CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

What is My Car Worth? Hedonic Price Analysis of the German Used Car Market

Master's thesis

Author: Bc. Radka Doležalová

Study program: Economics and Finance

Supervisor: Mgr. Petr Polák MSc.

Year of defense: 2020

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, July 30, 2020

Radka Dolezalova

Abstract

Valuation of used cars, affected by various technical attributes and information asymmetry, is the key objective of all agents operating on the automobile market. This thesis, focusing on a hedonic price analysis, aims to determine basic as well as additional attributes as determinants of a used car market price. In addition, the analysis sheds light upon novel attributes (service records, cigarette smoke pollution of a vehicle interior, selling channel factor in the e-commerce environment, and a German geographical division). The hedonic price research uses the unique data sample of the German used car market, extracted from the database of the e-commerce platform AutoScout24 comprised of almost 51 thousand vehicles and 57 attributes. The model selection is specified by the incorporation of the Bayesian model averaging approach. The research proves the complexity of a valuation of a used vehicle in a term of a substantial number of relevant variables. The most interesting innovative conclusions are non-significant effect of selling channels and small local price differences among two German regions. Remarkable are also the significant effect of the status of previous owners, bodywork colour, and smoke pollution. The estimated vehicle lifespan of 10 years shows that cars have shorter than generally accepted lifespan.

JEL Classification	D46, C11, L62, D12
Keywords	market for used cars; used car prices; hedonic price analysis; Bayesian model averaging; Germany
Title	What is My Car Worth? Hedonic Price Analysis of the German Used Car Market

Abstrakt

Ocenění ojetých vozů, ovlivněných různými technickými atributy a informační asymetrií, je klíčovým zájmem všech agentů působících na automobilovém trhu. Tato práce se zaměřuje na analýzu hedonických cen a jejím cílem je určit základní a doplňkové atributy, které jsou klíčovými faktory tržní hodnoty ojetého automobilu. Analýza dále osvětluje vliv nových atributů (servisních záznamů, zamoření interiéru vozidla cigaretovým kouřem, faktoru prodejního kanálu v prostředí e-commerce a zeměpisného rozdělení Německa). Hedonická cenová analýza využívá jedinečný vzorek dat německého trhu ojetých vozů extrahovaný z databáze e-commerce platformy AutoScout24, který obsahuje téměř 51 tisíc vozidel a 57 atributů. Výběr modelu je specifikován na základě aplikace Bayesovského modelu průměrování. Výzkum prokázal složitost oceňování ojetého vozidla z hlediska vysokého počtu relevantních proměnných. Mezi nejzajímavější inovativní závěry patří skutečnost, že prodejní kanál nemá vliv na tržní hodnotu a že se objevují menší cenové rozdíly mezi dvěma německými regiony. Pozoruhodný je také významný effekt statutu předchozích majitelů, barvy karoserie a znečištění vozidla cigaretovým kouřem. Odhadovaná desetiletá životnost vozidla ukazuje, že automobily mají kratší délku životnosti než je obecně předpokládáno.

Klasifikace JEL	D46, C11, L62, D12
Klíčová slova	trh ojetých vozů; cena ojetých vozů; hedonická metoda oceňování; Bayesovská metoda průměrování, Německo
Název práce	Jaká je hodnota mého vozu? Hedonická metoda oceňování německého trhu ojetých vozů

Acknowledgments

The author of this thesis would like to express a sincere appreciation to Mgr. Petr Polák MSc. for providing positive guidance during the writing that steered the heading of this thesis. In addition, the author is grateful to Rogier van Rooij, whose inspiring collaboration contributed to overcoming of all inconveniences connected to the AutoScout24 web scraping code. Special acknowledgments belong to the author's family and friends for being a support during the whole studies.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Dolezalova, Radka: *What is My Car Worth? Hedonic Price Analysis of the German Used Car Market*. Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 87. Advisor: Mgr. Petr Polák MSc.

Contents

List of Tables

List of Figures

Acronyms

BMA	Bayesian model averaging
cc	cubic centimetres
CNG	compressed natural gas
EU	European Union
hp	horse power
kW	kilowatt
LPG	liquefied petroleum gas
MPV	multi-purpose vehicle
PIP	posterior inclusion probability
PMP	posterior model probability
SUV	sport utility vehicle
US	United States of America
VDA	Verkehrsclub Deutschland
VW	Volkswagen

Master's Thesis Proposal

Author	Bc. Radka Doležalová
Supervisor	Mgr. Petr Polák MSc.
Proposed topic	The determinants affecting residual prices of used cars

Motivation The automobile industry evinced significant changes in recent years. With the inception of the economic crisis in 2008, customers started with searching for alternative products. Therefore a relatively high retail price of a car steered their seek to the second-hand market.

In the consequence of the increasing used car demand arose a question which factors influence the price of the second-hand car most significantly. Individuals offering their cars on the market, professional merchants selling pre-owned cars, households intending to buy a new car and resale it in the future, lease companies, manufacturers and many other car market players wonder how to value a given car. This thesis aims to answer this question and to appraise car characteristics as value drivers.

In recent years, the hedonic price classification has attracted a lot of researches attention and therefore the related literature is relatively miscellaneous. But some of the researches suffer from a limited number of observations or lesser number of car characteristics and from a missing regional effect. This thesis aims to evaluate a complete online database of used car comprised by hundreds of thousands advertisements and detail description of each automobile on the Internet platform Autoscout24.de.

Moreover, the researches vary according to used methodology significantly. Refer to Madigan and Raftery (1994) who oppose unsubstantiated model selection in literature, the hedonic price model in this thesis will be performed on the basis of alternative approach to model selection: Frequentist model averaging (FMA) and Bayesian model averaging (BMA).

Hypotheses

Hypothesis #1: What are the main car characteristics influencing a residual value of a used car?

Hypothesis #2: How the hedonic price factors differ in regions?

Methodology Cross-sectional data collected on the online platform Autoscout24.de will contain basic car characteristics (age, mileage, manufacturer, model, body type, transmission, fuel type, fuel consumption etc.) as well as features related to a car equipment and geographical specification. The effect of variables to an offered price will be analysed using appropriate linear models and machine learning approaches. The most inspiring researches that will serve as an examples are Gongqi *et al.* (2011) with their use of BP neural network and Pal *et al.* (2018) comparing linear regression and RF. The final model specification will be derived from Frequentist model averaging (FMA) and Bayesian model averaging (BMA).

Expected Contribution The target of this thesis is to complete a present literature concerning with the hedonic price analysis, to proffer a research based on a sufficient amount of observations and various car characteristics. Moreover, the accuracy of different used models in the context of BMA and FMA performing as appropriate model selections will be discussed. To conclude a practical application, both sellers and buyers (professional as well as individuals) can utilize expected results of this thesis. New information about the real value drivers could be able to mitigate an information asymmetry and an adverse selection problem.

Outline

1. Introduction

2. The used car market description. This section will focus on the second-hand car market and its specificity. The integral part of this introduction will be a differentiation of selling channels, product definition and contemplation about policy interest in the used car market.

3. European specificity. To put a subsequent research into perspective of nowadays evolution of European used car market, this section will describe participants, market channels and second-hand market environment in general in European Union.

4. The used car market from the microeconomic point of view. This section will refer to a basic microeconomic identification of the market and market players. A deeper analysis will expose roles of sellers and buyers, an information problem and the last part will be concerned with adverse selection, quality level and market for lemons phenomenon.

5. Literature overview. The literature overview section focuses on two integral questions: what are the most important factors affecting the hedonic price and whether good products are driven out by bad quality cars. Both questions

will be summarised by various recent studies differing in approaches. The second part will introduce key research papers concerning with BMA and FMA methods.

6. Methodology. Methodology used in the own research will be discussed in details referring to requirements, advantages and limitations.

7. Results. The overall results of this thesis will be evaluated and critically assessed. A potential future research questions will be defined.

8. Conclusion.

Core bibliography

Akerlof, G. A. (1970): "The market for "lemons": Quality uncertainty and the market mechanism." The Quarterly Journal of Economics 84(3): pp. 488-500.

Andrews, T. C. Benzing (2007): "The determinants of price in internet auctions of used cars." Atlantic Economic Journal 35: pp. 43-57.

Bauer, I., L. Zavolokina, G. Schwabe (2019): "Is there a market for trusted car data?" Electronic Markets 30(2): pp. 211-225.

Bond, E. W. (1982): "A direct test of the "lemons" model: The market for used pickup trucks." The Ameri can Economic Review 72(4): pp. 836-840.

Gongqi, S., W. Yansong, Z. Qiang (2011): "New model for residual value prediction of the used car based on BP neural network and nonlinear curve fit." Measuring Technology and Mechatronics Automation 2: pp. 682-685.

Hoeting, J. A., D. Madigan, A. E. Raftery, C. T. Volinsky (1999): "Bayesian model averaging: A tutorial." Statistical Science 14(4): pp. 382-401.

Kihm, A. C. Vance (2016): "The determinants of equity transmission between the new and used car markets: A hedonic analysis." Journal of the Operational Research Society 67: pp. 1250-1258.

Kim, J.-C. (1985): "The market for "lemons" reconsidered: A model of the used car market with asymmetric information." The American Economic Review 75(4): pp. 836-43.

Mishra, D. P., J. B. Heide, S. G. Cort (1998): "Information asymmetry and levels of agency relationships." Journal of Marketing Research 35(3): pp. 277-295.

Madigan, D and Raftery, A. E (1994). "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam'sWindow." Journal of the American Statistical Association. 89: 1535-1546.

Pal, N., P. Arora, P. Kohli, D. Sundararaman, S. S. Palakurthy (2018): "How much is my car worth? a methodology for predicting used carsâ€™ prices using random forest." Proceedings of the 2018 Future of Information and Communication Conference (FICC), Vol. 1.

Pudaruth, S. (2014): "Predicting the price of used cars using machine learning techniques." International Journal of Informatione and Computation Technology 4(7): pp. 753-764.

Wasserman L. (2000). "Bayesian Model Selection and Model Averaging." Journal of Mathematical Psychology, 44:92-107.

Wu, J.-D., C.-C. Hsu, H.-C. Chen (2009): "An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference." Expert Syst. Appl. 36(4): pp. 7809-7817.

Chapter 1

Introduction

Since transportation is an integral part of the daily life of today's society, at some point in time most households and companies face difficulties connected to a purchase of a car. Nevertheless, each vehicle on the market consists of tens of technical attributes that impede a customer's decision. In the process of a purchase of a used car, typical for various heterogeneous quality levels and substantial information asymmetry, the decision-making process can turn to an even more precarious problem. According to Bauer *et al.* (2019), the market offering used cars belongs to three markets with the lowest trust. For these reasons, a need for a precise determination of factors influencing a car value has arisen as an essential ability of all market players (private and professional buyers and sellers, leasing companies, insurance brokers and manufacturers).

In recent years, new aspects connected to the topic of the automobile industry appeared. The ecological perspective with a sustainability targeting, as well as the overpopulation in urban areas that produce heavy traffic and parking problems, mass and dynamic automotive development together with globalisation, the substantial effect of e-commerce and the Internet as an information source, are just a few of very intensively discussed topics nowadays. On the ground of these transformations, the relevance of older studies is becoming limited and a publication of new analyses concerning the topic of the used car market is required constantly. This thesis aims to detect the current drivers of a car value on the German used market.

An analysis of the German used car market is an optimal choice from different perspectives. Germany as one of the most important global automotive producers and a country with a high motorisation rate is specific for a large used car market and substantial (new and used car) export character. In ad-

dition, the German second-hand car market suffers from a lack of studies. The analysis presented in this thesis propounds a hedonic price analysis performed with the dataset derived from the database of the biggest European vehicle e-commerce platform AutoScout24. This database has never been utilized in the literature yet because of the need for an exacting data scraping procedure.

This study extends the current literature in various ways. Primarily it complements the deficiency of studied e-commerce databases. The biggest car market player with a huge number of visitors is used to provide a comprehensive sample with almost 51 thousand observations and a set of diverse explanatory variables (57 unique specificities). The selected attributes clarify or complement the existing knowledge about a market price formation. Is a buyer able to utilize regional price differences for a favourable purchase? Does a selling channel keep an ability to play an important role in price formation in the e-commerce environment? Does it pay off to attend all regular services? And to what extent are important for instance, colour of bodywork colour, previous owners' status and their attitude to smoke in a car? To answer (not only) these questions, the analysis sheds light also upon the less discussed characteristics and proposes an innovative geographical division of the local market in Germany, analysis of the effect of full service history, cigarette smoke pollution of a vehicle interior and selling channel effect.

Uniquely, in the context of the used automobile market, the approach of Bayesian model averaging (BMA) is used to deal with model uncertainty and the detection of redundant variables. The modeling of a price is based on the hedonic price theory that considers a good as a set of measurable qualities deriving a utility for a customer, according to an approach described by Lancaster (1966) and Griliches (1961). Besides the testing of the main attributes of the price formation, the thesis aims to describe the automobile market in a wider context of automotive production and the effect of information asymmetry. What is emphasized in this thesis is a comparison of individual features in the context of existing related literature in substantial detail.

This thesis is structured as follows. In Chapter 2 the used car market is introduced in the context of the basic economic theory, discussing also the problems of adverse selection and the "lemon" market phenomenon. Subsequently, in Chapter 3 the German used market is presented with details about European Union interconnectivity, automotive production and environmental dilemma. Chapter 4 deals with the hedonic price theory and Chapter 5 briefly summarizes the current literature knowledge. In Chapter 6 the crucial researches are

discussed in the context of the analyzed AutoScout24 dataset, consequent sampling procedure, a definition of variables, and expected results. Chapter 7 aims to describe the BMA approach and its application for the hedonic price model. Chapter 8 is dedicated to the discussion of the result.

Chapter 2

Used Car Market

2.1 Specificities of the Market

Year 2008 went to history as the gradual falling point of the worldwide economy and in the automobile industry, the effect of the crisis was observed immediately. The market of new cars, characteristic of relatively high priced products, experienced a lowering of demand (Kihm & Vance 2016). On the contrary, the used car market offering substituting goods evinced an increasing trend of sold products (Pal *et al.* 2018). Kihm & Vance (2016) showed that the average price of a German used car was by 7,4% lower than the price of a comparable new car in 2008. A very extraordinary example of a substitution effect between retail and second-hand car market was described by Pudaruth (2014) in Mauritius. In this eastern African island, the registration of cars grown by 234% but the sales of new cars recorded the fall by 8% from 2003 to 2013. An enormous increase in the used car market performed also China with an annual rise of 30% in 2011 (Gongqi *et al.* 2011). In recent years, the revolution of the business with used cars is based on e-commerce that is implemented by the majority of used-car dealers nowadays (Andrews & Benzing 2007).

From a perspective of microeconomics, the case of the used car market is a very remarkable topic because of its various phenomena. A car is a specific product with its relatively high initial value, long lifespan and a various possible maintenance levels (Gongqi *et al.* 2011). The used car market is characteristic of the huge quality heterogeneity that increases the uncertainty on the market (Akerlof 1970). The true quality of a good cannot be recognized before the realization of purchase and therefore a buyer faces an information asymmetry and an adverse selection problem (Akerlof 1970). Another inter-

esting dimension to study is a complementary relationship between vehicle use and fuel consumption on the market. By inspecting the market players, it is evident that the roles of a buyer and a seller can be swapped in the course of time (Kim 1985).

2.2 Classification of Selling Channels

The selling channels of the second-hand car market is constantly very diverse. The market could be easily divided into online and offline market forms. Moreover, it can be distinguished between goods sold via dealer channels or cars offered by private individuals. Proved by the analysis performed by Andrews & Benzing (2007), on average a dealer earns a higher price premium (the difference between the executed price and the actual price of a car) in comparison to an individual seller. The professionals are expected to have also more incentives and capabilities to offer better information or warranty for their goods (Kim 1985).

More specific channels are auctions and leases that deserve a deeper discussion. An online auction is a platform creating an environment to offer used cars for given ask prices. After an *ex ante* defined time period, a sale is realized for the highest bid price. The phenomenon of the online auction appears to be very scientifically attractive as an important source of data and observation of customers' behaviour in a very specific environment. On the ground of affordability, a lease of a car becomes more popular in developed countries nowadays (Pal *et al.* 2018). The relationship between a lessor and a lessee is based on fixed instalments and an option to repurchase a car at the end of the lease period (Pudaruth 2014). It is important to consider a lease as a generator of new products for the used car market. This attitude is supported by the fact that clients' repurchases are not usually realized and a car is offered in the second-hand market (Pudaruth 2014). To emphasize the importance of the proper definition of value drivers of used cars, the leasing provider problem should be described in more details. Both types of estimation errors are treacherous for this business: overestimation as well as underestimation of the residual value of a car at the end of a lease period. An underestimation leads to an excessively larger instalment causing an outflow of potential customers. An overestimation generates financial loss since an expected residual value is larger than a realized selling price (Pudaruth 2014).

2.3 Information Asymmetry and the Market for "Lemons" Phenomenon

The used car market, similarly as insurance services or a problem of employee recruiting, is one of the best examples of the problem of information asymmetry (Akerlof 1970). In general, the asymmetrical effect between a buyer and a seller is caused by an inability of the buyer to recognize the true quality of the good before the purchase (Mishra *et al.* 1998). Among the most popular trick of dishonest sellers belong for instance manipulation of an odometer. According to the research performed by European Parliament (2018), this practice is abused in up to 50% cases of used car offers. Furthermore, information asymmetry and an associated adverse selection problem lead to an enlargement of the opportunistic behaviour (Mishra *et al.* 1998). And on top of that, the adverse selection, jointly with the atmosphere of distrust can weaken the market efficiency (Bauer *et al.* 2019). One of the most surprising investigations of the customer study in Germany revealed that the automobile industry is one of three markets with the lowest customers' trust. Lowering the quality can reduce the volume of goods supply and the repetition of the cascading effect can steer the market to the collapse in the worst scenario (Bauer *et al.* 2019).

In recent years, many research papers were concerned with the information asymmetry between agents operating on the used car market and many of these papers struggled to offer an appropriate solution as well. In spite of all exertion, the main barrier to deal with this undesirable effect of information asymmetry was not perfectly uncovered. Akerlof (1970) put forward intervention of government or private institutions to solve the asymmetry. As an optimal approach, sellers' guarantees, brand name and licensing system (similar to health service or law) are proposed. Mishra *et al.* (1998) evaluate customer bonds or non-salvageable investment to signal the quality of offered goods. A very interesting experiment by Bauer *et al.* (2019) deserves attention. This analysis of the market for pickup trucks addresses the problem to distrust in information sources and their quality, limitations of central register and high costs associated with data collection. The authors proved that the use of their blockchain database mitigates the difference between the actual market value and the realized price and therefore reduces the information asymmetrical effect between agents on the used car market.

Kim (1985) argues that the quality of a vehicle is connected with the pre-

vious owner of a car because it reflects a maintenance level, service or also individual driving habits of a user. As already discussed, a buyer is not able to assess a true level of quality before the realization of purchase. Therefore a question should be asked whether a seller has an incentive to behave honestly and offer a high quality good on the market for used cars. To follow Akerlof (1970) and his explanation of the market evolution, it should be reminded that a price and a quality of a new commodity offered on the market responds to the average price and quality of goods already traded on the market. Gresham's law assumes that a commodity of a low level of quality drives out a good commodity from the market because an agent owning a higher than an average quality car does not intend to sell it for a set price. In the case of asymmetric information, a seller has no motivation to offer a high quality good.

The problem is described as the market for the "lemons" phenomenon (Akerlof 1970). An expression "lemon" in this context refers to an automobile with a hidden defect (Hofstede *et al.* 2008). The theoretical background of quality uncertainty was described by Akerlof (1970), for the application of these principles in the market research, the author of this thesis would like to refer to Bond (1982). Bond (1982) studied the difference between cars purchased directly from manufacturers and those traded on the second-hand market and observed no evidence for the "lemons" market phenomenon on the pre-owned vehicles market.

Chapter 3

German Automotive Industry and Used Car Market

3.1 Market Position of Germany

Europe, which gave birth to the first motor vehicles in the late 19th century, has shown that it still belongs to the biggest producers and innovators in the automobile market. In 2018, with the production of 21% of world automobile manufacturing, Europeans proved that they play an important role. Moreover, customers in the European Union possess approximately 268 million passenger cars and this number had an increasing trend in recent years. The popularity of automobile ownership is represented by the motorisation rate reporting that in some member countries of the European Union almost 70% of the population own a passenger car (ACEA.be 2019). Moreover, among the European automotive oriented countries stands out one key market player: Germany.

To properly understand the main characteristics of the used car market in Germany, it is necessary to map new changes in European society affecting the mobility as well as an overall situation with automotive production in Germany precisely. New cars characterized by their initial prices and features are potential future goods in the used car market as well as the current substitutes to the second-hand vehicles from the perspective of the buyer.

The German vehicle manufacturing industry represents significant production capacity as well as the crucial innovation background leading the direction of the further progress of the global automotive industry. A substantial number of iconic brands: Audi, BMW, Mercedes Benz, Porsche and Volkswagen, originate from Germany. Moreover, Germans produce 19 % of all brands

worldwide, and approximately 3,650 new patents associated with automobile production are listed in Germany every year (Verband der Automobilindustrie 2020). The position of the German vehicle industry is important for the worldwide production and progress in the same way as the vehicle industry is an irreplaceable segment of the German economy. This synergistic effect is even better illustrated with the scenario of the potential results of the negative issues of German automotive. Truett & Truett (2017) describes the unfavorable progress of the vehicle industry that could imply serious consequences for the whole German national economy. The car manufacturing industry is the biggest exporter and it represents 35% of all R&D within all German industries (Verband der Automobilindustrie 2020). Since the production of automobiles is not only technologically but also human capital demanding, the German labour market is significantly impacted by the automotive production (Spatz & Nunnenkamp 2002). The majority of all German industries engineers are employed to construct automotive components and the overall average number of workers in automotive oscillates around 785,000 individuals (Verband der Automobilindustrie 2020). On the contrary to other German industries that were decreasing the number of their employees in response to the globalization and formation of a new competitive environment since 1991, automotive was continuing to recruit (Verband der Automobilindustrie 2020).

In recent history, the German industry was facing national reunion, euro currency establishment and restructuring of industry in the 1990s (Truett & Truett 2017). Although Germans coped with all problems relatively successfully, globalization appears to be the crucial issue of the industry progress nowadays (Spatz & Nunnenkamp 2002). Currently, more than half of the production of all automotive components is realized outside of the German territory (Peters 2000). Whereas Germany, the USA and Japan were a byword for the car producers in history, the new trends shift the production to the South and Eastern Europe and developing Asian and Latin American countries (Spatz & Nunnenkamp 2002). Although the most exhausting challenges for German automotive are still coming from Japan as described in detail by Truett & Truett (2017), the new competitive manufacturers are according to Spatz & Nunnenkamp (2002) Mexico, Spain, the Czech Republic and Korea primarily for the sake of their intermediate automotive production. Contrarily, to assess the positive impact of growing globalization, the outsourcing of automotive intermediate products can be seen as a desired substitute to the relatively high labour costs is Germany (Truett & Truett 2017).

The question is, whether Germany will deal successfully with the new global automotive reallocation in the future to sustain the current volume of sales and avoid a huge negative impact on the national economy. From the used car market perspective, it is important to take into consideration the substituting effect on customers that opts for a new or second-hand car. A crucial solution could be established by a government regulation that could support new car demand by disadvantaging the used vehicles in the market to motive automotive production. Since the automobile industry lobbying is very substantial in Germany (as described properly by for instance Gössling & Metzler 2017), the main purposes of planned and already realized restriction on the older cars that are hidden behind the European ecological targeting are very dubious in this context.

3.2 Automobile Dilemma

The ecology as a new social topic hits the headlines nowadays. The overall discussion is supported by the change in customers' habits as well as the regulation on the national levels. How evident are those tendencies in Germany?

The German policy aspires to reach the most ambiguous ecological goals in the global comparison to ensure sustainable mobility with environmentally friendly character using the elimination of emissions (Verband der Automobilindustrie 2020). To regulate carbon dioxide, the legislator has opted for two main alternatives: an implementation of fuel taxes or support of the demand for fuel-efficient vehicles (Schmid 2017). Moreover, the regulation aims to restrict the entry of old (mainly diesel-fueled) cars to the centres of selected cities and towns.

On the contrary, the priority of society appears to hold opposing goals. Urbanization, the boom of mobility and an increasing need for goods flow are becoming the new trends of the developed and also emerging societies (Verband der Automobilindustrie 2020). Germany does not diverge. According to Gössling & Metzler (2017), transport volume increased by 13 % from 1999 to 2011 in Germany but sales of alternative fuel cars do not rise sufficiently fast to maintain the ecological balance. The vast majority of all cars are represented by passenger cars powered by diesel and gasoline fuels and only 1.6% of newly registered cars use electric, hybrid or gas power. As a consequence, the volume of exhaust fumes emissions from transport are growing and the transport sector produces up to 21% of all emissions in Germany. German unlimited speed on

the highways and the rising popularity of fuel-inefficient cars (i.e. sport utility vehicles (SUVs)) are the abstract last nails in a coffin of the German effort for ecological targeting. The reason behind an unconcern of the customers could steam from the fuel price elasticity. Gössling & Metzler (2017) published the analysis demonstrating that the increased fuel price by 1% can lower the fuel consumption by only 0.27%. Even more alarming is the disinterest of drivers of company vehicles to entertain with their driving habits influencing emissions. To conclude the discrepancies between the society's effort to eliminate emission systematically and the irresponsible behaviour of individuals, the most alarming fact is that Germany will not meet its own emissions targets in case that the complex consumer buying and driving habits won't be changed (Gössling & Metzler 2017).

To improve the availability of information on the ecological burdens of vehicles for the public, German automobile clubs endeavour to inform about ecological conditions of vehicles on the market. To provide customers with basic and clear information about the environmental burdens of a given car, the legislator implemented emission labels. The coloured plates with a scale from A+ (ecological) to G (non-ecological) inform the public about the ecological footprint on the basis of a level of carbon dioxide emissions. A very pertinent contemplation about this principle is proposed by Gössling & Metzler (2017) that insinuate the paradox that the lower fuel consumption could have an impact on the fuel price elasticity from the customer's perspective and it could lead to an even higher overall amount of fuel consumption. On the ground of these conclusions, Gössling & Metzler (2017) consider the emission labels as a counterproductive action.

3.3 Dieselgate Affair

Apart from the emission labels, buyers of used (and new) cars are exposed to another confusing fact concerning emissions: Dieselgate. A few years ago, the European automobile market regarded a diesel drive as a solution of a pollution crisis (Schmid 2017). The year 2015 and Dieselgate affair of the falsification of nitrogen oxide and nitrogen dioxide emission footprint caused by the concern Volkswagen revealed in the USA changed the European heading significantly (Bovens 2016). Volkswagen's vehicle software in some diesel engines was equipped with sensors detecting laboratory conditions for emission measuring and setting an automobile into a low-performance mode to deceive

controlling mechanism (Hotten 2015). These abusive practices are claimed to be in operation in diesel engines of Audi A3, Volkswagen Jetta, Volkswagen Beetle, Volkswagen Golf and Volkswagen Passat. At the end, the concern proceeds to recall 2.4 million defeat cars in Germany, 8.5 million vehicles in Europe and 11 million automobiles worldwide in general (Hotten 2015).

The potential long-term consequences of the scandal are expected to appear in various layers. The affair of falsification of emission production has a negative effect on the perception of diesel fuel in general. However, the sales of diesel used cars evinced only a negligible effect as was shown by Schmid (2017) in the study of the France second-hand car market. But the reputation of Volkswagen has weakened and the company makes effort to restore its goodwill intensively. In 2015 the Volkswagen corporation evinced expected fall of share prices by more than 30% in comparison to the previous year (Hotten 2015). Its financial balance was disconcerted by consequent fines and costs connected to the recall of the affected vehicles. Last but not least, the whole German automotive industry was expected to suffer from the scandal due to its close interconnection character in the long-term.

3.4 Recent Progress of the German New Car Market

Contrarily to pessimistic prospects, the German new car market is evincing the growing tendencies in recent years. In 2019 the sales outstripped the amount of the previous year. This led to an amount of 16 million new registered cars in Germany that appears to be an imposing number in comparison to 51 million cars that represent all cars in use in Germany in 2018 (for a better illustration, there were 308 million passenger cars recorded in the European Union in 2018) (ACEA.be 2019). Mass production is the factor that influences the high motorisation rate in Germany reaching 569 vehicles per 1,000 inhabitants and maintains this country on the fifth position in the European motorization rating (behind Luxembourg, Italy, Poland and Slovenia) (ACEA.be 2019).

On one hand, the proportion of diesel engine vehicles among newly registered cars has a decreasing trend as an expected consequence of the mentioned Dieselgate affair. On the other hand, the indication of the weakening of diesel fuel automobiles was evident already prior to the scandal investigation. But what does appears to be a relevant concern is the lowering of an export bal-

ance of the newly produced passenger cars as indicated by the Verband der Automobilindustrie (2020).

3.5 Recent Progress of the German Used Car Market

The used car market in the European Union is characteristic of free trade character and predominating single currency that supports the flow of goods among member states. European Commission, Justice and Consumers (2014a) discovered a significant flow of used cars from EU15 countries to the EU13 members.[1] Germany belongs not only to the exporter country (as a part of the EU15 group), but this country occupies the sovereign leading exporting position in the used car market. An astonishing amount of 42% of all imported pre-owned cars in the European Union (EU) states comes from Germany (European Commission, Justice and Consumers 2014b). This trend could be explained by the fact that the number of fraudulent practices in EU13 countries is significantly higher than in the EU15 states (European Commission, Justice and Consumers 2014a). According to the survey performed by the European Commission, Justice and Consumers (2014a), customer's general knowledge of car attributes is rather low and traders often do not provide buyers with the truthful crucial car characteristics such as mileage state, carbon dioxide emissions or accident history. In general, the used car market is assessed as the market with a low trust ratio and a high proposition of fraud manners (European Commission, Justice and Consumers 2014b). Over a long period, the European Parliament appeals for a cross-country legislative integrity that would mitigate the heterogeneity of used car market trust among the EU member states (European Parliament 2018). For these reasons, the German second-hand market should be observed in the wider geographical perspective as a part of the EU market. The low trust of customers as well as low-developed legislation in the EU13 countries could have a significant effect on the German second-hand vehicle market as well.

The agents operating in the German used automobile market can be divided into traders, associations of traders, customers, associations joining customers

[1]The EU15 is comprised of 15 countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom. In 2004 the European Union accepted 13 new countries that are known are EU13: Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia.

and automobile clubs (European Commission, Justice and Consumers 2014a). Among the most important automobile clubs belongs Verkehrsclub Deutschland (VDA) and Allgemeiner Deutscher Automobil-Club that renders to the public, beyond other services, the information for a purchase/sale of a second-hand car to mitigate an information barrier (ADAC.de 2019). Even though a seller is provided with an adequate amount of information on the official websites of these clubs and associations, the majority of agents intend to derive information from Internet car portals in addition as shown in the survey performed by the European Commission, Justice and Consumers (2014b). This survey also ascertained that the final decision of buyers is based on four vehicle characteristics: age, mileage, make (manufacturer) and mechanical characteristics. Moreover, an important role plays also the ecological operating of a vehicle. Gössling & Metzler (2017) state that pre-owned car buyers in Germany are becoming more "green" than customers in the new car market and that they prefer the attributes of a car indicating the low ecological burdens. Taking into consideration the heading of the ecological dilemma, the effect of the environmental operation of a vehicle is expected to be limited. However, the current literature does not provide sufficient researches dealing with this topic in Germany.

Various websites regarding the automobile details and sales are becoming even more important sources of information for buyers, but the question is whether the Internet has the power to change the used car market and its nature significantly. Approximately half of all commercial second-hand sales are performed via the Internet car portals nowadays. This relatively high proportion can be a response to a detailed description of vehicles characteristics that a buyer has at his disposal in the various used car selling portals (European Commission, Justice and Consumers 2014b). Andrews & Benzing (2007) point out that e-commerce plays an irreplaceable role on the used car market. Moreover, the Internet is expected to have a decisive impact on the heading of the second-hand car market in the future.

Chapter 4

Hedonic Price Analysis

4.1 Theoretical Background

To study a decision-making process of a customer, Lancaster (1966) successfully defended an insufficiency of the traditional theory for its disregarding of intrinsic characteristics. To refer to the paper by Lancaster (1966) and the theory of the consumer behavior formulated by Arrow Debreu, the hedonic price model went down to history.

As described by Lancaster (1966), the hedonic price approach is based on the idea of considering a valuation of a good using its characteristics. A customer does not perceive the direct utility from a good but a utility is derived from attributes of the desired good. Moreover, Lancaster (1966) shows that goods can possess multiple attributes and the combination of various characteristics of good is valued differently than an individual attribute separately. Durable goods (cars given as an example) are considered to be a complex of technical characteristics and the overall price of a good is derived from the value of a complex of various technical attributes. These facts are reflected in a customer's purchasing decision since customer's benefit is influenced directly by characteristics of a purchased good. The desired good in terms of a set of attributes is selected to meet a utility maximization assumption of a customer (Erdem & Şentürk 2009).

Citing Rosen (1974), pioneering this concept, it is a problem of spatial equilibrium with the dimensional allocation of purchasing and buying decisions of the market players based on a complex of implicit values. The allocation of the decision is according to Rosen (1974) in compliance with the assumption that:

- the equality of quantity of goods supplied by buyers and the quantity of goods demanded by customers is satisfied
- the behavior of both market players is motivated by maximization of their utility
- market price is set to perfectly match supplied and demanded quantities
- the option of improvement of the position of an individual market player is excluded and
- any optimum location is feasible

According to hedonic price theory described by Rosen (1974), each good can be described using n objectively measurable characteristics oriented in a vector z:

$$z = (z_1, z_2, \ldots, z_n)$$

The overall value of a good *p(z)* in terms of its characteristics thus can be expressed as:

$$p(z) = p(z_1, z_2, \ldots, z_n)$$

The hedonic price analysis can be viewed in this context as a micro - econometrics analysis (Goodman 1998). Greenstone (2017) showed that the econometric approach uncovering the intrinsic values of an attribute of interest consists of two steps: formulation of a hedonic price schedule that represents a locus between market value of a good and the characteristics it possesses, and function describing customer's willingness to pay for a given good (or alternatively a buyer's willingness to sell a given good). For the further theoretical details of the product differentiation please refer to Rosen (1974) and Greenstone (2017)

More theoretical background can be found also in Lancaster (1966). Lancaster is commonly considered to be the author of the hedonic price approach in the current literature. However, Goodman (1998) claims that hedonic price modeling can be even older methodology and that it is rooted to Court (1939) illustrating his innovative approach using an automobile market. To other papers that desire a reader's attention belong also papers implementing the hedonic price analysis for various topics for instance Griliches (1961) concerning to panel data of a car market, McFadden (1974) and his research of transportation demand or Halvorsen & Pollakowski (1981) demonstrating a choice of an

appropriate functional form of the hedonic price model using housing property dataset.

4.2 Functional Form of the Model

What makes a hedonic price analysis a though proposition is a determination of a proper functional form. A comprehensive discussion can be found for instance in Halvorsen & Pollakowski (1981) that disagree with the theoretical attitude to the functional form decision. Authors demonstrate the use of Box-Cox transformation and likelihood ratio tests and demur to often used functional forms with the lack of objectively tested appropriateness. This approach was used also for instance in van Dalen & Bode (2004), where the appropriateness of the *a priori* preferred a log-linear model (based on the skewed distribution of prices) was confirmed using the regression of Box-Cox transformed model against the alternative of a linear functional form). Other examples can be found in Erdem & Şentürk (2009) that advocate the convention of the use of the log-linear and the semi-log models for the simplicity of an interpretation. To the supporters of the theoretically grounded functional form determination belong for instance Ohta & Griliches (1976) using semi-log forms for modeling.

Chapter 5

Current Literature Review

5.1 Cornerstones of the Automobile Research

The first cornerstones of vehicles' market research were set up at the beginning of the 20th century (shortly after an automobile became a symbol of modern society). Hedonic price analysis by Court (1939) is one of the important papers dealing with this topic in the 1930s. In the 1960s, the attention attracted researches concerning a very limited segment of the market in the United States of America (US) by Griliches (1961) and Triplett (1969) and in the consequent decade was popular the analysis focusing on British sedans by Cowling & Cubbin (1972). In the following years, the topic of new and used vehicles' market value became more popular. Simultaneously, the scope of researches was enriched with various types of cars and more explanatory variables. When exploring the researches in the 20th and the 21st century, a very considerable shift of a good specification, as well as the change of the market environment and the requirements of customers can be observed. The relevance of the older studies for the description of the current market is therefore very limited.[1]

In general, the field of interest of literature concerning this topic can be divided into three groups according to the main hypotheses. Papers belonging to the first group use cross-sectional automobile data to test the hypothesis of biased price indexes as a consequence of the quality changes in vehicles. Researches by Triplett (1969), Griliches (1961), Cowling & Cubbin (1972) and van Dalen & Bode (2004) belong to the most cited ones. The second group of researchers concentrates on the behaviour of the customers on the car market

[1]The older studies often evaluate the effect of attributes that are a part of the basic equipment of all cars nowadays (for instance heater or power steering).

in the consequence of their trust. To the most interesting studies belongs the research by Andrews & Benzing (2007) explaining the decision-making process of buyers in auctions using eBay[2] dataset. The last and the biggest part of researches is oriented entirely to the explanation of vehicle value by the variation in measurable qualities. These researches differ in used methodology and observed car characteristics significantly.

To the most interesting studies dealing with a hedonic price belongs the research by Kihm & Vance (2016) focusing on the German market using data from Allgemeiner Deutscher Automobil-Club (the automobile association) and the paper by Pal *et al.* (2018) who have analysed the German eBay dataset. The German used car market was compared with the French, Spain and British markets by Prado (2009) who concentrated solely on the leasing vehicles. The relevance of car characteristics for a new and a used vehicle separately was analysed in detail by Ohta & Griliches (1976) using US market data. The US market dataset was used also in the research by Yerger (1996) focusing on the effect of purchase recommendation published in the Consumer Report. A comprehensive overview of the US studies is a part a meta-analysis published by Environmental Protection Agency (2018) including 52 relevant papers. The Turkish second-hand car market was an object of the analysis by Erdem & Şentürk (2009) using innovative purchase attributes. Requena-Silvente & Walker (2006) described a situation on the UK market using a large number of (mainly binary) characteristics and introduced an alternative view on the manufacturers (using origin countries effect or concern affiliation). Methodologically innovative research was performed by Gössling & Metzler (2017) that utilize data from Shanghai and by Wu *et al.* (2009) focusing on the Taiwan market.

5.2 Data Used for Analysis

The missing data problem and appropriateness of measure of vehicle attributes belong to the most common problems in the concerning literature (Environmental Protection Agency 2018). Besides others, many used-car market researches suffer from a limited number of observations in the datasets. A very courageous research of small size used car market was performed by Pudaruth (2014) that endeavour to perform a model with a substantially limited number of observations (comprised of only 97 vehicles) of the Mauritian used car market. The

[2]The famous American e-commerce provider.

majority of literature strives to the same problem, refer to Andrews & Benzing (2007) observing 600 cars in auctions, Erdem & Şentürk (2009) using 1074 observation and the analysis with 3046 vehicles available performed by Ohta & Griliches (1976). On the contrary, a decent number of observed vehicles can be found for instance in analyses by Requena-Silvente & Walker (2006) including almost 12 thousand cars, by Pal *et al.* (2018) including 37 thousand of observations and by Kihm & Vance (2016) having 371 thousand of vehicles in the dataset (combining new and used cars). The final dataset used for the research in this thesis contains 50, 999 unique observations extracted from the AutoScout24.de (2019) server.

The meta-analysis presented by Environmental Protection Agency (2018) emphasize also the problem of multicollinearity, correlation and omitted variables problem in the current literature. In general, it is indeed treacherous to assess whether an enlarged amount of included features is desired. Environmental Protection Agency (2018) warns against a large set of explanatory variables that increase the likelihood of model overfitting. For this reason, the research presented in this thesis eliminates redundant variables using BMA approach, as described in detail in the Chapter 7.

5.3 Methodological Approach

To shed light upon the methodology, the inspected papers used quite all expected types of models (linear regression along with machine learning techniques). Amazing diversity in modeling is a valuable source of inspiration. To support advanced methodologies, Goodman (1998) demonstrates that easy methods are not sufficient for this type of task. To follow previously described researches, Pudaruth (2014) was betting on comparison of linear regression, k-nearest neighbour, Naive Bayes and Decision Trees, Pal *et al.* (2018) showed decent classification accuracy of Random Forest (comparing to linear regression) and Gössling & Metzler (2017) proved the superiority of adaptive neuro-fuzzy inference system in a comparison with Back Propagation Neural Network. On the contrary, the majority of papers use Least Squares approaches.

Although this topic requires a substantial amount of explanatory variables and the problem of model uncertainty is therefore very pertinent for this subject, to the best knowledge of the author of this thesis, the BMA approach has never been used in any used-car-oriented researches yet. This thesis aims to

extend the existing literature with the demonstration of the appropriateness of an application of BMA model selection.

5.4 Results

Although the methodological approaches vary significantly, all described researches show almost the same and expected conclusions about basic car characteristics. In the vast majority of analyses, the most important variable negatively influencing the residual price of an automobile is vehicle age (as discussed in detail for instance by Kihm & Vance 2016). Moreover, also mileage and type of body are significant drivers of a used car price.

More interesting conclusions related to this topic were derived for instance by Yerger (1996) concluding that the consumer report recommendation has a very limited impact on the purchasing decision of a customer or by Erdem & Şentürk (2009) that successfully proved that cars sold in a highly urban area of Istanbul have a lower value for customers in comparison to other Turkish regions. Other research proposing valuable results is the analysis by Kihm & Vance (2016) proving that a speed characteristic and engine capacity is reflected on a used car value and a new car price equally. Contrarily, a fuel consumption value is more important for a second-hand buyer and therefore a consumer buying a new car with an intention to resale it in the future should focus on cars with economical working. Furthermore, in the research presented by Ohta & Griliches (1976) qualitatively comparable new and used cars appears to be perfect substitutes (however, this research produced primarily non-significant estimators of car characteristics or estimators with the opposite than expected sign). The study analysing the Chinese automobile market in a wider context by Gössling & Metzler (2017) came to the conclusion that the highest depreciation rate was detected for imported cars of price exceeding 350,000 CNY (the highest price level in the performed analysis). Furthermore, Gössling & Metzler (2017) proved that the best preservation of price is connected with normal use of a vehicle since overused cars are considered low remaining operating time and underused cars appear to be incredible for buyers. When considering studies focusing on the market agents' behaviour in the response to a trust, important factors for an executed price on the market are a buyer's reputation, a visual aspect of an advertisement with an overall car presentation and timing of an auction. Andrews & Benzing (2007) came to the conclusion that a good reputation has verifiable a positive impact on the probability of selling. But

what can be assessed as an unexpected result, is an effect of vehicles' pictures in an online advertisement. A large number of photos can probably show more defects or make a buyer uncertain. Therefore a number of pictures negatively influences a price premium of a buyer in auctions. On the other hand, good timing of a publication has a positive financial effect for a buyer (Andrews & Benzing 2007). The majority of researches focusing on the allowance for quality improvement for the presentation of consumer price indexes came to the clear conclusion that a given allowance is not sufficient.

Chapter 6

Car Price Attributes and Data Description

6.1 AutoScout24: E-commerce Platform

"The largest online car market in Europe" - this is the popular slogan identifying the online platform AutoScout24 offering used and new cars, motorcycles, commercial vehicles, other means of transport (for instance boats) and automotive components. The German company AutoScout24 GmbH, established in 1998, achieved its prominence thanks to its offers exceeding 2 million vehicles associated with complete consulting, insurance and financial services. According to official information published on the AutoScout24 websites, the business is present in 18 countries in Europe (Germany, France, Spain, Netherlands, Belgium, Austria, Luxembourg, Hungary, the Czech Republic, Croatia, Poland, Romania, Sweden, Turkey, Russia and Ukraine) (AutoScout24.de 2019).

The impressive amount of active advertisement and a phenomenal number of visitors per month that fluctuates around 10 million potential buyers were the main reasons for the analysis of the data of the platform in this thesis. What should be appraised is a very comprehensive and structured characteristic of each vehicle on the AutoScout24 platform. Regrettably, during the process of formation of this thesis, an original rating of the appropriateness of an advertised price comparing to a price of similar vehicles was terminated and to date, it is not presented on the platform AutoScout24. The rating evinced the potential to mitigate a buyer's risk associated with accurate price recognition. However, a description of the used methodology was missing.

6.2 Data Collection

The dataset used for the analysis presented in this thesis was collected using a web scrape extraction (created with the use of the programming language Python). The process used all active advertisements in Germany published on the AutoScout24.com websites in February 2020. The web scraping was restricted only for the most frequently traded car brands (the threshold was set to select only brands containing more than $5,000$ cars in the whole AutoScout24.com database). For the purposes of this thesis, vehicles with a price not exceeding $50,000$ euro were taken into consideration.

Since the platform displays only a maximum amount of 400 advertisements constantly (regardless to the exact number of advertisement corresponding to a setting of a filter)[1], the web scraping method was designed to utilize a very detailed changing filter in order to detect all published advertisements. Consequently, the random selection of 10% of the advertisements was set. The final dataset generated by the website extraction amounts to $70,975$ unique vehicles. After the consequent data clearing process, the dataset used in the analysis was reduced to almost 51 thousand vehicles.

6.3 Price Attributes Discussion

A pre-owned car is a specific object on the used market. A very complex and technologically advanced character with a specific use, maintenance demands, long life character and relatively high initial price differentiates vehicles from other sorts of goods on the market (Gongqi *et al.* 2011). Moreover, significant heterogeneity of goods quality on the second-hand car market can be observed (Kim 1985).

To describe the characteristics of a vehicle, the majority of researchers distinguish between **basic features** and **additional features** describing the quality of vehicles. When talking about the **basic characteristics** of an automobile, the attention is paid primarily on the production car qualities and its age. Among the **additional characteristics** of a car belong mainly attributes describing minor differences between vehicles in the market (car accessories,

[1]The question that should be asked is whether AutoScout24 displays the limited number of advertisements satisfying a user's filter randomly or if the platform favours some vehicles and utilizes information of its customers. For this reason, all advertisements were detected by web scraping to avoid potential data misleading presentation.

types of equipment and a previous use). Please refer to the Table 6.1 to see a few examples illustrating differences of these categories.

Following another perspective on cars' attributes, a vehicle can be described by **physical characteristics** and **performance characteristics**. Please refer to Table 6.2 for a few examples. Cowling & Cubbin (1972) categorize **physical attributes** to characteristics that are directly measurable and those that evince a relationship with other variables that can be explicitly measured. According to the research presented by Ohta & Griliches (1976), the set of **physical variables** plays more important role in evaluation of vehicle value than **performance attributes**.

This following discussion aims to cast light on the characteristics collected from AutoScout24 platform in the context of the German used car market, as well as in the context of the relevant current literature. Please refer to the Table 6.3 representing the overview of the car attributes used in the analysis. Note that the correlation between explanatory variables (and potential multicollinearity) was controlled.

Table 6.1: Basic and Additional Car Characteristics - Examples

Basic car characteristics
Make
Model
Age
Mileage
Body type
Fuel type
Fuel consumption
Engine power
Engine capacity
Speed
Transmission

Additional Car Characteristics
Submodel
Driven axle
Brakes system
Country of the first registration
Warranty length
Service history
Number of doors
Air condition
GPS
Body colour
Status of the previous owner

Table 6.2: Physical and Performance Car Characteristics - Examples

Physical car characteristics
Engine power
Number of cylinders
Weight

Performance car characteristics
Maximal speed
Acceleration
Fuel economy

Table 6.3: Hedonic Price Attributes in Analyzed Sample

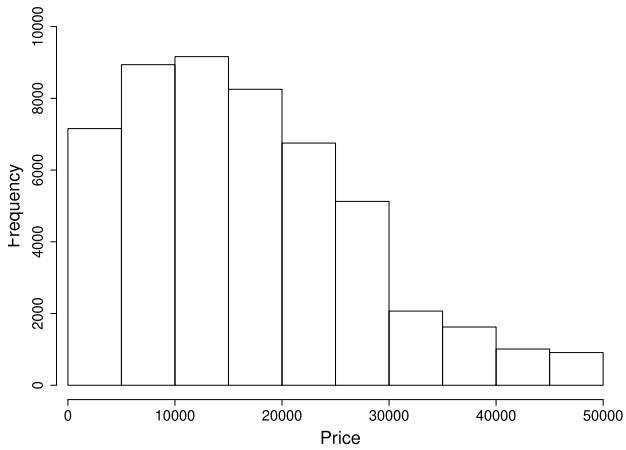
Car attributes	**Units/Categories/Values**
Age	months
Mileage	thousands km
Fuel type	Gasoline, Diesel
Costs on consumption	euro/100 km
Body type	Sedan, Station wagon, Off-road, Large-size, Compact, Small-size sports
Make	Alfa, Audi, BMW, Citroen, Dacia, Fiat, Ford, Honda, Hyundai, Jaguar, Jeep, Kia, Lancia, Land Rover, Mazda, Mercedes Benz, Mini, Mitsubishi, Nissan, Opel, Peugeot, Renault, Seat, Skoda, Smart, Suzuki, Toyota, Volvo, Volkswagen
Engine power	kW
Displacement	litre
Transmission type	Manual, Automatic
Status of the car	Used, Employee's, Demonstration
Selling channel	Commercial dealer, Private seller
Geographical location	PC0, PC1, PC2, PC3, PC4, PC5 PC6, PC7, PC8, PC9
Number of doors	Three-door version, Five-door version
Number of seats	Numerical values
Service history	1/0
Bodywork colour	Standard colours, Non-standard colours
Smoking vehicle	1/0

6.3.1 Price

The nature of the platform AutoScout24 does not allow tracking the formation of price on the ground of demand and supply as analyses of the auctions for instance performed by Andrews & Benzing (2007). The advertised price (ask price) is therefore assumed to be a final purchase price (a potential price negotiation between market participants or other price changes are disregarded). Nor the monitoring of the time between the publishing of an advertisement and trade is possible. A necessary assumption expects that a seller forms an advertised price on the ground of the relevant set of vehicle characteristics, sufficient knowledge of the market and optimal pricing. On the ground of these simplifications, the market is assumed to be cleared in a short period of time.

For the purposes of analysis, advertisements with a maximal price of 50,000 euro were selected. This approach prevents the presence of considerably luxurious vehicles that are unique and have often very specific characteristics. Similarly, the unusual low priced vehicles are not taken into consideration since those advertisements could contain immobile vehicles or can inspire customer's uncertainty and different attributes valuation (for instance, a customer can perceive a low odometer state of these vehicles as a potential fraud not as a positive characteristic). The final dataset omitted vehicles with a price lower than 990 euro.[2]

Figure 6.1: Histogram of Prices



[2]The original target was set to 1,000 euro, but taken into consideration a pricing policy of the majority of agents that advertise for not rounded prices, was adjusted to 990 euro.

When evaluating the current literature in the context of a vehicle price, an interesting perspective proposed by Andrews & Benzing (2007) that assembled a market value of a car by the Kelly Blue Book and a compared ask price in auction sales. The most commonly asked question in the concerning literature belong the form of the dependent variable. To draw inspiration, let us refer to Triplett (1969) or Yerger (1996) who are convincing a reader of a logarithmic form of price in models. On the basis of these papers, observed skewed distribution and discussion in Chapter 4, the dependent variable of the model analyzed in this thesis is in the logarithmic form.

6.3.2 Age

Vehicle age is defined using the date of the first registration. For the purposes of the analysis in this thesis, the base month February 2020 was used for the calculation. The value of age is stated in month units (similarly as in Prado 2009).

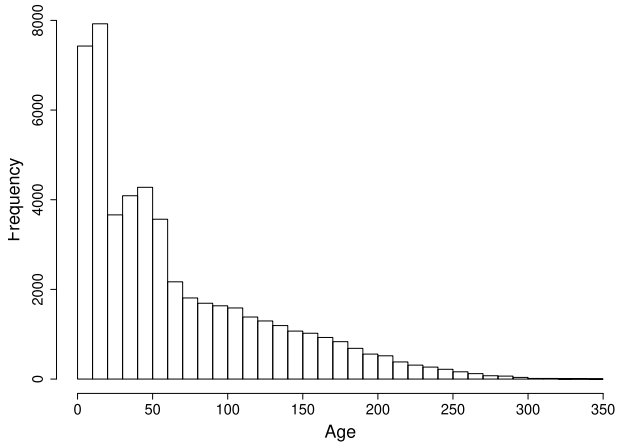
To follow the current knowledge in the literature, the question about the precise definition of the duration of a car life cycle is very often asked. In general, vehicles exceeding 15 years in operation are considered to reach the end of their life cycle. For instance, Pal *et al.* (2018) analyzing dataset from the German used car online platform eBay Kleinanzeige states that the lifespan of the vehicle varies from 5 to 15 years. A similar conclusion was drawn by Gongqi *et al.* (2011) focusing on the Shanghai used vehicle market. On the contrary, more than 10% of cars advertised via AutoScout24 platform with a price exceeding 4,000 euro are cars older than 15 years (classic cars excluded).[3] It could indicates that a lifespan of vehicles can differ (at least for some makes and models) from the universally considered upper limit and these cars have still a higher than a "scrap value". To take a look at the European market in numbers, the average age of the registered passenger cars varies between 8 to 17 years (classic cars included) in the individual countries in the EU. The lowest age of cars in use can be observed in the United Kingdom, Austria or Switzerland. Germany occupies the fifth upper position with 9.5 years old vehicles on average (classic cars included) (ACEA.be 2019).

In comparison, the diagram Figure 6.2 represents the frequency of the advertised cars via AutoScout24 according to its age. It can be summarized that traded cars on this platform have a predominantly relatively low working age.

[3]This information was derived from all advertisements in Germany in June 2020.

Approximately one-half of all cars are traded before they reach the age of 4 years (48 months). This trend can be caused by the sale of the leasing cars after the end of the leasing period, by sales of companies' cars after the end of the tax write-off period or potentially it can be speculated about an effect of individuals performing a nonofficial automobile resale business. The average vehicle offered via AutoScout24 platform has 5.5 years (66 months).

Figure 6.2: Histogram of Age in Months Units



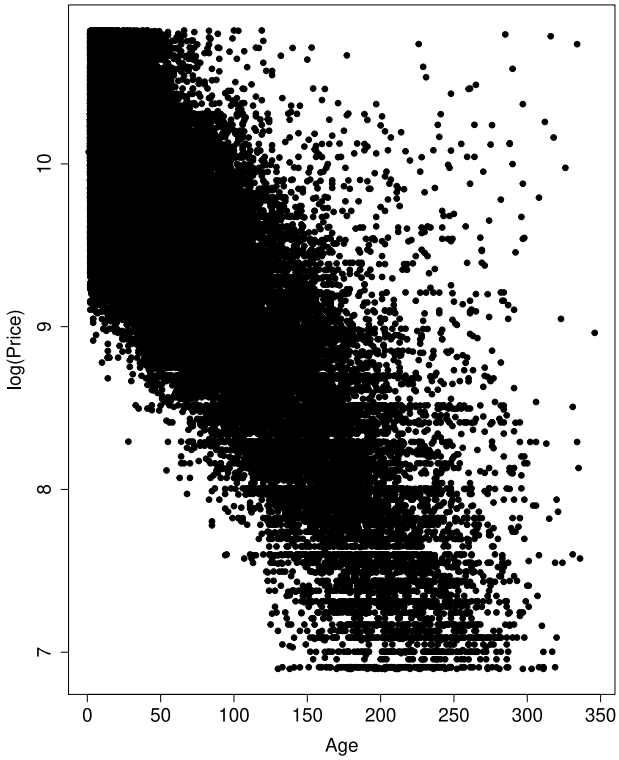
When analysing the current literature concerning to pre-owned car market, it can be concluded that a collection of vehicles to a dataset according to the age is miscellaneous. For instance, Bond (1982) uses for the research only vehicles manufactured between 1972 and 1976, Prado (2009) takes into consideration vehicles of age 1-10 years and Kihm & Vance (2016) excludes all cars older than 20 years and younger than 4 years from the dataset. Ohta & Griliches (1976), using vehicles up to 6 years in operation, draw a conclusion that older cars do not correspond to the usually traded goods on the organized used car market. On the ground of the previously described conclusions of the content of the AutoScout24 database, the analysis presented in this thesis takes into consideration cars older than 5 months up to age 30 years.

According to the German legislation, vehicles exceeding the age of 30 can be classified as classic cars (Gössling & Metzler 2017). The author of this thesis expects that the main aspects influencing a price of classic cars are sentiment or collectible value rather than physical or performance characteristics (for instance power, fuel consumption, maximal speed, comfort or safety). Moreover, the classic cars constitute a very heterogeneous group of goods with a very

individual price formation. For this reason, the hedonic price analysis of the classic cars could differ from the hedonic price of the rest of the used vehicles significantly. Therefor classic cars should be removed from the studied dataset.

Vehicle age as an explanatory variable is also discussed in the context of an appropriate functional form. On one hand, Gongqi *et al.* (2011) state that the age characteristics cannot be inserted in the linear form to the model since the depreciation ratio of a vehicle is unproportionally higher in the first years. Ohta & Griliches (1976) proves the hypothesis of the geometric depreciation of passenger vehicles and Prado (2009) uses a logarithmic form of age to model the depreciation effect. On the other hand, to the best knowledge of the author of this thesis, the majority of related literature inserts the age characteristic in the linear form. The scatter plot Figure 6.3, based on the AutoScout24 dataset, visualizes that the preferred functional form describing the relationship between a vehicle logarithmic price and age is also linear.

Figure 6.3: Scatter Plot of Vehicle Price and Age Characteristics



According to the economic theory, the age characteristic plays a role as an indicator of unobservable decreasing quality that reduces a vehicle market price (Bond 1982). For instance, the research performed by Kihm & Vance (2016) concluded that one additional year would decrease vehicle price by 7.4% *ceteris paribus*. In the research presented in this thesis, age characteristic is expected to be the main factor negatively impacting vehicles' values as a determinant of lowering of quality and reduction of remaining lifespan as well as an indicator of potential higher maintenance costs.

6.3.3 Mileage

The alternative view on a car age is a mileage perspective reporting a number of kilometres driven as recorded by an odometer. Some analysis such as the research performed by Prado (2009) constructs a variable of distance driven by month (by a proportion of the odometer state and age). This quantity signals the intensity of the car use and could indicate details about the owner or a use of a car (for instance, whether a car was used for business activities such as taxi service etc.).

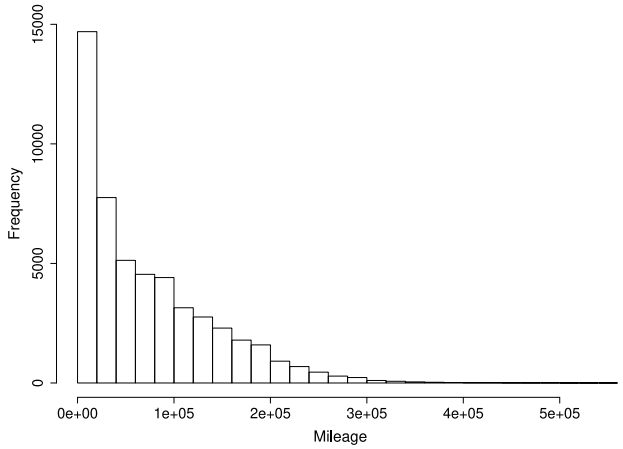
Taking the analysis by Prado (2009) as an inspiration, the proportion of mileage and age was used to eliminate advertisements with unrealistic values in the AutoScout24 dataset. The target for the plausible mileage state was set to be above 1,000 km per annum. The cleaning process was applied also for a maximum reasonable driven distance with a target of 200,000 km/year (setting a maximal total target of driven kilometers given by odometer was not necessary)[4]. For the analysis presented in this thesis, the value of mileage is stated in a unit of thousands of kilometres.

Gössling & Metzler (2017) state that the average German car increases its odometer status by 14,250 km annually. By inspection of an average vehicle advertised via AutoScout24.com that has driven $75,645$ km in total and its annual year odometer state increases by $14,032$ km annually on average (computed by adjusting the monthly odometer state increase), the corresponding results can be observed. The Figure 6.4 indicates that the vast majority of cars traded via AutoScout24 have driven less than 100 thousand km, moreover a share of

[4]Enormously used vehicles (for instance cars used for business conduction) were taken into account. This estimate was calculated for a car driven for 8 hours with an average speed 103 km/hour for 20 working days in a month. The average speed is an estimate for a trajectory with 50% time driven with an average speed 130 km/hour, 25% of time driven with 100 km/hour and 25% of time driven with 50 km/hour simulating an operation on highway roads, in non-urban areas and in cities.

vehicles with a relatively very low mileage state is noticeable. Approximately one-third of cars have driven less than 10 thousand km in general.

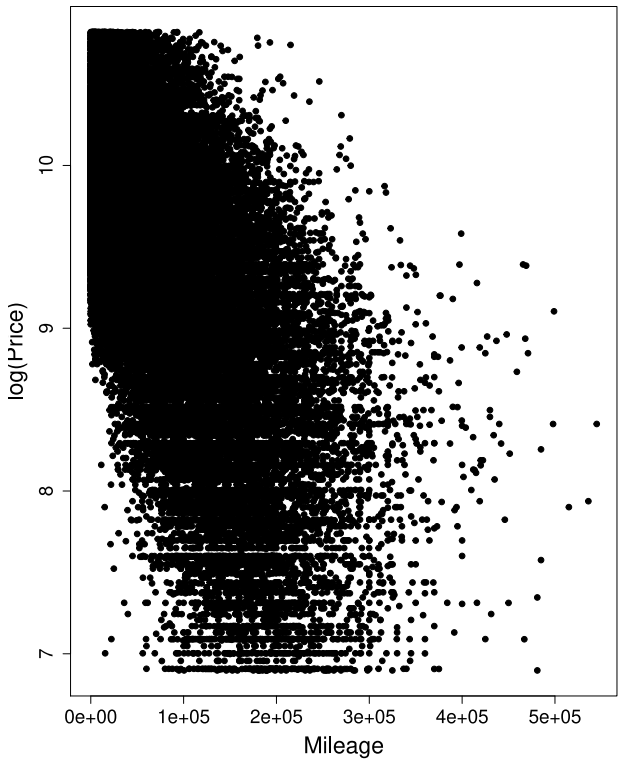
Figure 6.4: Histogram of Mileage



Similarly, as discussed for the age parameter, also the effect of the mileage characteristic on price is considered (by some researches) to have a non-linear nature. For instance Prado (2009) uses a logarithmic form of driven kilometres in the analysis. The recent analysis includes mileage attribute in linear form as follows from the scatter plot in Figure 6.5.

The expected effect of mileage to a vehicle market value is negative since it signals a decrease in the remaining lifetime of a vehicle. Pal *et al.* (2018) concludes that the number of kilometres driven by a car since the first registration is the most important factor influencing a selling price on the German market. Another German oriented study by Kihm & Vance (2016) ascertained that an increase in mileage has a negative and significant impact on price. The market value of a car decreases by 0.3% with every additional 1,000 km recorded by odometer considering other effects being fixed. A study by Gongqi *et al.* (2011), oriented for the differences in operating frequency, concludes that the overused and underused vehicles evince worse market value in comparison to those that are normally used.

Figure 6.5: Scatter Plot of Vehicle Price and Mileage Characteristics



6.3.4 Warranty Period

A manufacturer's warranty is an important characteristic eliminating buyer's uncertainty since lower maintenance costs related to the car being under a warranty are expected for an owner. The warranty is directly conditional on vehicle age and mileage characteristics and varies according to different conditions of manufacturers and different countries. The current literature holds very different attitudes to the question of whether it is necessary to include warranty attributed as independent variable in the hedonic price models. In the case of AutoScout24 dataset, missing warranty specification and often an incorrectly filled number of months of warranty was solved by omitting this variable from the dataset. The variant of manual completing of warranty information was rejected since this information should be derived separately for every manufacturer, every make and the country of origin of vehicle should be taken in consideration as well. The heterogeneity of the manufacturers' conditions for warranty and a high number of different brands in the dataset are the main barriers. Andrews & Benzing (2007) claim that warranty information has no additional value since it is already included in age and mileage char-

acteristics. To draw inspiration from Andrews & Benzing (2007), the research presented in this thesis assumes that buyers have an elementary knowledge about the warranty conditions of given make they wish to purchase and therefore the derivation of the desired information from the basic car characteristics is possible.

6.3.5 Fuel Type and Costs on Consumption

Nowadays, apart from traditional fuel power (such as fossil fuels), the automotive industry offers also cars fully powered by electric batteries or equipped with a very unconventional power. However, diesel and gasoline still remain the most frequently used fuels for passengers' vehicles. Undoubtedly, the high proportion of these alternatives is supported by fuel accessibility and the relatively lower initial price of a vehicle. On the contrary, electric-powered vehicles are perceived as goods with a high initial price and relatively low working costs. When analyzing the environmental benefits of electric cars (for instance a benefit of a lower amount of emissions produced in urban areas directly), the attention should be paid to the difference between solely electric cars and cars combining electric power with diesel or gasoline (well known as hybrid-powered vehicles). It is interesting to mention that hybrid fueled vehicles are expected to be more resistant to depreciation than cars solely fueled by fossil fuels (Pudaruth 2014). The terms liquefied petroleum gas (LPG) and compressed natural gas (CNG) denote a gas-based fuel of fossil origin. Their limits such as restricted entry to some car parks, storing the gas tank in the luggage compartment and limited replenish stations are the main reasons for a minority share on the automobile market. Nowadays, engineers entertain also with an idea of powering for instance by hydrogen, bioethanol[5] or biodiesel[6]. But the share of such powered vehicles on the market is negligible.

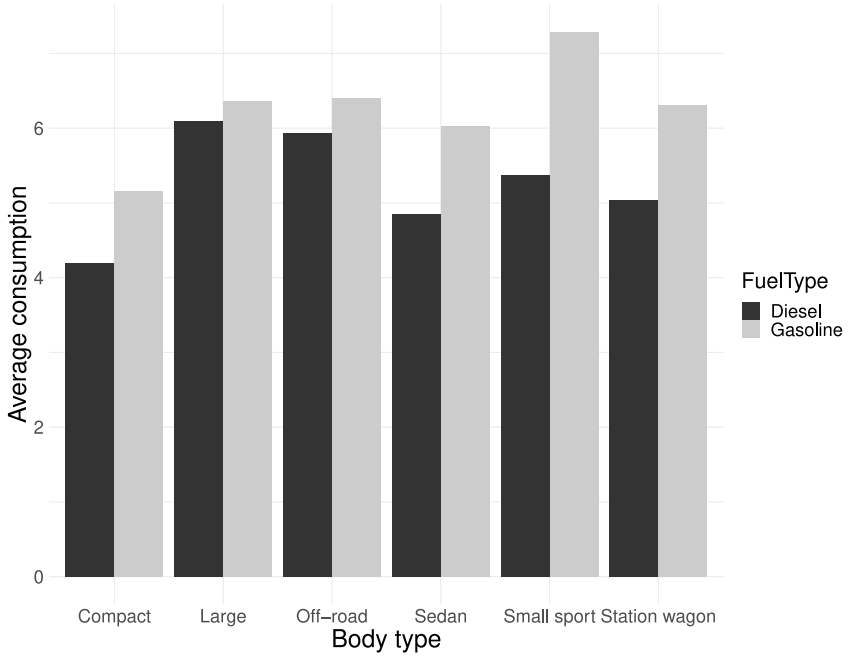
In Germany the majority of registered cars (54%) uses gasoline fuel, 42% of owners tank up diesel and the rest of vehicles are designed for electric (0.3%), gas (2.8%) or any rest of alternative powers (ACEA.be 2019). On the ground of the negligible share of alternatively fueled cars in the dataset of AutoScout24, the analysis presented in this thesis was performed solely for cars using diesel and gasoline for their operation.

[5]Bioethanol is a product of fermentation of biomass containing a high proportion of carbohydrates (Novosád 2009).

[6]Biodiesel is a diesel fuel blended with products of plants oils (Novosád 2009).

When comparing the gasoline-powered and diesel-powered cars, four aspects/stereotypes are often discussed by customers: a lower price of diesel fuel, a lower average fuel consumption of diesel engines comparing to the gasoline oriented ones and a longer lifespan of diesel vehicles and a higher initial value of diesel-powered vehicles. By analyzing cars on AutoScout24 platform, average fuel consumption reaches 5.4 l/100 km for diesel vehicles and the mean value of gasoline is by 0.67 l per 100 km higher. However, an important fact is often missed in the comparison of differently fueled vehicles. Diesel and gasoline engines cannot be seen as perfect substitutes on the market. Whereas diesel engine is more appropriate for larger vehicles and more frequently used vehicles (for instance cars for conduction of business), gasoline engines are more convenient for private cars or for vehicles with high-powerful engines (for instance sports cars). The related literature often emphasizes the correlation between fuel consumption and body type of a vehicle. A mean consumption is expected to be highest for larger vehicles and cars design for "drive enjoyment" often called sports cars. The Figure 6.6 summarizes, using a mean consumed value, that the most fuel-inefficient body types are small-size sports cars, off-road and large-size vehicles.

Figure 6.6: Comparison of Average Fuel Consumption in Litres According to a Body Type



The effect of fuel type on price is expected to be highly significant and of high magnitude. Vehicles equipped with diesel engines are expected to be of significantly higher value (similarly as in Prado 2009, observing the German eBay database). The higher price of new diesel-powered vehicles (reflecting technical requirements connected to a production of diesel engines) is expected to be transformed to a higher market value of used diesel vehicles on the used car market.

The server AutoScout24 displays a consumption value for "city", "country" and "combined" traffic as measured by manufacturers. Facing the same issue, Yerger (1996) decided to use the "city" value, whereas this analysis made a conclusion to consider "combined" fuel consumption to achieve a more realistic value of a car operation.[7] On the ground of the previously discussed problem with a general comparison of fuel consumption, the author of this thesis uses an artificially made variable: *costs on consumption*. The value of costs on consumption was calculated by a "combined" fuel consumption and price of diesel/gasoline fuel in February 2020 using data from Statista GmbH (2020). The quantity is presented in units euro/100 km.

In general, the price of a car is expected to be negatively influenced by fuel consumption from two different perspectives: costs on transportation and environmental burdens. The analysis presented by Kihm & Vance (2016) shows that the reduction by 1 liter of fuel per 100 km would increase a vehicle price by 813 euro *ceteris paribus*. This research also proved that new cars' prices are less elastic to changes in fuel consumption that the prices of used cars. It implies that a typical customer on the second-hand market appreciates the lower consumption costs (or lower environmental burdens) more than a buyer on the new car market. However, Environmental Protection Agency (2018) concludes that the feature of fuel savings is undervalued in the consumers purchasing decision. Other opposing view rebutting the conservatively expected negative relationship between vehicle price and fuel consumption was described in Chapter 3. The result could rely also on the customers' trust in consumption values measured under laboratory conditions.

[7] In this context, it is necessary to remind that the reported values of fuel consumption are computed under laboratory conditions by manufacturers and cannot be viewed as perfectly realistic for the regular operation.

6.3.6 Body Type and Market Segmentation

One of the factors that directly predetermine the character and use of the car is its body. The author of this thesis possess the view of Requena-Silvente & Walker (2006) that states that the body attribute is also a segmentation division of the car market. It can be distinguished between micro-size, sub-compact, compact, mid-size, large-family, large cars, compact SUV, sport, mini multi-purpose vehicle (MPV), large MPV, mid-size, mobile home and full-size SUV (Gössling & Metzler 2017). In the AutoScout24 dataset, nine different body types can be selected by an originator of an advertisement. It can be distinguished between compact, convertible, coupe, off-road/pick-up, sedan[8], station wagon, transporter and van body type. In addition, a seller can choose the denomination "other body type" for a vehicle. In this group, it is possible to find advertisements for motorcycles, trucks, boats and spare parts for vehicles. For the purposes of the analysis of cars, the "other body type" was omitted from the dataset. According to a small number of observations of coupe and convertible were observed together as sports small-size vehicles, transporter with vans are merged into a group of large-size vehicles. The research presented in this thesis uses therefore 6 different body categories. A sedan body shape is used as a comparative base (because of its high frequency in the dataset). The Figure 6.7 symbolizes the vehicle categories advertised via the platform AutoScout24. The dataset used for hedonic price analysis evinces that the most frequently traded car are sedans (almost 30% of the advertisements), followed by station wagons (with 23% of cars). To confirm the concerns expressed by environmentalists, off-road and pick-up cars (with high fuel consumption and high emission level) represent 20% of all advertised vehicles.

The current literature varies in a selection of observations according to body characteristics. Some researches do not even discuss the body selection, a certain group of researches study as many body styles as possible and some researchers consider only a few body types. For instance Triplett (1969) selected only sedans to simplify analysis. A similar example can be observed in Cowling & Cubbin (1972) or Ohta & Griliches (1976). Bond (1982), testing the hypothesis of the ability of "lemons" to drive out higher quality products out of the market, used only data about pick-up trucks.

To assess the overall effect of a body shape, Pal *et al.* (2018) showed that body type is one of the most important factors influencing a buyer's purchas-

[8]Sedan body type is also known as a saloon car in the British terminology.

Figure 6.7: Body Type Categories Advertised on AutoScout24 Platform

Source: AutoScout24.de (2019).

ing decision.[9] The analysis presented by Kihm & Vance (2016), oriented for comparison of effects of car characteristics on a price of new and used cars, showed that vehicle body is a more crucial characteristic for a buyer intending to purchase a pre-owned car than a customer desiring to purchase a new car. In the present study, the body type characteristics are expected to be highly significant. It is expected that *ceteris paribus* body types of off-road and sport small-size vehicles should be higher priced, similarly as on the new car market, comparing a base category sedan, since these shapes are constructed to attract moneyed customers. For the construction costs, the large-size vehicles are expected to be more expensive than sedan and small-size bodywork defined as low-cost market segment orientation (compact) is expected to be cheaper than sedan.

6.3.7 Make, Brand and Model

As a consequence of the active export policy of the manufacturers across the world and globalization, the German new and used car market offers a really huge amount of different automobile brands. However, the vehicles constructed

[9]The dataset used in Pal *et al.* (2018) contains not only passenger cars and therefore a huge effect was expected (for instance there is a substantial difference between the market segment for sports vehicles and buses). The role of the body in the AutoScout24 dataset presented in this thesis is expected to be slightly different from the research performed by Pal *et al.* (2018).

by manufacturers originating in Germany occupy the first positions on the domestic market permanently. This trend can be observed in the database filter of AutoScout24, containing more than 180 brands, as well. The highest proportion of advertisements is in the possession of Volkswagen constantly.

The current knowledge about the car make in the context of vehicle price is really broad. For instance, Pal *et al.* (2018) argues that the make and brand characteristics are the second most important explanatory variables affecting a market price. Gongqi *et al.* (2011), studying the Shanghai market, demonstrates that the imported vehicles loose their market value at the highest rate than domestically produced cars. Yerger (1996) came to the conclusion that the impact of maintenance on vehicle price is different for resale of a domestic and an imported vehicle. Kihm & Vance (2016) argue that luxury car brands tend to suffer from smaller conservation of value and therefore depreciate faster (the research was performed for selected models of BMW and Mercedes Benz). However, the number of studied makes and models is often very limited in some researches. For instance Prado (2009) uses only 2 manufacturers in his analysis. Cowling & Cubbin (1972) took into consideration solely British makes and Volkswagen. The slightly bigger number of makes was considered by Ohta & Griliches (1976) using 11 makes and 52 models. And the most courageous approach is described in van Dalen & Bode (2004) that analyzes approximately 11 thousand different car models. For the purposes of the analysis presented in this thesis, the selection process of advertisements was based solely on makes that contain more than 5,000 cars in the whole database of the platform AutoScout24 (Alfa, Audi, BMW, Citroen, Dacia, Fiat, Ford, Honda, Hyundai, Jaguar, Jeep, Kia, Lancia, Land Rover, Mazda, Mercedes Benz, Mini, Mitsubishi, Nissan, Opel, Peugeot, Renault, Seat, Skoda, Smart, Suzuki, Toyota, Volkswagen, Volvo).[10] Although an automobile aficionado considers makes in the context of the concern group, the current literature asses only the importance of brand characteristics and neglects the role of concern image. Researches based on the origin countries of the manufacturers are also very rare. The overview of the selected manufacturers for the research is summarized together with information about their origin and their current corporate affiliation in the table Table 6.4. Volkswagen (as the most frequent observation) is set as a base category in the analysis presented in this thesis.

The variable describing car make is expected to affect a hedonic price via its brand equity. A customer can perceive a brand as a guarantee of quality,

[10]It was controlled in December 2019 when the web scraping code was created.

Table 6.4: The Overview of Brands According to the Concern Affiliation and Country of Origin

Concern name	**Brand name**	**Origin country**
BMW group	BMW	Germany
	Mini	UK
Daimler AG	Smart	UK
	Mercedes Benz	Germany
Fiat Chrysler Automobiles	Jeep	USA
	Fiat	Italy
	Alfa Romeo	Italy
	Lancia	Italy
Ford Motor Co.	Ford	USA
Groupe PSA	Peugeot	France
	Dacia	Romania
	Opel	neovereno
	Citroen	France
Honda Motor Co.	Honda	Japan
Hyundai Motor Group	Hyundai	Korea
	Kia	Korea
Mazda Motor Corp.	Mazda	Japan
Renault-Nissan-Mitsubishi Alliance	Nissan	Japan
	Mitsubishi	Japan
	Renault	France
Suzuki Motor Corporation	Suzuki	Japan
Tata Motors	Jaguar	UK
	Land Rover	UK
Toyota Motor Corp.	Toyota	Japan
Volkswagen group	Volkswagen	Germany
	Audi	Germany
	Skoda	Czech Republic
	Seat	Spain
Zhejiang Geely Holding Group	Volvo	Sweden

Source: (Requena-Silvente & Walker 2006), official websites of the mentioned manufacturers

future maintenance, and service costs. In the present study, luxury-oriented brand as for instance Land Rover and Jaguar are expected to evince a higher market value compared to Volkswagen. On one hand, the market price is expected to be higher for original German brands (on the ground of the German automotive image as well as on the ground of the domestic production and accessibility of spare parts and service), on the other hand, some customers are expected to be attracted by imported brands because of their rare occurrence. Prado (2009) proved that in all studied countries (Germany, France, Spain, and Great Britain), German origin of a brand increases a price of a car sold in the second-hand market. The Volkswagen's competing brands: Seat and Skoda (as described by Requena-Silvente & Walker 2006) should be therefore slightly less priced in the analysis of the AutoScout24 dataset.

The model name was not used for the analysis in this thesis since the number of unique models exceeds the value of 150 and the market segmentation effect of the model is partially reflected in the body shape. The majority of the researches do not include the model attribute to the model for similar reasons neither.

6.3.8 Engine Power

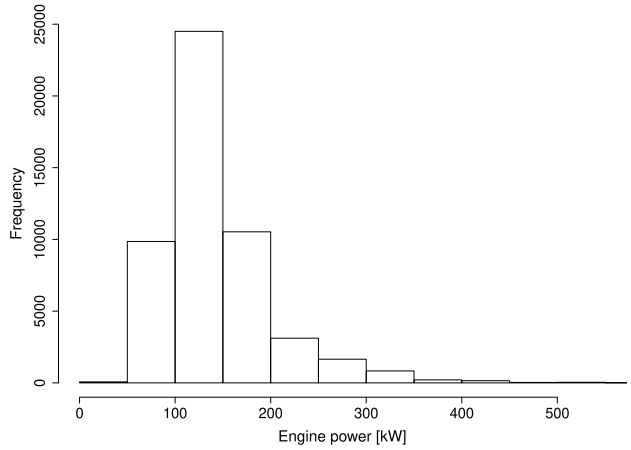
The vehicle power can be reported using horse power (hp) value or kilowatt (kW) value (for the hedonic price analysis presented in this thesis, kW units are selected). When talking about a power attribute, interesting is a relationship between engine power and displacement variable (varying among manufacturers) and determination of power by a type of fuel (the significant difference can be observed between diesel and gasoline engines). Gössling & Metzler (2017) discovered that the average registered passenger car in Germany is equipped with an engine of 78 kW in case it uses gasoline, 97 kW in case it is diesel-powered. Analyzed dataset of AutoScout24 presents slightly higher values (97 kW and 117 kW respectively). The data cleaning procedure is based on the omitting of vehicles exceeding 1,000 hp and cars with less than 27 hp.[11] The fixed conversion factor between hp and kW is 1.36. On the ground of the conversion factor, the test of the typing error of an advertiser was performed. Additional

[11]To the best of the author's knowledge, there is not a commonly traded passenger car below the value of 27 hp. Car enthusiasts would know the iconic French 27-horse powered Citroen 2 CV. To illustrate how low this value is in the context of the current automotive technology, note that for instance Ford vehicle introduced in 1901 (replacing the popular Quadricycle on the market) was equipped with 26-horsepower engine (Ford Motor Company 2020).

control of the dataset is based on the manual inspection of high values of power of brands that usually produce low-performed vehicles (Dacia, Fiat, Hyundai, Kia, Renault, Skoda, Smart, Suzuki, and Toyota). The inspiration was drawn from Pal *et al.* (2018) who used for filtering of unrealistic values a similar manual control. For the hedonic price analysis presented in this thesis, the variable in units of kW was selected. Please refer to the histogram Figure 6.8 presenting that many observations have similar power value, however the distribution is a slightly right-skewed.

Power is one of the most important factors affecting the comfort and safety of a ride and transportation ability. The majority of analysis studying the effect of engine power shows that engine power has a significantly positive effect on a vehicle price. The same result is expected in the analysis presented in this thesis.

Figure 6.8: Histogram of Power



6.3.9 Engine Capacity

Displacement (or engine capacity) characteristic measures the volume of engine cylinders in units cubic centimetres (cc) or litres (1,000 cc correspond to a volume of one liter). In the dataset analysed in this thesis, the units cc are used and all cars with an unrealistic value of displacement exceeding 10 thousand cc and below 500 cc were omitted. Although the engine capacity and power are expected to be positively correlated variables, the relationship between these varies for different manufacturers and models. The construction attitudes of

some manufacturers allow construction highly performed engines with maintaining a relatively low displacement. In the studied sample, the majority of researchers dealing with this topic holds the view that only one of these two attributes should be included in the model. On top of that, Prado (2009) explicitly suggests omitting either displacement or power in the research. However, the author of this thesis believes that the displacement variable is important to express a customer's favour to invest in lower capacity engine cars (keeping power fixed) in order to protect the environment or decrease operating costs.

6.3.10 Transmission Mechanism

A few years ago, an automatic transmission was a substantial attribute for an average American vehicle, whereas a symbol of a luxury passenger car in Europe. It represented one of the most striking differences between continental Europe and US standards. Nowadays, the technical solution of the automatic gearing mechanism becomes more popular and affordable. In the AutoScout24 dataset, the proportion between vehicles with manual and automatic transmission is 62% and 38% respectively. The manual variant is the base category in this hedonic price model.

As proved in the various previous studies (for instance Erdem & Şentürk 2009 or Yerger 1996), the market value of a car with automatic transmission is expected to be *ceteris paribus* higher than a vehicle possessing a manual transmission. An interesting question is whether the automatic transmission is still considered as a geographically differently valued attribute.

6.3.11 Status of a Previous Owner

A feature with the designation "car status" defined by AutoScout24 represents the information about the previous use of a car. A customer can distinguish between a vehicle that was used by individuals ("used car"), by companies ("employees car"), or put on display in a dealer's showroom ("demonstration car"). The overwhelming majority (of 83.4%) of observations in the studied sample belong to cars driven by private individuals, only 5.6% cars are employees vehicles and 11% has a status "demonstration car".[12]

A car used in a company by employees could be appraised as a vehicle with a lower maintenance level and therefore is expected to have a lower price in

[12]The cars from the platform AutoScout24 with the status "new car" or "preregistered car" are not part of the observed dataset.

comparison to the vehicle used by private individuals (that is considered as a base of the model). The "demonstration car" can be, in contrast to its worn in use, considered as a car with high maintenance level and could be therefore seen as the best substitute for a new car. The price a "demonstration car" is therefore expected to be higher than a price of a vehicle used by a private individual.

6.3.12 Selling Channel

As discussed in the previous sections, the used cars are offered via different channels. The majority of all advertised cars (more than 81%) are offered by commercial sellers, the rest (almost 19%) by private individuals. In the present analysis, a private seller is taken as a base category of the model. In the related literature, an effect of the selling channel is not often discussed and to the best knowledge of the author of this thesis, the factor of selling channel has not bee tested in the analysis of the German used car market yet. The majority of analysis does not generally take this factor into the consideration and some researches purposely avoid this effect by testing only cars from one selling channel. For instance, Pal *et al.* (2018) analyzes only vehicles offered by private sellers.

A selling channel feature is expected to have a contradictory effect. For a specific group of potential buyers, a commercial seller can be a guarantee of quality (for instance the correctness of an odometer status or car history). Contrarily, the trust of some customers in commercial dealers is very limited and some customers are not willing to pay commissions of the commercial seller reflected in the car price and therefore prefer privately sold vehicles. Andrews & Benzing (2007) showed that an average price premium is often higher for dealers, probably due to better selling know-how or advertising. In general, two potential results are expected, either a car sold via dealer channel is higher or the effect of the selling channel is outweigh by the e-commerce competitive environment with information accessibility and simplicity of product comparison.

6.3.13 Geographical Factor

A unique perspective can be studied using a variable describing the location of a seller. Buyers can opt between sellers in their surroundings and in other regions (where the market price of a car can be different). The region factor

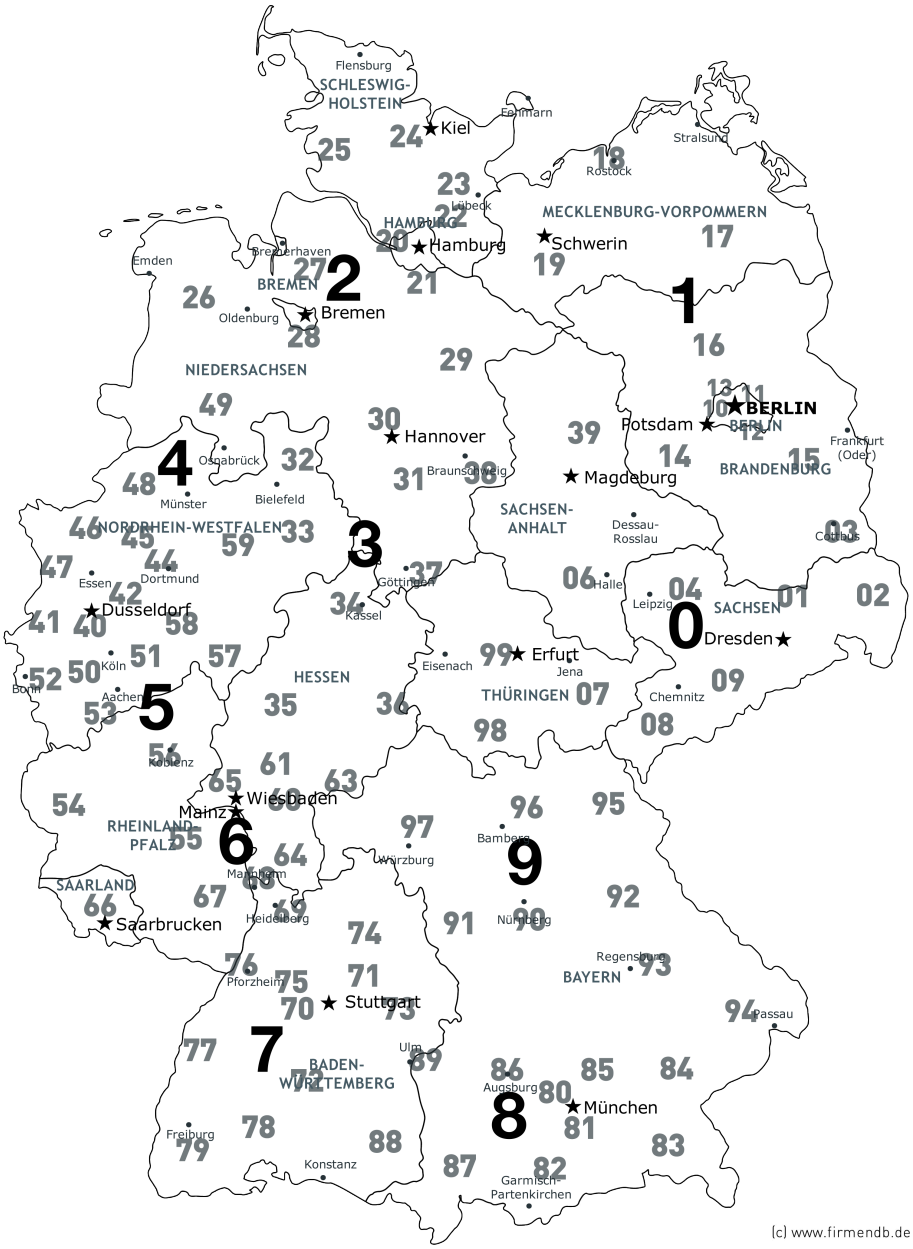
has the ability to indicate price differences among German regions and uncover an arbitrage factor.

The location effect is discussed mainly in US oriented studies. One of the specificities of the US car market is the Optimal Distribution of Auction Vehicles System determining an optimal location for vehicles' sale. Wu *et al.* (2009) point out that the location plays an important role in the realized sales of pre-owned vehicles and illustrates the statement with the US market data. The geographical effect is used also in other US-oriented study by Andrews & Benzing (2007) dividing the country to 4 regions. To the best knowledge of the author of this thesis, the analysis performed by Pal *et al.* (2018) is the only literature discussing the geographical division effect in Germany. Although the authors have the data of location (postcode data), they haven't included the effect to the model. The current literature does not offer any test of location effect on a price of used-cars on the German second-hand car market currently.

In this thesis, Germany is divided into ten regions using a postcode system that partially deflect from the federal boundaries (Germany consists of 16 federal states). Please refer to the map in the Figure 6.9 comparing the federal boundaries (represented by grey lines) and postcode regions (visualized by the corresponding numbers). For the purposes of the geographical division, the postal system appears to be a more appropriate approximation. The state boundaries demarcate unequally large territories (for instance the difference between the area of federal state Bremen, including only the city Bremen and the near surrounding, and Bavaria region, comprised from tents of thousands square kilometres, is enormous). The comparison of purely urban-based federal states (that implemented some ecological or car entry restrictions for instance) with regions including also rural areas would be misleading. Although the postal division system is more appropriate for the analysis presented in this thesis, it is necessary to mention that this system cannot divide Germany into absolutely comparable regions either.

The overall effect of the geographical characteristic of a buyer is expected to have a small or negligible impact on the hedonic price since no profitable arbitrage is expected. The hypothesis of the presence of persisting price inequality between the post-democratic and post-soviet part of Germany is not taken into consideration. The region number 4 (situated in the boundary of Lower Saxony and North Rhine-Westphalia) represents a base category in the model. The author does not expect any significant differences between the price of the cars sold in the base region and the price of the rest of vehicles.

Figure 6.9: The Difference Between the Federal State Division System and the Postcode Geographical Division System



Source: http://www.firmendb.de/.

6.3.14 Bodywork Colour

A bodywork colour appears as an additional feature in some studies. For instance, Erdem & Şentürk (2009), studying the Turkish used-car market, differentiates between black/grey coloured vehicles (the most popular locally) and the rest of colour spectrum and concludes the positive price effect of black and grey bodyworks. Draw the inspiration from Erdem & Şentürk (2009), the analysis presented in this thesis divided cars to those possessing "standard colours", that have more than 5% of representative vehicles in the dataset, and a mixture of the rest of colours. As "standard colours" were determined: black, grey, blue, red, white, and silver. To "non-standard colours" belong for instance, yellow, green, orange, etc.[13]

A set of "standard colours" is considered in the base of the model. From the theoretical perspective, the difference between a price of a "standardly" and "non-standardly" coloured cars is not obvious. A customer can be motivated to purchase an unusually coloured car because of a lower risk of theft (also in the context of vehicle history) or because of unique or eccentric vehicle appearance. On the contrary, customers on the second-hand market can be more motivated to seek a standard colour of bodywork since the spare parts in the basic colours are easily accessible (and therefore no additional costs for paintings are required) or since they prefer a conservative design.

6.3.15 Other Attributes

Among other vehicle attributes included in the analysis presented in this thesis belong a number of doors and seats, boolean feature of a complete service history[14] and the effect of a previous owner being a smoker. The attribute of the number of seats (included also for instance by Prado 2009) was controlled for unrealistic values. The number of doors is a peculiar issue since market agents consider this attribute differently.[15] The effect of doors was implemented into

[13]For the purposes of the analysis, the colours perceived by the vehicles owners were used instead of very misleading and huge scale colours officially presented by manufacturers.

[14]The full service history is not properly defined by the AutoScout24 platform and can, therefore, belong to misleading information. The author of this thesis incline to suppose that the status of having full service history is perceived by most market agents as full service electronic entries that are provided completely and responsibly most often by authorized service providers.

[15]Some customers count a boot lid as a door, some of them count only passenger doors. For this reason, the feature has to be appropriately adjusted (two and three-door body types are aggregated to a group labeled as a three-door version; four and five-door alternatives are merged in a term five-door version).

the model as boolean factors: three-door version and five-door version (keeping the five-door version in the base category). When talking about the effect of service history and smoker vehicle, a car without service history and with the smoker previous owner are base categories. It is expected that a positive effect will be proved for a number of seats attribute since it affects the number of passengers that can be transported. A five-door version is expected to be preferable by customers because of the comfort of getting in a vehicle. A non-smoking vehicle, in the current society accepting smoking to be injurious to health, should have a higher value (keeping other characteristics fixed) than a car driven by a smoker individual. The service history effect has not been studied in the current literature dealing with the German market yet. But it is expected that a vehicle with historical proof of services has a higher price in comparison to a car without these historical entries. The service history as an indicator of proper maintenance is considered to play an important role in a price determination of primary cars without warranty and older vehicles in general. Since AutoScout24 platform provides apparently quite a large proportion of vehicles with a warranty, the service history effect is expected to be of a relatively low magnitude.

Unfortunately, the present analysis could not utilize interesting features of emission class, country version of a vehicle, number of cylinders, drive chain, gears, the last inspection details, last service date, belt service, and information about the accident history and availability of advertised cars because of a substantial number of missing and incorrect data entries. The emission label was omitted from the dataset because more than 95% of observations possessed a green emission label.

In the existing literature, it is possible to find also a various number of other attributes, such as retail price and maximal speed (see Kihm & Vance 2016), breaks, acceleration, safety index, awards won by brands, sound system, and GPS navigation (refer to Pudaruth 2014), in older studies: length and weight (studied for instance by Triplett 1969), the reputation of a seller or a number of pictures and appropriateness of a title in an advertisement (see Andrews & Benzing 2007).

Chapter 7

Bayesian Model Averaging

7.1 Motivation

Considering an ample number of variables describing a car quality collected for the research presented in this thesis, the question of recognition of requisite and redundant variables arise. To recall the concerns of Hoeting *et al.* (1999), many statistical researches do not lay emphasis on the question of proper model definition sufficiently. As a consequence, the uncertainty of the model is often underestimated and biased models are produced (Raftery *et al.* 1997).

A felicitous solution of a proper selection among sets of competing models can be served by model averaging methods. These approaches are known as Frequentist Model Averaging and BMA estimation (Wang *et al.* 2009). The BMA approach grabbed the attention of researches in the 1990s and it is gaining popularity in the current literature nowadays. For instance Raftery *et al.* (1997) demonstrated that BMA evinces a better performance, on the ground of accuracy of the predictive performance, than any single model that can be formed from available variables. This thesis aims to manifest the use of BMA for addressing model uncertainty with the used-car dataset.

7.2 Basic Theory

The BMA approach proposes a mechanism of weighing all available models defined by the combination of all independent variables (Hoeting *et al.* 1999). Following the BMA principles described by Eicher *et al.* (2011), assume p candidate regressors $(X_1, \ldots, X_p)$ (in this research 57 features describing a used car) and a given dependent variable Y (in this case, Y represents a price of offered

vehicles in a logarithmic form). Then a researcher can opt between $K = 2^p$ alternative models $(M_1, \ldots, M_k)$ that are available (2^{57} different models in this research). Given $(X_1^{(k)}, \ldots, X_{p_k}^{(k)})$ that represent a subset of $(X_1, \ldots, X_p)$ and $\beta^{(k)} = (\beta_1^{(k)}, \ldots, \beta_{p_k}^{(k)})$ representing a vector of regression coefficients, the model M_k is defined as:

$$Y = \alpha + \sum_{j=1}^{p_k} \beta_j^{(k)} X_j^{(k)} + \epsilon \tag{7.1}$$

where ϵ is an error term, $\epsilon \sim N(0; \sigma^2)$.

The decision process of the appropriate model is based on the predictive performances of a models according to the cross validation. The weighting of all available models is performed on the basis of posterior model probability (PMP) of each model (Hoeting *et al.* 1999). More precisely (according to denomination by Eicher *et al.* 2011), denote D the data, (α, β, σ) the vector of parameters provided by the model M_k and θ the vector of these parameters in model M_k, likelihood function of model M_k: $pr(D|\theta_k, M_k)$ and prior density: $pr(\theta_k|M_k)$. Then the marginal likelihood $pr(D|M_k)$ is derived as:

$$pr(D|M_k) = \int pr(D|\theta_k, M_k) pr(\theta_k|M_k) d\theta_k \tag{7.2}$$

Using the derived marginal likelihood, the prior model probability expression of appropriateness of model M_k from the researcher's perspective *ex ante* $pr(M_k)$ and application of Baye's theorem, the term PMP $pr(M_k|D)$ is equal to:

$$pr(M_k|D) = \frac{pr(D|M_k) pr(M_k)}{\sum_{l=1}^{K} pr(D|M_l) pr(M_l)} \tag{7.3}$$

Posterior model probability can be viewed as an alternative of Adjusted R^2 in BMA estimation (Havránek *et al.* 2016). However, since an analysis can cover tents of features, the number of all alternative models can be enormous and the comparison of all models on the basis of PMP would be extremely time-consuming. Fernandez *et al.* (2001) states that mathematical difficulties start already with an amount of 25 explanatory variables in the analysis. For a model with a higher amount of aspiring regressors, a researcher has various tools to deal with the computation: coin-flip importance sampling algorithm, leaps-

and-bounds method by Raftery *et al.* (1997) and Markov Chain Monte Carlo method as described in details by Eicher *et al.* (2011). Markov Chain Monte Carlo sampler serves as the most popular one. For the purposes of this thesis, the computation was performed in R software for statistical computing using the package BMS. The BMS package developed by Zeugner (2011) proposes two Markov Chain Monte Carlo methods for approximation (Birth-death sampler and Reversible-jump sampler).

Citing Zeugner (2011), the process of approximation can be summarized as follows:

1. In the step i, model M_i is consider to be the "current best" candidate.
2. In the step $(i + 1)$, the aspiring model M_j is introduced and compared with model M_i on the basis of the PMP.
3. The candidate M_j can be either rejected or accepted as the new "current best" model.

Scenario A: In case M_j model is rejected, the process continues with the repetition of step 2 with the new aspiring model M_k that is again compared with model M_i. Scenario B: In case M_j model is accepted as the new "current best" candidate, the process continues with the repetition of step 2 with the new aspiring model M_k. In this scenario, the M_k alternative competes with the alternative model M_j. For more details about Markov Chain Monte Carlo approach incorporated in BMS R package please refer to Zeugner (2011) and Fernandez *et al.* (2001).

For the purposes of this thesis, the Markov Chain Monte Carlo sampler *Birth-death* sampler was incorporated. The model prior was assign to uniform model prior ($mprior =$ "$uniform''$) and the g prior was set to a default option of the bms function ($g =$ "UIP'').

By summing of posterior model probability for an explanatory variable of interest across all models $(M_1, \ldots, M_k)$ including the given variable, a researcher derives posterior inclusion probability (PIP). Following Jeffreys (1961), a predictor is identified on the ground of the value of PIP as:

- A decisive predictor if the corresponding PIP exceed a level of 0.99
- A strong predictor if the corresponding PIP belongs to the interval (0.95-0.99]

- A substantial predictor if the corresponding PIP belongs to the interval (0.75-0.95]
- A weak predictor if the corresponding PIP belongs to the interval (0.5-0.75]

7.3 BMA Application: Hedonic Price Analysis of Used Cars

For the model selection, a dataset containing 50,999 vehicles specified by 57 car attributes is used, please refer to the Table 7.1 for the overview of used vehicles specification. To avoid a perfect multicollinearity problem, the variables summarized in Table 7.2, representing the most common categories were removed from the sample. A car established as a base of the model can be therefore described as a gasoline-fueled sedan manufactured by Volkswagen, equipped with manual transmission and five-door "standardly coloured" bodywork. This vehicle was used by a private individual and a smoker. The advertisement is located in region 4 (according to a postcode division) and the full-service history is not available for this vehicle.

The Table 7.3 summarizes the results of the BMA. The candidate variables in the first column are ordered on the ground of their PIP value indicating the importance for explanation of data. Almost all car attributes (expect of *MakeAlfa*, *Displacement*, *NonStandardColors*, *CommerDealer* and regional attributes) has a one-unit PIP value meaning that these characteristics are included in 100% mass of models. The PIP values of *MakeAlfa*, *Displacement*, *NonStandardColors* and *PC8* are lower but still satisfy the frontier for decisive predictors. According to the division by Jeffreys (1961), all decisive predictors are used for the consequent analysis. The third column, reporting posterior mean, informs about the estimated coefficient derived as an averaged value across all alternative models, the fourth column states post standard deviation.

Table 7.1: Car Attributes Used for Model Selection Approach

Car attributes	**Data type**	**Car attributes**	**Data type**
Age	Numerical	*MakePeugeot*	Boolean
Mileage	Numerical	*MakeRenault*	Boolean
FuelDiesel	Boolean	*MakeSeat*	Boolean
CostsOnConsumption	Numerical	*MakeSkoda*	Boolean
BodyLargeSize	Boolean	*MakeSmart*	Boolean
BodySmallSport	Boolean	*MakeSuzuki*	Boolean
BodyOffroad	Boolean	*MakeToyota*	Boolean
BodyCompact	Boolean	*MakeVolvo*	Boolean
BodyStationwagon	Boolean	*EnginePowerkW*	Numerical
MakeAudi	Boolean	*Displacement*	Numerical
MakeBMW	Boolean	*TransmissionAutomatic*	Boolean
MakeFord	Boolean	*CarstatusEmployeescar*	Boolean
MakeMercedes	Boolean	*CarstatusDemonstration*	Boolean
MakeOpel	Boolean	*CommerDealer*	Boolean
MakeAlfa	Boolean	*PC0*	Boolean
MakeCitroen	Boolean	*PC1*	Boolean
MakeDacia	Boolean	*PC2*	Boolean
MakeFiat	Boolean	*PC3*	Boolean
MakeHonda	Boolean	*PC5*	Boolean
MakeHyundai	Boolean	*PC6*	Boolean
MakeJaguar	Boolean	*PC7*	Boolean
MakeJeep	Boolean	*PC8*	Boolean
MakeKia	Boolean	*PC9*	Boolean
MakeLancia	Boolean	*NrSeats*	Numerical
MakeLandRover	Boolean	*ThreeDoors*	Boolean
MakeMazda	Bolean	*FullService*	Boolean
MakeMini	Boolean	*NonStandardColours*	Boolean
MakeMitsubishi	Boolean	*NoSmoke*	Boolean
MakeNissan	Boolean		

Table 7.2: Car Attributes Omitted because of Multicollinearity

Base
FuelGasoline
BodySedan
MakeVolkswagen
TransmissionManual
CarstatusUsed
PC4
FiveDoors
NoFullService
StandardColours
Smoke

Table 7.3: Hedonic Price Analysis - BMA Results

Characteristics	**PIP**	**Post Mean**	**Post SD**
FullService	1.0000	0.0314	0.0025
NrSeats	1.0000	0.0324	0.0014
EnginePower-kW	1.0000	0.0049	0.00005
Mileage	1.0000	-0.0033	0.00003
FuelDiesel	1.0000	0.2153	0.0043
NoSmoke	1.0000	0.0142	0.0023
ThreeDoors	1.0000	-0.0372	0.0039
MakeAudi	1.0000	0.1272	0.0042
MakeBMW	1.0000	0.0925	0.0043
MakeFord	1.0000	-0.2023	0.0044
MakeMercedes	1.0000	0.1261	0.0045
MakeOpel	1.0000	-0.2108	0.0044
MakeCitroen	1.0000	-0.2582	0.0074
MakeDacia	1.0000	-0.4335	0.0107
MakeFiat	1.0000	-0.2917	0.0078
MakeHonda	1.0000	-0.0623	0.0115
MakeHyundai	1.0000	-0.2539	0.0068
MakeJaguar	1.0000	0.0882	0.0151
MakeJeep	1.0000	-0.1124	0.0134
MakeKia	1.0000	-0.2199	0.0075
MakeLancia	1.0000	-0.2377	0.0421
MakeLandRover	1.0000	0.1798	0.0128
MakeMazda	1.0000	-0.1208	0.0076
MakeMini	1.0000	0.0854	0.0075
MakeMitsubishi	1.0000	-0.2704	0.0110
MakeNissan	1.0000	-0.1776	0.0071
MakePeugeot	1.0000	-0.2403	0.0062
MakeRenault	1.0000	-0.2908	0.0056
MakeSeat	1.0000	-0.1262	0.0056
MakeSkoda	1.0000	-0.1056	0.0052
MakeSmart	1.0000	-0.2745	0.0094
MakeSuzuki	1.0000	-0.2167	0.0112
MakeToyota	1.0000	-0.1153	0.0081
MakeVolvo	1.0000	0.0610	0.0074
TransmissionAutomatic	1.0000	0.0964	0.0025
BodyOffroad	1.0000	0.1751	0.0033
BodyCompact	1.0000	-0.0677	0.0040
BodyStationwagon	1.0000	0.0500	0.0030
CarstatusEmployeescar	1.0000	-0.0386	0.0044
CarstatusDemonstration	1.0000	0.0573	0.0038
Age	1.0000	-0.0078	0.00003
CostsOnConsumption	1.0000	0.0316	0.0010
BodyLargeSize	1.0000	0.1335	0.0041
BodySmallSport	1.0000	0.2306	0.0051
MakeAlfa	0.9989	-0.0784	0.0160
Displacement	0.9976	-0.0211	0.0046
NonStandardColours	0.9946	0.0187	0.0041
PC8	0.9945	-0.0149	0.0032
PC9	0.9869	-0.0150	0.0034
PC7	0.0710	-0.0005	0.0021
PC0	0.0693	0.0006	0.0025
PC2	0.0276	-0.0002	0.0013
PC1	0.0094	0.00005	0.0006
PC6	0.0074	-0.000002	0.0002
PC3	0.0062	0.00002	0.0004
CommerDealer	0.0050	0.000008	0.0002
PC5	0.0042	0.000002	0.0002
(Intercept)	1.0000	9.1790	NA

Notes: PIP = posterior inclusion probability, Post SD = post standard deviation

Similarly, the Figure 7.1 represents the summarization of the models (represented by single columns) ordered horizontally according to their PMP that is demonstrated by the width of the columns. As can be seen, the width decreases from the left to the right, such as the most informative models are sorted in the left part of the figure. The row orientation is used to arrange the explanatory variables on the basis of their importance in the descending order from the top. Each model, represented by one column, is formed by the explanatory variables indicated in the rows. Every single cell in the table represents one feature in one particular model and on the ground of its colour it can be distinguished among a variable included and not included in the model. The white colour of the cell is reserved for variables that are not included in the model, the blue cell or lighter shaded cell in greyscale indicates an included explanatory variable with positive posterior mean and red or dark shaded cell is assigned to indicate an included variable with the negative posterior mean. On the ground of the consistency of the blue/red cells in the row, it can be concluded whether the signs varies in the models.

To check robustness of the BMA results, please refer to the Table 7.4. The previously described values can be compared to OLS estimators' coefficients (their magnitudes and signs) and their significance. When comparing the results, the BMA results appear to be consistent with the results stated in the OLS summary. The sings of coefficients, nor the information about the importance of variables for an explanation of data (stated by PIP in BMA and by p-value in OLS) varies in this comparison. Moreover, very similar magnitudes of coefficients were acquired.

Figure 7.1: Model Inclusion Based on Best 10,000 Models

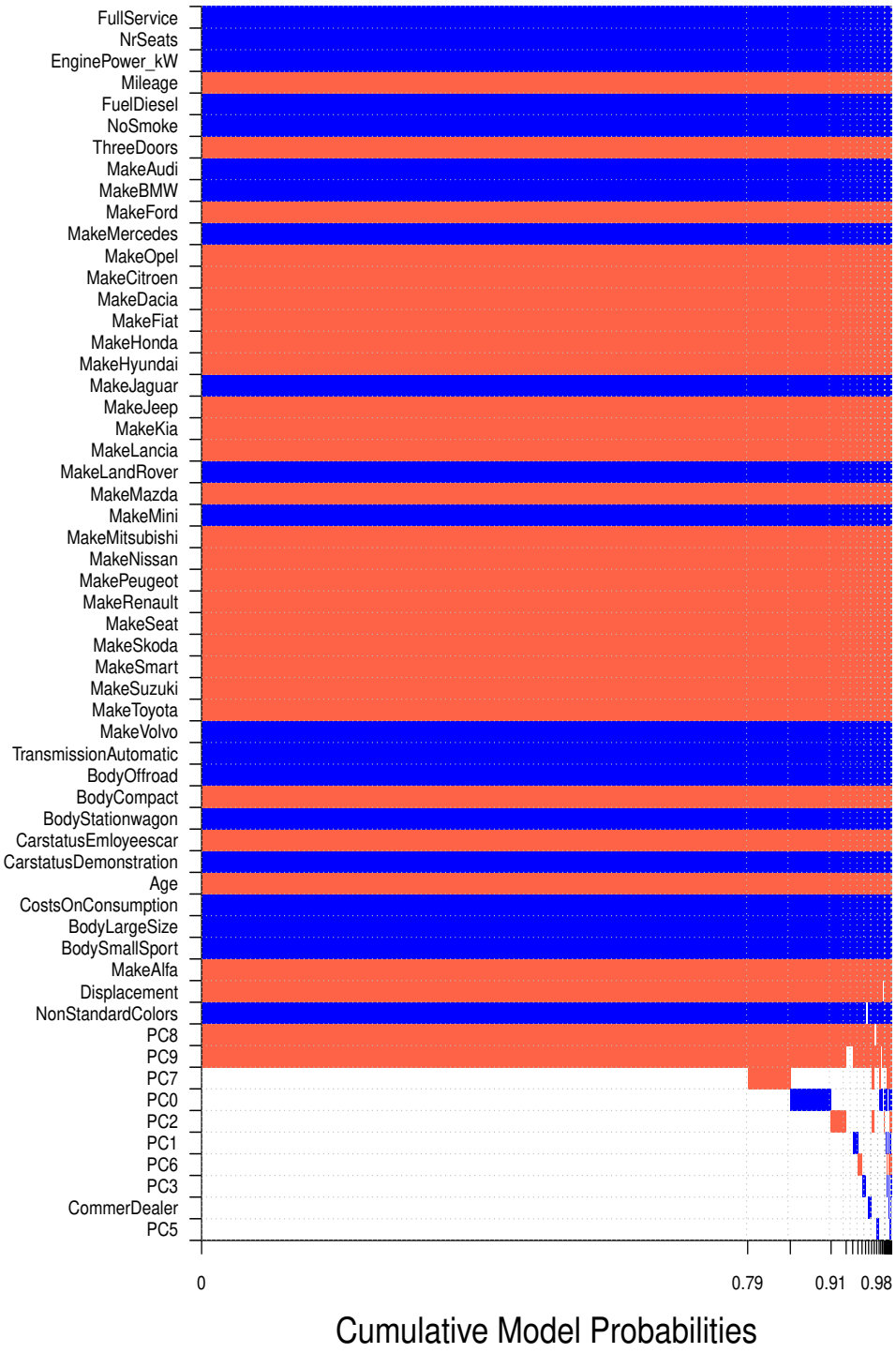


Table 7.4: Hedonic Price Analysis - Comparison Between BMA and OLS Results

	BMA			**OLS**	
	PIP	Post Mean	Post SD	Estimate	Std. Error
FullService	1.0000	0.0314	0.0025	0.0310	0.0025***
NrSeats	1.0000	0.0324	0.0014	0.0323	0.0014***
EnginePower-kW	1.0000	0.0049	0.00005	0.0051	0.00005***
Mileage	1.0000	-0.0033	0.00003	-0.0033	0.00002***
FuelDiesel	1.0000	0.2153	0.0043	0.2150	0.0043***
NoSmoke	1.0000	0.0142	0.0023	0.0146	0.0023***
ThreeDoors	1.0000	-0.0372	0.0039	-0.0372	0.0040***
MakeAudi	1.0000	0.1272	0.0042	0.1273	0.0043***
MakeBMW	1.0000	0.0925	0.0043	0.0924	0.0043***
MakeFord	1.0000	-0.2023	0.0044	-0.2019	0.0044***
MakeMercedes	1.0000	0.1261	0.0045	0.1262	0.0045***
MakeOpel	1.0000	-0.2108	0.0044	-0.2109	0.0044***
MakeCitroen	1.0000	-0.2582	0.0074	-0.2576	0.0074***
MakeDacia	1.0000	-0.4335	0.0107	-0.4332	0.0107***
MakeFiat	1.0000	-0.2917	0.0078	-0.2917	0.0078***
MakeHonda	1.0000	-0.0623	0.0115	-0.0618	0.0115***
MakeHyundai	1.0000	-0.2539	0.0068	-0.2534	0.0067***
MakeJaguar	1.0000	0.0882	0.0151	0.0890	0.0151***
MakeJeep	1.0000	-0.1124	0.0134	-0.1112	0.0134***
MakeKia	1.0000	-0.2199	0.0075	-0.2200	0.0075***
MakeLancia	1.0000	-0.2377	0.0421	-0.2376	0.0421***
MakeLandRover	1.0000	0.1798	0.0128	0.1805	0.0128***
MakeMazda	1.0000	-0.1208	0.0076	-0.1210	0.0076***
MakeMini	1.0000	0.0854	0.0075	0.0856	0.0075***
MakeMitsubishi	1.0000	-0.2704	0.0110	-0.2701	0.0110***
MakeNissan	1.0000	-0.1776	0.0071	-0.1772	0.0071***
MakePeugeot	1.0000	-0.2403	0.0062	-0.2398	0.0062***
MakeRenault	1.0000	-0.2908	0.0056	-0.2904	0.0057***
MakeSeat	1.0000	-0.1262	0.0056	-0.1269	0.0056***
MakeSkoda	1.0000	-0.1056	0.0052	-0.1062	0.0052***
MakeSmart	1.0000	-0.2745	0.0094	-0.2742	0.0094***
MakeSuzuki	1.0000	-0.2167	0.0112	-0.2174	0.0112***
MakeToyota	1.0000	-0.1153	0.0081	-0.1155	0.0080***
MakeVolvo	1.0000	0.0610	0.0074	0.0613	0.0074***
TransmissionAutomatic	1.0000	0.0964	0.0025	0.0967	0.0025***
BodyOffroad	1.0000	0.1751	0.0033	0.1752	0.0033***
BodyCompact	1.0000	-0.0677	0.0040	-0.0675	0.0041***
BodyStationwagon	1.0000	0.0500	0.0030	0.0507	0.0030***
CarstatusEmployeescar	1.0000	-0.0386	0.0044	-0.0391	0.0044***
CarstatusDemonstration	1.0000	0.0573	0.0038	0.0575	0.0038***
Age	1.0000	-0.0078	0.00003	-0.0078	0.00003***
CostsOnConsumption	1.0000	0.0316	0.0010	0.0316	0.0010***
BodyLargeSize	1.0000	0.1335	0.0041	0.1370	0.0041***
BodySmallSport	1.0000	0.2306	0.0051	0.2308	0.0051***
MakeAlfa	0.9989	-0.0784	0.0160	-0.0784	0.0160***
Displacement	0.9976	-0.0211	0.0046	-0.0211	0.0045***
NonStandardColours	0.9946	0.0187	0.0041	0.0187	0.000004***
PC8	0.9945	-0.0149	0.0032	-0.0149	0.0032***
(Intercept)	1.0000	9.1790	NA	9.1770	0.0093***

Notes: *** denotes that $p - value \leq 0.001$ and represents a statistical significance of an coefficient at the 99 percent confidence interval.

Chapter 8

Discussion of Results

8.1 Interpretation of the Model Results

When observing the BMA results, it is apparent that the vast majority of car attributes (48 out of 57 total candidate variables) are appraised as important variables for the valuation model. This finding supports the conviction of many researches (referring to for instance Lancaster 1966 or Gongqi *et al.* 2011) holding a view that a car should be studied as a very complex good and its price should be determined by a substantial number of features. To extend the current discussion, it can be summarized that an important role in the determination of a vehicle value play basic as well as additional characteristics. To the most crucial attributes marked as car value drivers belong primarily the variables representing vehicles make, body shape and fuel type. Contrarily, the majority of features representing German regions are appraised as redundant variables and the effect of the selling channel is not relevant for the hedonic price modeling neither. Almost all estimated coefficients of regressors in the BMA results have intuitive signs. As an exception can be determined only a positive sign of costs on consumption coefficient. And in contrast to the initial expectation, vehicle prices in the south part of Bavaria (region 8) vary from market values for the base region category (situated in the North-West of Germany) significantly.

To encapsulate the main results, several applications for agents intending to purchase a car and resell it in the pre-owned car market consequently can be formulated. Drivers can maximize the market value of their vehicles primarily by a responsible use in a term of mileage state, and by keeping an interior of a car unpolluted by smoking fumes. Contrarily, the effect of full service his-

tory does not evince to be an interesting strategy since the effect of historical service records increases a selling price to a very small extend (comparing to additional costs associated with primarily official service maintenance). Additionally, buyers can influence the selling price of the used car already by an appropriate purchase choice. As proved by the valuation model, all the basic attributes (make, model, body type, fuel type, engine power, engine capacity, transmission, etc.) as well as some additional attributes (for instance a bodywork colour, number of doors or number of seats) play an important role on the second-hand car market and agents could use the application of this knowledge for a new car purchasing decision. This application can be illustrated with an example of bodywork colour. Assuming, that manufacturers charge the "standard colours" and "non-standard colours" of bodywork identically, customers should opt for eccentrically coloured new vehicles to ensure a higher second-hand price of their cars. Concentrating on used car buyers, the regional price differencies are not excluded from a favourable purchase instruments but a determinative factor resides in the comparison of regional differences and transaction costs for a individual agent. Conclusions of the research presented in this thesis are relevant not only for private customers and private buyers offering their own cars but also for dealers with a portfolio of vehicles as well as for leasing companies or insurance brokers optimizing their business strategy. These results can be utilized in order to determine a proper valuation and pricing by all groups of agents on the pre-owned car market. Please refer to the following articles discussing the most interesting conclusions or to the appendix section of the thesis for other details of attributes.

Age and Mileage Predictably, the age and mileage characteristics have a negative effect on a vehicle price. Each additional month is, controlling for other attributes, associated with a 0.78% decrease in market value (one year is associated with approximately 9.36%). The results by Kihm & Vance (2016), representing the most relevant study for comparison (in terms of location and time), evince a slightly lower value (7.4%). To extend the current discussion about vehicle lifespan, a car is going to reach the end of its working age in approximately 10 years after the first registration and it is therefore in contrast to the 15-years life-cycle discussed in the existing literature and in contrast to initial expectations of this study. Unexpectedly, the ascertained lifespan is almost similar to the age of the average vehicle on the German roads (classic cars included).

The effect of each additional 1,000 kilometers recorded by an odometer are associated with a reduction by 0.33% (100,000 km with 33%) in vehicle value. The perfectly same result was obtained by Kihm & Vance (2016). On the ground of this finding, a car with an odometer state close to 300,000 km converges to a "scrap value".

Fuel Type The diesel-fueled vehicles are comparably more expensive than cars with gasoline power. This effect corresponds to the expectation since a transmission effect between new and used vehicle markets is taken into consideration. In the new car market, diesel-engine vehicles are priced more due to demanding manufacturing technology. A pre-owned vehicle using diesel for an operation is almost 22% more expensive than a comparable gasoline vehicle. To the best knowledge of the author, there is not sufficiently comparable German study suitable for the assessment of the related coefficient prior to the Dieselgate affair.

Costs on consumption As results indicate, the coefficient of costs on consumption (expressed in a term of a price for driving a distance of 100 km) has a positive sign. In general, the positive relationship is in stark contrast to results by for instance Kihm & Vance (2016) and Gössling & Metzler (2017) who concludes a (small but) negative price elasticity. On the contrary, Environmental Protection Agency (2018) states that the potential limited impact of fuel consumption on vehicle prices can be caused by a low ecological responsibility of customers. In this context, it is important to remind that studied fuel consumption values measured by manufacturers under laboratory conditions can represent partially biased information. It is not unlikely that a buyer could select an economically operating vehicle on the ground of other technical attributes than costs on consumption. As discussed in Chapter 6, the fuel economy is reflected not only in the variable costs on consumption but a customer can opt for an appropriate fuel type, engine capacity, and body shape in order to reduce operating costs or harmful emissions.

Geographical Factor Although no significant differences between local pricing were observed in most cases, cars advertised in the region number 8, predominantly situated in the federal state Bavaria, have *ceteris paribus* by approximately 1.5% smaller market value than those traded in the region number 4, situated in Lower Saxony and North Rhine-Westphalia. One of the potential

explanations would suggest that region 8 could have a specific substituting effect between new and used cars in comparison to other regions (the proportion of demand for used and new cars could be relatively higher in this region). The potential for arbitrage in Germany was not rejected. However, the regional price effect is rather small and can be for some buyers eliminated by transaction costs. For instance, taking into consideration the highest distance between two towns in regions number 4 and number 8 equal to 850 km, the regional price difference for an average vehicle is lower than transaction costs.[1]

Transmission Mechanism A car with an automatic transmission mechanism evinces a higher price by 9.6% in comparison to a manual geared vehicle. A higher added value (13%) of an automatic transmission was observed by Erdem & Şentürk (2009) on the Turkish market, however the US market values a car with this mechanism by only 4% higher than a car equipped with manual gears (Yerger 1996). For these reasons, the magnitude of the observed effect in the German second-hand market seems to be appropriate.

Status of a Previous Owner The status of the previous owner appears to play an important role. Comparing to a vehicle possessed by the private individual, an employee's vehicle is valued by 3.9% lower, however a car used as a demonstration vehicle evinces a price higher by 5.7%. Both values do not vary from the expectations.

Selling Channel The effect of the selling channel is not important for the data explanation, similarly as mentioned in the discussion of the expected feature impact. It is not unlikely that the negligible difference between a position of the commercial and private seller in a context of valuation effect has become unimportant as a response to the existence of an e-commerce environment that enables an easy vehicle comparison.

Service History The full service history is a factor that can mitigate the distrust of a customer. However, the research indicates that the market price of a vehicle with complete records of service is only 3.1% higher compared

[1]The computation is performed with an assumption of the average two-way costs on consumption being equal to 8 euro per 100 km, considering towing a purchased vehicle using another vehicle. For the purposes of the wage costs calculation, a two-day journey and minimal wage costs of 9.35 euro are taking into consideration. Depreciation and other costs are neglected. The expected profit (255 euro) is estimated according to an average price 17,008 euro of a used car in the dataset.

to vehicles without this proof of history. This effect could be explained by a large portion of cars under warranty and relatively low aged cars in the studied dataset (almost one-half of vehicles are younger than 4 years). In general, the responsible compliance of service limits does not evince to have an adequately high return in comparison to costs associated with (mainly official) service costs.[2]. The author of this thesis supposed that the service history effect could be more relevant for older vehicles. To the best knowledge of the author, the existing literature does not suggest any relevant comparison.

Bodywork Colour Although Erdem & Şentürk (2009) found out the higher value of vehicles with black and grey coloured bodyworks on the Turkish market, the present research indicates that "non-standardly coloured" cars (for instance, yellow, green, orange, pink, violet) are *ceteris paribus* valued higher (comparing to "standardly coloured" ones - black, grey, blue, red, white and silver) in the studied sample. The author of this thesis assumes that customers' tastes can geographically vary. As discussed, the popularity of non-standardly coloured vehicles could be caused by a lower risk of theft or a potential marketing effect of bold coloured bodyworks that can easily attract a customer or an effort of a customer to purchase an eccentric vehicle with a limited budget. The "non-standardly coloured" car is only by less than 2% more expensive comparing to a car with a "standard colour".

Cigarette Smoke Pollution A vehicle with a non-smoke identification is valued higher than a smoke polluted one on the used car market, controlling for other determinants. However, a magnitude of this effect is relatively small (non-smoker drivers can sell their cars at a higher price by 1.4% than smoker ones).

8.2 Second Hand Car Evaluation using Results

This section is concerned with a comparison of advertised prices with estimated market values in terms of vehicle attributes. For the purposes of the comparison, three vehicles were taken into consideration. In order to represent the most

[2]For an average priced vehicle in the dataset, the full service history would increase its value by 527 euro. Comparing to hundreds of euro invested in services annually, it is a lower than expected amount. However, the effect of services on the vehicle operation should be taken into consideration by a driver as well.

frequently offered car on AutoScout24 platform, two validation examples possess the base categories, discussed in Chapter 7. Vehicle A (Volkswagen Golf) is selected as a vehicle with the most frequent attributes, vehicle B (Volkswagen Polo) is selected as a car with a mean price value in the base sample. The third tested vehicle (Vehicle C) is selected on the out-of-sample basis (removed from the scraped dataset before the performance of the research) as a reference vehicle with the most similar attributes as Vehicle A. The comparison between prices based on the OLS model and observed prices, visualized in the Table 8.1 indicates that the estimated market vehicles values correspond to advertised prices for in-sample and out-of-sample observations.

Table 8.1: Comparison of Observed Price and Estimated Price in a Term of Vehicles Attributes

Car attributes	**Units**	**Vehicle A** **VW Golf**	**Vehicle B** **VW Polo**	**Vehicle C** **VW Golf**
Age	months	13	49	12
Mileage	km	5,818	61,134	5.800
Fuel type		Gasoline	Gasoline	Gasoline
Costs on consumption	euro/100 km	6.0	7.0	6.0
Body type		Sedan	Sedan	Sedan
Make		Volkswagen	Volkswagen	Volkswagen
Engine power	kW	85	44	85
Displacement	litre	0.999	0.999	0.999
Transmission type		Manual	Manual	Manual
Status of the car		Used	Used	Used
Selling channel		Private seller	Private seller	Commer. seller
Geographical location		PC4	PC4	PC8
Number of doors		5	5	5
Number of seats		5	5	5
Service history		No	No	Yes
Bodywork colour		Standard	Standard	Standard
Smoking vehicle		Yes	Yes	No
Advertised price		18,900	9,950	17,990
Estimated price		18,410	9,711	18,827
Percentage error		-2.59%	-2.40%	4.65%

Chapter 9

Conclusion

In recent years, the increasing importance of transportation of passengers and goods generated a rising number of agents that strive to quantify the determinants of car prices. At the same time, the existing research is becoming obsolete with dynamic social and technological changes. This thesis exposes a hidden potential of e-commerce as a data source to study (including a substantial number of advertisements with very detailed vehicle descriptions). As an extension of current knowledge about the used car market, this study concentrates on the analysis of the German e-commerce database of the biggest European online car platform AutoScout24 that have not been used for previous research yet. It utilizes a considerably large sample comprised of almost 51 thousand observations with 57 vehicle characteristics that differ from the predominating small-extend existing analyses.

The novelty of this thesis resides also in vehicle features selection on the basis of a detailed discussion of the current literature approaches and the technical relationships of attributes individually. Since the problem of model selection is not often properly discussed and solved in the related literature this thesis focuses primarily on this deficiency. Innovatively, the method of the model definition BMA is utilized to eliminate redundant features and avoid model uncertainty. To explain the price as a set of vehicle characteristics deriving the utility for a car owner, the concept of the hedonic price model is used.

Besides the testing of basic car characteristics, the study extends the existing literature additionally with an analysis of novel features. The research aims to test a unique geographical division of the German used car market. Often disregarded features are introduced, for instance, a selling channel effect, a refund of invested costs to regular service maintenance, and cigarette smoke

pollution of a vehicle interior. For this reason, the results can shed light on diverse unique perspectives of the vehicle valuation and has a capability to uncover a new car value drivers. An agent intending to buy and resell a car can use these conclusions to maximize the selling price. A rational choice of a new car in terms of its attributes, primarily basic features or for instance a bold bodywork colour, evinces an ability to generate a higher resell price on the used car market. The regular services (primarily by official service providers) do not belong to the appropriate strategies for an insufficient increase of a used car price. But what evinces an ability to increase a vehicle market value is a decision of a driver keeping a car interior unpolluted by smoking fumes. A buyer intending to purchase a used vehicle at a favourable price would not achieve a goal by choosing between sellers on the selling-channel basis. The conclusions on the selling channel could indicate a response to a highly competitive and transparent e-commerce environment. But for some buyers, an interesting utilization of the arbitrage effect is not rejected. Surprising is also conclusion on the estimated lifespan of a vehicle reaching 10 years that contrast to higher expected values in general. The diesel-powered vehicles retained a position of a highly valued alternative compared to gasoline consuming cars. Last but not least, the analysis proves that the price of a pre-owned vehicle is a complex task requiring a substantial amount of explanatory variables. The conclusions of this thesis can be utilized by also for other agents striving to evaluate a car market value (for instance, lease companies or dealers).

The analysis presented in this thesis sheds light upon a broad variety of determinants of a vehicle price. However, various other characteristics, for instance, emission labels/class effect, accident history, type of upholstery, or GPS navigation, deserve a further analysis. Moreover, the joint effect of attributes on a vehicle price can be clarified in detail. In the opinion of the author of this thesis, the analysis using a substantial number of variables selected on the ground of model averaging methods would be a very remarkable also for a study of transmission effect between new and used vehicles in Germany. The analysis taking into consideration the speed of the trade realization as a response of basic and alternative vehicle qualities would be a beneficial extension of nowadays knowledge as well.

Bibliography

ACEA.BE (2019): [Online; Accessed December, 2019]. Available from: `https://www.acea.be/`.

ADAC.DE (2019): [Online; Accessed December, 2019]. Available from: `https://www.adac.de/`.

AKERLOF, G. A. (1970): "The market for "lemons": Quality uncertainty and the market mechanism." *The Quarterly Journal of Economics* **84(3)**: pp. 488 – 500.

ANDREWS, T. & C. BENZING (2007): "The determinants of price in internet auctions of used cars." *Atlantic Economic Journal* **35**: pp. 43–57.

AUTOSCOUT24.DE (2019): [Online; Accessed December, 2019]. Available from: `https://www.autoscout24.de/`.

BAUER, I., L. ZAVOLOKINA, & G. SCHWABE (2019): "Is there a market for trusted car data?" *Electronic Markets* **30(2)**: pp. 211–225.

BOND, E. W. (1982): "A direct test of the "lemons" model: The market for used pickup trucks." *The American Economic Review* **72(4)**: pp. 836–840.

BOVENS, L. (2016): "The ethics of Dieselgate." *Midwest Studies in Philosophy* **40(1)**: pp. 262–283.

COURT, A. T. (1939): "Hedonic price indexes: With automotive examples." *The Dynamics of Automobile Demand* pp. 99–119.

COWLING, K. & J. CUBBIN (1972): "Hedonic price indexes for United Kingdom cars." *The Economic Journal* **82(327)**: pp. 963–978.

VAN DALEN, J. & B. BODE (2004): "Quality-corrected price indices: The case of the Dutch new passenger car market, 1990-1999." *Applied Economics* **36(11)**: pp. 1169–1197.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26**: pp. 30–55.

Environmental Protection Agency (2018): "Consumer willingness to pay for vehicle attributes: What is the current state of knowledge?" *Technical report*, U.S. Environmental Protection Agency, Washington, D.C.

Erdem, C. & I. Şentürk (2009): "A hedonic analysis of used car prices in Turkey." *International Journal of Economic Perspectives* **3**: pp. 141–149.

European Commission, Justice and Consumers (2014a): "Consumer market study on the functioning of the market for second-hand cars from a consumer perspective." *Technical report*.

European Commission, Justice and Consumers (2014b): "Study on the second hand cars market." *Technical report*.

European Parliament (2018): "Fighting mileage fraud on used cars." [Online; Accessed December, 2019]. Available from: `https://www.europarl.europa.eu/news/en/headlines/society/20180525STO04312/fighting-mileage-fraud-on-used-cars`.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Model uncertainty in cross-country growth regressions." *Journal of Applied Econometrics* **16**: pp. 563–576.

Ford Motor Company (2020): [Online; Accessed March, 2020]. Available from: `https://corporate.ford.com/history.html`.

Gongqi, S., W. Yansong, & Z. Qiang (2011): "New model for residual value prediction of the used car based on BP neural network and nonlinear curve fit." *Measuring Technology and Mechatronics Automation* **2**: pp. 682–685.

Goodman, A. C. (1998): "Andrew Court and the invention of hedonic price analysis." *Journal of Urban Economics* **44(2)**: pp. 291–298.

Gössling, S. & D. Metzler (2017): "Germany's climate policy: Facing an automobile dilemma." *Energy Policy* **105**: pp. 418 – 428.

Greenstone, M. (2017): "The continuing impact of Sherwin Rosen's "Hedonic prices and implicit markets: Product differentiation in pure competition"." *Journal of Political Economy* **125(6)**: pp. 1891 – 1902.

Griliches, Z. (1961): "Hedonic price indexes for automobiles: An econometric of quality change." In "The price statistics of the Federal Government," pp. 173–196. National Bureau of Economic Research.

Halvorsen, R. & H. O. Pollakowski (1981): "Choice of functional form for hedonic price equations." *Journal of Urban Economics* **10(1)**: pp. 37 – 49.

Havránek, T., D. Herman, & Z. Irsova (2016): "Does daylight saving save energy? A meta-analysis." *IES Working Paper 24/2016*, Charles University in Prague, Institute of Economic Studies (IES), Prague.

Hoeting, J. A., D. Madigan, A. E. Raftery, & C. T. Volinsky (1999): "Bayesian model averaging: A tutorial." *Statistical Science* **14(4)**: pp. 382–401.

Hofstede, G. J., T. Verwaart, & C. M. Jonker (2008): "Lemon car game." *Technical report*, ISAGA.

Hotten, R. (2015): "Volkswagen: The scandal explained." *BBC* [Online; Accessed March, 2020]. Available from: `https://www.bbc.com/news/business-34324772`.

Jeffreys, H. (1961): *The Theory of Probability.* Oxford University Press, 3rd edition.

Kihm, A. & C. Vance (2016): "The determinants of equity transmission between the new and used car markets: A hedonic analysis." *Journal of the Operational Research Society* **67**: pp. 1250–1258.

Kim, J.-C. (1985): "The market for "lemons" reconsidered: A model of the used car market with asymmetric information." *The American Economic Review* **75(4)**: pp. 836–43.

Lancaster, K. J. (1966): "A new approach to consumer theory." *The Journal of Political Economy* **74(2)**: pp. 132–157.

McFadden, D. (1974): "The measurement of urban travel demand." *Journal of Public Economics* **3(4)**: pp. 303–328.

Mishra, D. P., J. B. Heide, & S. G. Cort (1998): "Information asymmetry and levels of agency relationships." *Journal of Marketing Research* **35(3)**: pp. 277–295.

Novosád, J. (2009): *Alternativní pohon automobilů*. Master's thesis, Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav automobilního a dopravního inženýrství, Brno. pages 85. Advisor: doc. Ing. Zdeněk Kaplan, CSc.

Ohta, M. & Z. Griliches (1976): "Automobile prices revisited: Extensions of the hedonic hypothesis." In "Household Production and Consumption," pp. 325–398. National Bureau of Economic Research.

Pal, N., P. Arora, P. Kohli, D. Sundararaman, & S. S. Palakurthy (2018): "How much is my car worth? a methodology for predicting used cars' prices using random forest." Proceedings of the 2018 Future of Information and Communication Conference (FICC), Vol. 1.

Peters, J. (2000): "Buyer market power and innovative activities: Evidence for the German automobile industry." *Review of Industrial Organization* **16(1)**: pp. 13–38.

Prado, S. M. (2009): "The European used-car market at a glance: Hedonic resale price valuation in automotive leasing industry." *Economics Bulletin* **29(3)**: pp. 2086–2099.

Pudaruth, S. (2014): "Predicting the price of used cars using machine learning techniques." *International Journal of Informatione and Computation Technology* **4(7)**: pp. 753–764.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* **92(437)**: pp. 179–191.

Requena-Silvente, F. & J. Walker (2006): "Calculating hedonic price indices with unobserved product attributes: An application to the UK car market." *Economica* **73(291)**: pp. 509–532.

Rosen, S. (1974): "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy* **82(1)**: pp. 34–55.

Schmid, C. U. (2017): "Germany." In M. F. di Rattalma (editor), "The dieselgate: A legal perspective," pp. 27–46. Springer International Publishing.

Spatz, J. & P. Nunnenkamp (2002): "Globalization of the automobile industry: traditional locations under pressure?" *Kiel Working Paper 1093*, Aussenwirtschaft, University of St. Gallen, School of Economics and Political Science, Swiss Institute for International Economics and Applied Economics Research.

Statista GmbH (2020): "Durchschnittlicher Preis für einen Liter Diesel in Deutschland in den Monaten Juni 2019 bis Juni 2020." [Online; Accessed March, 2020]. Available from: `https://de.statista.com/statistik/daten/studie/1691/umfrage/`.

Triplett, J. E. (1969): "Automobiles and hedonic quality measurement." *Journal of Political Economy* **77(3)**: pp. 408–417.

Truett, L. J. & D. B. Truett (2017): "The German motor vehicle industry: Costs and crisis." *International Review of Economics and Finance* **48**: pp. 49–55.

Verband der Automobilindustrie (2020): [Online; Accessed March, 2020]. Available from: `https://www.vda.de`.

Wang, H., X. Zhang, & G. Zou (2009): "Frequentist model averaging estimation: a review." *Journal of Systems Science and Complexity* **36(4)**: pp. 7809–7817.

Wu, J.-D., C.-C. Hsu, & H.-C. Chen (2009): "An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference." *Expert Syst. Appl.* **36(4)**: pp. 7809–7817.

Yerger, D. (1996): "Used car markets: reliability does matter, but do consumer reports?" *Applied Economics Letters* **3(2)**: pp. 67–70.

Zeugner, S. (2011): "Bayesian Model Averaging with BMS." *Tutorial to the R-package BMS* .

Appendix

Results of the Hedonic Price Analysis - Additional Discussion

Body Type Each body shape evinces *ceteris paribus* a higher market value than a sedan, except for a compact type of small size and no sport, nor luxury appearance. A compact vehicle is therefore almost by 7% cheaper than a comparable sedan car. A sports small-size vehicles (convertible and coupe) are *ceteris paribus* by approximately 23% more expensive than a sedan shape. An off-road version is almost by 18%, a large-size version by 13% and station wagon shape nearly by 5% more valued by customers than a sedan car.

Make All manufacturers evince expected signs and magnitudes. Keeping other attributes fixed, for instance, Audi vehicle (a luxury-oriented brand) has a higher market value by almost 13% than Volkswagen. Comparable results evinced Mercedes Benz, slightly lower values are observed for Mini and Jaguar and the highest coefficient can be seen for Land Rover. Volkswagen's competitors (Seat and Skoda) are *ceteris paribus* slightly cheaper (by 12.6% and 10.6% respectively). The further research oriented for the difference between domestic and imported brands (connected with an idea of affordability of spare parts) could be interesting.

Engine Power The effect of engine power on a car value is positive as expected. With an additional 1 kW, a price of a used car increases by 0.5% keeping other effects fixed (similarly as proved by Requena-Silvente & Walker (2006) for a new car). However, Kihm & Vance (2016) dealing with the German used car data concluded the increase by only 0.25%.

Engine Capacity As expected, the car attribute of displacement has a negative impact on a vehicle market price. Every liter of engine capacity lowers the market value of a vehicle by 2%. From the perspective of a customer, a high displacement value separately should represent an undesirable attribute because of an indirect impact on operating costs and emissions.

Number of Doors, Number of Seats The positive effect of seats and a relatively higher price of a five-door alternative comparing to a three-door body style were proved. With every additional seat, the price of a vehicle (keeping other attributes fixed) will increase by 3%. This information is important for distinguishing between differently sized cars and cars created for different purposes (passenger and goods transportation cars) within one market segment (body shape).