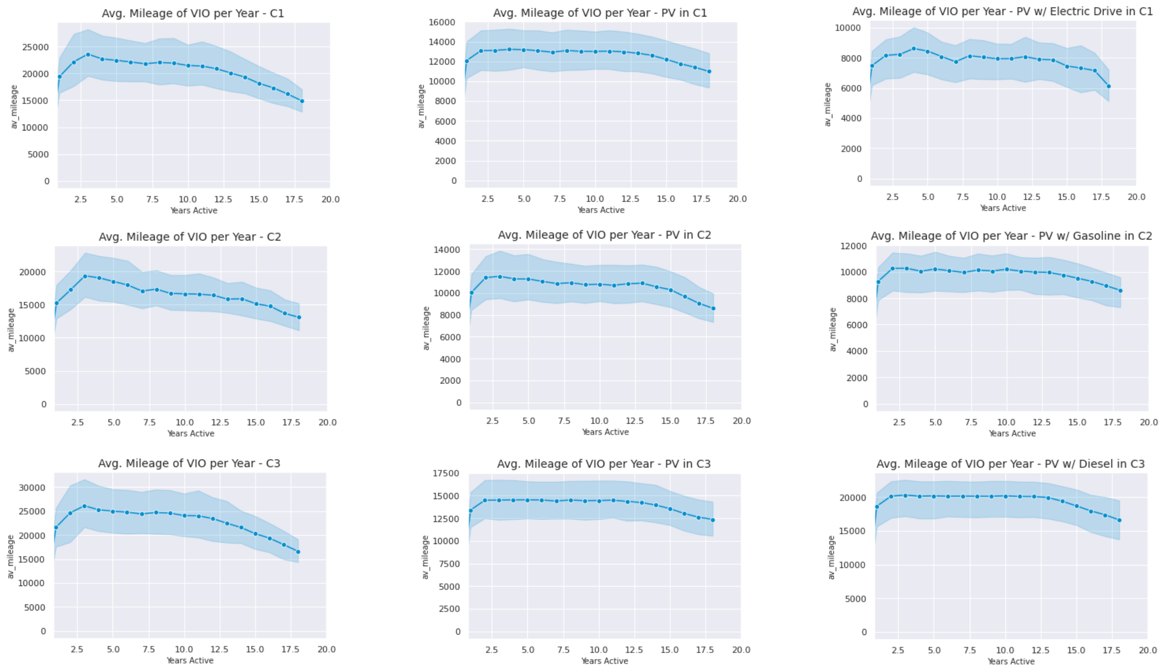


Figure E 35 Average Mileage Distribution for VIOs per Vehicle Type

To obtain mileages underlying the failure rate, the average values per active year are accumulated and translated to the respective age of the vehicle population.

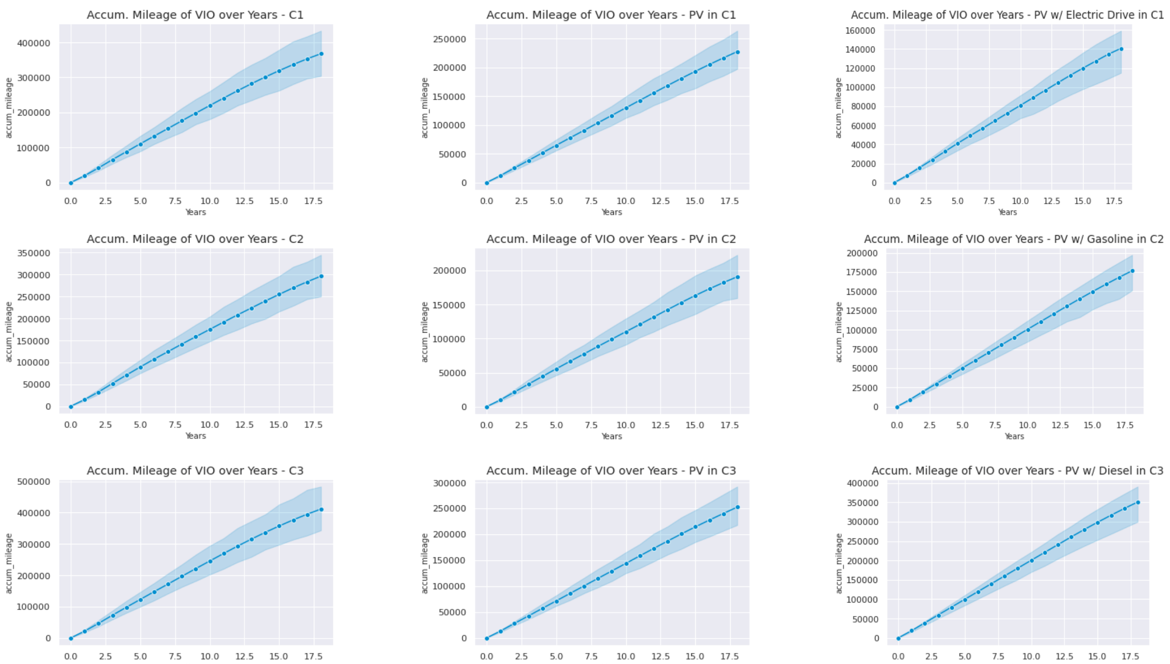


Figure E 36 Accumulated Mileage Distribution for VIOs per Vehicle Type

Appendix 14: Extrapolation of Main Features

Table E 48: Extrapolation of time_varying_unknowns

	Feature	Extrapolation Method	Sample Size and Horizon
1.	'item_category3_id_sum_item_cnt_month'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
2.	'item_category3_id_sum_item_cnt_month_lag_1' ... lags = [1 ,2 ,3 ,4, 5, 6]	Derived from extrapolated feature 'item_category3_id_sum_item_cnt_month'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
3.	'item_category2_id_sum_item_cnt_month'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
4.	'item_category2_id_sum_item_cnt_month_lag_1' ... lags = [1 ,2 ,3 ,4, 5, 6]	Derived from extrapolated feature 'item_category3_id_sum_item_cnt_month'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
5.	'item_category1_id_sum_item_cnt_month'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
6.	item_category1_id_sum_item_cnt_month_lag_1 ... lags = [1 ,2 ,3 ,4, 5, 6]	Derived from extrapolated feature 'item_category1_id_sum_item_cnt_month'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
7.	'item_category3_id_avg_item_cnt_month'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
8.	'item_category3_id_avg_item_cnt_month_lag_1' ... lags = [1 ,2 ,3 ,4, 5, 6]	Derived from extrapolated feature 'item_category3_id_avg_item_cnt_month'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
9.	'item_category2_id_avg_item_cnt_month'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
10.	'item_category2_id_avg_item_cnt_month_lag_1' ... lags = [1 ,2 ,3 ,4, 5, 6]	Derived from extrapolated feature 'item_category3_id_avg_item_cnt_month'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
11.	'item_category1_id_avg_item_cnt_month'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
12.	'item_category1_id_avg_item_cnt_month_lag_1' ... lags = [1 ,2 ,3 ,4, 5, 6]	Derived from extrapolated feature 'item_category3_id_avg_item_cnt_month'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
13.	'YearlySalesQuantity'	Simple Exponential Smoothing trend = 'additive' seasonal = None	No of Samples = 1 Horizon = 1 year (12 months)
14.	'unique_cus_month'	Holt-Winters trend = 'additive' seasonal_periods = 12	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)

		seasonal = 'additive'	
15.	'unique_cus_year'	Simple Exponential Smoothing trend = 'additive' seasonal = None	No of Samples = 1 Horizon = 1 year (12 months)
16.	'topcus_c'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
17.	'b'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
18.	'c'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
19.	'e'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
20.	'b_pct'	Derived from extrapolated feature 'b'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
21.	'csl1'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
22.	'csl1_lag_1'	Derived from extrapolated Feature 'csl1'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
23.	'csl1_lag_2'	Derived from extrapolated Feature 'csl1'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
24.	'ippingc_perpc'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
25.	'ippinlc_perpc'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
26.	'exratelcgc_ma2cent'	MA2cent	Sample Size = 1 Horizon = 12 Months
27.	'pct_change_exratelcgc_ma2cent'	Derived from extrapolated Feature 'exratelcgc_ma2cent'	Sample Size = 1 Horizon = 12 Months
28.	'seasonality_pct'	Holt-Winters trend = 'additive' seasonal_periods = 12 seasonal = 'additive'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)
29.	'is_seasonal'	Derived from extrapolated Feature 'seasonality_pct'	No of Samples = 6 RollingWindow = 2 Horizon = 5 (4, 2)

Appendix 15: Dimensionality Reduction of Segmentation Features

The approach is derived from empirical analysis and also follows a statement by Backhaus et. al. (2021: 495) that redundant information may lead to distortions in the cluster analysis. Accordingly, an algorithm is developed by which both ranks are derived, and subsequent eliminations of highly correlated SFs are achieved.

Given:

The input data is highly dimensional with SF = $\{SF_1,\ SF_2, \ldots, SF_S\}$ extracted and built for various MECE categories $\{SalesGroup,\ Product,\ SalesRegion\}$ on different levels $\{meta\ features, statistical\ features\}$.

Algorithm in Pseudo Code**:**

within MECE category $\{SalesGroup\}$

for each SF_S in SF:

determine	$r_{s_o s_u}$	
determine	$count$ per SF_S where $r_{s_o s_u} < \ \lvert 0.5 \rvert$[82]	
determine	$\sum_1^{count} r_{s_o s_u}$ per SF_S	
sort	SF_S in SF according to their $counts$ in descending order	
select	SF_S with $\max(count)$ for the prioritized SF-set.	
	if	$count$ of SF_{S_o} == $count$ of SF_{S_u}
	select	SF_S where $\min(\sum_1^u r_{s_o s_u})$
stop	if	$count$ of SF_{S_o} == $count$ of SF_{S_u} and
	if	$\sum_o^u r_{s_o s_u}$ of $SF_{S_o} \geq \sum_o^u r_{c_o c_u}$ of SF_{S_u}

within MECE categories $\{Product,\ SalesRegion\}$

for each SF_S in SF:

determine	$r_{s_o s_u}$	
determine	$count$ per SF_S where $r_{s_o s_u} < \ \lvert 0.5 \rvert$	
determine	$\sum_1^{count} r_{s_o s_u}$ per SF_S	
sort	SF_S in SF according to their $counts$ in descending order	
select	SF_S with $\max(count)$ for the prioritized SF-set.	
	if	$count$ of SF_{S_o} == $count$ of SF_{S_u}

[82] The threshold value for the absolute correlation r is set at 0.5. This is based on the work of Asuero, Sayago and González (2006: 47), who define values smaller than 0.5 representing a low correlation.

$$\begin{array}{lll} & \text{select} & SF_s \text{ where } \min(\sum_1^u r_{s_o s_u}) \\ \text{stop} & \text{if} & count \text{ of } SF_{s_o} == count \text{ of } SF_{s_u} \text{ and} \\ & \text{if} & \sum_o^u r_{s_o s_u} \text{ of } SF_{s_o} \geq \sum_o^u r_{c_o c_u} \text{ of } SF_{s_u} \end{array}$$

Output:

The output is a low dimensional data set with a reduced number of SF = $\{SF_1,\ SF_2, \ldots, SF_{S-s}\}$ for various MECE categories on different levels $\{meta\ attributes, statistical\ features\}$.

Table E 49: Prioritized Feature Set to be Applied for Time Series Segmentation

	META FEATURES		STATISTICAL FEATURES	
Level	Key Attributes	Descriptive Attributes	Business-induced Statistical Features	Structural Time Series Characteristics
SG	-	-	'GrowthRate_SA_PMC' 'Mean_SA_Year' 'APV_PMC' 'CV_SA_PMC'	'benford_correlation' 'quantile_0.5' 'number_peaks__n_10' 'variance_larger_than_ 'standard_deviation' 'approximate_entropy__m2_r_0.1' 'approximate_entropy__m2_r_0.7' 'linear_trend__attr_rvalue' 'linear_trend__attr_pvalue' 'linear_trend__attr_slope' 'partial_autocorrelation__lag_2' 'autocorrelation__lag_4' 'autocorrelation__lag_6'
SKU	'PartType ' 'BU' 'PG2'	'PPC-P AA_Mat10'	'Seasonality_PG2' 'APV_SKU' 'APFR_Year_SKU' 'CV_SA_SKU'	-
SR	'Dom_CS'	-	'GrowthRate_SA_SR' 'Mean_SA_Year_SR' 'AvAPV_SR'	-

In a second step, a *wrapper* method i.e., sequential selection is applied, testing each SF_S for its impact on the number of clusters K, the size of clusters (i.e., number of within-cluster members) and evaluation metrics. For its inherent advantages, k-means is used for Choose-Features Experiments.

Appendix 16: Hyperparameters

Table E 50: Most Relevant Hyperparameters for Neural Nets

Parameter	Description	Set vs. Tunable	References	Proposed Search Space
ARCHITECTURE				
NUM_LAYERS	The LSTM-cell architecture is shaped by the parameter NUM_LAYERS/LSTM_LAYERS, which is a common parameter of RNNs. It defines how many LSTM-cells are stacked over each other. This increases the complexity of the model and its ability to learn patterns	Tunable	Stathakis (2009)	heuristics or pruning with [1, 2, …]
HIDDEN_SIZE	It defines the hidden state size and is common across the whole model. It is also used for each LSTM encoder and decoder cell in the LSTM-GRN to set the number of MFs in the hidden state. In the GRN, it defines the size for the input, hidden and output layer, the size of the hidden layer in the attention heads as well as in the GLU, the variable selection network and the input size of the additive and normalization layers. Too few neurons result in underfitting, too many neurons in overfitting. Also, time consumed for training can increase to a point when it is impossible to adequately train the network.	Tunable	Stathakis (2009) Lim et. al (2020)	Rule of Thumb: • HIDDEN_SIZE = number of MFs in the input data set; sometimes one additional neuron is added for a bias term • HIDDEN_SIZE = number of training samples • HIDDEN_SIZE should be in-between the number of input nodes and the number of output nodes • HIDDEN_SIZE = (number of input nodes + number of output nodes)/2 • HIDDEN_SIZE = 2/3 * number of input nodes + number of output nodes • HIDDEN_SIZE < 2 * number of input nodes resulting search space: [8, 16, 32, 64, 128, 256, 512]
MIN_ENCODER_LENGTH MAX_ENCODER_LENGTH	MIN_/MAX_ENCODER_LENGTH sets the minimal/maximal horizon to look back in the history of the time series. It is specifically important for multi-head attention to restrict the input data.	Tunable Reference to ACF/PACF analysis	Lim et. al (2020)	Rule of Thumb: • MIN_ENCODER_LENGTH ≤ number of bins of shortest time series in the model • MAX_ENCODER_LENGTH = number of bins of shortest time series in the model resulting search space: [3; 6; 12; 24; 36; 48]
DROPOUT	A simple approach to reduce the danger of overfitting: Every	Tunable	Srivastava et. al. (2014)	resulting search space: [0.01 - 0.3]

	neuron within the NN has a probability to be ignored during a training step. The number of neurons dropped can be specified with this parameter.			
ATTENTION_HEAD_SIZE	The multi-head attention employs multiple heads, each for a different subset, to enhance context learning.	Tunable	Lim et. al. (2020)	resulting search space: [1; 2; 3; 4]
TARGET_NORMALIZER	• MinMax: [0;1] with data ≠ normally distributed • Standard: [-1;1] with data = normally distributed • RobustScaler (using quantiles 0.25-0.75)	Set	-	None MinMax StandardScaler RobustScaler
OUTPUT_SIZE	It defines the size of the output layer i.e., the number of quantile or point forecasts to be made.	Set	-	dependence on the number of output values
TRAINING				
ACTIVATION	Activation functions are used to forget or remember specific information of the input vector i.e., via output values between 0 and 1 elements, that are not needed, are eliminated. Activation functions can be linear or non-linear. For neural nets predominantly non-linear functions are to be used	Tunable	Glorot et al., (2011) Keras API Reference (2022)	resulting search space: • 'sigmoid' (logistic) • 'tanh' (hyperbolic tangent) • 'relu' (rectified linear unit) • 'elu', • 'selu', • 'softmax', • 'softplus', • 'exponential'
OPTIMIZER	OPTIMZER helps to minimize LOSS_FUNC. • adam = Adaptive Moment Estimation • sgd = Stochastic Gradient Descent • adamax = variant of Adam based on the infinity norm	Tunable	Sutskever et. al. (2013) Kingma and Ba (2015) Keras API Reference (2022)	resulting search space: • 'adam', • 'sgd', • 'adamax'
LOSS_FUNC	The purpose of the model's internal loss function, that is calculated during training, is to compute the quantity that a model should seek to minimize during training.	Set	Keras API Reference (2022)	options: • MAE(), • MSE(), • RMSE() • MAPE(), • SMAPE(), • QuantileLoss() • NegativeBinomialDistributionLoss()
LEARNING_RATE	It decides how strong a batch or whole epoch affects the network's weights.	Tunable	Keras API Reference (2022)	resulting search space: [1e-6 - 1e-1]
LEARNING_RATE_ DECAY_FACTOR	If the model optimization is stagnating, decreasing the learning rate could help . The DECAY_FACTOR is the step size by which to decrease the learning rate.	Tunable	Lim et. al. (2020)	resulting search space: [0 - 1]

EARLY_STOPPING / CALLBACK	The parameter is used to prevent overfitting: Either loss or accuracy values can be MONITORed by the early stopping callback function. If loss values are chosen, training comes to halt when there is an increment observed in loss values If accuracy values are chosen, training comes to halt when there is decrement observed in accuracy values.	Set	Prechelt (2002) Lim et. al. (2020) Keras API Reference (2022)	definition of early stopping function.
MONITOR	The value to be monitored by the function needs to be assigned.	Set	Keras API Reference (2022)	options: • validation loss with MODE 'min' • validation accuracy with MODE 'max'
MODE	The mode in which the change in the quantity monitored should be observed.	Set	Keras API Reference (2022)	options • 'max' • 'min' • 'auto' = function automatically monitors with the suitable mode
MIN_DELTA	MIN_DELTA is the increment of improvement that should at least be achieved when monitoring the values.	Set	Keras API Reference (2022)	resulting search space: [0.0001 – 0.001]
PATIENCE	The patience to observe before • reducing the learning rate • activating the callback function	Tunable	Keras API Reference (2022)	resulting search space: [5 – MAX_EPOCHS]
MAX_EPOCHS	The parameter MAX_EPOCHS defines how many times the full training dataset is passed through the net for learning purposes. A high number of epochs could end up in overfitting while a low number underfits the model. Hence, to enable generalization capacity and to mitigate overfitting, the model should be trained on the optimal number of epochs.	Tunable: in combination with Early Stopping	Prechelt (2002)	resulting search space: [20 – 100]
CLIP_GRADIENT / GRADIENT_CLIP_VAL	CLIP_GRADIENT denotes the maximum value of gradient, i.e., the gradient is clipped if it is too large.	Tunable	Lim et. al. (2020)	resulting search space: [0.01;100]

Table E 51: Most Relevant Hyperparameters for Gradient Boosting Regressor

Parameter	Description	Set vs. Tunable	References	Proposed Search Space
ARCHITECTURE				
BOOSTER	BOOSTER defines which booster to use. gbtree and dart use tree-based models while gblinear uses linear functions.	Set	XGBoost (2022)	• gbtree • gblinear • dart default= gbtree
MAX_DEPTH	MAX_DEPTH denotes the maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit.	Set	Chen and Guestrin (2016) Zhuang et. al. (2022)	resulting search space: [0, ∞[default=6
MIN_CHILD_WEIGHT	If the tree partition step results in a leaf node with the sum of instance weight less than MIN_CHILD_WEIGHT, then the building process will give up further partitioning. The larger MIN_CHILD_WEIGHT is, the more conservative the algorithm will be.	Tunable	XGBoost (2022)	resulting search space: [0, ∞[default=1
SUBSAMPLE	SUBSAMPLE is the subsample ratio of the training instance. Setting it to 0.5 means that XGBoost would randomly sample half of the training data prior to growing trees and this will prevent overfitting.	Tunable	XGBoost (2022)	resulting search space: [0, ∞[default=1
SAMPLING_METHOD	SAMPLING_METHOD defines the method to use to sample the training instances. It is important to consider that this sampling method is only supported when TREE_METHOD is set to gpu_hist; other tree methods only support uniform sampling	Set	XGBoost (2022)	• uniform: each training instance has an equal probability of being selected. • gradient_based: the selection probability for each training instance is proportional to the regularized absolute value of gradients
COLSAMPLE_BYTREE, COLSAMPLE_BYLEVEL COLSAMPLE_BYNODE	This is a family of parameters for subsampling of columns. Differentiation according to • subsample ratio of columns for each level • subsample ratio of columns for each node (split)	Tunable	XGBoost (2022)	resulting search space:]0, 1] default=1
REG_LAMBDA REG_ALPHA	L2 regularization term on weights. Increasing this value will make model more conservative. L1 regularization term on weights. Increasing this value	Tunable	XGBoost (2022)	default = 1 default = 0

	will make model more conservative. default = 0			
TREE_METHOD	XGBoost supports auto, exact, approx, hist and gpu_hist for distributed training	Set	XGBoost (2022)	• auto = heuristic (exact greedy) • exact: Exact greedy algorithm which enumerates all split candidates for small data sets • approx: Approximate greedy algorithm using quantile sketch and gradient histogram for medium-sized data sets • hist: Faster histogram optimized approximate greedy algorithm for large data sets • gpu_hist: GPU implementation of hist algorithm.
TRAINING				
OBJECTIVE	OBJECITVE enables the specification of the learning task and the corresponding objective function. It is to be differentiated if it is a regression or classification task.	Set	XGBoost (2022)	• reg:linear respectively reg:squarederror = regression with squared loss • reg:logistic = logistic regression • reg:gamma = gamma regression with log-link • reg:tweedie = tweedie regression with log-link with con-comitant definition of tweedie_variance_power
EVAL_METRIC	EVAL_METRIC defines the evaluation metrics for validation data. A default metric will be assigned according to objective. Besides the default value, more metrics can be added.	Set	XGBoost (2022)	rmse for regression
ETA / LEARNING_RATE	ETA denotes a step size shrinkage parameter to prevent overfitting. After each boosting step, the weights of new features sets are obtained, and eta shrinks the feature weights to make the boosting process more conservative.	Tunable	XGBoost (2022)	resulting search space: [0, 1] default=0.3
GAMMA / MIN_SPLIT_LOSS	GAMMA denotes the minimum loss reduction required to make a further partition on a leaf node of the tree. The larger it is, the more conservative the algorithm will be.	Tunable	XGBoost (2022)	resulting search space: [0, ∞[default=0
BASE_SCORE	BASE_SCORE denotes the global bias i.e., the initial	Set	XGBoost (2022)	default = 0.5

Appendix 17: List of Experiments

Control

- Experiment 1: all static and time_varying_known MFs are used as input [$base$]
- Experiment 2: all transactional and portfolio related MFs are used as input [D_0]
- Experiment 3: all static and time_varying_known MFs plus transactional and portfolio related MFs are used as input [$base$] + [D_0]
- Experiment 4: all MFs are used as input excluding lag-features [$base\_all\_D$]

Input

[$base$] **+ Customer-related features** [D_1]

- Experiment 5: all customer-related information is used as input
- Experiment 6: unique_cus_month is used as input
- Experiment 7: unique_cus_month and topcus_c are used as input
- Experiment 8: unique_cus_month and customer segment related features are used as input

[$base$] **+ Product-related features** [D_2]

- Experiment 9: all product-related information is used as input
- Experiment 10: customer service level and its lagged information are used as input
- Experiment 11: customer service level, its lagged information and *life_cycle_age_AA* are used as input

[$base$] **+ Price-related features** [D_3]

- Experiment 12: all price related MFs including ex-rate features are used as input
- Experiment 13: global pricing information is used as input (production costs (*PPC-P AA_Mat10*), invoiced price (*IPPinGC*) and bonus information); local pricing information is removed completely
- Experiment 14: global pricing information is used as input, but limited to production costs (*PPC-P AA_Mat10*) and invoiced price (*IPPinGC*) as input feature
- Experiment 15: global pricing information is used as input but limited to invoiced price (*IPPinGC*) and bonus information
- Experiment 16: local pricing information is used as input; global pricing information is removed completely
- Experiment 17: global pricing information and local pricing information is used as input except for ex-rate features
- Experiment 18: global pricing information and ex-rate features are used as input; IPPinLC is excluded

[$base$] **+ Time-related features** [D_4]

- Experiment 19: all time-related information is used as input
- Experiment 20: seasonal information is removed completely; *is_holiday* is used as input

- Experiment 21: time-related binary information is used as input: *is_seasonal*; *is_holiday*
- Experiment 22: all time-related information is used as input except of seasonal information for growing products

Model XGB with [$base\_all\_D$] **+ Lag features**

- Experiment 23: all MFs including lag-features are used as input

Appendix 18: Results of Experiments

Table E 52: Control Experiments - Perfect Information

COM $[base]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	1.204	1.195	1.213
	MAAPE_ARC2	0.919	0.912	0.926
A	SMAPE_200	0.905	0.888	0.922
	MAAPE_ARC2	0.681	0.669	0.694
B	SMAPE_200	1.103	1.086	1.119
	MAAPE_ARC2	0.847	0.834	0.859
C	SMAPE_200	1.365	1.353	1.377
	MAAPE_ARC2	1.043	1.033	1.053
$[D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.967	0.957	0.976
	MAAPE_ARC2	0.815	0.808	0.822
A	SMAPE_200	0.578	0.564	0.592
	MAAPE_ARC2	0.537	0.526	0.548
B	SMAPE_200	0.745	0.730	0.760
	MAAPE_ARC2	0.667	0.655	0.678
C	SMAPE_200	1.219	1.207	1.231
	MAAPE_ARC2	0.991	0.982	0.999
$[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.346	0.339	0.353
	MAAPE_ARC2	0.263	0.257	0.268
A	SMAPE_200	0.216	0.206	0.226
	MAAPE_ARC2	0.187	0.179	0.195
B	SMAPE_200	0.303	0.292	0.315
	MAAPE_ARC2	0.247	0.238	0.256
C	SMAPE_200	0.415	0.404	0.427
	MAAPE_ARC2	0.299	0.290	0.307
TP $[base]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	1.283	1.274	1.293
	MAAPE_ARC2	1.030	1.022	1.037
A	SMAPE_200	1.042	1.018	1.065
	MAAPE_ARC2	0.770	0.752	0.788
B	SMAPE_200	1.170	1.153	1.187
	MAAPE_ARC2	0.953	0.939	0.967
C	SMAPE_200	1.384	1.372	1.396
	MAAPE_ARC2	1.118	1.109	1.128
$[D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.746	0.737	0.754
	MAAPE_ARC2	0.562	0.555	0.569
A	SMAPE_200	0.638	0.620	0.656
	MAAPE_ARC2	0.549	0.535	0.562
B	SMAPE_200	0.752	0.740	0.765
	MAAPE_ARC2	0.540	0.531	0.550
C	SMAPE_200	0.782	0.767	0.797
	MAAPE_ARC2	0.617	0.606	0.628
$[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.264	0.258	0.270
	MAAPE_ARC2	0.179	0.175	0.183
A	SMAPE_200	0.191	0.181	0.202
	MAAPE_ARC2	0.159	0.151	0.167
B	SMAPE_200	0.270	0.262	0.278
	MAAPE_ARC2	0.174	0.169	0.180
C	SMAPE_200	0.285	0.275	0.296
	MAAPE_ARC2	0.200	0.192	0.207

Table E 53: Choose Feature Experiments - Perfect Information

COM				
$[base\_D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.704	0.696	0.712
	MAAPE_ARC2	0.589	0.583	0.595
A	SMAPE_200	0.537	0.524	0.549
	MAAPE_ARC2	0.502	0.492	0.512
B	SMAPE_200	0.653	0.640	0.665
	MAAPE_ARC2	0.581	0.572	0.591
C	SMAPE_200	0.791	0.779	0.803
	MAAPE_ARC2	0.624	0.615	0.634
$[base\_D_0\_D_1]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.332	0.326	0.339
	MAAPE_ARC2	0.262	0.257	0.267
A	SMAPE_200	0.218	0.209	0.228
	MAAPE_ARC2	0.199	0.192	0.207
B	SMAPE_200	0.299	0.288	0.310
	MAAPE_ARC2	0.253	0.244	0.262
C	SMAPE_200	0.391	0.380	0.402
	MAAPE_ARC2	0.290	0.282	0.298
$[base\_D_0\_D_2]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.598	0.591	0.604
	MAAPE_ARC2	0.511	0.505	0.516
A	SMAPE_200	0.527	0.516	0.539
	MAAPE_ARC2	0.493	0.484	0.503
B	SMAPE_200	0.616	0.605	0.628
	MAAPE_ARC2	0.551	0.543	0.560
C	SMAPE_200	0.615	0.604	0.625
	MAAPE_ARC2	0.497	0.489	0.505
$[base\_D_0\_D_3]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.658	0.651	0.666
	MAAPE_ARC2	0.542	0.537	0.548
A	SMAPE_200	0.520	0.508	0.532
	MAAPE_ARC2	0.488	0.478	0.497
B	SMAPE_200	0.639	0.628	0.651
	MAAPE_ARC2	0.566	0.557	0.575
C	SMAPE_200	0.719	0.707	0.730
	MAAPE_ARC2	0.551	0.542	0.560
$[base\_D_0\_D_4]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.557	0.550	0.564
	MAAPE_ARC2	0.483	0.478	0.488
A	SMAPE_200	0.485	0.473	0.496
	MAAPE_ARC2	0.454	0.445	0.464
B	SMAPE_200	0.580	0.568	0.591
	MAAPE_ARC2	0.525	0.516	0.534
C	SMAPE_200	0.572	0.562	0.583
	MAAPE_ARC2	0.473	0.465	0.481
TP				
$[base\_D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.577	0.569	0.585
	MAAPE_ARC2	0.460	0.454	0.466
A	SMAPE_200	0.547	0.536	0.558
	MAAPE_ARC2	0.417	0.408	0.426
B	SMAPE_200	0.573	0.557	0.589
	MAAPE_ARC2	0.513	0.500	0.526
C	SMAPE_200	0.645	0.631	0.659
	MAAPE_ARC2	0.532	0.521	0.543
$[base\_D_0\_D_1]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.254	0.248	0.259
	MAAPE_ARC2	0.172	0.168	0.175
A	SMAPE_200	0.186	0.175	0.196
	MAAPE_ARC2	0.158	0.150	0.165
B	SMAPE_200	0.259	0.252	0.267
	MAAPE_ARC2	0.166	0.160	0.171
C	SMAPE_200	0.273	0.262	0.284
	MAAPE_ARC2	0.192	0.185	0.199

$[base\_D_0\_D_2]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.546	0.539	0.554
	MAAPE_ARC2	0.450	0.443	0.456
A	SMAPE_200	0.516	0.505	0.527
	MAAPE_ARC2	0.407	0.398	0.416
B	SMAPE_200	0.562	0.547	0.577
	MAAPE_ARC2	0.509	0.497	0.522
C	SMAPE_200	0.608	0.595	0.621
	MAAPE_ARC2	0.517	0.507	0.528
$[base\_D_0\_D_3]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.571	0.563	0.579
	MAAPE_ARC2	0.455	0.449	0.461
A	SMAPE_200	0.540	0.529	0.551
	MAAPE_ARC2	0.411	0.402	0.419
B	SMAPE_200	0.564	0.548	0.580
	MAAPE_ARC2	0.504	0.491	0.517
C	SMAPE_200	0.644	0.630	0.658
	MAAPE_ARC2	0.531	0.520	0.542
$[base\_D_0\_D_4]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.473	0.466	0.480
	MAAPE_ARC2	0.398	0.393	0.404
A	SMAPE_200	0.472	0.462	0.482
	MAAPE_ARC2	0.379	0.371	0.387
B	SMAPE_200	0.523	0.508	0.538
	MAAPE_ARC2	0.474	0.461	0.487
C	SMAPE_200	0.569	0.556	0.582
	MAAPE_ARC2	0.493	0.482	0.503

Table E 54: Control Experiments - Realistic Information

COM $[base]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	1.361	1.357	1.364
	MAAPE_ARC2	1.092	1.089	1.095
A	SMAPE_200	1.079	1.072	1.086
	MAAPE_ARC2	0,767	0.762	0.772
B	SMAPE_200	1.165	1.158	1.171
	MAAPE_ARC2	0.943	0.938	0.948
C	SMAPE_200	1.559	1.555	1.563
	MAAPE_ARC2	1.283	1.280	1.286
$[D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.751	0.748	0.754
	MAAPE_ARC2	0.590	0.587	0.592
A	SMAPE_200	0.638	0.633	0.644
	MAAPE_ARC2	0.572	0.568	0.576
B	SMAPE_200	0.699	0.694	0.705
	MAAPE_ARC2	0.591	0.587	0.595
C	SMAPE_200	0.817	0.812	0.822
	MAAPE_ARC2	0.596	0.592	0.599
$[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.629	0.626	0.632
	MAAPE_ARC2	0.481	0.479	0.483
A	SMAPE_200	0.516	0.510	0.521
	MAAPE_ARC2	0.463	0.459	0.467
B	SMAPE_200	0.574	0.569	0.580
	MAAPE_ARC2	0.485	0.482	0.489
C	SMAPE_200	0.697	0.692	0.701
	MAAPE_ARC2	0.485	0.482	0.488
TP $[base]$	METRIC		CI_low:	CI_up:

All	SMAPE_200	1.431	1.428	1.435
	MAAPE_ARC2	1.085	1.082	1.087
A	SMAPE_200	1.216	1.207	1.225
	MAAPE_ARC2	0.798	0.792	0.805
B	SMAPE_200	1.278	1.272	1.285
	MAAPE_ARC2	0.980	0.975	0.986
C	SMAPE_200	1.545	1.541	1.550
	MAAPE_ARC2	1.193	1.190	1.196
$[D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.801	0.794	0.808
	MAAPE_ARC2	0.602	0.597	0.608
A	SMAPE_200	0.737	0.724	0.750
	MAAPE_ARC2	0.611	0.602	0.621
B	SMAPE_200	0.810	0.799	0.821
	MAAPE_ARC2	0.568	0.560	0.577
C	SMAPE_200	0.832	0.820	0.844
	MAAPE_ARC2	0.648	0.640	0.657
$[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.568	0.562	0.574
	MAAPE_ARC2	0.442	0.437	0.446
A	SMAPE_200	0,473	0.464	0,481
	MAAPE_ARC2	0.345	0.338	0.351
B	SMAPE_200	0.639	0.628	0.649
	MAAPE_ARC2	0.513	0.505	0.520
C	SMAPE_200	0.680	0.667	0.692
	MAAPE_ARC2	0.556	0.547	0.564

Table E 55: Choose Feature Experiments - Realistic Information

COM $[base\_D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.700	0.697	0.703
	MAAPE_ARC2	0.543	0.541	0.545
A	SMAPE_200	0.617	0.611	0.622
	MAAPE_ARC2	0.540	0.536	0.544
B	SMAPE_200	0.676	0.671	0.681
	MAAPE_ARC2	0.560	0.556	0.563
C	SMAPE_200	0.741	0.737	0.746
	MAAPE_ARC2	0.536	0.532	0.539
$[base\_D_0\_D_1]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.640	0.637	0.643
	MAAPE_ARC2	0.495	0.493	0.498
A	SMAPE_200	0.522	0.516	0.527
	MAAPE_ARC2	0.475	0,471	0.480
B	SMAPE_200	0.578	0.573	0.584
	MAAPE_ARC2	0.495	0.491	0.499
C	SMAPE_200	0.713	0.709	0.717
	MAAPE_ARC2	0.503	0.500	0.506
$[base\_D_0\_D_2]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.774	0.771	0.777
	MAAPE_ARC2	0.640	0.637	0.642
A	SMAPE_200	0.632	0.627	0.638
	MAAPE_ARC2	0.568	0.564	0.572
B	SMAPE_200	0.706	0.700	0.711
	MAAPE_ARC2	0.612	0.608	0.616
C	SMAPE_200	0.860	0.855	0.864
	MAAPE_ARC2	0.679	0.675	0.683
$[base\_D_0\_D_3]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.716	0.713	0.719
	MAAPE_ARC2	0.547	0.545	0.550
A	SMAPE_200	0.665	0.659	0.671
	MAAPE_ARC2	0.555	0.551	0.559
B	SMAPE_200	0.697	0.692	0.703
	MAAPE_ARC2	0.566	0.562	0.570

C	SMAPE_200	0.744	0.739	0.748
	MAAPE_ARC2	0.536	0.532	0.539
$[base\_D_0\_D_4]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.804	0.801	0.807
	MAAPE_ARC2	0.649	0.646	0.651
A	SMAPE_200	0.625	0.620	0.631
	MAAPE_ARC2	0.545	0.541	0.549
B	SMAPE_200	0.707	0.702	0.712
	MAAPE_ARC2	0.592	0.588	0.596
C	SMAPE_200	0.916	0.911	0.921
	MAAPE_ARC2	0.714	0.711	0.718
TP $[base\_D_0]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.686	0.680	0.692
	MAAPE_ARC2	0.505	0.500	0.510
A	SMAPE_200	0.648	0.639	0.658
	MAAPE_ARC2	0.437	0.430	0.444
B	SMAPE_200	0.688	0.676	0.701
	MAAPE_ARC2	0.563	0.554	0.571
C	SMAPE_200	0.741	0.731	0.752
	MAAPE_ARC2	0.570	0.562	0.577
$[base\_D_0\_D_1]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.558	0.552	0.564
	MAAPE_ARC2	0.434	0.430	0.439
A	SMAPE_200	0.451	0.443	0.459
	MAAPE_ARC2	0.332	0.326	0.338
B	SMAPE_200	0.629	0.619	0.639
	MAAPE_ARC2	0.508	0.501	0.516
C	SMAPE_200	0.696	0.683	0.709
	MAAPE_ARC2	0.556	0.547	0.565
$[base\_D_0\_D_2]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.678	0.672	0.685
	MAAPE_ARC2	0.532	0.527	0.537
A	SMAPE_200	0.638	0.627	0.648
	MAAPE_ARC2	0.480	0.471	0.488
B	SMAPE_200	0.699	0.687	0.712
	MAAPE_ARC2	0.574	0.564	0.583
C	SMAPE_200	0.725	0.714	0.736
	MAAPE_ARC2	0.585	0.576	0.593
$[base\_D_0\_D_3]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.593	0.587	0.599
	MAAPE_ARC2	0.476	0.471	0.480
A	SMAPE_200	0.504	0.495	0.512
	MAAPE_ARC2	0.383	0.376	0.390
B	SMAPE_200	0.663	0.653	0.674
	MAAPE_ARC2	0.550	0.542	0.558
C	SMAPE_200	0.692	0.680	0.704
	MAAPE_ARC2	0.576	0.567	0.585
$[base\_D_0\_D_4]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.681	0.675	0.687
	MAAPE_ARC2	0.497	0.493	0.502
A	SMAPE_200	0.636	0.626	0.645
	MAAPE_ARC2	0.425	0.418	0.432
B	SMAPE_200	0.738	0.728	0.748
	MAAPE_ARC2	0.563	0.555	0.570
C	SMAPE_200	0.698	0.685	0.710
	MAAPE_ARC2	0.565	0.556	0.573

Table E 56: Segmentation Experiments - Realistic Information

COM per ABC $[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.792	0.785	0.800
	MAAPE_ARC2	0.591	0.587	0.596
A/B	SMAPE_200	0.741	0.731	0.752
	MAAPE_ARC2	0.670	0.662	0.677
INTERMITTENT	SMAPE_200	1.106	1.090	1.122
	MAAPE_ARC2	0.728	0.717	0.738
NON-INTERMIT	SMAPE_200	0.700	0.691	0.710
	MAAPE_ARC2	0.554	0.548	0.560
per ABC and PLC $[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.791	0.780	0.801
	MAAPE_ARC2	0.551	0.545	0.557
A/B	SMAPE_200	0.722	0.712	0.733
	MAAPE_ARC2	0.538	0.533	0.544
INTERMITTENT	SMAPE_200	1.053	1.037	1.069
	MAAPE_ARC2	0.572	0.564	0.580
NON-INTERMIT	SMAPE_200	0.691	0.679	0.703
	MAAPE_ARC2	0.531	0.527	0.536
per ABC, PLC and currency area $[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.787	0.773	0.801
	MAAPE_ARC2	0.562	0.552	0.572
A/B	SMAPE_200	0.688	0.660	0.716
	MAAPE_ARC2	0.536	0.520	0.553
INTERMITTENT	SMAPE_200	0.893	0.868	0.919
	MAAPE_ARC2	0.617	0.608	0.627
NON-INTERMIT	SMAPE_200	0.638	0.635	0.641
	MAAPE_ARC2	0.500	0.484	0.517
TP per ABC $[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.762	0.748	0.777
	MAAPE_ARC2	0.664	0.652	0.677
A/B	SMAPE_200	0.762	0.744	0.781
	MAAPE_ARC2	0.652	0.638	0.666
INTERMITTENT	SMAPE_200	0.898	0.881	0.915
	MAAPE_ARC2	0.701	0.686	0.716
NON-INTERMIT	SMAPE_200	0.742	0.712	0.772
	MAAPE_ARC2	0.576	0.555	0.598
per ABC and PLC $[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.739	0.731	0.747
	MAAPE_ARC2	0.551	0.543	0.559
A/B	SMAPE_200	0.725	0.719	0.732
	MAAPE_ARC2	0.507	0.503	0.512
INTERMITTENT	SMAPE_200	0.806	0.800	0.812
	MAAPE_ARC2	0.649	0.638	0.661
NON-INTERMIT	SMAPE_200	0.717	0.707	0.727
	MAAPE_ARC2	0.488	0.482	0.493
per ABC, PLC and currency area $[base\_all\_D]$	METRIC		CI_low:	CI_up:
All	SMAPE_200	0.722	0.693	0.730
	MAAPE_ARC2	0.569	0.549	0.590
A/B	SMAPE_200	0.699	0.685	0.713
	MAAPE_ARC2	0.458	0.447	0.468
INTERMITTENT	SMAPE_200	0.758	0.743	0.772
	MAAPE_ARC2	0.625	0.621	0.629
NON-INTERMIT	SMAPE_200	0.679	0.659	0.699
	MAAPE_ARC2	0.401	0.388	0.414

Comment on clustering by lifecycle phases:

In the dominant cluster (mature products), there is little impact on the quality of forecasts and metrics when disregarding invidual phases. However, most trends that are predominantly visible in growth and decline phases are not recognized: Here, forecasts tend to be too low (growth phase / increasing trend) or too high (decline phase / decreasing trend).

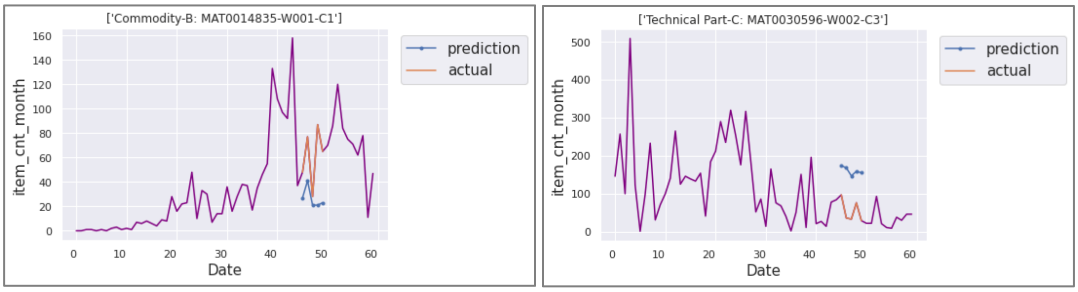


Figure E 37 Demand Evolutions and Predictions for Growing and Declining Phases

After clustering and reconfiguring the parameters accordingly, the forecasts improve within the growing and declining products. However, it is still evident that error metrics and forecasts for mature products are significantly lower and more accurate than for the other two clusters, remains.

Table E 57: Extreme Values in Evaluation Metrics - Perfect Information

	METRIC	EXTREMES [%]
COM		
[*base*]	SMAPE_200 > 1.8	0.201
	SMAPE_200 > 1.6	0.282
	MAAPE_ARC2 > 1.4	0.222
	MAAPE_ARC2 > 1.2	0.319
$[D_0]$	SMAPE_200 > 1.8	0.134
	SMAPE_200 > 1.6	0.178
	MAAPE_ARC2 > 1.4	0.125
	MAAPE_ARC2 > 1.2	0.229
[*base_all_D*]	SMAPE_200 > 1.8	0.024
	SMAPE_200 > 1.6	0.033
	MAAPE_ARC2 > 1.4	0.024
	MAAPE_ARC2 > 1.2	0.041
TP		
[*base*]	SMAPE_200 > 1.8	0.258
	SMAPE_200 > 1.6	0.361
	MAAPE_ARC2 > 1.4	0.314
	MAAPE_ARC2 > 1.2	0.451
$[D_0]$	SMAPE_200 > 1.8	0.034
	SMAPE_200 > 1.6	0.068
	MAAPE_ARC2 > 1.4	0.021
	MAAPE_ARC2 > 1.2	0.075
[*base_all_D*]	SMAPE_200 > 1.8	0.002
	SMAPE_200 > 1.6	0.007
	MAAPE_ARC2 > 1.4	0.002
	MAAPE_ARC2 > 1.2	0.007
COM		
$[base\_D_0]$	SMAPE_200 > 1.8	0.028
	SMAPE_200 > 1.6	0.046
	MAAPE_ARC2 > 1.4	0.021
	MAAPE_ARC2 > 1.2	0.063
$[base\_D_0\_D_1]$	SMAPE_200 > 1.8	0.013
	SMAPE_200 > 1.6	0.022

	MAAPE_ARC2 > 1.4	0.012
	MAAPE_ARC2 > 1.2	0.030
$[base\_D_0\_D_2]$	SMAPE_200 > 1.8	0.008
	SMAPE_200 > 1.6	0.017
	MAAPE_ARC2 > 1.4	0.005
	MAAPE_ARC2 > 1.2	0.027
$[base\_D_0\_D_3]$	SMAPE_200 > 1.8	0.014
	SMAPE_200 > 1.6	0.028
	MAAPE_ARC2 > 1.4	0.009
	MAAPE_ARC2 > 1.2	0.040
$[base\_D_0\_D_4]$	SMAPE_200 > 1.8	0.008
	SMAPE_200 > 1.6	0.018
	MAAPE_ARC2 > 1.4	0.005
	MAAPE_ARC2 > 1.2	0.028
TP		
$[base\_D_0]$	SMAPE_200 > 1.8	0.013
	SMAPE_200 > 1.6	0.033
	MAAPE_ARC2 > 1.4	0.009
	MAAPE_ARC2 > 1.2	0.038
$[base\_D_0\_D_1]$	SMAPE_200 > 1.8	0.002
	SMAPE_200 > 1.6	0.005
	MAAPE_ARC2 > 1.4	0.002
	MAAPE_ARC2 > 1.2	0.005
$[base\_D_0\_D_2]$	SMAPE_200 > 1.8	0.011
	SMAPE_200 > 1.6	0.032
	MAAPE_ARC2 > 1.4	0.010
	MAAPE_ARC2 > 1.2	0.038
$[base\_D_0\_D_3]$	SMAPE_200 > 1.8	0.012
	SMAPE_200 > 1.6	0.032
	MAAPE_ARC2 > 1.4	0.009
	MAAPE_ARC2 > 1.2	0.038
$[base\_D_0\_D_4]$	SMAPE_200 > 1.8	0.005
	SMAPE_200 > 1.6	0.012
	MAAPE_ARC2 > 1.4	0.0053
	MAAPE_ARC2 > 1.2	0.0204

Table E 58: Extreme Values in Evaluation Metrics - Realistic Information

	METRIC	**EXTREMES [%]**
COM		
$[base]$	SMAPE_200 > 1.8	0.326
	SMAPE_200 > 1.6	0.441
	MAAPE_ARC2 > 1.4	0.410
	MAAPE_ARC2 > 1.2	0.515
$[D_0]$	SMAPE_200 > 1.8	0.044
	SMAPE_200 > 1.6	0.072
	MAAPE_ARC2 > 1.4	0.031
	MAAPE_ARC2 > 1.2	0.078
$[base\_all\_D]$	SMAPE_200 > 1.8	0.023
	SMAPE_200 > 1.6	0.039
	MAAPE_ARC2 > 1.4	0.011
	MAAPE_ARC2 > 1.2	0.035
TP		
$[base]$	SMAPE_200 > 1.8	0.384
	SMAPE_200 > 1.6	0.504
	MAAPE_ARC2 > 1.4	0.389
	MAAPE_ARC2 > 1.2	0.504
$[D_0]$	SMAPE_200 > 1.8	0.054
	SMAPE_200 > 1.6	0.084
	MAAPE_ARC2 > 1.4	0.042
	MAAPE_ARC2 > 1.2	0.095
$[base\_all\_D]$	SMAPE_200 > 1.8	0.020
	SMAPE_200 > 1.6	0.034
	MAAPE_ARC2 > 1.4	0.012
	MAAPE_ARC2 > 1.2	0.037
COM		
	SMAPE_200 > 1.8	0.035

$[base\_D_0]$	SMAPE_200 > 1.6	0.060
	MAAPE_ARC2 > 1.4	0.022
	MAAPE_ARC2 > 1.2	0.061
$[base\_D_0\_D_1]$	SMAPE_200 > 1.8	0.023
	SMAPE_200 > 1.6	0.041
	MAAPE_ARC2 > 1.4	0.012
	MAAPE_ARC2 > 1.2	0.038
$[base\_D_0\_D_2]$	SMAPE_200 > 1.8	0.041
	SMAPE_200 > 1.6	0.075
	MAAPE_ARC2 > 1.4	0.041
	MAAPE_ARC2 > 1.2	0.105
$[base\_D_0\_D_3]$	SMAPE_200 > 1.8	0.038
	SMAPE_200 > 1.6	0.064
	MAAPE_ARC2 > 1.4	0.022
	MAAPE_ARC2 > 1.2	0.061
$[base\_D_0\_D_4]$	SMAPE_200 > 1.8	0.053
	SMAPE_200 > 1.6	0.089
	MAAPE_ARC2 > 1.4	0.050
	MAAPE_ARC2 > 1.2	0.118
TP		
$[base\_D_0]$	SMAPE_200 > 1.8	0.018
	SMAPE_200 > 1.6	0.027
	MAAPE_ARC2 > 1.4	0.009
	MAAPE_ARC2 > 1.2	0.025
$[base\_D_0\_D_1]$	SMAPE_200 > 1.8	0.021
	SMAPE_200 > 1.6	0.035
	MAAPE_ARC2 > 1.4	0.013
	MAAPE_ARC2 > 1.2	0.039
$[base\_D_0\_D_2]$	SMAPE_200 > 1.8	0.019
	SMAPE_200 > 1.6	0.030
	MAAPE_ARC2 > 1.4	0.013
	MAAPE_ARC2 > 1.2	0.034
$[base\_D_0\_D_3]$	SMAPE_200 > 1.8	0.017
	SMAPE_200 > 1.6	0.027
	MAAPE_ARC2 > 1.4	0.008
	MAAPE_ARC2 > 1.2	0.023
$[base\_D_0\_D_4]$	SMAPE_200 > 1.8	0.028
	SMAPE_200 > 1.6	0.047
	MAAPE_ARC2 > 1.4	0.027
	MAAPE_ARC2 > 1.2	0.067