

2017

## Developing Statistical Models to Assess Productivity in the Automotive Manufacturing Sector

Amir Abolhassani

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Abolhassani, Amir, "Developing Statistical Models to Assess Productivity in the Automotive Manufacturing Sector" (2017). *Graduate Theses, Dissertations, and Problem Reports*. 5021. <https://researchrepository.wvu.edu/etd/5021>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

**Developing Statistical Models to Assess Productivity in the Automotive  
Manufacturing Sector**

**Amir Abolhassani**

**Dissertation**

**Submitted to the  
Benjamin M. Statler College of Engineering and Mineral Resources at  
West Virginia University**

**in partial fulfillment of the requirement for the degree of  
Doctor of Philosophy  
in  
Industrial Engineering**

**Bhaskaran Gopalakrishnan, Ph.D., P.E., Chair**

**E. James Harner, Ph.D., Co-Chair**

**Majid Jaridi, Ph.D.**

**Kenneth R. Currie, Ph.D., P.E.**

**Ashish D. Nimbarte, Ph.D.**

**Department of Industrial and Management Systems Engineering**

**West Virginia University  
Morgantown, West Virginia**

**2017**

**Keywords: Productivity improvement, Automotive Industry, Effective Factors,  
Production Process, Hours per Vehicle**

**Copyright 2017 Amir Abolhassani**

## **Abstract**

# **Developing Statistical Models to Assess Productivity in the Automotive Manufacturing Sector**

**Amir Abolhassani**

The purpose of this study is to identify the most important activity in a value chain, effective factors, their impact, and to find estimation models of the most well-known productivity measurement, Hours per Vehicle (HPV), in the automotive industry in North American manufacturing plants. HPV is a widely recognized production performance indicator that is used by a significant percentage of worldwide automakers. During a comprehensive literature review, 13 important factors that affect HPV were defined as launching a new vehicle, ownership, car segment, model types, year, annual available working days, vehicle variety, flexibility, annual production volume, car assembly and capacity (CAC) utilization, outsourcing, platform strategy, and hourly employee's percentage.

Data used in this study was from North American plants that participated in the Harbour's survey from 1999 to 2007. Data are synthesized using a uniform methodology from information supplied by the plants and supplemented with plant visits by Harbour Consulting auditors. Overall, there are 682 manufacturing plants in the statistical sample from 10 different multinational automakers.

Several robust and advanced statistical methods were used to analyze the data and derive the best possible HPV regression equations. The final statistical models were validated through exhaustive cross-validation procedures. Mixed integer distributed ant colony optimization (MIDACO) algorithm, a nonlinear programming algorithm, that can robustly solve problems with critical function properties like high non-convexity, non-differentiability, flat spots, and even stochastic noise was used to achieve HPV target value.

During the study period, the HPV was reduced 48 minutes on the average each year. Annual production volume, flexible manufacturing, outsourcing, and platform strategy improve HPV. However, vehicle variety, model types, available annual working days, CAC, percentage of the hourly employees, and launching a new model penalize HPV. Japanese plants are the benchmark regarding the HPV followed by joint ventures and Americans. On average, the HPV is lower for Japanese and joint ventures in comparison to American automakers by about 1.83 and 1.28 hours, respectively. Launching a new model and adding a new variety in body styles or chassis configurations raises the HPV, depending on the car class; however, manufacturing plants compensate for this issue by using platform sharing and flexible manufacturing strategies. While launching a new vehicle common platform sharing, flexible manufacturing, and more salaried employees (lower hourly) strategies will help carmakers to overcome the effect of launching new vehicles productivity penalization to some extent.

The research investigates current strategies that help automakers to enhance their production performance and reduce their productivity gap. The HPV regression equations that are developed in this research may be used effectively to help carmakers to set guidelines to improve their productivity with respect to internal and external constraints, strengths, weaknesses, opportunities, and threats.

## **DEDICATION**

*To my loving wife Maryam and our wonderful sons  
Hooman and Mehrad. Their support, encouragement, and prayers  
helped me to overcome the difficulties and eased the pressure that I  
faced during my PhD study.*

## **Acknowledgements**

I would like to express my deepest appreciation to the committee chair, Professor Bhaskaran Gopalakrishnan, for his outstanding guidance and support during the process of researching and writing this dissertation. I would also like to thank my co-advisor, Professor James Harner, for his guidance, help, and support throughout my research. It has been such a great opportunity for me to work with him over this period. He has been a great advisor and a person of integrity. Without his guidance and persistent help this dissertation would not have been possible.

I would like to express my sincere appreciation to Professor Majid Jaridi, who for many years guided and encouraged me during the course of this research. Professor Jaridi is a wonderful mentor and he always has thought provoking questions to make me think beyond the confinements of my study. I would like to thank my committee members: Professors Kenneth Currie and Ashish Nimbarte for being my committee members and accepting the heavy task of going through my dissertation. I would like to extend my special thanks to Professor Gale Boyd at Duke University for his invaluable insights and supports, and Mr. Arnaud Billet and Mr. Maxime Boniteau from Renault, for their help and support during the course of this research. I was lucky to work with all of them and learn from their experiences and skills in the field of research.

I am grateful to all of my professors, my colleagues, and the staff of the Industrial Engineering Department for their help and support. I am deeply indebted to my wife, Maryam Amini, who provided me with the love, time, support, inspiration, and motivation needed to finish this dissertation. Finally, my thanks and appreciation are to my parents and family for their support, help, and encouragement. I am blessed to have such a supportive and loving family.

## Table of Contents

List of Figures .....	ix
List of Tables .....	xi
Nomenclature .....	xiii
1 : Introduction.....	1
1.1. Worldwide and USA Automotive Industry Highlights.....	1
1.2. Significant Movements in the Automotive Industry .....	2
1.3. Future Challenges and Opportunities in Automotive Industry .....	3
1.4. Research Motivation .....	4
1.5. Research Objectives .....	5
1.6. Research Organization .....	6
1.7. Conclusion.....	6
2 : Literature Survey .....	7
2.1. Introduction .....	7
2.2. Origin of the Term Productivity.....	7
2.3. Automotive Industry History and Production Process.....	9
2.4. Porter’s Value Chain .....	11
2.5. Effective Factors on Productivity.....	13
2.5.1. Quality Management Systems (QMS) .....	13
2.5.2. Outsourcing of the Activity.....	14
2.5.3. Common Platform Application .....	14
2.5.4. Participatory Ergonomics in Product Development.....	15
2.5.5. Robots’ Application in Manufacturing Lines.....	15
2.5.6. Information Technology (IT), Enterprise Resource Planning (ERP), and Simulation ...	16
2.5.7. Human Resource Management (HRM) .....	17

2.5.8. Modularization .....	17
2.5.9. Lean Manufacturing and Quality Tools Deployment .....	18
2.5.10. Just in Time (JIT) .....	19
2.5.11. Total Productive Maintenance (TPM).....	19
2.5.12. 5S.....	20
2.6. The Main Activity of Porter’s Value Chain.....	20
2.7. Labor Productivity Measures .....	22
2.8. Studies that Have Been Done to Improve Productivity .....	26
2.9. Conclusion.....	33
3 : Research Objectives and Approach .....	34
3.1. Introduction .....	34
3.2. Data Collected for This Study.....	34
3.3. Important Factors on HPV .....	35
3.4. Approach for the Statistical Analysis.....	55
3.5. Conclusion.....	58
4 : Current Work and Preliminary Results.....	59
4.1. Introduction .....	59
4.2. ANOVA for all the Variables.....	59
4.3. Tukey’s Comparison for Different Car Segments .....	61
4.4. Tukey’s Comparison for Different Years.....	63
4.5. Tukey’s Comparison for Different Models .....	63
4.6. Hypotheses Testing and Statistical Analysis.....	64
4.6.1. Variable Types.....	65
4.6.2. Data Preparation and Cleaning .....	66
4.6.3. Generic Multiple Linear Regression Model .....	69
4.6.4. Criterion for Choosing the Optimal Model .....	70

4.6.5. Initial Multiple Linear Regression Model .....	71
4.6.6. Base Model with Considering Year as a Categorical Variable .....	73
4.6.7. Base Model with Considering Car Model Types as a Categorical Variable .....	74
4.6.8. Model Adequacy Checking .....	75
4.6.9. Comparing all three Models and Hypotheses Testing .....	79
4.6.10. Model Selection Among the three Models .....	85
4.7. Conclusion.....	87
5 : Model Improvement .....	88
5.1. Introduction .....	88
5.2. Transformations .....	88
5.3. Variables Interaction .....	92
5.4. Model Validation.....	92
5.5. Conclusion.....	93
6 : Different Statistical Analysis .....	94
6.1. Introduction .....	94
6.2. Robust Regression.....	94
6.2.1 M-Estimate .....	95
6.2.2 MM-Estimate.....	99
6.2.3 SMDM-Estimate.....	100
6.2.4 Least Trimmed Squares Robust (LTS) regression .....	102
6.2.5 Quantile regression .....	103
6.2.6 Principal component regression (PCR) .....	104
6.3. Shrinkage Methods.....	106
6.3.1 Ridge Regression.....	106
6.3.2 Lasso (least absolute shrinkage and selection operator) regression .....	109
6.3.3 Elastic net regression .....	111



6.3.4 Selecting the Tuning Parameter.....	113
6.3.5 Shrinkage Methods Assumptions .....	113
6.4. Hybrid Methods.....	115
6.5. Variables Interaction .....	117
6.6. Analyzing the Relevant Outperforming Residuals.....	120
6.7. Analyzing the Plants That Were not Productive .....	124
6.8. Overview of North American Car Manufacturers.....	125
6.9. Conclusion.....	127
7 : Mixed Integer Programming.....	129
7.1. Introduction .....	129
7.2. Mixed Integer Programming .....	129
7.3. Defining the Business Strategies to Find a Desired HPV Value.....	130
7.3.1. Branch and Bound Method.....	131
7.3.2. Stochastic Metaheuristics .....	133
7.3.3. Ant Colony Optimization (ACO) .....	133
7.3.4. Mixed Integer Distributed Ant Colony Optimization (MIDACO).....	134
7.4. Conclusion.....	136
8 : Conclusion, Contributions, Limitations, and Future Research .....	137
8.1. Conclusion.....	137
8.2. Contributions.....	138
8.3. Limitations .....	139
8.4. Future Research.....	139
Literature Cited .....	141
Appendix.....	149

## List of Figures

Figure 1.1. U.S. Automotive Sales and Forecast, 2007-2018 (K. Hill, Menk, &Cooper, 2015) .....	2
Figure 2.1. Model T .....	9
Figure 2.2. Automotive manufacturing process .....	10
Figure 2.3. Porter’s value chain .....	11
Figure 2.4. Porter’s Value Chain .....	21
Figure 2.5. Average vehicles produced per worker from 1963 to 1996.....	23
Figure 2.6. A schematic view of the people who are involved for this measure .....	24
Figure 2.7. Worldwide car makers who attends in the Harbour survey.....	26
Figure 3.1. HPV variation in the study period for American, Japanese, and Joint venture companies .....	35
Figure 3.2. HPV variation in the study period for different brands .....	36
Figure 3.3. CAC trend during the study period for American, Japanese, and joint venture plants.....	38
Figure 3.4. HPV versus model type trend during the study period.....	39
Figure 3.5. Number of platform/models type trend during the study period for American, Japanese, and joint venture plants.....	40
Figure 3.6. Vehicle variety trend during the study period for American, Japanese, and joint venture plants .....	42
Figure 3.7. Platform sharing strategy trend during the study period for American, Japanese, and joint venture plants.....	44
Figure 3.8. Production volume trend during the study period for American, Japanese, and joint venture plants.....	45
Figure 3.9. Assembly line strategy to fit maximum demand (adopted from Diffner, 2011) .....	47
Figure 3.10. Dedicated assembly line to maximize utilization (adopted from Diffner, 2011) .....	48
Figure 3.11. Mixed model assembly line application to maximize the assembly utilization (adopted from Diffner, 2011) .....	48

Figure 3.12. The hourly employees' percentage trend during the study period for American, Japanese, and joint venture plants.....	52
Figure 3.13. Different phases and steps of the research.....	57
Figure 4.1. Density function for all variables versus different segment .....	62
Figure 4.2. HPV variation in the study period for different model types per plant .....	64
Figure 4.3a. Scatterplot of the variables .....	67
Figure 4.4b. Scatterplot of the variables .....	68
Figure 4.5. Model adequacy checking for the base model.....	76
Figure 4.6. Model adequacy checking for the base model that year is categorical.....	77
Figure 4.7. Model adequacy checking for the base model that car model types is categorical .....	78
Figure 4.8. All regressors coefficients trends when new variables are added to the model.....	81
Figure 4.9. Scatter plot of HPV versus TCR, DART, DAFWII .....	84
Figure 5.1. Plotting different transformations on APV (the blue line is LOESS Curve Fitting (Local Polynomial Regression)).....	90
Figure 5.2. Plotting different transformations on CAC (the blue line is LOESS Curve Fitting (Local Polynomial Regression)).....	90
Figure 6.1. Coefficients of all variables VS different quantiles.....	105
Figure 6.2. Geometric interpretation of ridge regression (adapted from(PSU, 2016)) .....	107
Figure 6.3. The unstandardized ridge regression coefficients.....	108
Figure 6.4. Geometric interpretation of lasso regression (adapted from (James et al., 2013)) .....	110
Figure 6.5. The unstandardized lasso regression coefficients.....	111
Figure 6.6. The unstandardized elastic-net regression coefficients .....	112
Figure 6.7. Mean-Squared error of ridge and lasso regressions for different values of $\lambda$ .....	114
Figure 6.8. The trend of HPV for the plants that had the highest HPV over the study period.....	124
Figure 6.9. Worldwide profitability Pre-tax profit per vehicle for North American car manufactures ....	126

## List of Tables

Table 2.1. Summary of the word productivity’s origin (adopted from (Tangen, 2005)).....	8
Table 2.2. Contributions and limitations of the existing literature .....	30
Table 3.1. Frequency of data for various automakers’ in study period.....	34
Table 3.2. Number of plants in each segment during the study period.....	37
Table 3.3. Flexible manufacturing index during the study period .....	49
Table 3.4. Outsourcing percentage during the study period .....	50
Table 3.5. Percentage of plants that were involved with a new product launch .....	51
Table 3.6. Annual available working days for different car makers .....	51
Table 3.7. List of all hypotheses .....	55
Table 4.1. Analysis of variance for all the variables for American, Japanese, and Joint venture companies .....	60
Table 4.2. Tukey’s Pairwise comparisons for different segments .....	61
Table 4.3. Tukey’s Pairwise Comparisons of HPV for different years.....	63
Table 4.4. Tukey’s Pairwise Comparisons for different Model types per plant .....	64
Table 4.5. Different variables types .....	66
Table 4.6. ANOVA and relevant statistics for the base model .....	72
Table 4.7. ANOVA and relevant statistics for the base model with year as categorical .....	73
Table 4.8. ANOVA and relevant statistics for the base model with car model’s types as categorical .....	74
Table 4.9. Comparison of all three models about different variables .....	79
Table 4.10. Hypotheses result.....	82
Table 4.11. Pearson Correlation for the Ownership.....	85
Table 4.12. Discussed statistical model performance characteristic .....	86
Table 5.1. Modified Tukey’s ladder of transformation.....	89
Table 5.2. Different Model Transformation comparison .....	91

Table 6.1. Objective and weight functions for least-squares, Huber, and Tukey’s bisquare estimators (adapted from Fox and Weisberg (2010)).....	96
Table 6.2. Huber and Tukey’s bisquare model comparisons .....	98
Table 6.3. MM and SMDM model comparisons .....	101
Table 6.4. Least-trimmed squares coefficients .....	103
Table 6.5. Comparison of the resulted different statistical models.....	116
Table 6.6. Different statistical models comparison.....	118
Table 6.7. The Final Model Coefficients .....	119
Table 6.8. The mean and median for the flexibility, hourly.total, and platformstrategy .....	120
Table 6.9. Outperforming and not outperforming residuals comparison.....	122
Table 6.10. Outperforming and not outperforming residuals comparison for the ownership and car segment .....	123
Table 6.11. Labor and benefit cost per vehicle for Big Three .....	125
Table 7.1. Desired factors values for HPV of 10H – CPLEX solver.....	132
Table 7.2. Desired factors values for HPV of 10H – MIDACO Algorithm .....	134
Table 7.3. The optimization model adapted for all car segments and ownerships .....	136

## Nomenclature

AIC	Akaike Information Criterion
BIC	Baysian Information Criterion
CAC	Car Assembly and Capacity Utilization
DAFWII	Days Away from Work
DART	Days Away, Restricted, And Transfer
EDA	Exploratory Data Analysis
FM	Flexible Manufacturing
HPV	Hours Per Vehicle
HVAC	Heating, Ventilating, And Air Conditioning
IEA	International Ergonomic Association
LOOCV	Leave One Out Cross Validation
LTSR	Least Trimmed Squares Regression
MILP	Mixed Integer Linear Programing
MTM	Methods Time Measurement
OEM	Original Equipment Manufacturer
OSHA	Occupational Safety and Health Administration
RO	Research Objective
RSS	Residual Sum of Squared
SME	Small or Medium Enterprise
TCR	Total Case Rate
TPF	Total Productivity Factor
TPS	Toyota Production System
TQM	Total Quality Management
VPE	Vehicles Per Employee
WRMC	World Market Research Centre
OLS	Ordinary Least Squares
PCR	Principal Component Regression
MIP	Mixed Integer Programming
MILP	Mixed Integer Linear Programing
MINLP	Mixed Integer Nonlinear Programing
PSO	Particle Swarm Optimization
ABC	Artificial Bee Colony optimization
PSO	Particle Swarm Optimization

ABCO	Artificial Bee Colony Optimization
ACO	Ant colony optimization
MIDACO	Mixed Integer Distributed Ant Colony Optimization
BB	Branch and Bound
PCR	Principal component regression
MAD	Median Absolute Deviation

# **1 : Introduction**

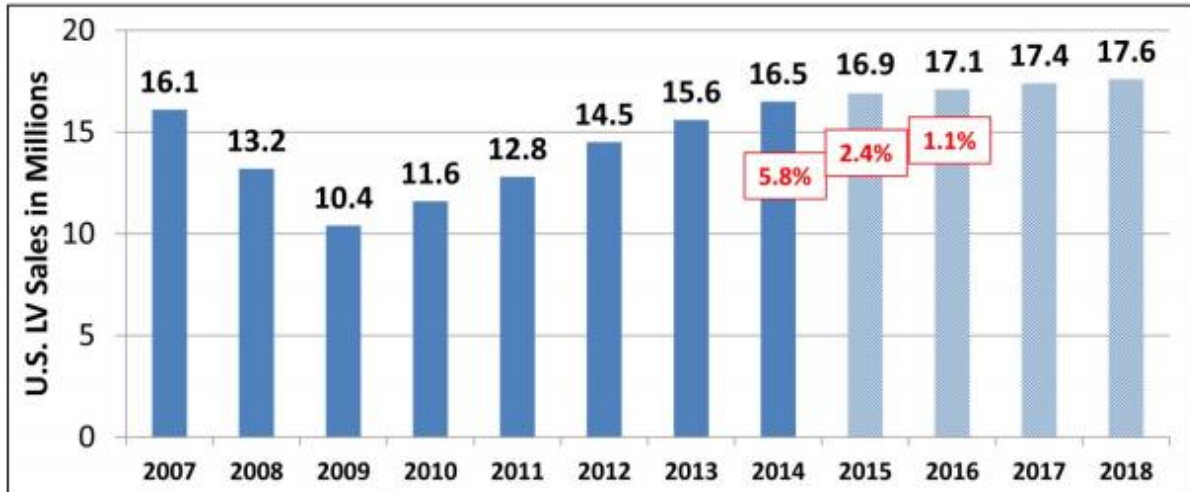
## **1.1. Worldwide and USA Automotive Industry Highlights**

The total number of vehicles produced worldwide in 2014 surpassed 88.24 million units. This is a record, and represents an increase of 3% over 2013 production. Automakers' gross revenues reached almost \$2.5 trillion. Building 88 million vehicles requires the employment of more than 10.5 million people directly in making the vehicles and the parts that go into them. This is more than 6.7% percent of the world's total manufacturing employment. In addition to these direct employees, about five times more are employed indirectly in related manufacturing and service provision, such that an estimated more than 63 million people earn their living from cars, trucks, buses and coaches. Sales in Asia accounted for more than half of the total production, and China ranked at the top with about 23% of worldwide sales (OICA, 2015). In 2014 Volkswagen and Toyota Motor Group ranked 11<sup>th</sup> and 13<sup>th</sup>, respectively, in revenue among all Fortune Global 500 companies, and 1<sup>st</sup> and 2<sup>nd</sup> among all international automakers (Fortune, 2014).

Auto industry executives and experts in the U.S. market, forecast annualized sales in North America in the near term of relatively 17 million cars, up from 13 million in 2008. However, the outlook in Europe is much weaker as the region is emerging fitfully from a six-year sales slump. Sales have plunged in Russia and South America - they were down by about 25 percent and 15 percent, respectively, in August 2014 year-over-year. Meanwhile, the Indian market's performance has been inconsistent. Growth in China - the world's largest vehicle market — has slowed, even though investments by most original equipment manufacturers (OEMs), which are betting big on future demand, continue to ramp up. While these markets have been unstable, the American market is projected to remain stable and growing, reacting strategically to these demand shifts will be an absolute priority for industry leaders (McKinsey, 2013).

Meanwhile, the total number of vehicles sold in the USA in 2014 surpassed 16.5 million units. Sales are forecast to increase by approximately 12.8 percent, from 2013 to 2018. Figure 1.1 displays historical and forecasted sales for the U.S. automotive industry from 2007 to 2018. The forecast suggests that automobile sales over the next several years will continue to increase, returning to the long-term trend from 16.9 to 17.6 million units annually (K. Hill, Menk, & Cooper, 2015).





Source: Center for Automotive Research, January 2015

**Figure 1.1. U.S. Automotive Sales and Forecast, 2007-2018 (K. Hill, Menk, & Cooper, 2015)**

The overall carmakers’ revenue that was generated in the U.S. was approximately \$500 billion in 2014. In total, the auto industry was responsible for 7.25 million private sector jobs. Auto manufacturers, suppliers, and dealers themselves employ more than 1.5 million people and directly contributed to the creation of another 5.7 million jobs. Jobs related to the auto industry go far beyond designing, building, and selling vehicles. Automakers in the U.S. are also among the largest purchasers of aluminum, iron, steel, lead, copper, rubber, plastics, textiles, vinyl, and computer chips. For that reason, the automotive industry has a strong multiplier effect, the Center for Automotive Research (CAR) researchers found: each job for an auto manufacturer in the United States creates nearly 7 other positions in industries across the economy (K. Hill, Menk, Cregger, et al., 2015).

The American market is a popular one across the globe, considering both production volume and the variety of models; for instance, some 320 models with 2,308 variations were reported in this country in 2004 (Biesebroeck, 2007). Considering all of the above reasons, it was decided to study this niche market.

## 1.2. Significant Movements in the Automotive Industry

There are two important moments in the automotive industry: mass production and lean manufacturing.

- Mass production

The era of mass production was sometimes called “the second industrial revolution”. Mass production was popularized in the early 20<sup>th</sup> century by Henry Ford at Ford Motor Company. Mass production of assemblies typically used electric-motor-powered moving conveyor belts to move complete or semi-complete products to workers, who performed repetitive and standardized tasks (Duguay, 1997).

- Lean manufacturing

The aim of lean manufacturing is continuous elimination of all waste in the production process. These sources of waste include over-processing, overproduction, excessive inventory, waiting (production line stoppage), unnecessary transportation, unessential motion, defects, rework, and knowledge disconnection. The main benefits of lean manufacturing are lower production costs, increase in output, product quality improvement, enhanced operator safety, and shorter production lead times. Toyota, as a pioneer of lean manufacturing, has employed many aspects of this method in the Toyota Production System (TPS) since the 1950s (Herron & Hicks, 2008).

### **1.3. Future Challenges and Opportunities in Automotive Industry**

Apart from sales volume growth after the 2008 recession, several challenges will shape the automotive industry in the short and medium-term (Hirsh et al., 2015; McKinsey, 2013). These include:

1. Complexity and cost pressure: The increase in national and international regulations with respect to environmental and safety standards will raise costs but also increase complexity, as they need to be managed apart from domestic markets. For example, U.S. Corporate Average Fuel Economy (CAFÉ) standards that will be effective in 2016 are projected to add as much as \$1,000 to the production cost of a vehicle, according to the National Automobile Dealers Association. Only a minority of auto buyers are willing to pay for more environmentally friendly choices such as electric vehicles, so the cost pressure falls largely on original equipment manufacturers (OEMs). On the other hand, OEMs have to develop alternative powertrain technologies for lower-emission vehicles without knowing what will end up being the prevailing technology of the future, which requires significant investment. Given all these

pressures, the automakers must be looking for some innovative ways to boost the productivity while maintaining the quality and meeting the domestic and international standards.

2. Customer differentiation: Customers seem to prefer many options among vehicles that fill an increasing number of finely differentiated market niches. Particularly in the U.S. market, the number of different vehicles available to consumers has increased dramatically over the years. The growing number of derivatives serving different vehicle segments and markets raises investment and complexity. It will be more difficult for OEMs to differentiate themselves with new features while extracting economic value from these forces. Hence, the automakers have to look at strategies to increase customer satisfaction by creating differentiation through different cars' variation while being profitable.

3. Increased electronics and software content: The cost of electronics and software content in autos has been reduced by 20 percent of the total cost from a decade ago. More importantly, electronics systems continue to contribute more than 90 percent of innovations and new features. All major OEMs are targeting traditional product areas such as quality and safety; infotainment provides a way for OEMs and suppliers to differentiate their products.

The increasing importance of infotainment and telematics systems is disruptive for OEMs and traditional suppliers, putting a premium on innovation and changing the ways that industry players design and develop new products and services. Software breakthroughs are becoming as critical as hardware innovation, and competition is increasingly coming from nontraditional players. Ever more vital software content also has accelerated the pace of change in products and features. Whereas the time frame for new vehicle launches is typically three to four years, the cycle for new software iterations, often driven by interactivity with mobile devices, is measured in months.

#### **1.4. Research Motivation**

Considering the above challenges, it was decided to investigate the ways that automakers can improve productivity and utilize flexible manufacturing to create customer differentiation during this competitive era and reduce the new product launch lead time. Although considerable research has been devoted to identifying factors that affect production performance, less attention has been paid to a comprehensive practical study on these factors for American, Japanese, and

joint venture companies. Therefore, the researcher decided to identify the important productivity factors and measures in an automotive generic value chain.

### **1.5. Research Objectives**

The main objective of this research is to define the strategies and techniques to assist automotive manufacturers to enhance and improve the productivity of the production processes. Here is a list of detailed research objectives (RO):

- RO1: Analysis of the product, process, and system level factor that impact productivity in the automotive sector.
- RO2: Analysis and determination of productivity measures commonly used in the automotive sector of manufacturing as well as the productivity measures associated with critical activities on the basis of Porter's value chain.
- RO3: Analysis and determination of the factors that most effectively influence the commonly used productivity measures and their variation in terms of car ownership demographics.
- RO4: Developing statistical models to estimate the productivity measures associated with process variables in terms of car ownership demographics.
- RO5: Validate the statistical models and determine the most effective method in terms of estimation potential.
- RO6: Analyzing the relevant outperforming residuals in the statistical models to determine the existence of specific pursuant strategies that they adhere to.
- RO7: Developing a Mixed Integer Programming (MIP) model to specify target values for the productivity measures determined for focused research
- RO8: Developing a software program that utilizes the developed statistical models and is synergistically integrated with a Mixed Integer Programming (MIP) model to specify target values for the productivity measures determined for focused research.

However, some of these research objectives (RO1-RO2), per se, aren't the final aims of the research but help to build a systematic and robust guideline to move forward. Since it's not clear which organization's activity has the highest importance in productivity improvement, the

common effective factors on productivity improvement are identified and mapped in Porter's value chain. Then the most important activity that can improve productivity is defined and common productivity measures for this activity will be explored, accordingly.

These research objectives will be addressed during the literature survey, exploratory data analysis, and final data analysis.

## **1.6. Research Organization**

This dissertation is organized as follows: chapter 2 reviews the existing literature, chapter 3 addresses research objectives and approaches, chapter 4 presents current work and preliminary results, and chapter 5 presents model improvements, chapter 6 discusses different statistical analysis, chapter 7 addresses mixed integer programming, and chapter 8 discusses conclusion, contributions, limitation, and future research.

## **1.7. Conclusion**

The automotive industry is one of the largest industry sector worldwide and domestically in terms of annual turnover and number of employees. There have been two main revolutionaries in automotive industry as mass production and lean manufacturing. Production activity is one the most crucial processes in any generic value chain, and hence potentially could be a good area to target for productivity improvement. There are eight research objectives that will be addressed in this study. Although considerable research has been devoted to identifying strategies and techniques that affect production performance, less attention has been paid to a comprehensive study on these factors. This study determines the factors that have a statistically significant effect on the production process productivity and presents a framework that can help automakers enhance their productivity.

## **2 : Literature Survey**

### **2.1. Introduction**

In this chapter the origin of the word productivity, automotive industry history and its production process, Porter's value chain, and effective factors on productivity are briefly discussed. Then, the effective factors on productivity are mapped in Porter's value chain to define what activity has the most significant impact on productivity improvement. Then, through a comprehensive literature survey the production labor productivity measures for automotive industry are reviewed.

The first two research objectives (presented in chapter one) are addressed accordingly. Furthermore, through a robust procedure the most common labor productivity is selected and discussed in detail. Finally, the important factors on the most common production productivity indicator and the current research about it are reviewed.

### **2.2. Origin of the Term Productivity**

The word productivity is used often, however; there is a lot of debate about its definition and how vitally important it is. The literature indicates some vagueness and how the topic is important.

The term productivity was used for the first time by Quesnay in 1766 (Steiner, 1987). Since then it has been utilized in many different contexts and areas. Chew defined productivity as the relation of output (such as produced goods) to input (such as consumed resources and labors) in the manufacturing operation (Chew, 1988). Bernolak (1997) presented a useful explanation of productivity for a manufacturing facility. He defined productivity as "how much and how well we produce from the resources used". Increased productivity requires producing superior goods (in quality or in number) from the same resources. The same effect can be achieved by using fewer resources to produce the same amount or quality of goods. By resources, Bernolak means "all human and physical resources" that play a role in the production process. These include machinery (fixed and moving), labor, raw materials, tools, inventories, and assets like land and buildings (Bernolak, 1997).

The summary of the word productivity's origin has been illustrated in Table 2.1.

**Table 2.1. Summary of the word productivity’s origin (adopted from (Tangen, 2005))**

Productivity definition	Reference
Productivity = faculty to produce	(Littre´, 1883)
Productivity is what man can accomplish with material, capital and technology. Productivity is mainly an issue of personal manner. It is an attitude that we must continuously improve ourselves and the things around us	(Japan Productivity Centre, 1958)
Productivity = units of output/units of input	(Chew, 1988)
Productivity = actual output/expected resources used	(Sink & Tuttle, 1989)
Productivity = total income/cost þ goal profit	(Fisher, 1990)
Productivity = value added/input of production factors	(Aspen et al., 1991)
Productivity is defined as the ratio of what is produced to what is required to produce it. Productivity measures the relationship between output such as goods and services produced, and inputs that include labor, capital, material and other resources	(T. Hill, 1993)
Productivity (output per hour of work) is the central long-run factor determining any population’s average of living	(Thurow et al., 1993)
Productivity = the quality or state of bringing forth, of generating, of causing to exist, of yielding large result or yielding abundantly	(Koss & Lewis, 1993)
Productivity means how much and how well we produce from the resources used. If we produce more or better goods from the same resources, we increase productivity. Or if we produce the same goods from lesser resources, we also increase productivity. By “resources”, we mean all human and physical resources, i.e. the people who produce the goods or provide the services, and the assets with which the people can produce the goods or provide the services	(Bernolak, 1997)
Productivity is a comparison of the physical inputs to a factory with the physical outputs from the factory	(Kaplan & Cooper, 1998)
Productivity = (output/input) × quality factor = efficiency × utilization × quality	(Al-Darrab, 2000)
Productivity is the ability to satisfy the market’s need for goods and services with a minimum of total resource consumption	(Moseng & Rolstadås, 2001)

Mistereck et al. (1992) stated that productivity improvements can be basically caused in five different relationships (Mistereck et al., 1992):

1. Managed growth: Output increases faster than input; the increase in input is proportionately less than the increase in output.

2. Working smarter: More output from the same input
3. Ideal: More output with a reduction in input (which is not easy to achieve)
4. Greater efficiency: Same output with fewer inputs
5. Managed decline: Output decreases, but input decreases more; the decrease in input is proportionately greater than the decrease in output

### **2.3. Automotive Industry History and Production Process**

The main highlights in the history of the auto industry are reviewed in this section. The earliest vehicles, powered by sails and methods other than animals and humans were recorded as early as the 1600s. However, the invention of the steam engine in the 1700s and the gas and gasoline engine a century later led to the development of the engine that would power the first true cars. These engines were later improved to include four-stroke internal combustion. The auto industry in Europe and North America was established because of these advancements (D. Bradley et al., 2005).

Technological development post 1890, such as steering wheels and accelerator pedals, increased ease of use for the consumer. Furthermore, developments such as airbags and ABS brakes, create a safer product. These advancements fostered the growth of the auto industry. At the same time, infrastructure enhancements such as roads, service stations, and regulations like traffic laws and driver's license made driving safer that was compatible with the products being made. The first famous vehicle, Ford's Model T, was developed and popularized at 1900s (D. Bradley et al., 2005). Model T is illustrated in Figure 2.1.

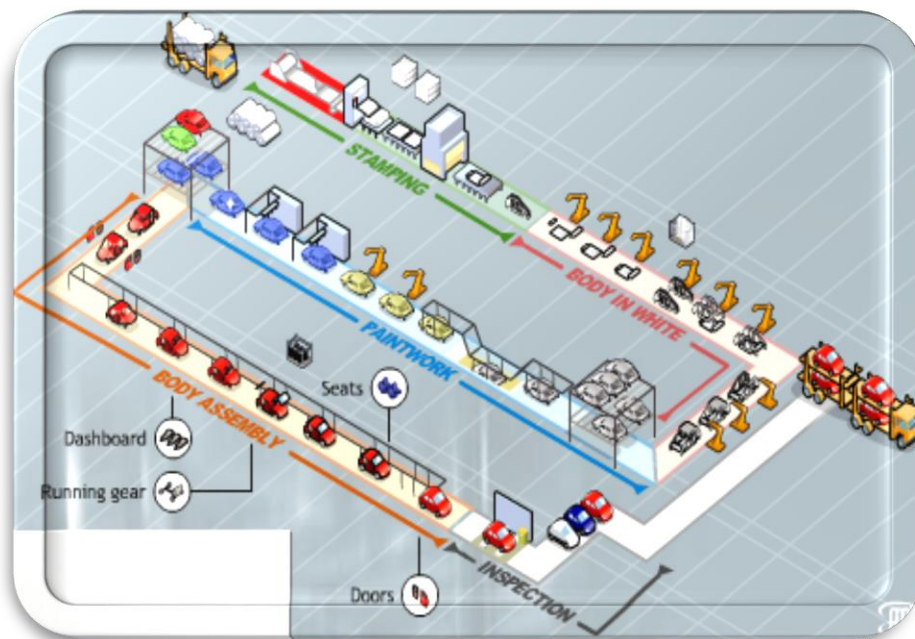


**Figure 2.1. Model T**



Any car manufacturer irrespective of their products (such as light cars, trucks, vans, and sport cars) have almost a common production process. These are as,

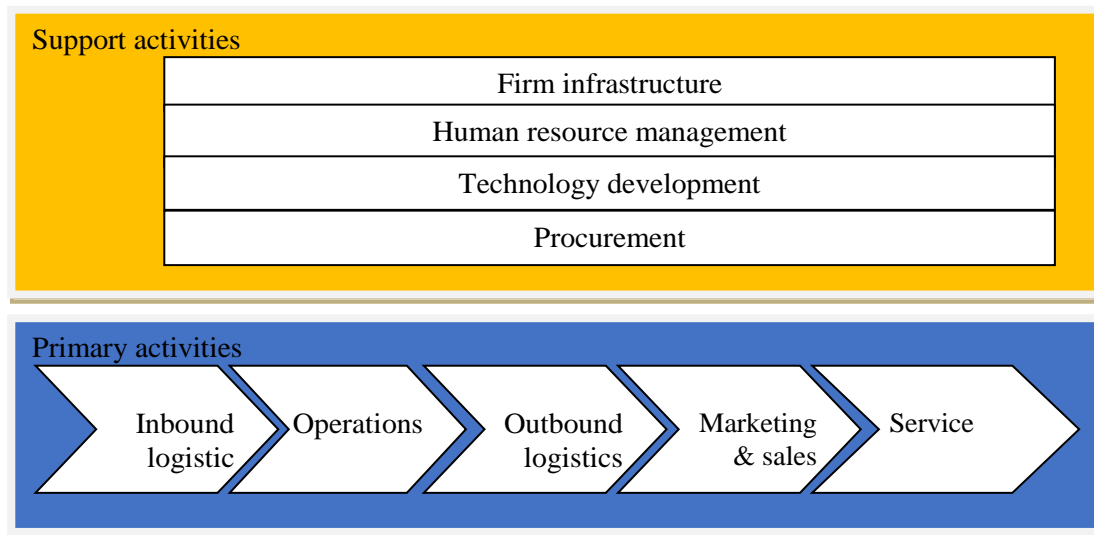
1. Pressing: Steel sheets are processed in a pressing machine. A variety of large and small automotive parts are created in rapid succession.
2. Body shop: After press processing steel sheet components is welded and the framework of a highly-accurate automobile body is created.
3. Paint shop: Several layers of paint are applied to the assembled body. The body is given a lustrous and beautiful finish.
4. Resin molding: Using injection molding equipment, the bumper, fuel tank, instrument panel, etc. are created from resin.
5. Assembly shop: Once painting has been finished for the body, several thousand parts are attached including interior components, instruments, electrical wiring, engine, and tires. The end result is a finished and drivable car.
6. Inspection: Each completed car is subject to a rigorous and multifaceted inspection for various parts and their functionality such as brakes, headlights, emissions, etc. The cars are then shipped as completed vehicles with outstanding quality. A schematic view of an automotive manufacturing process is illustrated in Figure 2.2.



**Figure 2.2. Automotive manufacturing process**

## 2.4. Porter's Value Chain

Each company (no matter if it's a production or service organization) has a collection of activities such as marketing, product design, production, delivery, and product support. All these activities can be presented using the company's value chain. The concept was first introduced by Michael Porter (1985). He considered each company to have two types of activities: support and primary (Porter, 1985). Different primary and supportive activities categories have been shown in Figure 2.3.



**Figure 2.3. Porter's value chain**

Here are the detailed of primary and support activities:

- **Primary activities**

There are five basic categories of primary activities. Each category could be divisible into a number of distinct activities that depend on the particular industry and organizational strategy. Here are some explanations about each activity:

1. **Inbound logistics:** Activities associated with receiving, storing, and disseminating inputs to the product, such as material handling, warehousing, inventory control, vehicle scheduling, and returns to suppliers.
2. **Operations:** Activities associated with transforming inputs into the final product, such as machining, assembly, equipment maintenance, testing, facility operations, and packaging.

3. **Outbound logistics:** Activities associated with collecting, storing, and physically distributing the product to buyers, such as finished goods warehousing, material handling, delivery vehicle operation, order processing, and scheduling.
4. **Marketing and sales:** Activities associated with providing a means by which buyers can purchase the product and inducing them to do so, such as advertising, promotion, sales force, quoting, channel selection, channel relations, and pricing.
5. **Service:** Activities associated with providing service to enhance or maintain the value of the product, such as installation, repair, training, parts supply, and product adjustment.

- **Support activities**

Support activities can be divided into four basic categories. As with primary activities, each category of support activities is divisible into a number of distinct value activities that are specific to a given industry and organizational strategy. Here are some explanations about each activity:

1. **Firm infrastructure:** Firm infrastructure consists of a number of activities including general management, planning, finance, accounting, legal, government affairs, and quality management. Infrastructure, unlike other support activities, usually supports the entire chain and not individual activities. Depending on whether a firm is diversified or not, firm infrastructure may be self-contained or divided between a business unit and the parent corporation. In diversified firms, infrastructure activities are typically split between the business unit and corporate levels (e.g., financing is often done at the corporate level while quality management is done at the business unit level). Many infrastructure activities occur at both the business unit and corporate levels.
2. **Human resource management:** Human resource management consists of activities involved in the recruiting, hiring, training, development, and compensation of all types of personnel. Human resource management supports both individual primary and support activities (e.g., hiring of engineers) and the entire value chain (e.g., labor negotiations).
3. **Technology development:** Every value activity embodies technology, be it know-how, procedures, or technology embedded in process equipment. The array of technologies employed in most firms is very broad, ranging from those technologies used in preparing documents and transporting goods to those technologies embedded in the product itself.

4. Procurement: Procurement refers to the function of purchasing inputs used in the firm's value chain, not to the purchased inputs themselves. Purchased inputs include raw materials, supplies, and other consumable items as well as assets such as machinery, laboratory equipment, office equipment, and buildings.

Nowadays, outsourcing of these activities is a common approach. Many corporations do a diagnostic analysis on and outsource some activities that are not their core competitive advantages and keep the most sensitive ones inside.

## **2.5. Effective Factors on Productivity**

First, the effective factors on the productivity will be defined in the automotive sector based on the literature survey. The first research objective (RO1) will be addressed in this section accordingly. Finally, these factors were mapped into Porter's value chain in order to define the firm's most important activity on productivity improvement.

### **2.5.1. Quality Management Systems (QMS)**

Two main areas for this category can be considered as,

- **Total Quality Management (TQM), ISO TS 16949, and ISO 9001**

Golhar and Deshpande (1999) compared productivity performance of the Canadian (n=43) and American (n=95) automotive parts manufacturers that implemented total quality management (TQM) philosophy (by conducting a survey). They identified three different categories of productivity measures such as internal business, financial, and customer related. They found the organizations that implemented TQM improved their performance on various productivity measures; however, there was a significant difference between the two countries on some of those measures. The scholars found that there was a significant relation between customer orientation and internal business process improvement. Unlike American manufacturers there wasn't statistical evidence for the existence of this relation for the Canadian manufacturing facilities (Golhar & Deshpande, 1999).

Quazi and Jacobs conducted a survey of 28 Singaporean companies in both production and service categories. In this study, 47% of production and 78% of service companies believed that the main reason for improving arose from the implementation of quality management systems (ISO 9001:2000) (Quazi & Jacobs, 2004). Furthermore, 95% of respondents agreed

that implementing the quality management system improved the customer relationship management, and 83% of respondents agreed that implementing the quality management system improved their competitive position. Finally, 65% of respondents believed that deployment of these systems enhanced the job training programs.

- **Six Sigma**

Soković et al. (2006) studied the application of six sigma tools and methodology (define, measure, analyze, improve, and control (DMAIC)) within an automotive parts manufacturing (Soković et al., 2006). The scholars have pointed out several significant achievements such as,

- Reducing tool expenses by 40%
- Reducing the costs of poor quality (COPQ) by 55%
- Reducing labor expenses by 59%
- Reducing production time by 38%
- Reducing the index cost/volume by 31%

Hence, systematic application of Six Sigma DMAIC tools and methodology in a manufacturing facility result in similar achievements.

### **2.5.2. Outsourcing of the Activity**

Direction (2005) stated that Chrysler outsourced most of its manufacturing process and reduced the warranty costs by 34% during three years. Customer surveys also have shown that the quality of products improved remarkably and even exceeded Mercedes-Benz's products in some cases (Direction, 2005).

Another empirical study conducted by Jiang et al. (2006), based on a sample of 51 publicly traded firms that outsourced parts of their operations between 1990 and 2002, found no evidence that outsourcing will improve a firm's productivity and profitability. However, the research provided evidence that outsourcing can improve a firm's cost-efficiency (Jiang et al., 2006).

### **2.5.3. Common Platform Application**

Gopal et al. (2013) used data from North American automotive plants from 1999-2007. They found that a new product launch penalized the plants, resulting in a productivity loss of 12-15%.

However, they also found that body shop flexibility and using the common platform strategy could compensate for this (Gopal et al., 2013).

Tay stated that using platforms in design of a new product reduced the product development time sometime between 30% to 50%, depending on the company and product (Tay, 2003). Therefore, this method can be an efficient way to reduce the manufacturing facility cost while developing a new product; however, the quality standards can be met also.

#### **2.5.4. Participatory Ergonomics in Product Development**

Nowadays, participatory ergonomics (PE) has become a widespread field of research and practice. Dimensions of PE were identified as designing equipment or tasks, designing jobs, teams or work organizations, and formulating policies or strategies. Participatory ergonomics programs seek to maximize the involvement of the workers in the process based on the simple fact that a worker is an expert on his/her job (Broberg et al., 2011).

Sundin et al. (2004) found several significant improvements that have been achieved by developing a participatory ergonomics in a bus factory assembly line. Assembly time of cables on the chassis reduced from 55.6 to 14.3 minutes by using computer simulation and operators' ergonomic status improvement. On the other hand, the assembly time of the cooling system on the chassis was decreased from 3.31 to 1.3 minutes. Furthermore, the raw material handling difficulties were also reduced by 15% (Sundin et al., 2004).

#### **2.5.5. Robots' Application in Manufacturing Lines**

Robots have been used broadly in several complex and sensitive manufacturing process, particularly in the body and paint shops. Robots' application can create a broad flexibility while enhancing the process quality and productivity.

Using seven robots instead of manual operations in BMW's Regensburg paint shop, the rate of reworking was dropped from 30% to 10%. Meanwhile, automation considerably reduces paint shop employees complaint (Kochan, 2005).

In another study, glass-integrated curved seal extrusion (CSE) technology was introduced. Utilizing the robots for the automobile-window mole demonstrated of a 50% reduction in manufacturing costs in one of Toyota's plants (Sakai & Amasaka, 2007). The authors stated that using this technology brought high productivity and quality assurance in a global production.

Bogue (2013) investigated the application of robots in the automotive industry and explored its advantageous. He found that robotic vision technology was playing an important role to improve quality and productivity in the automotive industry (Bogue, 2013).

#### **2.5.6. Information Technology (IT), Enterprise Resource Planning (ERP), and Simulation**

“Information technology” can be defined in various ways. Among the most common is the BEA’s (U.S. Bureau of Economic Analysis) category “Office, Computing and Accounting Machinery (OCAM) which consists primarily of computers. Some researchers look specifically at computer capital, while others consider the BEA’s broader category, “Information Processing Equipment (IPE).” IPE includes communications equipment, scientific and engineering instruments, photocopiers and related equipment. Software and related services are sometimes included in the IT capital. Recent studies often examine the productivity of information systems staff, or of workers who use computers at work (Brynjolfsson & Yang, 1996).

An experimental study was conducted on 60 American companies by Kini; it was found that there was a positive correlation between the investment in IT infrastructure and performance improvement. Other benefits of IT implementation were enhancing the quality of products, capability of process to produce the lower tolerances, production yields, production capacity, and economies of scale. It also increased employees’ participation in knowledge acquisition (Kini, 2002).

May and Carter (2001) used a novel advanced telecom system among 40 engineers who were in different countries such as Italy, Germany, England, and France. This system reduced the processing time (from 10% to 50%) in different stages of a new product launch (May & Carter, 2001). Meanwhile, the new product development cost was cut by 20% and £90 million was saved. Eventually, the total savings was about £1 billion which was a significant achievement; however, the researcher is not sure if all this savings, originated only by utilizing advanced telecom system.

Kuo (2014) conducted a case study and expert interviews for a Taiwanese manufacturing company to reveal the tangible (financial structure, operation capability, short-term solvency, profitability, and cash flow) and intangible (financial effectiveness, sales and marketing effectiveness, production efficiency, business strategy effectiveness, and human resources) effects of ERP implementation. His study had various categories such as firms’ reputations, productivity, product

quality, and decision quality (Kuo, 2014). He found that tangible effects and business processes have been significantly improved after ERP implementation.

El-Khalil (2015) utilized a discrete simulation model to improve productivity at one of the original equipment manufacturer's (OEM's) body shops for one of the "Big Three" in North America. He proposed two model scenarios: the first scenario considered changing the model mix percentage leading to a cost improvement of \$1.6 million/annually USD while the second focused on the top bottleneck processes flexibility with a return on investment (ROI) of 497 percent. His proposed system, based on simulation data, increased the body shop's throughput by 2.4 jobs per hour (JPH), a 2.6% overall improvement, reduced scrap by 0.3%, and improved uptime to 79.5%.

### **2.5.7. Human Resource Management (HRM)**

Labor cost usually consists of 12% to 15% of the total cost to produce each car (Dorf, 1998). Edward Dynson estimated that 68% of U.S. economic growth for the years 1929 to 1982 was because of labor productivity improvement. In Japan, annual labor productivity growth was 6% (yearly) from 1950 to 1981. Therefore, labor productivity can play a key role in the area of productivity improvement.

Macduffie (1996) sought to explore if innovative HR practices affect performance, not individually but as interrelated elements in an internally consistent HR bundle or system. He analyzed data from a 89-90 survey of 62 automotive assembly plants testing three indices representing distinct bundles of human resource and manufacturing practices (J. P. MacDuffie, Sethuraman, K., Fisher, M. L., 1996). Macduffie concluded that the plants employed flexible production with high HR practices commitment, and low inventory outperformed plants using mass production methods.

### **2.5.8. Modularization**

Baldwin and Clark (2009) defined modularity as the process of "building a complex product or process from smaller subsystems that can be designed independently yet function together as a whole". The authors defined how Mercedes-Benz utilized the modular concept for their sport-utility assembly plant in Alabama, USA. Mercedes-Benz used a particular structure module containing heating and air-conditioning systems, airbags, the steering column and the wiring harness, and the instrument cluster to illustrate the flexibility associated with the modular approach to car assembly (Baldwin & Clark, 2009).



Doran (2003) studied the modularization deployment within the automotive sector supply chain for the UK firms. The study focused on the modularization implementation for the first-tier suppliers and defining the requirements for doing so (Doran, 2003).

Voordijk et al. (2006) considered the applicability of modularity in three pillars such as product, process, and supply chain architecture through a case study for three contractors' companies. The authors found the three-dimensional modularity concept works well, it means that the degree of modularity in the final output product has a one-to-one correspondence with the degree of modularity in transformation processes and supply chains (Voordijk et al., 2006). The research was an exploratory study without any mathematical or statistical analysis.

### **2.5.9. Lean Manufacturing and Quality Tools Deployment**

Lean manufacturing could be considered as one of the miracles' of the 20<sup>th</sup> century for manufacturing operations. Many organizations have tried its deployment along with the organizational cultural change.

The North East Productivity Alliance (NEPA) has disseminated some selected lean manufacturing tools and techniques into companies in the North East of England. The aim of NEPA was to help companies improve productivity through applying lean manufacturing management practices and knowledge. Herron and Hicks (2008) studied 15 companies in the UK. Initial results showed that the savings were eight times greater than total costs (Herron & Hicks, 2008). They found that the ability of the change agents and management commitment were the key determinants of success.

Brondo and Baba (2010) studied General Motors' Lansing Grand River Assembly (LGRA) plant as a case study for three years (2003-2005). This plant was the first GM vehicle assembly plant in North America that implemented GM's Global Manufacturing System (GMS) that was built based on lean manufacturing. The authors explored cultural phenomena to study the LGRA's perceived initial success with lean manufacturing. The authors found how organizational processes beyond the control of a single plant may lead to breakdown in participatory structures. It could create some risks for sustainability of lean manufacturing for the long-term (Brondo & Baba, 2010).

Trimble et al (2013) conducted a survey through 161 UK automotive suppliers to study the level of understanding and use of 33 continuous improvement tools in an organization and identifying the barriers that cause the failure to achieve the expected results. The authors explored the elements of continuous practices methodology selection, training, and methodologies and tools implemented. The research found the main failures were lack of communication across the

organization from the top level to shop floor level employees, lack of existence of non-strategic approaches, and a lack of appropriate resources (Trimble et al., 2013).

Lean manufacturing dimensions could be considered as continuous/one-piece flow, eliminate waste (kaizen), error proofing/poka-yoke, new process equipment/technology, preventative maintenance (total productive maintenance), pull-based production/Kanban, quick changeover, reduce cycle time, reduce inventory (just in time), reduce lot size, reduce setup time, remove bottlenecks, single supplier focus, and 5S. Here are some selected important lean tools that have been often applied more frequently in the manufacturing plants.

#### **2.5.10. Just in Time (JIT)**

Gonzalez-Benito and Spring (2000) studied 150 Spanish suppliers and found that by involving the suppliers in the new product design phase the product quality improved; however, the product development time was reduced also. They also found that just in time (JIT) implementation cut the cost of administration, inspection, and material handling even though, the quality and productivity were improved (Gonzalez-Benito & Spring, 2000).

By studying 15 Canadian automotive parts manufacturers that practice JIT and non-JIT methods, researchers concluded that organizations that used a variety of productivity measures experienced greater efficiency and profitability, regardless of plant employed (Callen et al., 2005). Furthermore, they showed if the productivity measures were business-driven the efficiency and profitability impacted more significantly. Finally, they found that productivity measurements facilitate the JIT practices and performance outcomes.

#### **2.5.11. Total Productive Maintenance (TPM)**

As many organizations aim to become world class, the importance of facility maintenance management in the manufacturing environment has increased rapidly. For many organizations, it is important to respond to the global competitive pressure by seeking to increase the productivity, maximizing the overall equipment effectiveness, and pursuing an effective and efficient maintenance program.

Cholasuke et al. (2004) conducted a pilot survey through 18 UK manufacturing facilities to explore the maintenance opportunity improvement. They found that one-third of the organizations seriously follow good maintenance management practices and realized its full benefits. Their study

revealed that the involvement of the management is a crucial element to give the right guidance and direction to the maintenance function (Cholasuke et al., 2004).

#### **2.5.12. 5S**

The 5S method is a Japanese tool that developed by Osada and Hirano. The 5S provide a fertile ground for the proliferation of continuous improvement and TPS, its pillars are sort (seiri ), set in order (seiton), shine (seiso), standardize (seiketsu), and sustain (shitsuke).

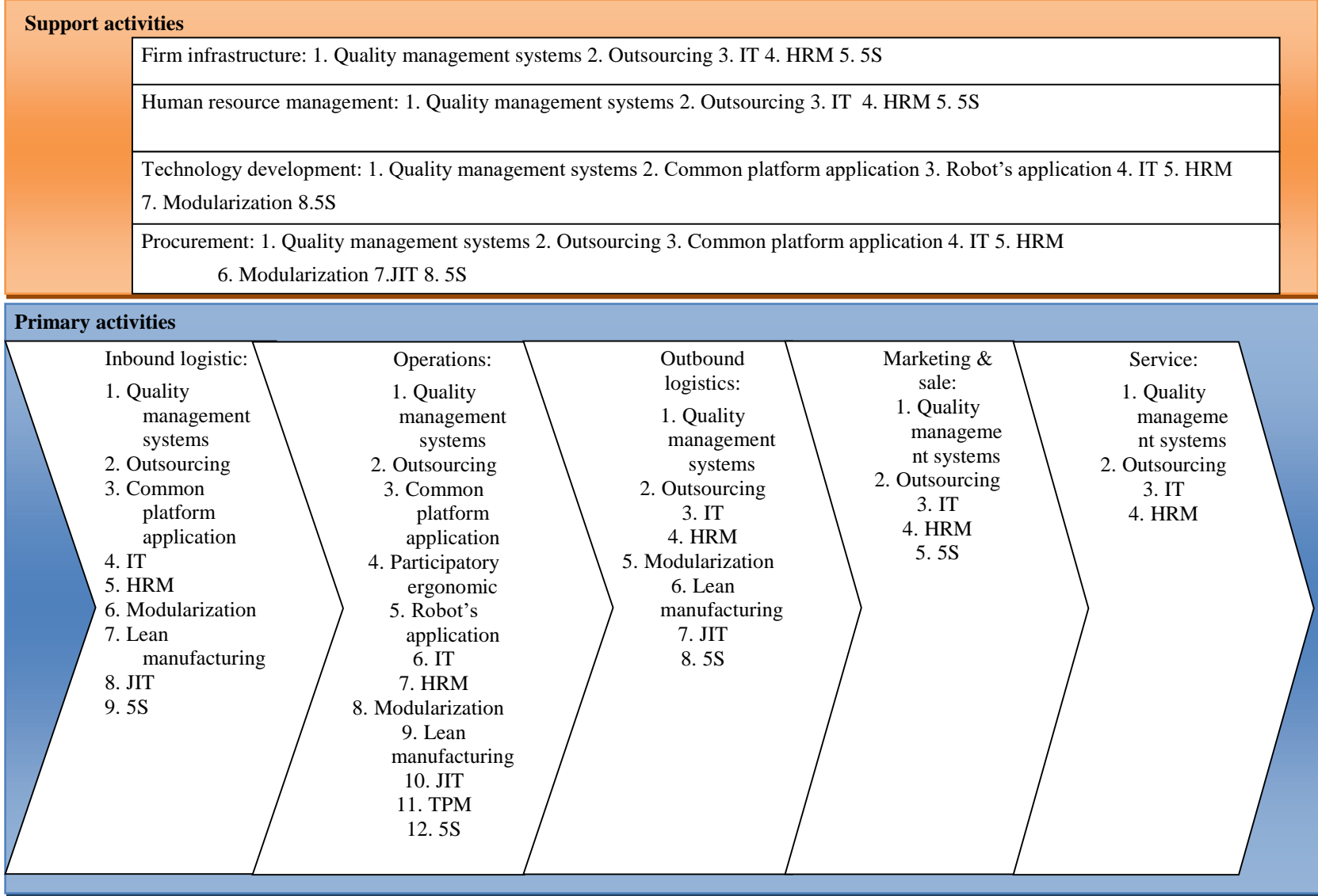
Bayo-Moriones et al. (2010) conducted a questionnaire survey of 203 Spanish manufacturing plants to explore the relationship between 5S use and performance. The results showed a positive association with some operational performance measures, particularly those referring to productivity and quality (Bayo-Moriones et al., 2010). On the other hand, the scholars found a positive correlation between the 5S application and some contextual factors such as plant's size, the plants' integration in a multinational group, the firms' quality program, and utilized technology.

#### **2.6. The Main Activity of Porter's Value Chain**

One of the goals of this research is defining the most effective activity based on Porter's value chain, either support or primary activities, to improve productivity for the automotive industry. So far the first research objective (RO1) has been addressed and important effective factors on in the automotive industry's productivity were reviewed.

In order to address the second research objective (RO2), it's needed to define which activity has the highest significance on productivity improvement. Hence, the effective factors on productivity have been defined and mapped in Porter's value chain in order to reveal which activity has the highest impact on the productivity improvement. The summary of this mapping has been shown in Figure 2.4.

Figure 2.4 shows that the most important activity on productivity improvement is operation activity by 12 effective factors (all the studied factors) and followed by inbound logistics, outbound logistics, and procurement activities by nine, eight, and eight effective factors, respectively. The quality management systems, outsourcing, IT, HRM, and 5S impact all the supportive and primary activities.



**Figure 2.4. Porter's Value Chain**

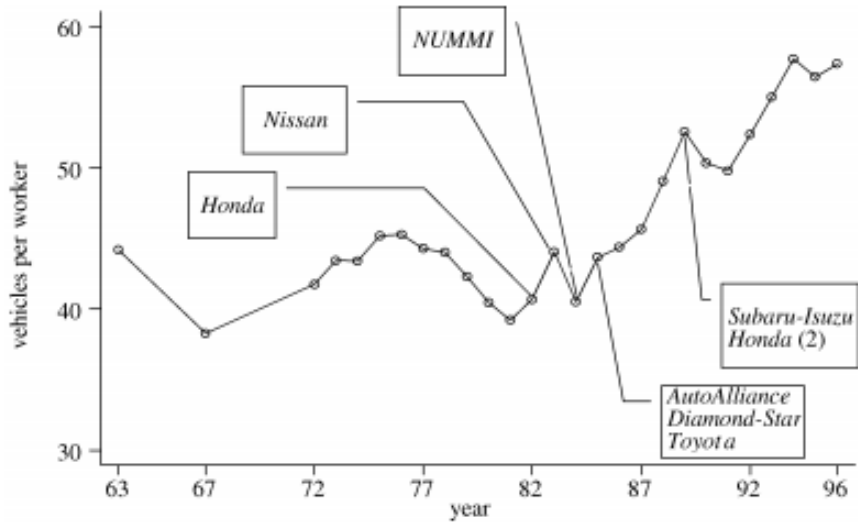
## 2.7. Labor Productivity Measures

So far operation/production activity has been defined as the most effective activity on productivity improvement. Hence, to address the second research objective (RO2) various productivity measures for production activity are studied in this section.

Labor is a key cost element that has to be addressed by the original equipment manufacturers (OEMs) after direct material purchases (Uludag & Weisenstein, 2005). Labor cost usually consists of 12% to 15% of the total cost to produce each car (Dorf, 1998). Four key labor-related indicators that are commonly used are reviewed below.

- **White collar labor force percentage (WCLFP)** measures the percentage of the white collar employees. White collar employees at “Big Three” auto companies (General Motors, Ford, and Chrysler) include: (1) executives in the top management positions; (2) middle management at the divisional and plant levels; (3) professional staffs such as technical professionals; (4) nonprofessional staff such as clerical employees; and (5) first-line supervisors at the plant line (J. P. MacDuffie, 1996). Pucik (1984) compared the white collar labor force percentage among the U.S. and Japanese automobile industries. He found that white collar employees percentage ranged from 26-33% and 29-33% for American and Japanese car makers, respectively (Pucik, 1984). The author found the middle management (12-14%) and professional staff (14%) have the highest rate of white collar employees for American and Japanese companies, respectively.
- **Vehicles per employee (VPE)** measures the number of cars produced in the manufacturing facility per production employee in the data year. In 2002, World Market Research Centre, (WMRC) initially measured this indicator for European automakers on an annual basis. There is a similar indicator that measures the number of vehicles sold per employee of any type.

Average number of vehicles produced per worker for U.S. car makers is illustrated in Figure 2.5, from 1963 to 1996 (J. Van Biesebroeck, 2003). Figure 2.5 shows this measure has been improved after the 1980's, right after the Japanese arrival in the U.S. car industry. It can be concluded that the entry of more productive plants caused the U.S. automotive industry to move significantly towards productivity.



**Figure 2.5. Average vehicles produced per worker from 1963 to 1996**

Nissan's Sunderland car plant surpassed all other European automakers in this measure in 2000 by almost 20%. The improvement at Nissan was achieved as the company succeeded in slightly reducing its total workforce while at the same time adding a third model, the Almera, and increasing output by 7% (WMRC, 2002). General Motors (GM) led the other automakers with 42 cars sold per worker; however, Volkswagen (VW) finished unexpectedly far behind at just 18 units per worker (Shunk, 2011).

- **Car assembly and capacity (CAC) utilization** is the total number of vehicles produced per design capacity line rate in December of the data year. CAC can be seen as an indicator of equipment and investment productivity (Harbour, 2002). The annual capacity and utilization can be calculated as,

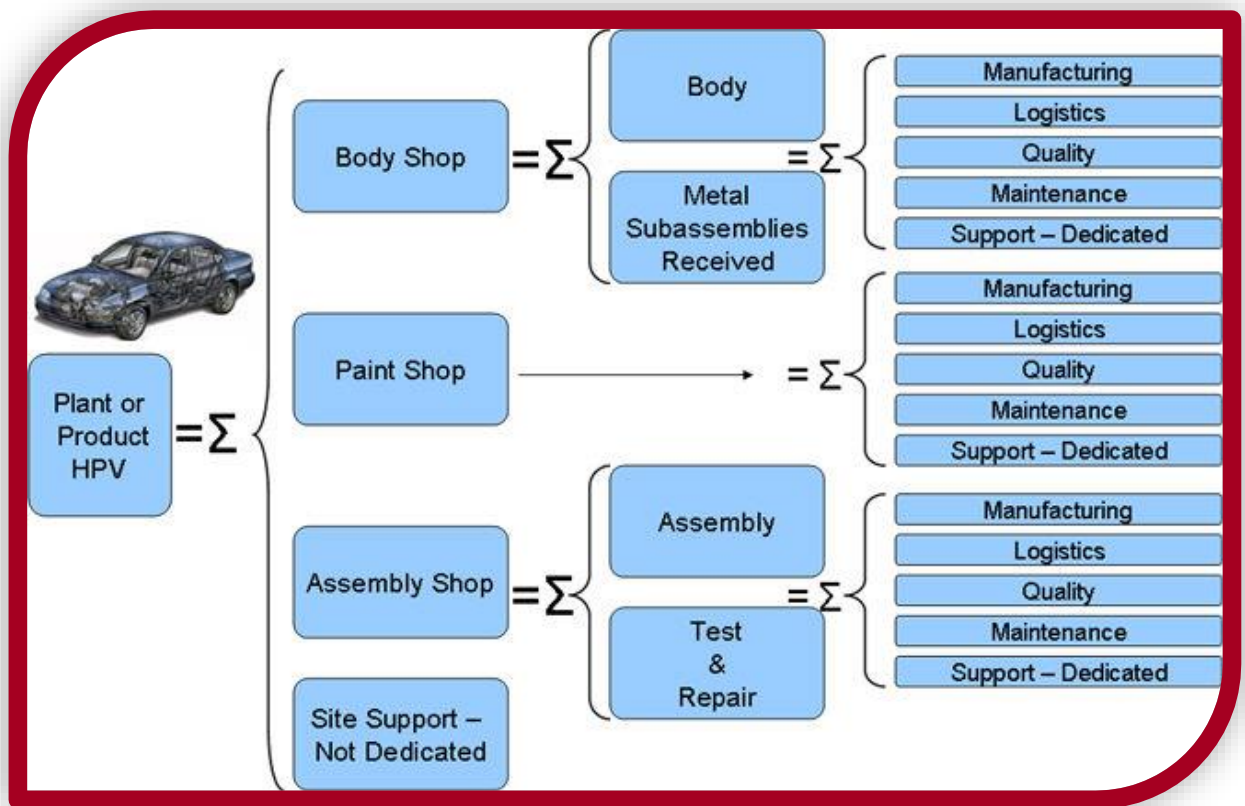
$$\text{Annual capacity} = \text{Capacity line rate} \times 16 \text{ hours per vehicle} \times 235 \text{ days per year}$$

Capacity line rate is based on December unless otherwise noted

$$\text{CAC} = (\text{Actual Production} / \text{Annual Capacity}) \times 100\%$$

The CAC can be higher than 100% if the plant is working in three shifts or operating more than 235 days. For more information regarding the CAC indicator, readers are referred to Harbour Institute Reports.

- Hours per vehicle (HPV)** includes all direct, indirect, hourly, and salary (direct and indirect) work associated with the production and processing of cars through body, paint, chassis, trim, final assembly, blanking, stamping, engine/transmission machining, and engine/transmission assembly (Harbour, 2002). This indicator is the inverse of labor productivity in the automobile industry, which means a lower HPV indicates a more productive plant. The employees who are considered for this measure are categorized into five core areas: manufacturing, logistic, quality, maintenance, and support. A schematic view of the people who are involved on this measure is illustrated in Figure 2.6



**Figure 2.6. A schematic view of the people who are involved for this measure**

The category definitions for the five core areas are:

- Manufacturing** – This includes all direct, indirect, hourly and salary (direct and indirect) work associated with the manufacturing and processing of vehicles through body, paint, trim, chassis, final assembly, blanking, stamping, engine/transmission machining and

assembly. More than 50% of a worker's time must be dedicated to manufacturing for the hours to be included

- **Logistics** - This includes all hourly and salary (direct and indirect) labor associated with material handling and logistics. More than 50% of a worker's time must be dedicated to manufacturing for the hours to be included.
- **Quality** – This includes all hourly and salary (direct and indirect) labor associated with quality and rework processes. More than 50% of a worker's time must be dedicated to manufacturing for the hours to be included.
- **Maintenance** – This includes all hourly and salary (direct and indirect) labor associated with the maintenance process. More than 50% of a worker's time must be dedicated to manufacturing for the hours to be included.
- **Support** - This includes all hourly and salary (direct and indirect) labor dedicated to body, paint, assembly, stamping or machining. More than 50% of a worker's time must be dedicated to manufacturing for the hours to be included.

For more detailed information regarding HPV, readers are again referred to Harbour Institute Reports. The HPV has been measured annually by the Harbour Institute since 1978 for American automakers and since 1995 for European automakers. Some 21 international companies that produced 84.34% of all assembled vehicles in 2014 have participated in the Harbour Institute's survey every year. The current worldwide car makers who are participating in this annual survey (from North America/South America/Europe/Asia) are illustrated in Figure 2.7. Moreover, Uldag and Weisenstein (2005) reported that a large OEM saved 800 million USD in a year by improving its HPV by 13 hours, or 20%.





**Figure 2.7. Worldwide car makers who attends in the Harbour survey**

So far as HPV is the most important production productivity measure, it was decided to study the direct/indirect effective factors on this measure based on the literature review and researchers experience and knowledge. Therefore, the RO2 is addressed and four labor productivity measures defined based on the literature review in which the HPV is the most common production productivity measure.

### **2.8. Studies that Have Been Done to Improve Productivity**

Uldag and Weisenstein (2005) reported that a large OEM saved 800 million USD in a year by improving its HPV by 13 hours, or 20%. Researchers reported the main influential factors were parts modularization, close collaboration between engineering and manufacturing departments throughout all phases of vehicle design, automation, a continuous improvement culture, and some lean manufacturing practices such as just in time, work standardization in shop floor and support activities, and quality tools (Uludag & Weisenstein, 2005).

Taj and Berro (2006) examined a business case study in an automotive body shop at the rear-pan line. The scholars demonstrated how constraint management and lean manufacturing could work together to improve productivity, quality, and efficiency. The authors successfully reduced the

production cycle times from 65 seconds to 55 seconds by removing the idleness of unclamped movements, eliminating non-essential welds, and reducing in the idleness of robot-robot interference wait time (Taj & Berro, 2006).

Biesebroeck (2007) introduced an HPV model that includes outsourcing, variety of vehicles, flexibility, interaction of those variables, scale of economy, and years of production factors (J. Van Biesebroeck, 2007). He used the data from Harbour's survey for 10 years of time horizon (from 1994 to 2004) to estimate the HPV based on the aforementioned factors. He used 840 observations to build his model without considering the car segmentation, plant ownership (American, Japanese, or joint venture), platform sharing strategy, number of available working days, or new product launch effects. He used insourcing as the distance from the North American and country-specific industry centers. However, it could also be defined as the percentage of activities that are handled inside the plant.

Weyer (2011) presented a detailed methodology for calculating and controlling the HPV. He considered some variables which influenced HPV and introduced an effective policy to improve productivity. Variables used in Weyer's model include factory shutdowns, level of personnel absenteeism, manufacturing flexibility, product variability, and the number of working days in a year (Weyer, 2011). Unfortunately, neither mathematical modeling nor statistical analysis was performed on the available data.

Almström and Kinnander (2011) developed a productivity potential assessment (PPA) method to measure the productivity opportunities at the shop-floor level in the Swedish manufacturing industry. The method was applied and tested through a one-day assessment over 70 cases during five years. Two certified analysts evaluated each company's productivity potentials through a standardized work process. The evaluators used a structured checklist including 40 questions in 11 areas as strategic goals, work methods, maintenance, competence, cleanliness and order, material handling, change over, continuous improvements, calculations, planning, and quality (Almström & Kinnander, 2011). The authors found that the PPA is a practical method that is based on a systems perspective, but it lacks practical recommendations for improving a company's productivity status at the national and global levels.

Gupta and Prasad (2011) examined a survey of 76 Indian-Japanese and Indian manufacturing facilities to find the human resource management (HRM) factors that affect productivity and

overall firms' performance. The authors used factorial analysis to examine the various HRM factors that affect the organizations' productivity. They found that talent planning and engagement, talent motivations, and assessment training needs are common factors that affect productivity in both Indian-Japanese and Indian facilities. However, service training and talent acquisition are only significant for Indian- Japanese and Indian, respectively (Gupta &Prasad, 2011). Even though, surveys have several advantages, their validity can sometimes be a real issue. Survey questions are usually standardized so that a broad range of people will understand them. Therefore, survey results may not be as valid as results obtained by using methods of data collection that allow a researcher to more comprehensively examine the topic being studied.

Alden et al. (2006) studied the ways to increase the General Motors Corporation (GMC) production lines throughput performance by initiating a long-term project. GM quantified throughput performance enhancement in the three main areas in the design and operations of its manufacturing systems. They included: (1) developing the algorithms for estimating throughput performance, defining bottlenecks, and optimizing buffer allocation, (2) defining some real-time plant-floor data-collection systems to capture the dynamic data, (3) establishing common processes to identify opportunities and implementing performance improvements using more manufacturing flexibility. Following these activities and some organizational learning, GM saved over \$2.1 billion in 10 countries and more than 30 vehicle plants (Alden et al., 2006).

Zahoor Sarwar et al. (2012) studied the productivity status and trend in Indus Motors and Honda Atlas over a span of ten years, from 2000 to 2010. The authors used Sumanth (1994) and Cobb-Douglas (1928) studies to estimate partial and total factor productivity (TFP), respectively (Cobb &Douglas, 1928; Sumanth, 1984). The scholars used different data such as the number of employees, total man-hours, wages of these employees, fixed capital input, working capital input, cost of energy consumed, cost of materials used, cost of other expenses including taxes, travel expenses, and all other overhead costs. Then, they calculated partial productivities such as labor productivity, material productivity, capital productivity, and total productivity. They found low levels of labor productivity and capital productivity causing huge losses and no growth in those firms (Zahoor Sarwar et al., 2012). Since only two automotive manufacturing facilities were studied (Indus Motors and Honda Atlas), the findings may not be generalized to many other automotive plants.

Gopal et al. (2013) used data from North American automotive plants from 1999-2007. They found that a new product launch penalized the plants, resulting in a productivity loss of 12-15%. However, they also found that body shop flexibility and using the common platform strategy could compensate for this (Gopal et al., 2013). The authors only focused on the plants' productivity during the launch phase and the reduction of its influence in the body shop.

El-Khalil (2014) studied managing and improving the robot spot welding efficiency for North American automotive body shops. He visited different body shop labs and interviewed the facility engineering and managerial personnel over the span of two years. He found that there was a great opportunity to improve HPV at the "Big Three" (General Motors, Ford, and Chrysler) by utilizing flexible systems, common vehicle body parts, and improved processes in the body shop (R. El-Khalil, 2014). In another study, El-Khalil (2015) utilized a discrete simulation model to improve productivity at one of the OEM's body shops for one of the "Big Three" in North America. His proposed system, based on simulation data, increased the body shop's throughput by 2.4 jobs per hour (JPH), a 2.6% overall improvement, reduced scrap by 0.3%, and improved uptime to 79.5%. The research was concentrated on the management decision on buffer sizes, batch size, and the top bottlenecks work station alternative (R. El-Khalil, 2015). The labor productivity, robots uptime, scrap root cause, and maintenance response to downtime were excluded from his study. The study adopted a detailed level view at the body assembly shop only, rather than a global view for all automaker shops, which is of interest in this study.

Al-Tahat and Jalham (2015) investigated the relationships between lean manufacturing and quality and productivity improvement through conducting a survey on 300 Jordanian companies. The scholars considered eight dimensions for lean manufacturing as variability reduction, visual control, poka-yoke, quality at the source, kaizen, 5S, root cause analysis, and total quality management (Al-Tahat & Jalham, 2015). They checked the relation between the aforementioned eight lean production dimensions and quality and productivity improvement. The proposed conceptual model was developed and tested in Jordanian companies which calls its generality into question.

Gunasekaran and Cecille (1999) performed a case study of Valeo, a SME French company located in the United Kingdom that produced wiper systems. The company was a job shop company and produced a high variety of parts with low volume. The authors studied two specific work stations, one assembling jets with hoses and the other assembling jets and hoses with other

devices for head lamp cleaning systems. The authors studied the implementation of three types of productivity improvement strategies including: tools improving, kanban system deployment, and development of an autonomous cell (Gunasekaran & Cecille, 1998). The scholars found the critical successful implementation factors included top management support, using cross functional teams who had a good communication system and clear objectives, workstations' operator effective involvement during the project, and using preventive maintenance for the tools and machines.

Rathilall and Singh (2011) conducted a survey in Durban, South Africa to study the existing lean production tools in an automotive company. The authors' objectives were to define the strengths and weaknesses of existing a lean manufacturing system- considering nine different dimensions such as eliminate the waste, continuous improvement, zero defects, just in time, multifunctional teams, decentralized responsibility, integrated functions, vertical information systems, pulled instead of push system- from employees' point of view (Rathilall & Singh, 2011). Different analysis methods employed for collected data such as frequencies, means, and gap values for descriptive analysis and factor analysis, communalities, and hypotheses testing for inferential statistics. They found that a significant gap existed between the actual lean production deployment at the shop level and the documented standards. The lean manufacturing misconceptions and its deployment inconsistency were a common problem throughout the entire the manufacturing facility.

Table 2.2 was created to present the contributions as well as the limitations of the existing work and demonstrate the need for this research.

**Table 2.2. Contributions and limitations of the existing literature**

<b>Author(s)</b>	<b>Methodology</b>	<b>Research Contributions</b>	<b>Limitations</b>
Uldag and Weisenstein (2005)	Case study	A large original equipment manufacturer (OEM) saved \$800 million by improving its HPV by 20% (13 hours)	No mathematical analysis that could be of help to other car makers were presented

Taj and Berro (2006)	Business case study	The authors considered how constraint management and lean manufacturing could work together to improve productivity, quality, and efficiency	No mathematical analysis that could be of help to other car makers were presented
Biesebroeck (2007)	Statistical analysis on the Harbour data from 1994 to 2004	His model included factors such as outsourcing, variety of vehicles, flexibility, interaction of them, scale of economy, and years of production factors	Segmentation, Ownership, platform sharing strategy, number of available working days, hourly/total employees' percentage, CAC, number of models, and new product launch effects were not considered
Weyer (2011)	Quite descriptive research in nature	Factory shutdowns, level of personnel absenteeism, flexible manufacturing, product variability, and number of working days in a year were studied	Neither mathematical modeling nor statistical analysis were performed
Almström and Kinnander (2011)	Case study on 70 Swedish organizations	They developed a productivity potential assessment (PPA) method to measure the productivity	It lacked practical recommendations for the companies' productivity status at the national and global levels and how come it can be improved

		opportunities at the shop-floor level	
Gupta and Prasad (2011)	Examination of a survey of 76 Indian-Japanese and Indian facilities and performing a factor analysis	They found that talent planning and engagement, talent motivations, and assessment training needs are common factors that affect overall performance of organizations in both Indian-Japanese and Indian firms	In general, the validity of carrying out a survey in a scientific research could be a real issue
Zahoor Sarwar et al. (2012)	Statistical analysis on two Pakistani car makers	Effective utilization of technology could significantly enhance the manufacturing plants' productivity	Since only two automotive manufacturing facilities were used, the findings may not be generalized
Gopal et al. (2013)	Statistical analysis on the Harbour data from 1997 to 2007	He found new product launch penalized the plants resulting in productivity loss of 12% to 15%	The authors only focused on the plants' productivity during the launch phase and the way of diminishing its influence only in the body shop
El-Khalil (2014)	Interviewed with different body personnel and visited several Big Three plants	He found utilizing flexible systems, common vehicles body parts, and improved processes in the body shop would	No mathematical analysis that could be of help to other car makers was presented

		improve body shop productivity	
El-Khalil (2015)	Business case study and doing simulation	His proposed system increased the body shop throughput by 2.4 JPH (2.6 percent overall improvement), reduced scrap by 0.3 percent, and improved uptime to 79.5 percent	Labor productivity factor was ignored in the simulation model

Examination of Table 2.2 reveals that there is a gap in the state of the art concerning the use of a robust statistical method to estimate the HPV and to find the desired settings of the independent variables using an optimization technique such as mixed integer linear/nonlinear programming (MILP/MINLP) in order to achieve a target value for the HPV.

## 2.9. Conclusion

In this chapter, different factors that have some sort of influence on productivity were reviewed. These factors were mapped into the Porter value chain and it was found that the operation activity has the highest potential for productivity improvement. Then, four productivity measures were studied and HPV, the most common productivity measure, was chosen for further analysis. Finally, several studies that have been done about HPV were reviewed to pinpoint the drawbacks and potential improvement areas.



### 3 : Research Objectives and Approach

#### 3.1. Introduction

In this chapter data collected for this study, important factors that affect HPV, exploratory data analysis (EDA) about the important factors and statistics such as their trend during the study period mean, median, minimum, maximum, and the range of them are reviewed. Furthermore, the third research objective (RO3) and related hypotheses will be discussed accordingly. Finally, the approach to address the hypotheses will be presented.

#### 3.2. Data Collected for This Study

Data used in this study was for all North American plants that participated in the Harbour Institute's survey from 1999-2007. All statistics were synthesized using a uniform methodology from information supplied by the manufacturing plants, supplemented with plant visits by representatives of Harbour Consulting auditors. There are 86 different plants in the statistical sample from 12 automakers' brands, including Chrysler, Ford, GM, Honda, joint venture companies (such as Cami, Nummi, and Auto Alliance), Mitsubishi, Nissan, Toyota, Hyundai, and Volkswagen. The frequency of data for various automakers during the study period is shown in Table 3.1. However, Hyundai and Volkswagen plants will be excluded from further analysis because of very few (total of 3) number of plants.

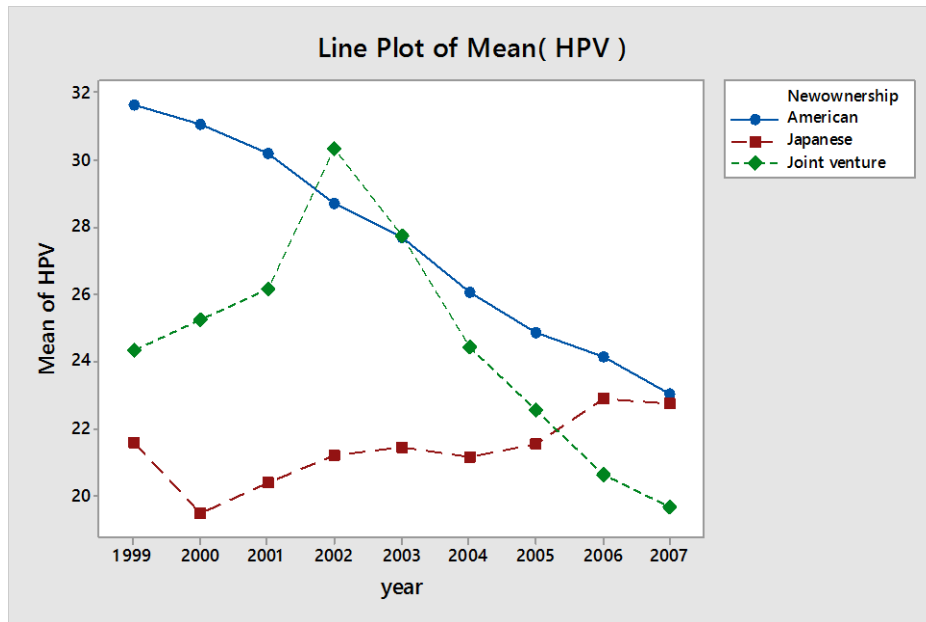
**Table 3.1. Frequency of data for various automakers' in study period**

<b>Brands</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>Total</b>
<b>Chrysler</b>	14	14	15	15	14	13	13	13	13	<b>124</b>
<b>Ford</b>	22	22	21	21	21	21	17	18	16	<b>179</b>
<b>GM</b>	29	30	30	30	28	28	25	22	21	<b>243</b>
<b>Honda</b>	2	2	2	3	3	3	5	3	6	<b>29</b>
<b>Hyundai</b>	0	0	0	0	0	0	0	0	1	<b>1</b>
<b>Joint venture</b>	5	5	4	4	4	4	5	4	4	<b>39</b>
<b>Mitsubishi</b>	1	2	1	1	1	1	1	0	0	<b>8</b>
<b>Nissan</b>	2	2	2	2	2	2	4	0	6	<b>22</b>
<b>Toyota</b>	4	4	4	2	4	5	5	5	5	<b>38</b>
<b>Volkswagen</b>	0	0	0	0	0	0	0	1	1	<b>2</b>
<b>Total</b>	79	81	79	78	77	77	75	66	73	<b>685</b>

### 3.3. Important Factors on HPV

According to previous works and the Harbour Institute’s data, a number of factors are defined or developed for further study. Then, the exploratory data analysis (EDA) is done for each variable separately. The goal of EDA is to discover the data pattern and allow the researcher to develop the hypotheses based on the data characteristic (Behrens, 1997). Tukey often called EDA detective work and believed that “exploratory data analysis can never be the whole story, but nothing else can serve as the foundation stone” (J. W. Tukey, 1977). Some preliminary steps of the EDA are done for the following variables that are considered in this study:

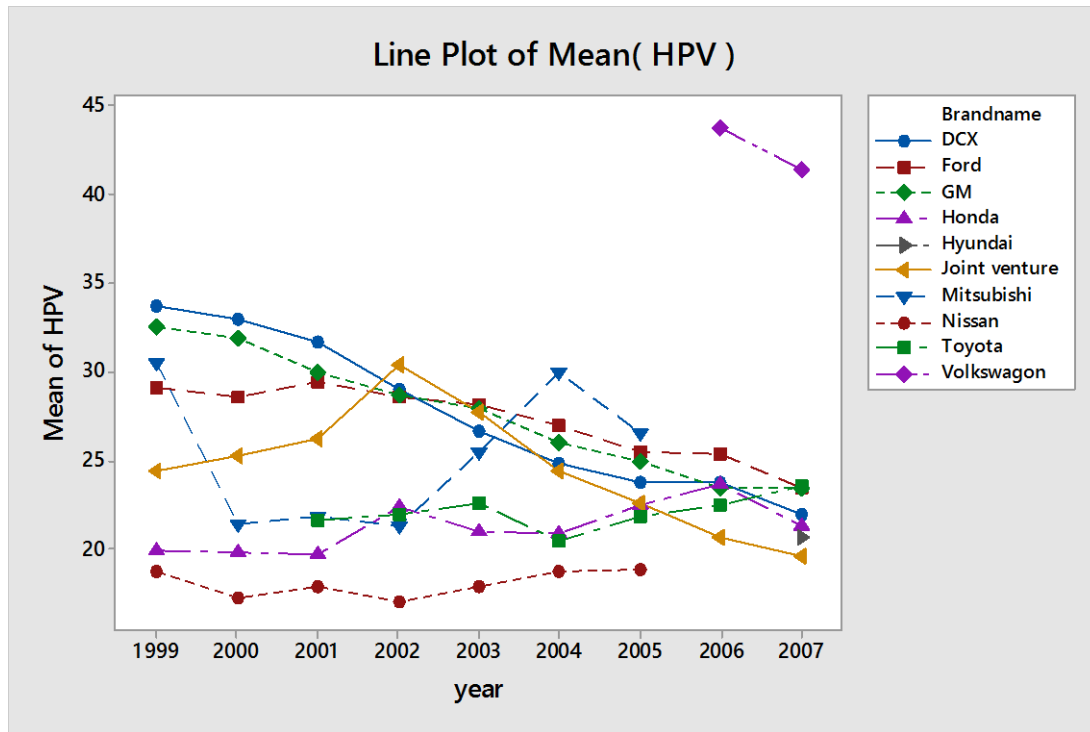
1. **Hours per vehicle (HPV)** has been previously discussed. The HPV variation during the study period for American, Japanese, and joint venture plants is illustrated in Figure 3.1.



**Figure 3.1. HPV variation in the study period for American, Japanese, and Joint venture companies**

Up to 2005, Japanese companies on average had the best value of this indicator among automakers located in North America (the effect of the car classes were ignored). After 2005, joint venture companies beat the Japanese and American. However, American plants reduced their productivity gap with Japanese plants and were almost close to Japanese plants in 2007.

The general trend of all carmakers is shown in Figure 3.2. Nissan ranks as number one (on average) among all the automotive brands, but the best HPV during the study period (13.57 hrs/vehicle) has been achieved by Chrysler Toledo Supplier Park in 2007. Volkswagen is at the other end of the spectrum and is the least productive plant with a huge gap with its competitors.



**Figure 3.2. HPV variation in the study period for different brands**

According to Figures 3.1 and 3.2., the following hypotheses could be considered for further analysis:

- H 3.1: The HPV variation for American, Japanese, and joint venture companies is statistically different
- H 3.2: HPV is correlated with the year of production

2. **Vehicle segment** is defined as a car's class. Fourteen car segments were used in the Harbour Institute's survey of the North American companies. These segments are compact, medium duty, full-size pickup, full-size sport utility vehicle (SUV), large van, luxury, midsize, large (midsize crossover), midsize SUV, minivan, small pickup, small SUV, sports car, and subcompact.

Here are two examples of car segments,

Subcompact cars are popular for their low prices and ability to squeeze out more miles from every gallon of gas. They're oftentimes basic transportation, offering first time car buyers an entry point into the new car market.



Full size vehicles feature supersized V-8 engines paired with three-speed automatic transmissions. Interiors were always roomy, with cavernous trunks and long hoods to match. Sold as two- and four-door sedans, some models also included station wagons.



The number of plants for each car segment, mean, median, SE of mean, standard deviation, minimum, maximum, and range of HPV for each segment has been shown in Table 3.2.

**Table 3.2. Number of plants in each segment during the study period**

Segment	N	Mean	Median	SE Mean	StDev	Minimum	Maximum	Range
COMPACT	75	28.07	25.43	1.09	9.46	17.53	66.58	49.05
FULL-SIZE PICKUP	108	28.509	25.82	0.967	10.049	17.72	82.2	64.48
FULL-SIZE SUV	47	28.33	27.5	0.747	5.12	21.06	50.53	29.47
LARGE (MIDSIZE CROSSOVER)	18	25.45	23.38	1.5	6.37	19.25	46.3	27.05
LARGE VAN	16	29.25	26.4	1.63	6.51	22.83	41.35	18.52
LUXURY	20	<b>38.37</b>	36.81	2.61	11.69	24.49	69.33	44.84
MEDIUM DUTY	13	<b>47.65</b>	48.76	2.17	7.82	31.73	58.87	27.14
MIDSIZE	95	<b>20.067</b>	19.06	0.367	3.581	15.18	35.89	20.71
MIDSIZE SUV	66	24.417	23.585	0.548	4.455	13.57	38.79	25.22
MINIVAN	53	28.649	28.06	0.633	4.611	22.44	46.08	23.64
SMALL PICKUP	43	22.976	22.44	0.445	2.921	18.97	30.94	11.97
SMALL SUV	27	25.36	22.1	1.63	8.47	18.29	54.91	36.62
SPORTS CAR	20	34.41	30.2	2.55	11.41	21.54	56.01	34.47
SUB COMPACT	81	24.287	23.09	0.576	5.186	17.09	38.16	21.07
<b>Total</b>	<b>682</b>	-	-	-	-	-	-	-

It can be seen from Table 3.2 that the medium duty, luxury, and sport car segments have the highest HPV with 47.65, 38.37, and 34.41, respectively. However, the midsize segment has the lowest HPV and ranked as the most productive segment in all car segments.

Hence another hypothesis could be considered as the relation of HPV and car segment.

H 3.3: The car segment has a significant impact on the HPV

3. **Car assembly and capacity utilization (CAC)** has also been discussed before. Toyota was the benchmark among North American Automakers in 2007, an average value of 101% was achieved for all of Toyota’s plants in this region. On the other hand, Mitsubishi with an average value of 67% placed at the other end of the spectrum. The CAC has been increasing by an average of 2.4% and 1% annually for the joint venture and Japanese companies in sample data during the study period, respectively. Therefore, joint venture companies had the best improvement of this indicator among North American Automakers during the study period. Ford, GM, and Chrysler (American plants) had the widest range, which varied from 132%, 130%, and 126%, respectively. Toyota with 43% had the lowest CAC variability during the study period. The trend of CAC for American, Japanese, and joint venture companies is shown in Figure 3.3.

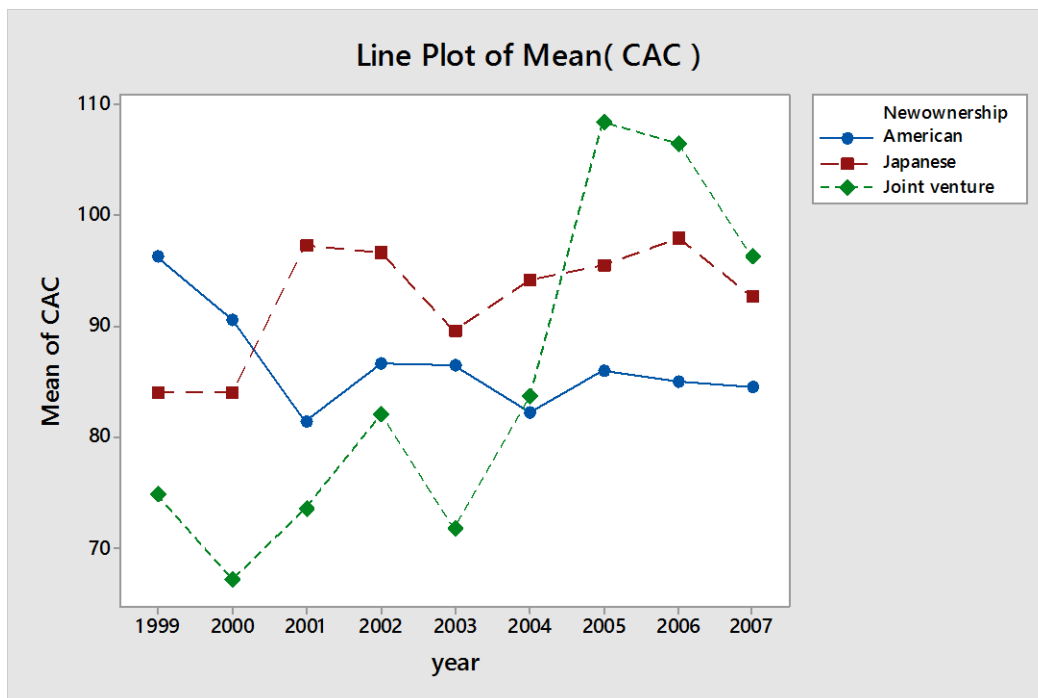
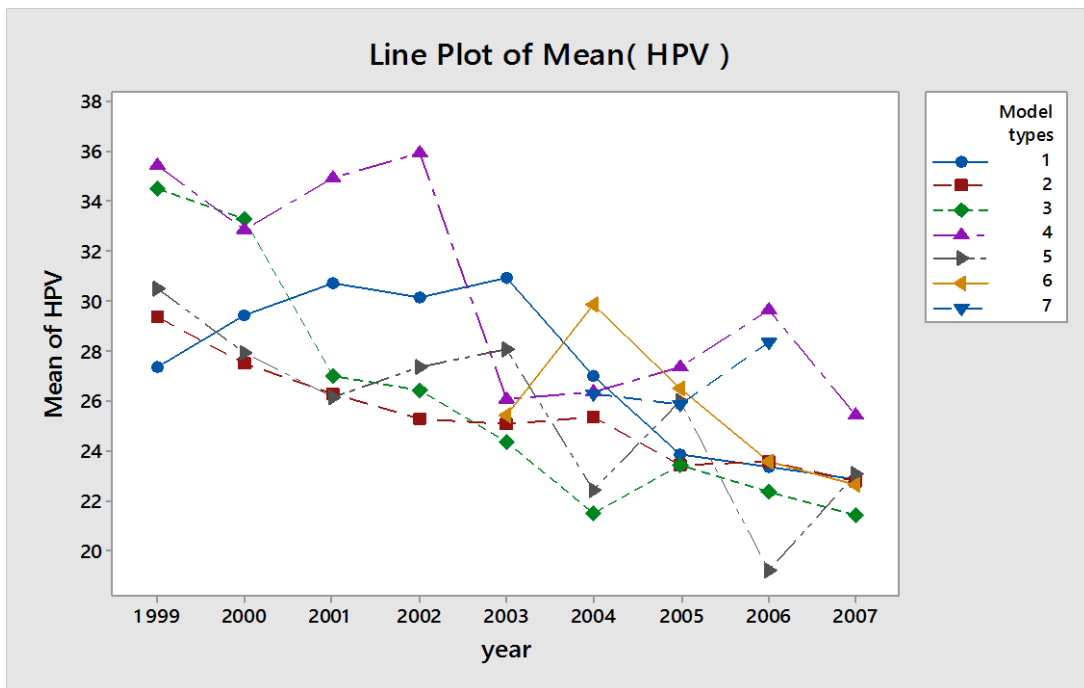


Figure 3.3. CAC trend during the study period for American, Japanese, and joint venture plants

This variable could be considered as another potential effective factor on HPV and is selected for further analysis.

H 3.4: HPV is correlated with CAC

- Number of models** refers to the number of different models produced in a facility. Mitsubishi built on average 5.25 models in each plant, the highest among all brands. Toyota and joint venture plants ranked at the other end of the spectrum with 1.88 and 1.92 models in each plant, respectively. The trend of HPV versus various model types during the study period is shown in Figure 3.4.

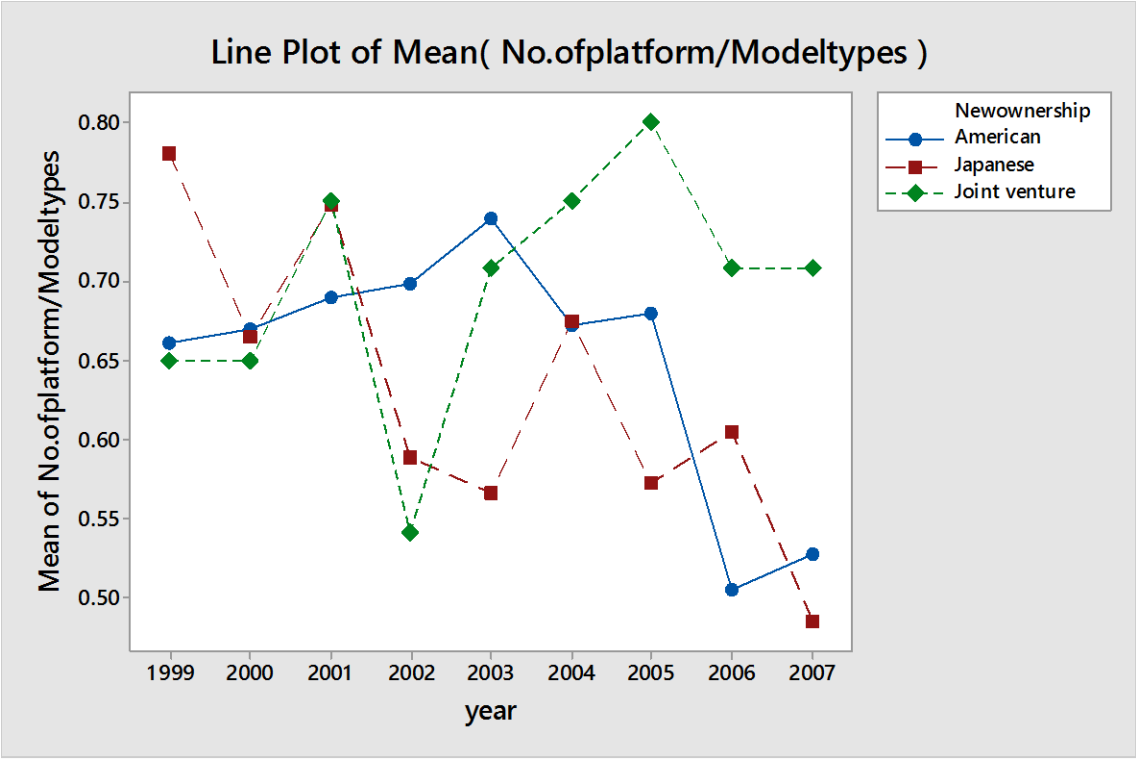


**Figure 3.4. HPV versus model type trend during the study period**

Figure 3.4 indicates that the plants that are assembling a larger number of models (especially four different model types) usually have higher HPV. Hence, another hypothesis could be of the relation between HPV and model types.

H 3.5: HPV is correlated with the number of models

5. **Number of platforms** refers to the discrete number of platforms built in a plant. A platform is determined either as the welded or framed underbody a car is built with and rides on (Harbour, 2002). American and Japanese companies moved toward reducing the number of platforms over model types during the study period by producing more types of models based on fewer platforms. Joint venture plants had very sporadic fluctuations during the study time. But, they slightly increased the number of platforms over the model types from 0.65 to 0.7 (from 1999 to 2007). Mitsubishi was the benchmark here, as it on average assembled two models on each platform in every separate plant. The trend of number of platforms over the number of models that was produced by different ownership brands is illustrated in Figure 3.5.



**Figure 3.5. Number of platform/models type trend during the study period for American, Japanese, and joint venture plants**

Nissan decreased the total number of platforms from 24 in 1997 to 6 in 2005 across their plants worldwide. More recently Volkswagen, with its MQB platform, has become the world benchmark for the application of common platforms, and has achieved the highest number of vehicles produced on a single platform.

The average number of platforms assembled in each separate plant for the luxury segment was 2.43, and it ranked lowest in terms of platform sharing. However, the average number of platforms assembled for the large van segment was 1.0 and it placed at the low end of the spectrum.

6. **Number of body styles** refers to the types of variations in body type, which include 2-door, 3-door, 4-door, 5-door, convertible, wagon, passenger, and cargo van (Harbour, 2002). Many plants build one platform with different body styles. All North American companies except the original American plants of Chrysler, Ford, and GM slightly increased this index to create differentiation for their customers.

In 2007, Ford assembled 3.13 different body styles on average in each plant, ranking first on this matter. However, Toyota assembled 1.8 different body styles on average in each plant, and it ranked lowest. The large van segment had the largest number of body styles among various segments at 7.38 per each plant. However, the average number of body styles assembled for the large segment was 1.39 and it was positioned at the other end of the spectrum.

7. **Number of chassis configurations** refers to each separate chassis configuration produced in the plant and variations in chassis. Potential variations include front-wheel drive, rear-wheel drive, and all-wheel drive (Harbour, 2002).

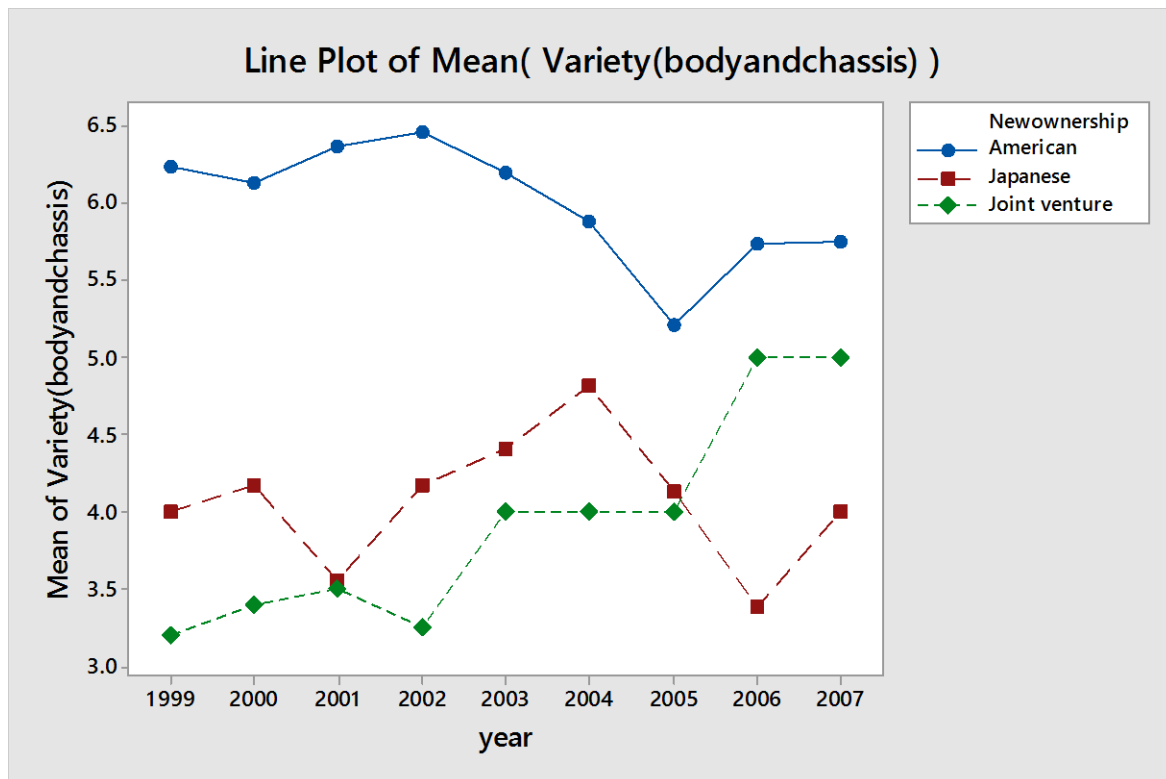
American and Japanese plants slightly decreased the average number of chassis configuration per each plants (from 3.5 to 2.9 and 2.6 to 2.05, respectively); however, joint venture plants significantly increased it from 1.8 to 3 on average per each plant. In 2007, Ford assembled 3.06 different chassis configurations on average in each plant, and it ranked highest on this matter. However, Toyota assembled 1.4 different chassis configurations on average in each plant, and it ranked lowest. The full-size pickup segment had the largest number of chassis configurations among various segments at 5.25 per plant. Though, the average number of chassis configurations assembled for the large segment was much lower, at 1.17 per plant.

8. **Vehicle variety** refers to the total number of body styles and chassis configurations. American plants slightly decreased the average number of vehicle variety per each plant from 6.2 to 5.6; however, joint venture plants significantly increased it from 3.2 to 5 on average per each plant. On average, the vehicle variety among Japanese plants didn't have a perceptible change during



the study period. The trend of vehicle variety for American, Japanese, and joint venture companies is shown in Figure 3.6.

In 2007, Ford assembled 6.375 different vehicle varieties on average in each plant, and it ranked the highest. However, Toyota assembled 3.2 different vehicle varieties on average in each plant, and it ranked the lowest. The range of total body styles and chassis configurations assembled for Ford was 12, and ranked highest among the North American brands. However, Toyota and Honda with 3 vehicle variety range ranked at the other end of the spectrum.



**Figure 3.6. Vehicle variety trend during the study period for American, Japanese, and joint venture plants**

The large van segment had the largest vehicle variety among various segments at 11.375 per each plant. However, the average number of vehicle varieties assembled for the large segment was 2.5, and it was at the other end of the spectrum.

Assembling higher variety of cars in each plants usually requires multi skill operators, more flexible equipment, more advanced logistical systems etc. These issues may penalize the productivity in a plant and increase HPV. Therefore, another interesting variable that might affect HPV is vehicle variety.

### H 3.6: HPV is correlated with the vehicle variety

9. **Platform sharing strategy** is defined as the number of platforms over the number of models assembled in a plant.

Automakers use platform sharing to minimize development costs, take advantage of manufacturing economies of production volume, maximize product value, improve quality, and coordinate a well-executed product launch. Wayman (2006) indicated that by using a platform sharing strategy “our clients have realized sales increases of 5-10% at the same levels of expenditure, or have achieved 10-15% expenditure reductions in marketing with little or no adverse revenue impact. Overall, gross profit improvements of 5% are typical” (Wayman, 2006).

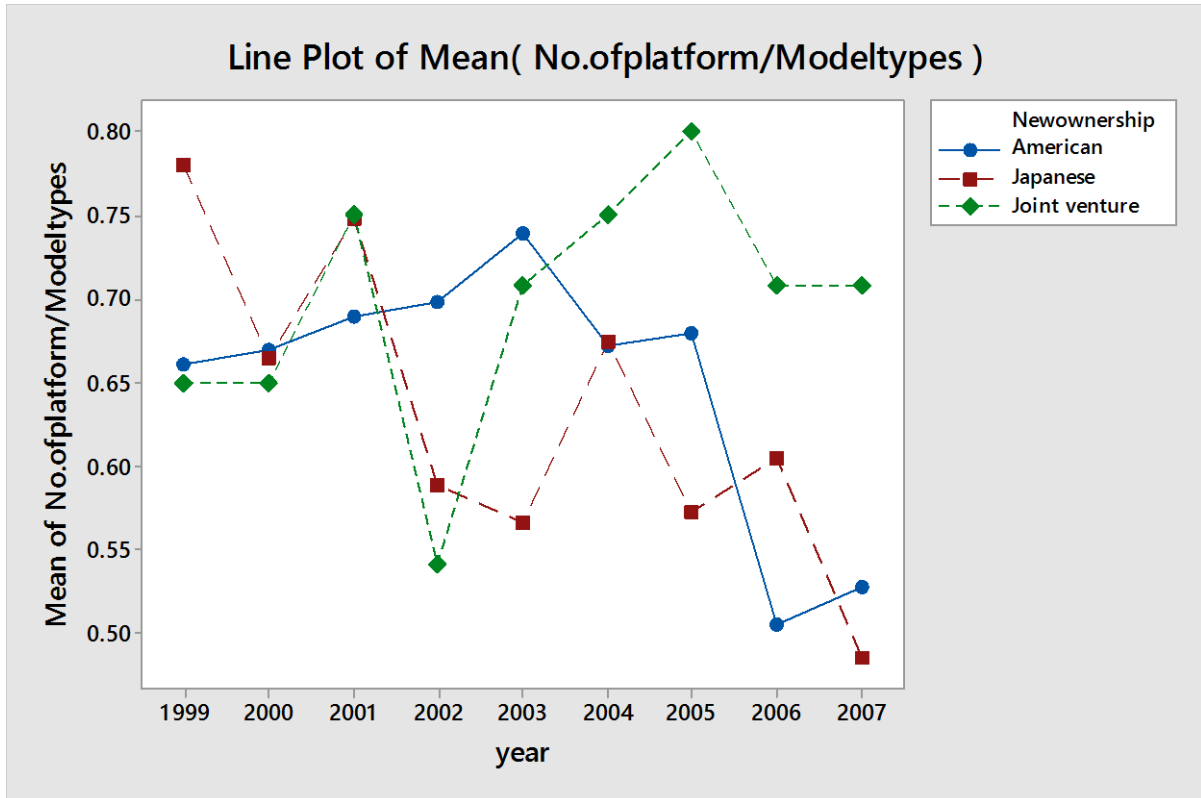
Sehgal and Gorai (2012) estimated that by 2020 the 10 major OEMs (General Motors, Ford, Toyota, Nissan, Honda, Daimler, Volkswagen, Peugeot Citroen, Renault, and Fiat) will reduce their number of platforms by about a third, from over 175 platforms in 2010. These OEMs will focus on mass production across a few core platforms. For instance, GM announced that it is considering reducing its platforms from 30 in 2010 to 14 in 2018 and to four flexible platforms by 2025. It is expected that GM stands to save 1 billion USD per year, primarily due to the implementation of a platform sharing strategy. The resulting complexity increases costs somewhat, but the additional expense is outweighed by savings from the sharing of common components between cars and platforms, and increased volume (Sehgal & Gorai, 2012).

The common platforms adoption will also lead to suppliers’ consolidation that will result in a smaller number of large but global suppliers. Recently, Ford announced that it will reduce its current suppliers from 1,150 to 750 (Hirsh et al., 2015).

In general, each brand name other than joint venture plants moved towards a platform sharing strategy during the study period. Figure 3.7 shows the trend of platform sharing strategy during the study period. Japanese plants had the most significant change (from 0.878 to 0.4853), followed by American.

In 2007, Nissan assembled 0.417 platforms over number of models on average in each plant, and it ranked the best brand on this matter. However, Toyota assembled 0.708 platforms over number of models on average in each plant, and it ranked lowest.

The full-size SUV segment had the largest utilized platform sharing strategy among various segments at 0.4052 per plant. However, the average number of platform sharing strategy for the medium duty was 1.308, and it was at the low end of the spectrum. Platform strategy means using fewer platforms, adding more standardization at the production shop floor level, and using lower product variety that might improve productivity and consequently HPV.



**Figure 3.7. Platform sharing strategy trend during the study period for American, Japanese, and joint venture plants**

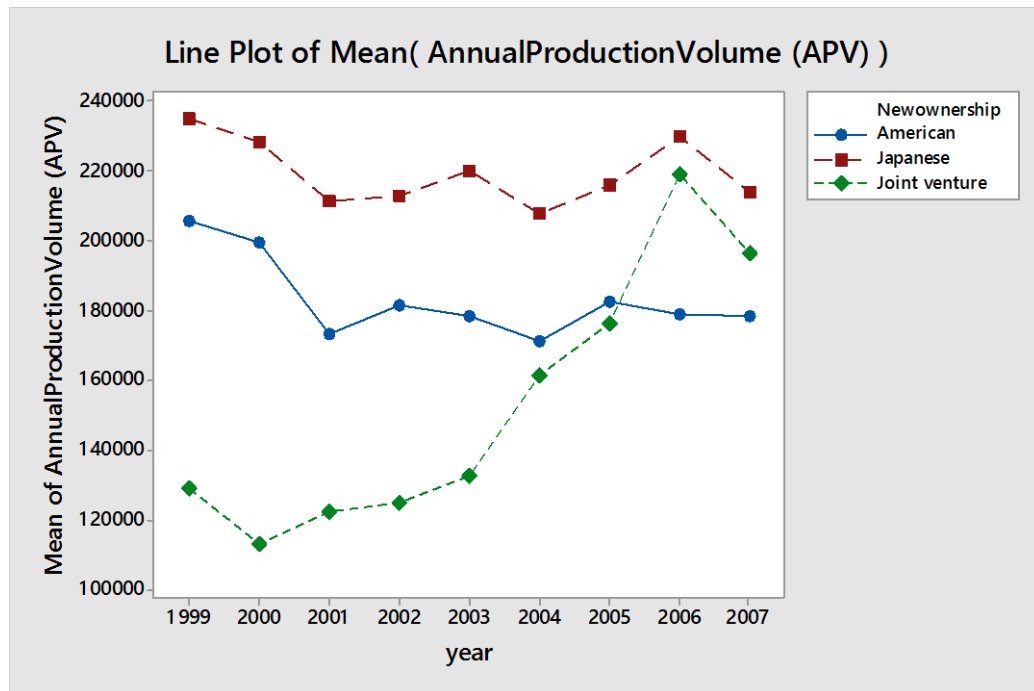
Hence, another interesting variable that might affect HPV is platform strategy,

H 3.7: HPV is correlated with Platform strategy

10. **Production volume** is the total number of vehicles produced by the end of the year. Joint venture plants increased the production volume by 52.02% during the study period, the highest growth rate among North American automakers. However, American and Japanese plants on average have a constant production volume during the study period and Japanese plants placed

the highest on this issue. The trend of production volume for American, Japanese, and joint venture plants is shown Figure 3.8.

In 2007, Honda produced 234,340 vehicles on average in each plant and had the highest volume while Ford had the lowest with 165,394 cars. The midsize segment had the largest production volume among various segments with 246,306 units per plant. However, the average units assembled for the medium duty segment was 24,442 units and placed at the low end of the spectrum.



**Figure 3.8. Production volume trend during the study period for American, Japanese, and joint venture plants**

Annual production could be a main player in productivity measurement; hence, it may have a significant impact on HPV.

H 3.8: HPV is correlated with annual production volume

11. **Flexible manufacturing** is the use of common processes, tools, and locator points to increase flexibility and help manufacturers build better vehicles at lower costs while maintaining quality. More flexibility usually means greater use of common equipment and tooling, as well as the reuse of that common equipment and tooling, which can provide significant cost savings.

Flexible manufacturing is becoming more crucial in a continually unpredictable environment and competitive world, as there is no reliable way for automakers to predict where customers will go to purchase their next vehicle. The companies that win in the marketplace will be the ones with the flexibility to best match cost, quality, and productivity with customer demand. Flexible manufacturing plants are often presented as the miracle solution to produce greater variety, but they usually come at a steep price. Any company that is working to improve flexibility must realize that flexible systems are one of the keys to effectively making products according to customer demand (Harbour, 2004).

Studies have shown that plants with greater flexibility generally have had more successful new vehicle launches, including faster launches at lower costs and at higher quality levels than previously achieved. By reducing the amount of change at launch, high quality can also be achieved at a faster pace. In addition, if an improvement is made, it can be applied immediately across all plants, as all the processes are the same (Harbour, 2004).

Moreno and Terwiesch (2015) studied the relation between production flexibility and pricing decisions on the North American automotive industry from 2002-2009. They found that the manufacturing facilities that are flexible can annually benefit 200-700 million USD, which is a significant savings. On the other hand, the scholars found the plants could benefit from the utilization augmentation (up to 6%) while deploying flexibility (Moreno & Terwiesch, 2015). The flexible manufacturing (FM) index used in this study can be calculated as follows:

$$FM = \text{Ln} (F_{Eq} \times F_V \times F_M \times F_U)$$

Where,

$F_{Eq}$  = Equipment flexibility which is defined as,

$$\frac{\text{No. of body styles}}{\text{No. of body lines}} \times \frac{\text{No. of body styles}}{\text{No. of paint lines}} \times \frac{\text{Vehicle variety}}{\text{No. of assembly lines}}$$

$F_V$  = Volume flexibility

$F_M$  = Mix flexibility

$F_U$  = Utilization flexibility

Equipment flexibility is defined as the flexibility for the equipment. Number of body styles, chassis configurations, body lines, paint lines, and assembly lines are reported by the Harbour

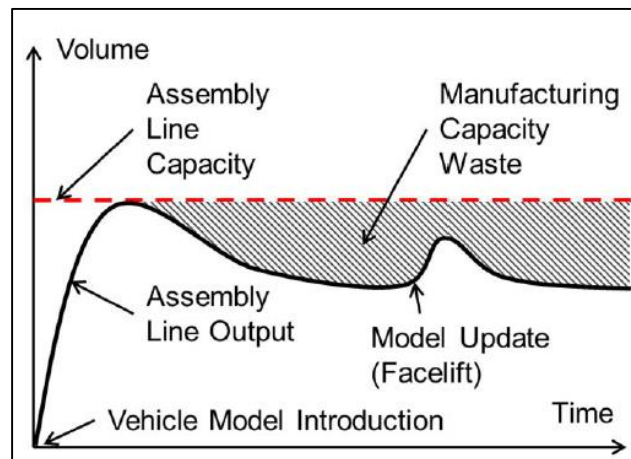
Institute. For instance, this index is 31.5 for a plant which has 3 body styles, 4 chassis configurations, 1 body line, 2 paint lines, and 1 assembly line.

Volume flexibility is usually related with the range of production volumes in which an organization is profitable over a given time period (Sethi & Sethi, 1990). By increasing the scale of economy, volume flexibility will increase also. Hence, the annual production volume could be considered as an appropriate volume flexibility symbol.

Moreno and Terwiesch (2015) studied the US automotive industry from 2002-2009 to investigate any potential relation between volume flexibility and discount flexibility. They found the plants who have more volume flexibility could save 10% of the total average discounts provided in the industry (\$200 to \$700 between year 2002-2009). Moreover, the flexible plants could benefit from facility's utilization increment (Moseng & Rolstadås, 2001).

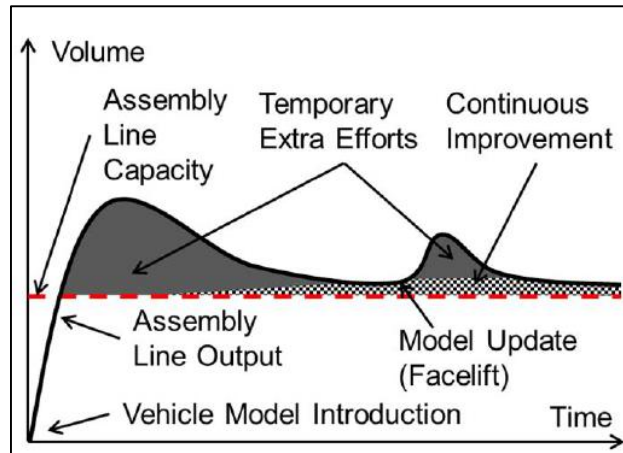
Mix flexibility is considered the number of models that are produced in a facility. If a plant assembles various numbers of model its flexibility is higher than a facility that assembled one model only.

One of the heritage that most plants got from Henry Ford's assembly line was fitting the capacity of assembly line with the maximum market demand (based on the forecasts). The main drawback of this strategy is after a few years or probably months the demand may decline for a vehicle model and there would be some capacity waste, Figure 3.9.



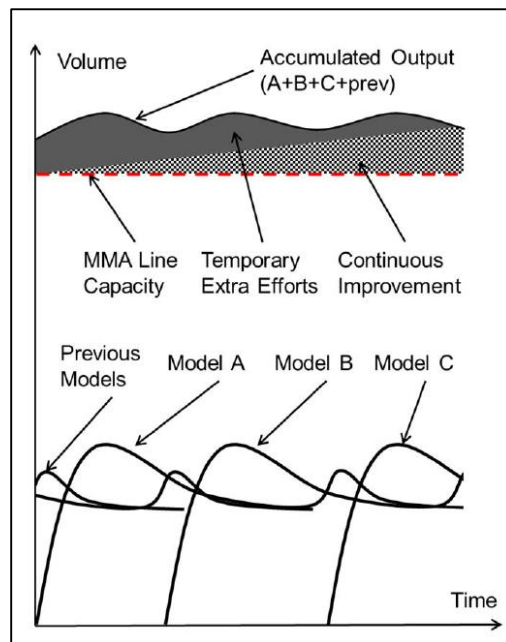
**Figure 3.9. Assembly line strategy to fit maximum demand (adopted from Diffner, 2011)**  
Another approach is to set the capacity line slightly lower than the minimum forecasted demand. Peaks in demand could be handled by temporary efforts such as extra operators, overtime, and shift. By concentrating on continuous improvement, the assembly line output can be increased and

respond to unforeseen demand shortly after the new vehicle advent phase as described in Figure 3.10.



**Figure 3.10. Dedicated assembly line to maximize utilization (adopted from Diffner, 2011)**

The Mixed Model Assembly (MMA) can be a key asset in a market fluctuation when the carmakers are not sure which model will be successful to what degree. Figure 3.11 shows how a MMA approach brings a smoother output in different lifecycles.



**Figure 3.11. Mixed model assembly line application to maximize the assembly utilization (adopted from Diffner, 2011)**

Utilization flexibility is simply the car capacity and utilization of the plant was discussed before. The combination of mix flexibility and utilization flexibility will help the manufacturing facility to minimize the risk of unused capacity.

The average FM indices of the different brands shown in Table 3.3. Mitsubishi achieved a score of 22.26 on average and was the most flexible brand among North American manufacturing facilities while Ford had the highest variation. Toyota and Joint venture had the lowest score of 19.19 and 19.28, respectively. However, the flexibility indices for joint ventures had the highest increase during the study period (from 18.5 in 1999 to 20.6 in 2007). Moreover, American plants and Japanese slightly increase this index throughout the study period.

**Table 3.3. Flexible manufacturing index during the study period**

Brand name	N	Mean	SE Mean	StDev	Min.	Median	Max.	Range
<b>Chrysler</b>	124	19.87	0.197	2.197	16.12	19.99	24.48	8.358
<b>Ford</b>	179	20.04	0.185	2.473	13.34	19.8	25.81	12.47
<b>GM</b>	243	20.43	0.149	2.325	11.82	20.61	25.8	13.98
<b>Honda</b>	29	19.78	0.275	1.479	16.84	19.43	23.01	6.172
<b>Joint venture</b>	39	19.28	0.236	1.474	15.49	18.97	22.18	6.689
<b>Mitsubishi</b>	8	22.26	0.376	1.064	20.5	22.64	23.58	3.081
<b>Nissan</b>	22	19.88	0.344	1.613	16.96	20.15	22.73	5.775
<b>Toyota</b>	38	19.19	0.237	1.463	16.72	18.69	21.59	4.872

Therefore, moving towards FM might theoretically improve productivity, and flexible plants might have lower HPV.

H 3.9: HPV is correlated with flexibility

12. **Outsourcing** refers to the percentage of activities which are handled outside the plant. According to HPV methodology, these activities include: bumper assembly; head liner assembly; instrument panel (I/P) assembly; door trim assembly; closure panel assembly; exhaust system assembly; suspension assembly; fuel tank assembly; engine and transmission assembly; heating, ventilating, and air conditioning (HVAC) assembly; wheels/tire assembly; front end module; door inner assembly; bumper mold; bumper paint; frame weld; fuel tank manufacturing; instrument panel mold; seat manufacturing; suspension manufacturing, and wheel manufacturing (Harbour, 2002).

Chrysler decided to outsource 60% of the production responsibility for a new Jeep Wrangler to its suppliers. Joann Muller indicated “the entire auto industry is moving to outsource bigger chunks of vehicles, but progress has been slow because of union opposition. Signs of improvement are everywhere at Chrysler. Company officials say warranty costs have dropped



34% in three years, while a recent consumer survey gave Chrysler’s vehicles a higher score on initial quality than Mercedes-Benz’s’’ (Harbour, 2005).

The outsourcing percentages for different brand names are shown in Table 3.4. Toyota had the lowest outsourcing activity with 24.1%, but GM has the highest outsourcing with 69.57%. Toyota also had the lowest range and standard deviation among the North American facilities with 22.38% and 5.49%, respectively. In general, the outsourcing percentage for each brand name was constant during the study period. However, American companies unlike the Japanese and Joint venture companies, were usually willing to outsource their activities more.

**Table 3.4. Outsourcing percentage during the study period**

Brand name	N	Mean	SE Mean	St Dev	Min.	Median	Max.	Range
<b>Chrysler</b>	124	68.37	1.07	11.91	46.85	68.18	100.00	53.15
<b>Ford</b>	179	64.01	0.84	11.29	23.08	63.64	96.15	73.08
<b>GM</b>	243	69.57	0.67	10.39	45.11	68.88	88.46	43.36
<b>Honda</b>	29	52.60	1.33	7.14	41.26	53.85	65.73	24.48
<b>Joint</b>	39	41.26	3.51	21.90	15.38	32.17	83.97	68.59
<b>Mitsubishi</b>	8	49.91	4.76	13.46	38.46	42.13	68.88	30.42
<b>Nissan</b>	22	50.41	1.38	6.48	31.47	49.65	57.69	26.22
<b>Toyota</b>	38	24.10	0.94	5.49	16.08	23.08	38.46	22.38

Outsourcing the functions will probably reduce the complexity of the organizational structure and the number of supportive personnel. These facts lead to the hypothesis that outsourcing improves productivity.

H 3.10: HPV is correlated with outsourcing

13. **New product launch** requires a lot of effort by engineering, logistical, quality, and production personnel to overcome launch phase difficulties, which may cause HPV to increase significantly during this period. Most plants try to compensate for this increase by moving toward flexibility and outsourcing some functions. During the study period, it was quite obvious that the plants had some difficulties maintaining HPV while new products were being launched; however, the Japanese plants managed new product launches better than the Americans overall. The percentage of the plants that were involved with a new product launch is shown in Table 3.5.

**Table 3.5. Percentage of plants that were involved with a new product launch**

Brand name	N	Mean	SE Mean	St Dev
<b>Chrysler</b>	124	0.2258	0.0377	0.4198
<b>Ford</b>	179	0.1788	0.0287	0.3842
<b>GM</b>	243	0.2181	0.0265	0.4138
<b>Honda</b>	29	0.1379	0.0652	0.3509
<b>Joint venture</b>	39	0.1538	0.0585	0.3655
<b>Mitsubishi</b>	8	0.125	0.125	0.354
<b>Nissan</b>	22	0.2727	0.0972	0.4558
<b>Toyota</b>	38	0.1176	0.0561	0.327

Table 3.5 indicates that Nissan has the highest number of plants that were involved with a new product launch, almost 27.3%. Toyota positioned itself at the other end of the spectrum with only 11.7%.

While launching a new product, operators must be trained with new product assembly methods, and machines and equipment have to be tuned. Hence, the possibility of mismanagement between different departments and of a quality crisis will increase and productivity may drop.

H 3.11: HPV is correlated with launching a new vehicle

14. **Annual available working days** is the number of straight-time working days planned in the data year. This includes unplanned shutdowns, like market adjustments, weather-related closures, and launch down-days. Not included in the annual available working days are planned vacations and recognized holidays (e.g. summer shutdowns and Christmas holidays) (Harbour, 2002).

The annual available working days for different brands is illustrated in Table 3.6. Table 3.6 indicates that for different brand names, the average annual available working days was almost constant during the study period and within the range of 230-245 days. The Japanese plants had the lowest range and Americans had the highest range of annual available working days.

**Table 3.6. Annual available working days for different car makers**

Brand name	N	Mean	SE Mean	StDev	Min.	Max.	Range
<b>Chrysler</b>	124	240.3	1.07	11.92	152	251	99
<b>Ford</b>	179	231.97	2.67	35.78	39	336	297
<b>GM</b>	243	234.6	2.04	31.82	35	333	298
<b>Honda</b>	29	241.59	0.729	3.92	234	245	11

Brand name	N	Mean	SE Mean	StDev	Min.	Max.	Range
Joint venture	39	238.49	1.38	8.63	215	261	46
Mitsubishi	8	238.75	0.648	1.83	235	241	6
Nissan	22	237.95	0.692	3.24	234	243	9
Toyota	38	241.97	0.446	2.6	235	246	11

Given this data, one can hypothesize that if the production is routine and stable i.e. the annual available working days is high, the productivity will improve. Therefore,

H 3.12: HPV is correlated with annual available working days

15. **Hourly employees' percentage** is the number of hourly employees over the total on-roll employment for the plant in December, comprised of hourly (direct and indirect), salary, and normal daily total absenteeism (controllable and uncontrollable) for all on-site manufacturing (strategic and nonstrategic) (Harbour, 2002).

The hourly employees' percentages for different ownerships are shown in Figure 3.9. It can be seen that Japanese plants had the highest salaried employees' percentage; however, American plants had the lowest. During the study period, the hourly employees' percentage was almost constant for all the plants and ranged between 80-100%.

Ford has the highest hourly employees' percentage with 93% and Honda has the lowest with 83%.

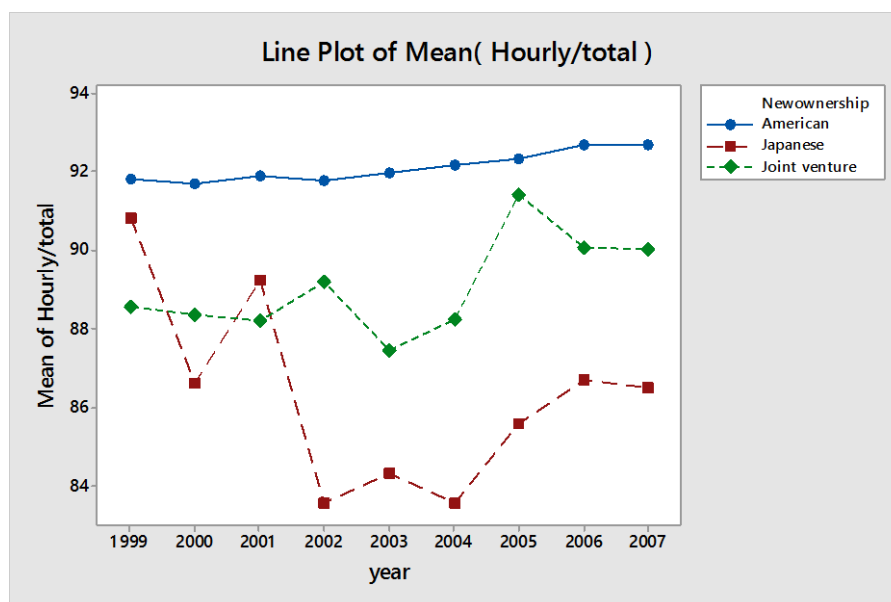


Figure 3.12. The hourly employees' percentage trend during the study period for American, Japanese, and joint venture plants

One can theorize that the organization loyalty will increase if the hourly employees' percentages are lower, and thus the productivity will improve.

H3.13: HPV is correlated with hourly employees' percentage

The range of all discussed variables based on the available data during the study period is shown in Table AI (appendix).

**16. Ergonomics** is a scientific discipline, which is concerned with improving the productivity, health, safety and comfort of people, as well as promoting effective interaction among people, technology, and the environment in which both must operate (IEA, 2016).

Baraldi and Paulo (2011) studied the benefits of workstations ergonomic improvements in a Brazilian automotive assembly shop. The authors compared the performance of two assembly lines of an assembly shop (a new and an old ones) through six internal indicators such as quality of product, total assembly line time, normal of standard (NOS), customer satisfaction index, total warranty costs, and workers absenteeism index during a 12-month period (Baraldi & Paulo, 2011). The methods time measurement (MTM) was chosen to compare the total assembly line time. They found the necessary task time on new assembly line is 42% less than the old assembly line. This achievement was because of more automation in the new line and workstation ergonomic improvement. However, the study presented other advantages because of the ergonomic workstations improvement such as reduction of absenteeism, prevention of illnesses, and improvement in productivity, product quality, and workers' quality of life.

Thun et al. (2011) conducted a survey to study the benefits of ergonomics improving in a shop floor and its social and economic impacts. The sample data included 55 German companies in the automotive industry sector. The participants were asked about the harmful tasks in the production process and their impacts on the ergonomic and health areas (Thun et al., 2011). The research focus was on two main areas such as worker-oriented and work-oriented. The authors used factor and cluster analysis for the performance effects comparison of work and worker oriented ergonomic practices. The performance effects included: increased productivity, higher flexibility, fewer mistakes and defects, less days of absence, reduced fluctuation, increased health of shop floor workers, reduction of workloads, increased safety/fewer accidents, increased work motivation and satisfaction, more comfort at work, and

relief of corporate welfare systems. In general, they found that the companies implementing both work and worker related practices have higher performance effects.

Battini et al. (2011) studied the relation between ergonomics and assembly design techniques through assessing the concurrent engineering approach. They used technological variables (like those related to work methods and time), environmental variables (such as absenteeism, staff turnover, and work force motivation), and ergonomic evaluations (like human diversity) to build a comprehensive method (Battini et al., 2011). They used EM-plant (a taxonomic simulation) tool to simulate, measure the current and future assembly process, and validate the utilized method. Several achievements have been obtained by redesigning the work station that resulted from the simulation in their study. They improved productivity by 15% and considerable reduction in fatigue levels and number of injuries.

Based on reviewing the above studies, it was revealed that workstation ergonomic status may have a significant impact on the productivity and operator's absenteeism or number of manufacturing facilities injuries. The Occupational Safety and Health Administration (OSHA) collected work-related injury and illness data from employers within specific industry and employment size specifications from 1996 through 2011. The data provided is used to calculate and establish specific injury and illness incidence rates throughout the U.S. The database contains a table with the name, address, industry, and associated total case rate (TCR), days away, restricted, and transfer (DART) case rate, and the days away from work (DAFWII) case rate for the establishments that provided OSHA with valid data for calendar years 1996 through 2011 (OSHA, 2011).

An incidence rate of injuries and illnesses is computed from the following formula:

$$\text{Incidence rate} = (\text{Number of injuries and illnesses} \times 200,000) / \text{Employee hours worked}$$

The readers are referred to “How to Compute a Firm's Incidence Rate for Safety Management”<sup>1</sup> for further information.

Hence, the researcheres decided to consider the association of ergonomics and productivity through OSHA database and Harbour data for the common plants. It means that if the total case rate or days away, restricted, and transfer are low probably the HPV is low.

---

<sup>1</sup> <http://www.bls.gov/iif/osheval.htm> - accessed March 20th 2016

H3.14: HPV is correlated with the ergonomics indices provided by OSHA

All the stated hypotheses are summarized in Table 3.7.

**Table 3.7. List of all hypotheses**

Hyp.	Research Hypotheses
H3.1	The HPV variation for American, Japanese, and joint venture companies is significantly different
H3.2	HPV is correlated with the year of production
H3.3	The car segment has a significant impact on the HPV
H3.4	HPV is correlated with CAC
H3.5	HPV is correlated with the number of models
H3.6	HPV is correlated with the vehicle variety
H3.7	HPV is correlated with Platform strategy
H3.8	HPV is correlated with annual production volume
H3.9	HPV is correlated with flexibility
H3.10	HPV is correlated with outsourcing
H3.11	HPV is correlated with launching a new vehicle
H3.12	HPV is correlated with annual available working days
H3.13	HPV is correlated with hourly employees' percentage
H3.14	HPV is correlated with the ergonomics indices provided by OSHA

In the next section the approach to address variables differences in terms of car ownership demographics, hypotheses testing, and developing the statistical models will be presented.

### 3.4. Approach for the Statistical Analysis

As mentioned earlier, the impact of ownership brand, year of measuring HPV, vehicle segment, CAC, number of models, production variety, platform sharing strategy, annual production volume, flexible manufacturing, outsourcing, whether or not the plant was involved with a new product launch, annual available working days, percentage of hourly employees, and ergonomics indices factors on HPV would be of interest to this research. Hence, the following steps will be then to address the third research objective:

- Studying the differences among American, Japanese, and joint venture companies regarding the stated factors which will be answered by analysis of variance (ANOVA).
- Developing statistical models to check the previously stated hypotheses and estimate the HPV accordingly.

Different phases and steps of the research are illustrated in Figure 3.10; however, the third phase will be discussed in more depth in chapters 5, 6, 7, and 8.

### First Phase: Literature survey

Defining the generic important factors on productivity in automotive sector

Defining the most important activity in a generic value chain in automotive industry

Defining the most common productivity measure for production process

Defining the important factors that improve HPV

### Second Phase: Research objectives and approach

Hypotheses development (14)

ANOVA for all the variables based on the ownerships

Developing three Tukey test to define the best initial model

Selecting the base model for further analyses

### Third Phase: Enhancing the statistical model and model validation

Utilizing and developing different statistical methods and discussion

Developing a computer based software program base on the best statistical method and MIP

Recommendations, implications, limitations, and future research

**Figure 3.13. Different phases and steps of the research**



### **3.5. Conclusion**

In this chapter, data collected for the study and 14 potential variables that may affect HPV (the third research objective) were reviewed to identify trends during the study period. Accordingly, 14 hypotheses were formed that would be addressed in the next chapter.

## **4 : Current Work and Preliminary Results**

### **4.1. Introduction**

In this chapter ANOVA will be performed first to check whether or not the mean of all variables defined in chapter 3 are the same for American, Japanese, and joint venture ownerships. Then the previously defined hypotheses are statistically checked (third research objective) and the preliminary results for the fourth research objective are shown and discussed later.

### **4.2. ANOVA for all the Variables**

Analysis of variances (ANOVA) for all variables of American, Japanese, and joint venture plants are shown in Table 4.1.

The analysis of variances (ANOVA) indicates the HPV mean is statistically different within different ownerships (p-value is 0.000). Grouping information using the Tukey's pairwise comparisons indicates that there are two main groups for HPV, as illustrated in Table 4.1.

Table 4.1 indicates that the mean of all variables, except launching a new vehicle and annual available working days, within various groups of American, Japanese, and joint venture are statistically different ( $p \leq 0.054$ ). This is because different companies' ownerships (American, Japanese, and joint venture) are trying to have a tradeoff between launching the new products and creating customer differentiation. This means they launch some new products but not many because of the huge investment required to launch new autos. However, number of available working days normally must be very close to each other for any ownership provided, there is not any plant shutdown or differences in various countries' working calendar such as America, Canada, and Mexico.

Table 4.1 shows the Tukey and Fisher pairwise comparison for all variables based on three different ownerships. Table 4.1 indicates that for each variable (other than launching a new vehicle, annual available working days, and CAC) there are at least two groups, and for hourly employees' percentage three groups. Hence, different ownerships have different strategies for the varieties of vehicle, flexibility, model types, annual production volume, CAC, outsourcing, platform strategy,

and hourly percentage employee, and the mean of these variables are statistically different for American, Japanese, and joint ventures. As expected, the Tukey comparison is more conservative other than Fisher. For instance for CAC we have one group for Tukey pairwise comparison; however, it is two groups for the Fisher comparison.

**Table 4.1. Analysis of variance for all the variables for American, Japanese, and Joint venture companies**

Variables	Ownership	P-Value	Tukey Pairwise Comparisons			Fisher Pairwise Comparisons		
			N	Mean	Gr	N	Mean	Gr
HPV	American	0.000	546	27.733	A	546	27.733	A
	Joint venture		39	24.517	A B	39	24.517	B
	Japanese		93	21.525	B	97	21.525	B
Launch a vehicle	American	0.463	546	0.207	A	546	0.207	A
	Japanese		93	0.1613	A	97	0.1613	A
	Joint venture		39	0.1538	A	39	0.1538	A
AWD	Japanese	0.158	93	240.624	A	97	240.624	A
	Joint venture		39	238.49	A	39	238.49	A
	American		546	235.03	A	546	235.03	A
Variety (body & chassis)	American	0.000	546	6.018	A	546	6.018	A
	Japanese		93	4.097	B	97	4.097	B
	Joint venture		39	3.897	B	39	3.897	B
Flexibility	American	0.001	546	140	A	546	140	A
	Japanese		93	27.86	B	97	27.86	B
	Joint venture		39	26	A B	39	26	B
Model types	Japanese	0.054	93	2.452	A	97	2.452	A
	American		546	2.2912	A B	546	2.2912	A B
	Joint venture		39	1.923	B	39	1.923	B
Annual Production Volume APV)	Japanese	0.000	93	216,978	A	97	216,978	A
	American		546	183,335	B	546	183,335	B
	Joint venture		39	151,545	B	39	151,545	C
CAC	Japanese	0.000	93	93.37	A	97	93.37	A
	American		546	86.7	A	546	86.7	B
	Joint venture		39	84.8	A	39	84.8	A B
Outsourcing	American	0.000	546	67.475	A	546	67.475	A
	Japanese		93	41.44	B	97	41.44	B
	Joint venture		39	41.26	B	39	41.26	B
Platform strategy	Japanese	0.000	93	0.3481	A	97	0.3481	A
	Joint venture		39	0.3317	A B	39	0.3317	A
	American		546	0.27913	B	546	0.27913	B
Hourly/total	American	0.000	546	92.0945	A	546	92.0945	A
	Joint venture		39	89.103	B	39	89.103	B
	Japanese		93	85.911	C	97	85.911	C

### 4.3. Tukey’s Comparison for Different Car Segments

Analysis of variance (ANOVA) shows that the mean HPV within various segments is significantly different (p-value is 0.000). As earlier mentioned, it seems that car segmentation plays an important role in evaluating the HPV.

Grouping information using the Tukey method indicates that there are six significant groups (A-F), as shown in Table 4.2. Some car segments overlap, but for ease of calculation three main groups were used based on the HPV values. Groups A and B were categorized in such a way that HPV mean ranges from 20-25 for group A and 25-30 for group B. Sports, luxury, and medium duty were categorized as group C, midsize SUV, subcompact, small pickup, and midsize as group A, and the remainder as group B. Hence, in this study, three main groups were considered for the car segment factor (see Table 6.7 for the segment footnote).

**Table 4.2. Tukey’s Pairwise comparisons for different segments**

Segment	N	Mean	Proposed	Used	Frequency
Medium Duty	13	47.65	A	C	53
Luxury	20	38.37	B	C	
Sports Car	20	34.41	B	C	
Large Van	16	29.25	C, D	B	344
Minivan	53	28.649	C, D	B	
Full-Size Pickup	108	28.509	C	B	
Full-Size SUV	47	28.33	C, D	B	
Compact	75	28.07	C, D	B	
Large	18	25.45	C, D, E	B	
Small SUV	27	25.36	D, E	B	
Midsize SUV	66	24.417	E	A	285
Sub Compact	81	24.287	E	A	
Small Pickup	43	22.976	E	A	
Midsize	95	20.067	F	A	
<b>Total</b>	682	-	-	-	682

The parametric density function for all variables versus different segments have been illustrated in Figure 4.1. It could be depicted that, HPV, vehicle variety, flexibility, annual production volume, CAC, outsourcing, and hourly employees’ percentage variables are significantly different for various segments.

# Histogram of HPV, Vehicle Laun, AWD, Variety(body, Ln(Flexibili, ... Normal

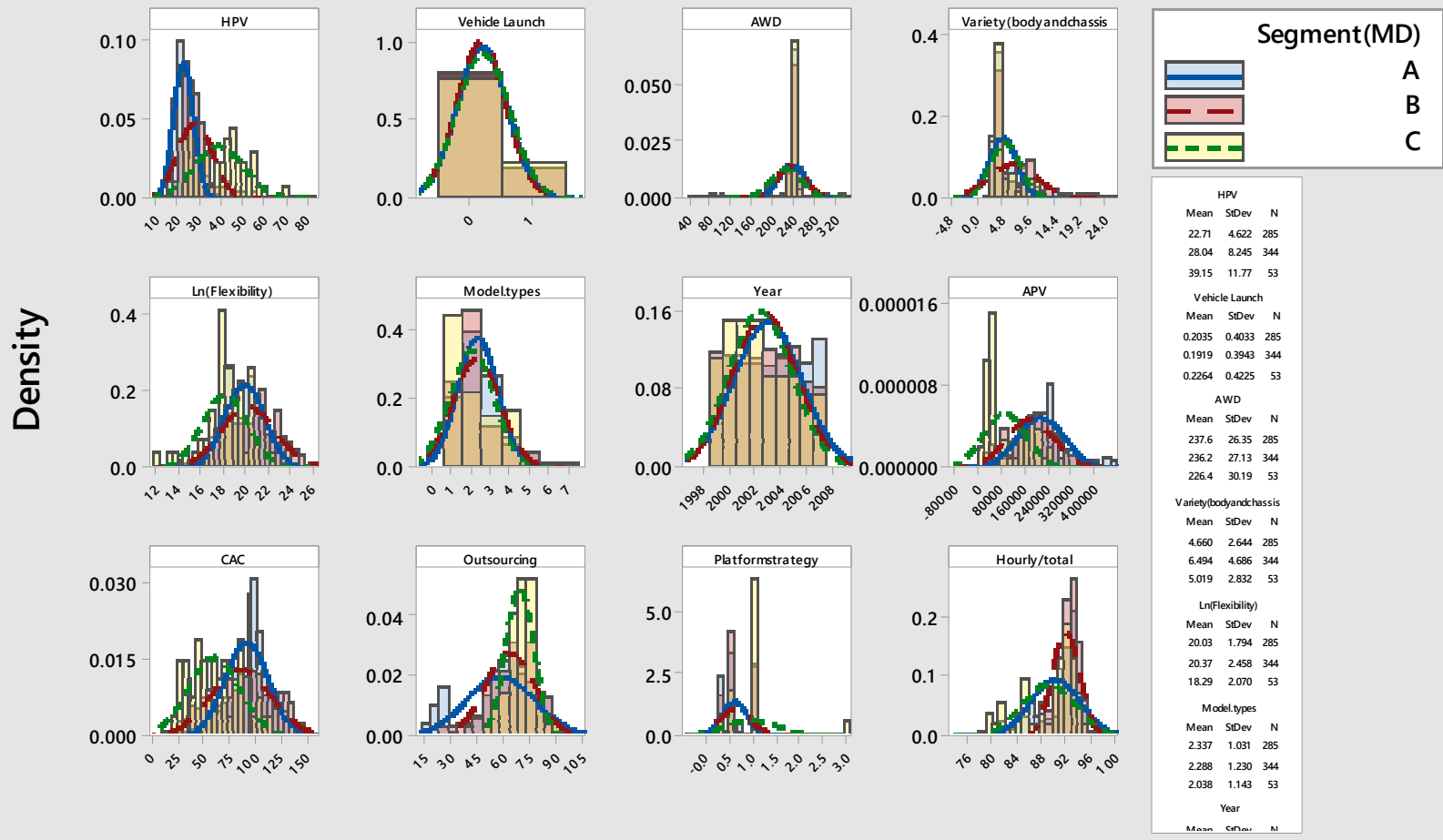


Figure 4.1. Density function for all variables versus different segment

The parametric density function for all variables versus different ownerships have been illustrated in Figure AI (appendix). It could be depicted that, HPV, AWD, vehicle variety, flexibility, model types, annual production volume, CAC, outsourcing, and hourly employees' percentage variables are significantly different for various ownerships.

#### 4.4. Tukey's Comparison for Different Years

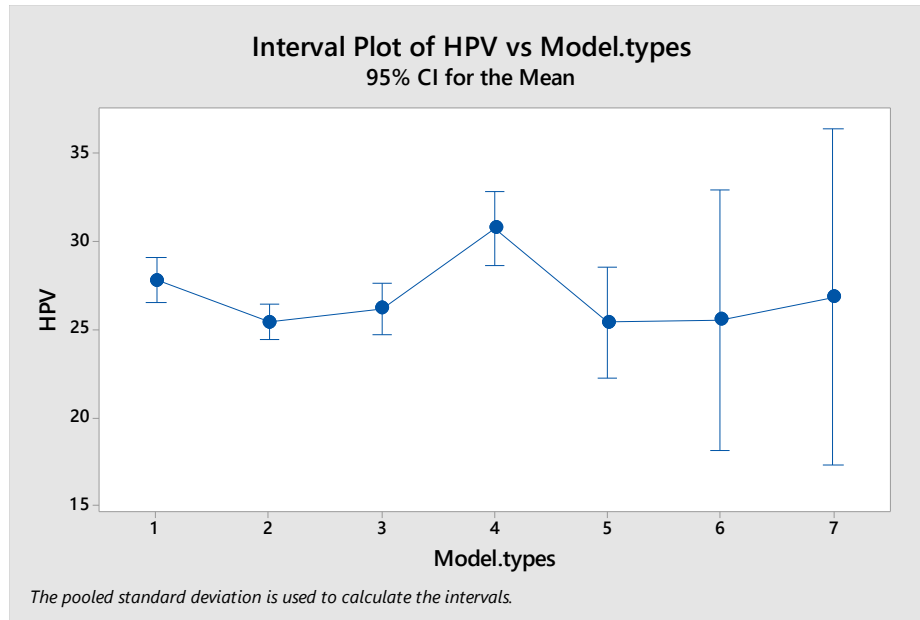
The analysis of variances (ANOVA) indicates the HPV mean is statistically different within different years (p-value is 0.000). Grouping information using the Tukey's pairwise comparisons indicates that there are four main groups (A-D), as illustrated in Table 4.3. It shows that 1999, 2000, and 2007 were statistically placed in one group each; however, the other years had some overlapping groups. Hence, another approach could be considering year as a categorical variable in three different levels for all nine years as 1999-2001 (group 1), 2002-2004 (group 2), and 2005-2007 (group 3).

**Table 4.3. Tukey's Pairwise Comparisons of HPV for different years**

Year	N	Mean	Proposed Grouping	Used Grouping
1999	79	30.01	A	1
2000	81	29.34	A	1
2001	79	28.95	A, B	1
2002	78	27.988	A, B, C	2
2003	77	26.858	A, B, C, D	2
2004	77	25.261	B, C, D	2
2005	75	24.021	C, D	3
2006	65	23.771	C, D	3
2007	71	22.758	D	3
Total	682	-	-	-

#### 4.5. Tukey's Comparison for Different Models

The HPV interval plot significantly varies for different model types, as seen in Figure 4.2. It seems that the model types' variations is increasing as the number of models increased. It was decided to group information using Tukey's pairwise comparisons to check if the variability of the obtained statistical model will be reduced. Table 4.4 shows the plants that assembled 2 and 3 car models and 4 car models were statistically placed in one group each (B and A), respectively; however, the other plants had some overlapping data.



**Figure 4.2. HPV variation in the study period for different model types per plant**

Since there are some overlapping groups for the plants that assembled 1, 5, 6, and 7 car models, it was decided to use another kind of grouping. The plants that assembled 1, 2, 3 models (group one), 4 (group two), and greater than 4 (group three). The results are shown in Table 4.4.

**Table 4.4. Tukey's Pairwise Comparisons for different Model types per plant**

Model types	N	Mean	Grouping	Used Grouping
4	61	30.8	A	2
1	171	27.838	A, B	1
7	3	26.873	A, B	3
3	127	26.201	B	1
6	5	25.59	A, B	3
2	287	25.456	B	1
5	28	25.435	A, B	3
Total	682	-	-	-

#### 4.6. Hypotheses Testing and Statistical Analysis

In this section, data preparation, determination of the initial multiple linear regression model, hypotheses testing, initial multiple linear regression model adequacy checking, and a discussion of the findings are presented.

Linear modeling addresses two broad goals. Some researchers use linear models for the purposes of predicting dependent variables along with measuring the uncertainty in the prediction. Some researchers use linear modeling to understand the relationship between the variables and try to develop an explanation for the data in hand (Faraway, 2005). In this research, both approaches are considered.

#### **4.6.1. Variable Types**

There are different types of variables that are used in the literature. The utilized variables definition are followed as,

- **Categorical variable (or nominal variable):** It is an independent or predictor variable, usually containing values indicating membership in one of several possible categories. E.g., car segment which can be in group A, B, or C.
- **Interval-scaled variable:** That is when a measurement difference between two values is meaningful. For instance, the difference between plants that assemble two and three models is the same as between four and five models.
- **Continuous Variable:** It is a variable that has an infinite number of possible values. In other words, any positive value is possible for the variable such as HPV (Indiana, 2016).

All the variables and their types are shown in Table 4.5. As earlier mentioned, it was decided to consider the year and model types as categorical variables and to check if the obtained statistical models with this approach will be better than the initial statistical model using year and car model types as interval-scale variable.



**Table 4.5. Different variables types**

No.	Variables	Categorical (nominal)	Interval - scale	Continuous	Categorical – Another approach
1	HPV			X	
2	Launch a New vehicle	X			
3	Car segment	X			
4	Ownership	X			
5	Year		X		X
6	AWD		X		
7	Variety (body & chassis)		X		
8	Flexibility			X	
9	Model types		X		X
10	Annual Production Volume (APV)		X		
11	CAC			X	
12	Outsourcing			X	
13	Platform strategy			X	
14	Hourly/total			X	

#### 4.6.2. Data Preparation and Cleaning

Once the data were collected from Harbor Institute reports, edit checks were performed and the HPV was plotted versus all 13 individual calculated variables to detect extreme outliers and linearity of regressors with respect to the response. The scatter plot of all the variables is shown in Figures 4.3 and 4.4. This figures indicate that some predictors may not be linear with respect to the HPV, which must be taken into account later. On the other hand, it also shows that there are some outliers which have to be taken care of also.

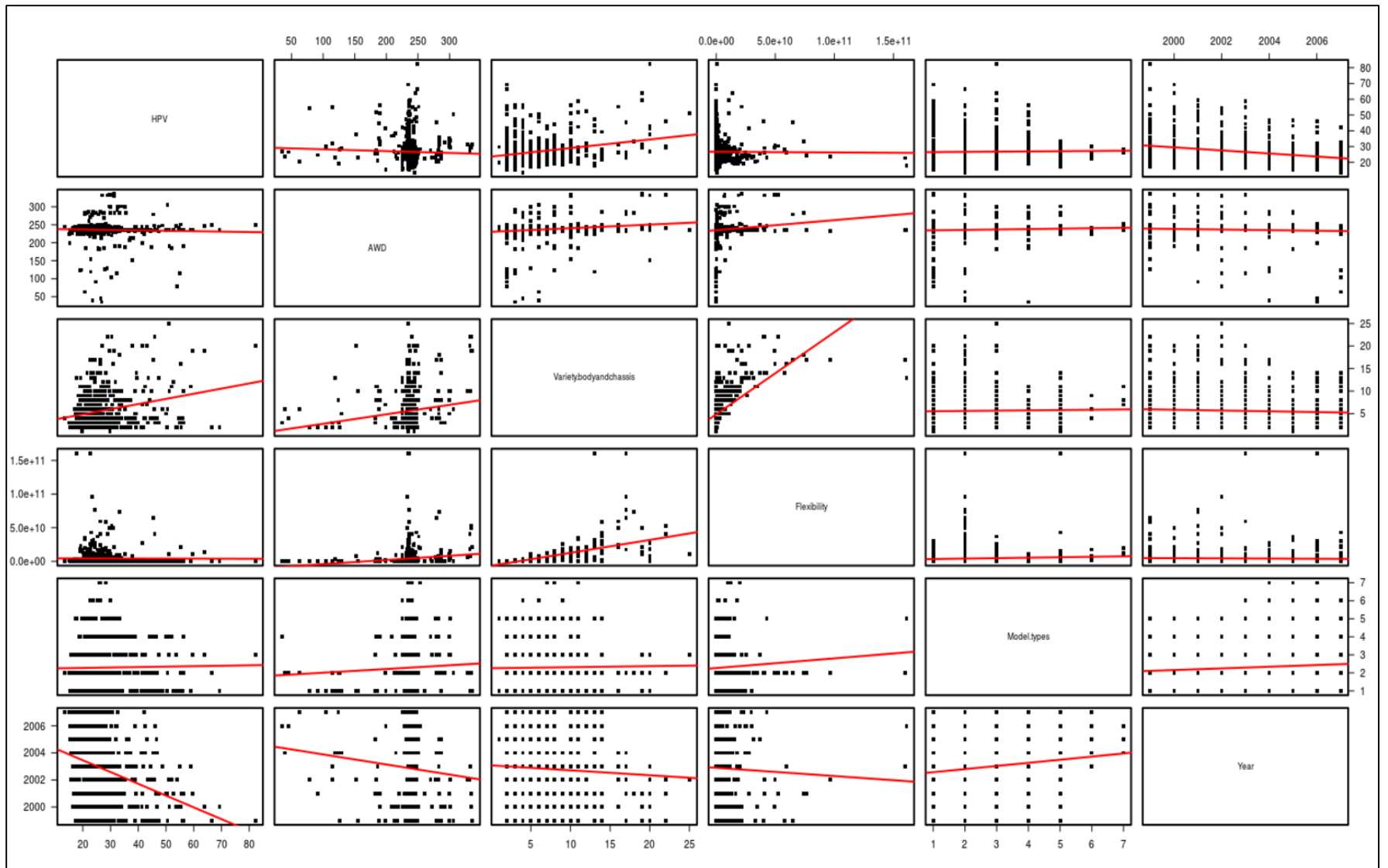


Figure 4.3. Scatterplot of the variables

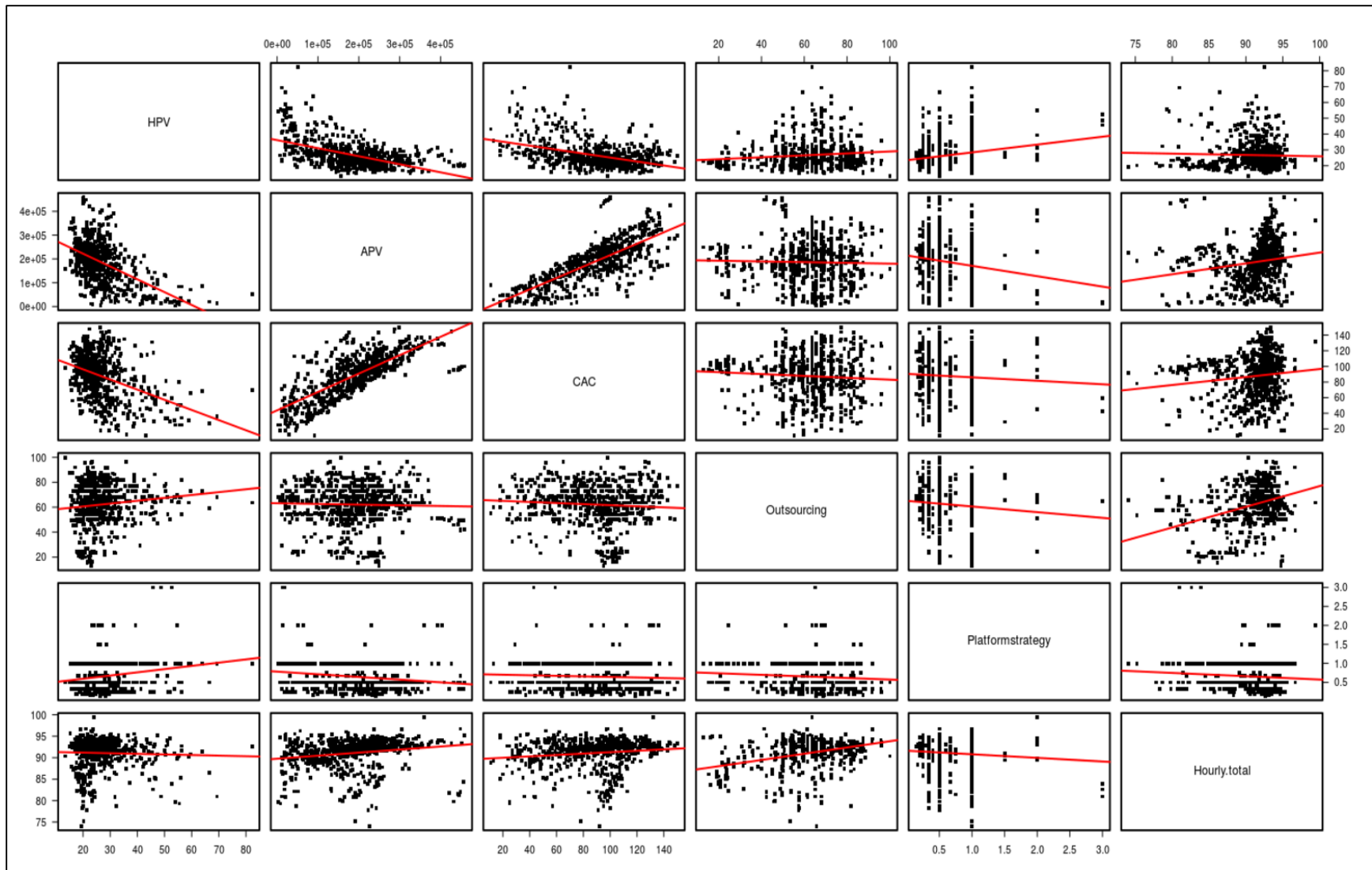


Figure 4.4. Scatterplot of the variables

### 4.6.3. Generic Multiple Linear Regression Model

A multiple linear regression is developed based on 13 independent variables (all variables other than ergonomic), and HPV is considered as the dependent variable. The generic form of the model includes all variables and their interaction; however, no transformation and interaction are considered at this step.

The regression equation, 4.1, for HPV is as follows:

$$E(\text{HPV}) = \beta_0 + \sum_{i=1}^{13} \beta_i F(x_i) + \sum_{i \neq j}^{13} \beta_{ij} F(x_i) F(x_j) \quad (4.1)$$

Where,

E(HPV)	= The expected value of hours per vehicle
F(x)	= Identity function or real transformation or indicator for categorical variable (real transformation may include the power third, square, square root, one over the power third, log normal and inverse of each original continuous each original continuous variables CAC, annual production volume, flexibility, outsourcing, annual available working days, and platform sharing strategy)
X <sub>1</sub>	= Car segment, (categorical variable in three level)
X <sub>2</sub>	= New product launch
X <sub>3</sub>	= Ownership, (American, Japanese, and joint venture)
X <sub>4</sub>	= Year
X <sub>5</sub>	= Car Assembly and Capacity utilization (CAC)
X <sub>6</sub>	= Vehicle variety
X <sub>7</sub>	= Number of models
X <sub>8</sub>	= Annual production volume
X <sub>9</sub>	= Flexibility
X <sub>10</sub>	= Outsourcing
X <sub>11</sub>	= Annual available working days
X <sub>12</sub>	= Platform sharing strategy
X <sub>13</sub>	= Hourly employees' percentage

Note: The car segment was defined from the outcome variable (HPV), however, it was explained 5% to 8% of the total variations, depending statistical methods. Hence, it was decided to keep in the regression model.

#### 4.6.4. Criterion for Choosing the Optimal Model

There are couple of criteria that help to choose the most appropriate statistical regression model like Mallorw's  $C_p$ , the Akaike information criteria (AIC), the Bayesian Information Criterion (BIC), and the adjusted  $R^2$ . Here are the definitions of different criterion (James et al., 2013).

Mallow's  $C_p$  is computed as;

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} + 2p - n \quad (4.2)$$

Where,

$RSS_p$  = Residual sum of squares of models with p variables

$\hat{\sigma}^2$  = An estimate of the variance of the error, associated with each response measurement for the model  $\hat{\sigma}^2 = \frac{RSS_p}{(n - p)}$

n = Number of observations

P = Number of parameters (excluding  $\beta_0$ )

Models with the lowest  $C_p$  value will be chosen.

The Akaike model selection procedure entails calculating AIC for each model under consideration and selecting the model with the minimum value of AIC as the preferred. In the context of selecting among regression models, a "best" model can be selected for each different size subset of predictors as well as overall. The AIC criterion is given by,

$$AIC = -2 \text{ maximum log likelihood} + 2p \quad (4.3)$$

Where,

$$-2 \text{ maximum log likelihood} = n \times \log(RSS_p/n) \text{ which is known as deviance}$$

AIC criterion has the advantage of generality and can be applied far beyond normal linear models.

The BIC criterion is given by,

$$BIC = -2 \text{ maximum log likelihood} + \log(n) \times p \quad (4.4)$$

Similar to  $C_p$  and AIC, the BIC will tend to take on a small value for a model with a low test error, and so generally the model that has the lowest BIC value will be selected. If  $n > 7$ , then the penalty term for BIC exceeds that of AIC.

The adjusted  $R^2$  statistic is calculated as,

$$\text{Adjusted } R^2 = 1 - \frac{RSS/(n-p-1)}{TSS/n-1} \quad (4.5)$$

Where,

TSS = Total sum of squares

A large value of the adjusted  $R^2$  indicates a model with a small test error or equivalent we could minimize  $RSS/(n-p-1)$ .

#### 4.6.5. Initial Multiple Linear Regression Model

A regression model with many explanatory variables may be difficult to maintain, and models with a limited number of explanatory variables are easier to implement. A stepwise regression is among the several methods for selecting an appropriate regression model. In this study, the ordinary least squares stepwise regression (at this step of the research without variables interaction) with AIC was used to find the best subset of independent variables that adequately describes the relationship between themselves and the HPV. AIC tries to select the model that most adequately describes an unknown and high dimensional reality. Moreover, AIC is aimed to find the best approximating model to the unknown data generating process. Additionally, AIC is best for prediction as it is asymptotically equivalent to cross-validation with leave one (LOO) strategy.

The base model was built using the stepwise regression method for all independent variables. Three statistical models will be developed; a base model with all 13 predictors, another model that year is categorical, and lastly the car model type as categorical.

The analysis of variance and overall test of the statistical models are obtained as follows (Kutner et al., 2005):

$$\begin{cases} H_0: \beta_1 = \beta_2 = \dots = \beta_p \\ H_A: \beta_j \neq 0 \text{ at least for one } j, j=1, \dots, p \end{cases}$$

The F-test is utilized to test the aforementioned hypothesis. Rejection of  $H_0$  implies that at least one of the regressors,  $X_1, X_2, \dots, X_p$ , contributes significantly to the model.

Under the null hypothesis,  $SSR/\sigma^2 \sim \chi_p^2$  and  $SSE/\sigma^2 \sim \chi_{n-(p+1)}^2$  are independent. Therefore,

$$F_0 = \frac{SSR/p}{SSE/(n-p-1)} = \frac{MSR}{MSE} \sim F(p, n-p-1) \quad (4.6)$$

Where,

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$SSR = SSR(x_1) + SSR(x_2|x_1) + \dots + SSR(x_p|x_{p-1}, x_{p-2}, \dots, x_1)$$

The resulting statistical model and the 11 significant variables (at the 10% significance level) that were selected by the stepwise procedure are shown in Table 4.6 for the base model.

**Table 4.6. ANOVA and relevant statistics for the base model**

Variable	Estimate	Std. Error	t value	Pr(> t )	
<b>(Intercept)</b>	1.77E+03	1.70E+02	10.429	< 2e-16	***
<b>AWD</b>	2.73E-02	8.62E-03	3.166	0.001617	**
<b>Variety body and chassis</b>	5.75E-01	7.08E-02	8.128	2.12E-15	***
<b>Flexibility</b>	-8.60E-11	2.13E-11	-4.044	5.87E-05	***
<b>Model.types</b>	5.92E-01	1.93E-01	3.064	0.002274	**
<b>Year</b>	-8.64E-01	8.47E-02	-10.192	1.79E-12	***
<b>APV</b>	-3.28E-05	4.10E-06	-7.989	5.95E-15	***
<b>CAC</b>	-2.41E-02	1.24E-02	-1.945	0.05219	.
<b>Hourly.total</b>	-1.76E-01	8.53E-02	-2.066	0.039225	*
<b>New ownership Japanese</b>	-2.94E+00	8.29E-01	-3.547	0.000417	***
<b>New ownership Joint venture</b>	-2.46E+00	9.76E-01	-2.526	0.011777	*
<b>Vehicle Launch</b>	3.90E+00	5.43E-01	7.186	1.79E-12	***
<b>Segment.MD.B</b>	2.90E+00	4.88E-01	5.941	4.55E-09	***
<b>Segment.MD.C</b>	9.99E+00	9.26E-01	10.78	< 2e-16	***
<b>Signif. Codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>					
<b>Residual standard error: 5.562 on 668 degrees of freedom</b>					
<b>Multiple R-squared: 0.5863, Adjusted R-squared: 0.5783</b>					
<b>F-statistic: 72.83 on 13 and 668 DF, p-value: &lt; 2.2e-16</b>					
<b>AIC = 2,354.42</b>					

Findings from this model include:

- The global F-test (p-value = 0.000) indicates that the model is significant for predicting the HPV based on a group of independent variables included in the model.
- Out of 13 independent variables, all except for outsourcing and platform strategy were found to be statistically significant at the 10% significance level.
- The adjusted R-squared value is 0.578, which means approximately 58% of the variation in HPV is explained by the set of 11 independent variables chosen.

- The final AIC was 2,354.42

#### 4.6.6. Base Model with Considering Year as a Categorical Variable

In this section another statistical model is developed using a year as a three-level-categorical variable as discussed in section 4.4.

The resulting statistical model and the 12 significant variables (at the 10% significance level) that were selected by the stepwise procedure are shown in Table 4.7.

**Table 4.7. ANOVA and relevant statistics for the base model with year as categorical**

Variable	Estimate	Std. Error	t value	Pr(> t )	
<b>(Intercept)</b>	4.49E+01	8.40E+00	5.35	1.20E-07	***
<b>Vehicle Launch</b>	3.95E+00	5.52E-01	7.15	2.20E-12	***
<b>New ownership Japanese</b>	-4.06E+00	1.00E+00	-4.04	6.00E-05	***
<b>New ownership Joint venture</b>	-3.31e+00	1.10E+00	-3.01	0.0027	**
<b>Segment.MD.B</b>	3.08E+00	4.95E-01	6.23	8.50E-10	***
<b>Segment.MD.C</b>	1.02E+01	9.38E-01	10.86	< 2e-16	***
<b>Model.types</b>	6.17E-01	1.98E-01	3.11	0.0019	**
<b>Year code.B</b>	-3.06E+00	5.21E-01	-5.87	6.90E-09	***
<b>Year code.C</b>	-4.89E+00	5.42E-01	-9.01	< 2e-16	***
<b>AWD</b>	2.82E-02	8.71E-03	3.24	0.0013	**
<b>Variety bodyandchassis</b>	5.56E-01	7.26E-02	7.67	6.30E-14	***
<b>Flexibility</b>	-8.87E-11	2.15E-11	-4.13	4.10E-05	***
<b>APV</b>	-3.23E-05	4.15E-06	-7.77	2.90E-14	***
<b>CAC</b>	-2.21E-02	1.25E-02	-1.76	0.0783	.
<b>Outsourcing</b>	-2.90E-02	1.80E-02	-1.61	0.1083	
<b>Hourly.total</b>	-2.10E-01	8.69E-02	-2.42	0.016	*
<b>Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>					
<b>Residual standard error: 5.62 on 666 degrees of freedom</b>					
<b>Multiple R-squared: 0.579, Adjusted R-squared: 0.57</b>					
<b>F-statistic: 61.2 on 15 and 666 DF, p-value: &lt; 2e-16</b>					
<b>AIC = 2,370</b>					

Findings from this model include:

- The global F-test (p-value = 0.000) indicates that the model is significant for predicting the HPV based on a group of independent variables found in the model.
- Out of 13 independent variables, all other than platform strategy were found to be statistically significant at the 10% significance level and outsourcing is significant at 10.08%.



- The adjusted R-squared value is 0.57, which means approximately 57% of the variation in HPV is explained by the set of 11 independent variables chosen, without outsourcing variable.
- The final AIC was 2,370

#### 4.6.7. Base Model with Considering Car Model Types as a Categorical Variable

In this section another statistical model is developed based on car model's type studied as a three-level-categorical variable as discussed in section 4.5.

The resulting statistical model and the 11 significant variables (at the 10% significance level) that were selected by the stepwise procedure are shown in Table 4.8 for the initial model.

**Table 4.8. ANOVA and relevant statistics for the base model with car model's types as categorical**

Variable	Estimate	Std. Error	t value	Pr(> t )	
<b>(Intercept)</b>	1.72E+03	1.69E+02	10.18	< 2e-16	***
<b>Vehicle Launch</b>	3.93E+00	5.40E-01	7.27	1.00E-12	***
<b>New ownership Japanese</b>	-2.94E+00	8.28E-01	-3.55	4.10E-04	***
<b>New ownership Joint.venture</b>	-2.59E+00	9.74E-01	-2.65	0.00813	**
<b>Segment.MD.B</b>	2.83E+00	4.88E-01	5.8	1.00E-08	***
<b>Segment.MD.C</b>	9.64E+00	9.29E-01	10.38	< 2e-16	***
<b>Year</b>	-8.39E-01	8.45E-02	-9.93	< 2e-16	***
<b>Model.Types.code.B</b>	2.68E+00	7.64E-01	3.5	4.90E-04	***
<b>Model.Types.code.C</b>	5.89E-01	9.73E-01	0.61	0.54512	
<b>AWD</b>	2.96E-02	8.63E-03	3.44	0.00063	***
<b>Variety body and chassis</b>	5.66E-01	7.08E-02	7.99	6.00E-15	***
<b>Flexibility</b>	-7.94E-11	2.13E-11	-3.72	2.10E-04	***
<b>APV</b>	-3.23E-05	4.08E-06	-7.91	1.00E-14	***
<b>CAC</b>	-2.50E-02	1.23E-02	-2.03	0.04282	*
<b>Hourly.total</b>	-1.89E-01	8.54E-02	-2.21	0.02755	*
<b>Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>					
<b>Residual standard error: 5.55 on 667 degrees of freedom</b>					
<b>Multiple R-squared: 0.588, Adjusted R-squared: 0.58</b>					
<b>F-statistic: 68 on 14 and 667 DF, p-value: &lt; 2e-16</b>					
<b>AIC = 2,353</b>					

Findings from this model include:

- The global F-test (p-value = 0.000) indicates that the model is significant for predicting the HPV based on a group of independent variables found in the model.
- Out of 13 independent variables, all were found to be statistically significant at the 10% significance level except outsourcing and platform strategy. However, Model Types code.c was not statistically significant also.
- The adjusted R-squared value is 0.58, which means approximately 58% of the variation in HPV is explained by the set of 11 independent variables chosen.
- The final AIC was 2,353

#### **4.6.8. Model Adequacy Checking**

Examination of the residuals is a crucial part of constructing any general linear model. If the model is adequate, the residuals should be structureless and they should contain no obvious patterns (Montgomery, 2012). Analysis of residuals for the purpose of model adequacy checking includes the following four steps:

1. The normality assumption: check whether or not the residuals are normally distributed, NID  $(0, \sigma^2)$ .
2. No correlation between the residuals in time sequence: plotting of residuals in time order will help to find if there is any special time sequence for the errors.
3. Residuals versus fitted value: the plotting of residuals versus fitted values should be structureless and should not follow any specific pattern.
4. Plots of residuals versus other variables: plotting the residuals against each variable in the final model should indicate that the variances are constant.

Residuals versus fitted values, normal plot of residuals, residuals scale location, and standardized residuals versus the leverage are shown in Figures 4.5-4.7 for all three statistical models. A careful examination of these Figures show that there is no severe abnormality among the residuals, and the above conditions are met. However, four observations (129, 409, 416, and 491) have the high potential of being outliers.

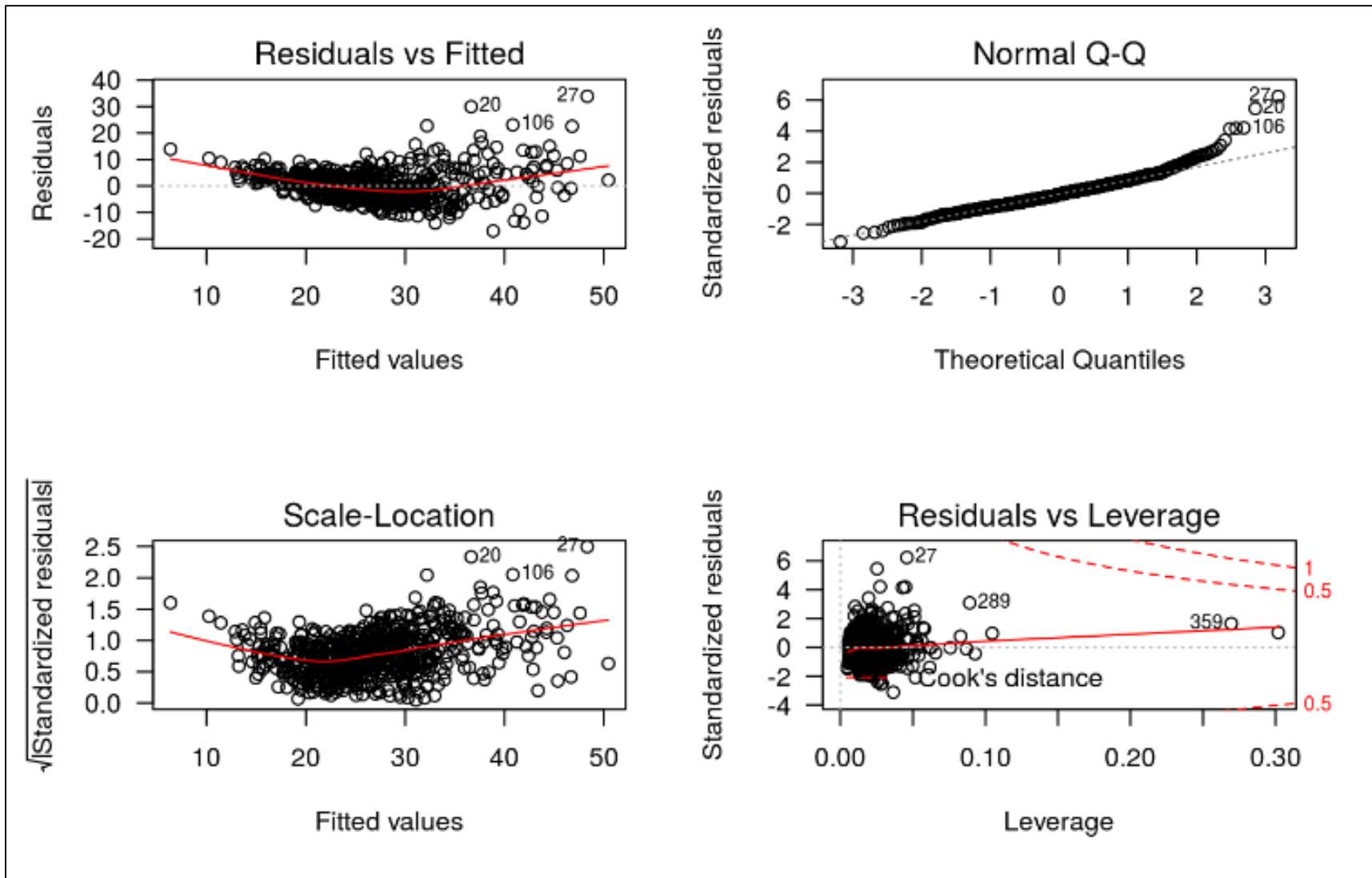


Figure 4.5. Model adequacy checking for the base model

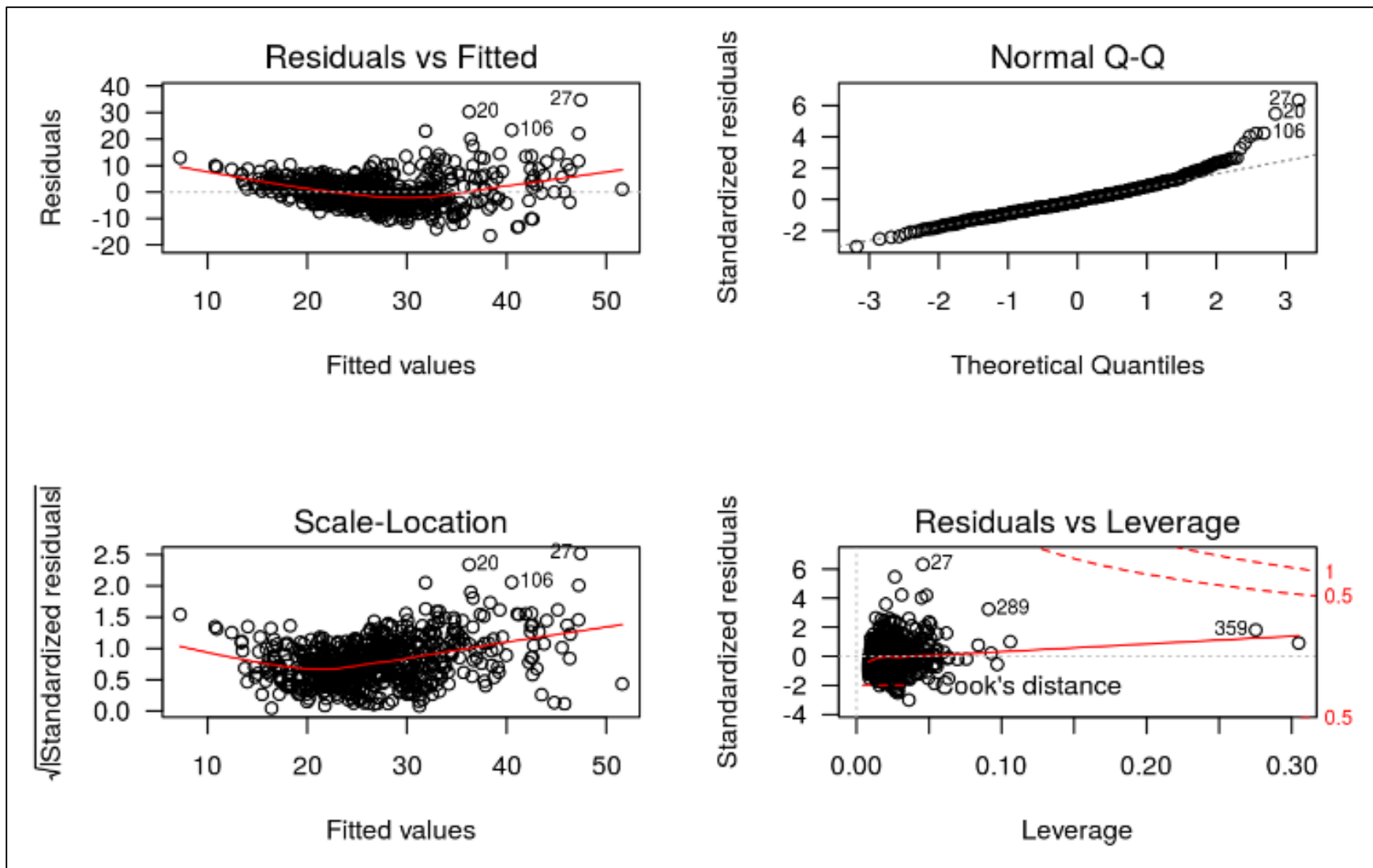


Figure 4.6. Model adequacy checking for the base model that year is categorical

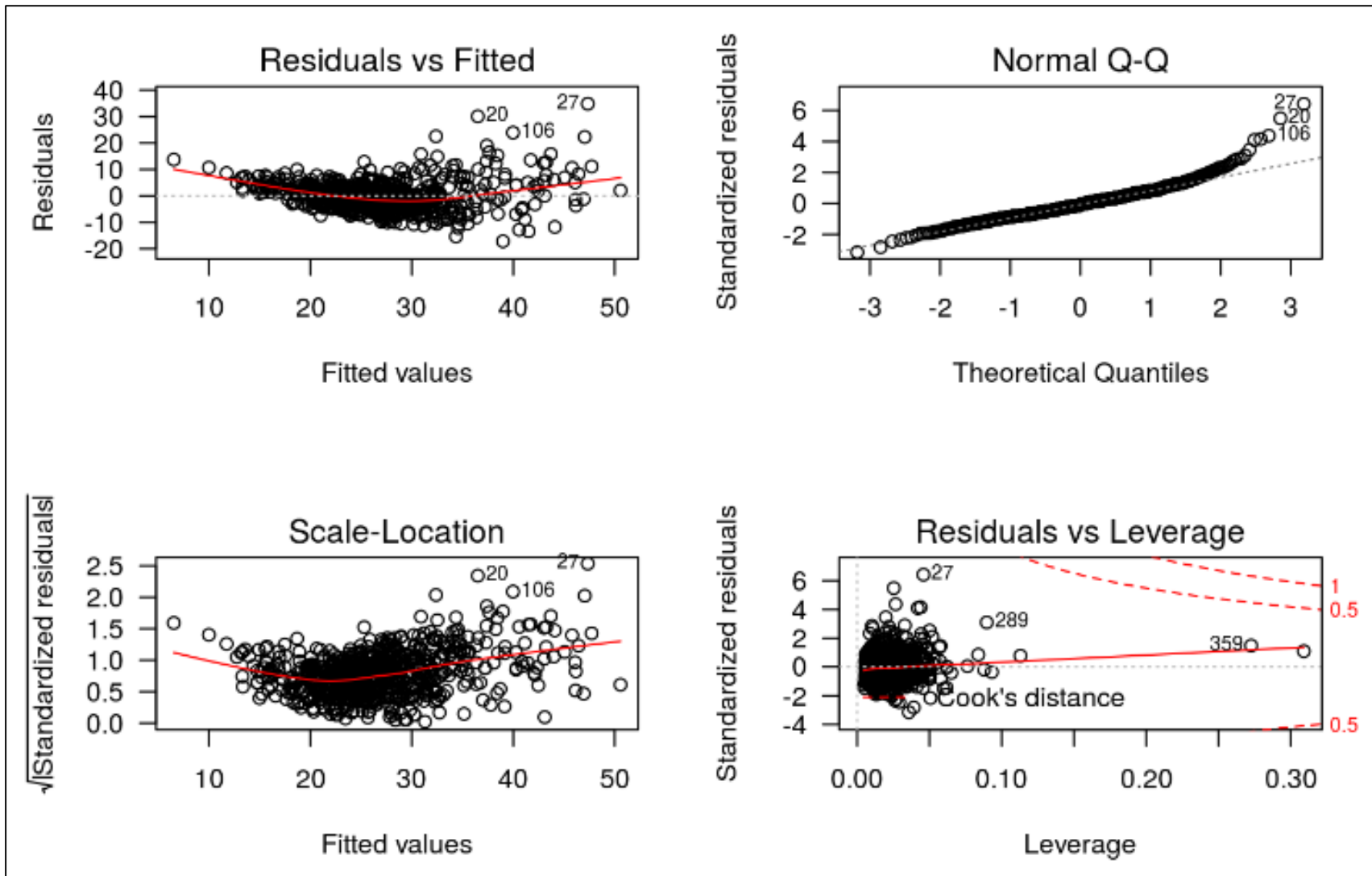


Figure 4.7. Model adequacy checking for the base model that car model types is categorical

#### 4.6.9. Comparing all three Models and Hypotheses Testing

A comparison regarding the different variables' interpretability for three discussed models is shown in Table 4.9.

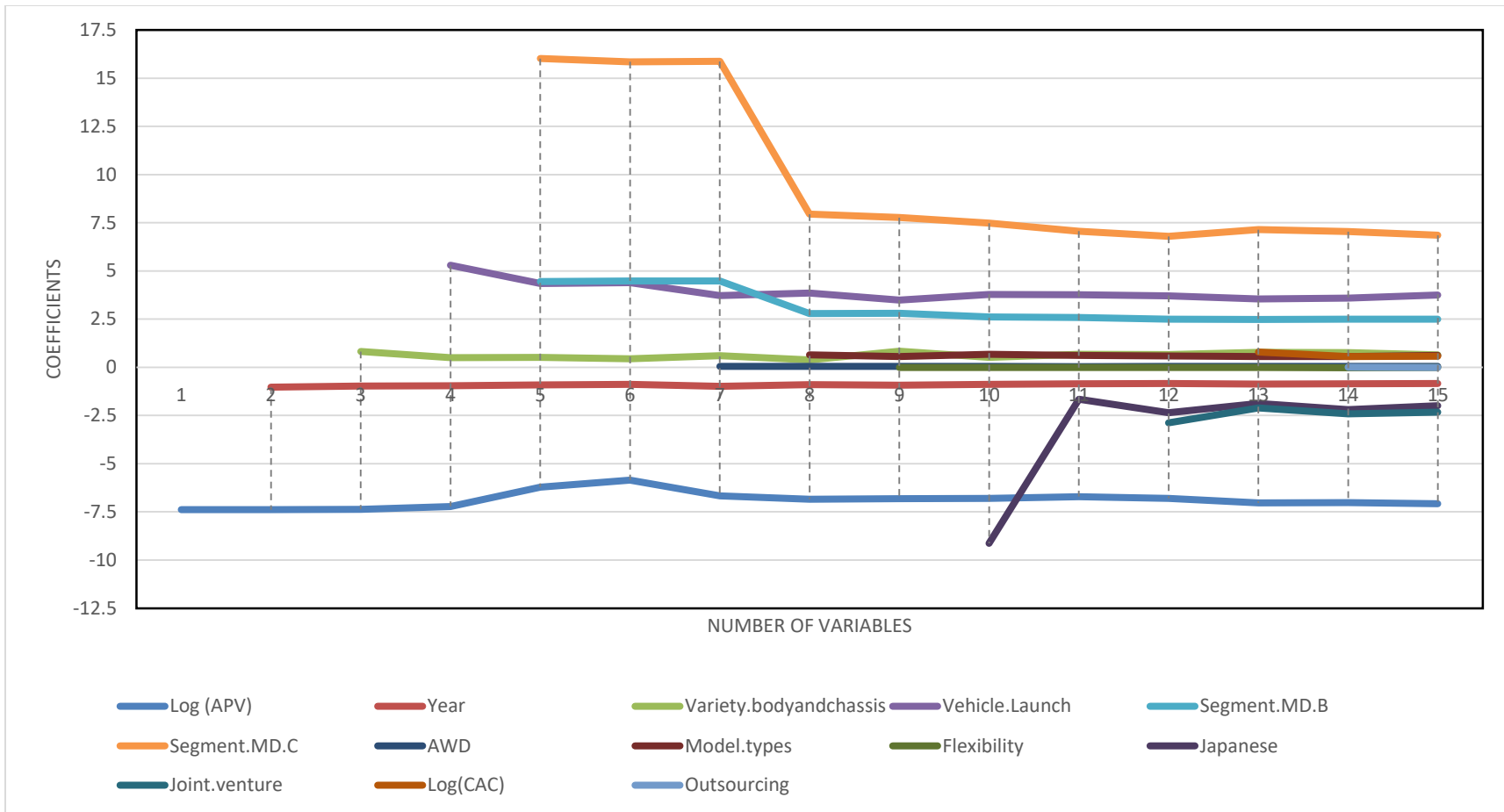
**Table 4.9. Comparison of all three models about different variables**

Variable	Base Model		Year as categorical		Car model types as categorical		Result
	Estimate	Pr(> t )	Estimate	Pr(> t )	Estimate	Pr(> t )	
(Intercept)	1.77E+03	< 2e-16	4.49E+01	1.20E-07	1.72E+03	< 2e-16	
AWD	2.73E-02	0.001617	2.82E-02	0.0013	2.96E-02	0.00063	+
Variety (body & chassis)	5.75E-01	2.12E-15	5.56E-01	6.30E-14	0.566	60E-15	+
Flexibility	-8.60E-11	5.87E-05	-8.87E-11	4.10E-05	-7.94E-11	0.00021	-
Model types	5.92E-01	0.002274	6.17E-01	0.0019	NA		+
Year	-8.64E-01	1.79E-12	NA		-8.39E-01	< 2e-16	-
APV	-3.28E-05	5.95E-15	-3.23E-05	2.90E-14	-3.23E-05	1.00E-14	-
CAC	-2.41E-02	0.05219	-2.21E-02	0.0783	-2.50E-02	0.04282	-
Platform strategy	NS		NS		NS		NS
Hourly/total	-1.76E-01	0.039225	-2.10E-01	0.016	-1.89E-01	0.02755	-
New ownership Japanese	-2.94E+00	0.000417	-4.06E+00	6.00E-05	-2.94E+00	4.10E-04	-
New ownership Joint venture	-2.46E+00	0.011777	-3.31E+00	0.0027	-2.59E+00	0.00813	-
Vehicle Launch	3.90E+00	1.79E-12	3.95E+00	2.20E-12	3.93	1.00E-12	+
Segment MD.B	2.90E+00	4.55E-09	3.08E+00	8.50E-10	2.83E+00	1.00E-08	+
Segment MD.C	9.99E+00	< 2e-16	1.02E+01	< 2e-16	9.64E+00	< 2e-16	+
Outsourcing	NS		-2.90E-02	0.1083	NS		- (NS)
Year code.B	NA		-3.06E+00	6.90E-09	NA		-
Year code.C	NA		-4.89E+00	< 2e-16	NA		-
Model Types code.B	NA		NA		2.68E+00	4.90E-04	+
Model Types code.C	NA		NA		5.89E-01	0.54512	

Signif. Codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 and NA means not applicable

NS: not significant, NA: not applicable

The impact of each variable on the HPV can be checked by Table 4.9. However, the results of Table 4.9 are reliable if the correlational structure is robust. Each individual predictor was chosen at a time with respect to its importance (through the best subset selection) and then the other variables were added to the models to check whether or not the correlational structure is robust. If the correlational structure is robust there mustn't be a significant change in the coefficients value and their signs. The coefficients trend for all regressors when new variables are added to the statistical models are shown in Figure 4.8. Figure 4.8 indicates that all the coefficients (values and signs) are almost consistent in any statistical models. On the other hand, the sign of the coefficients is identical with respect to Table 4.9 (the only difference is the CAC which in Figure 4.6  $\ln(\text{APV})$  was used). Hence, the results of Table 4.9 and Figure 4.6 are completely in agreement to each other.



**Figure 4.8. All regressors coefficients trends when new variables are added to the model**



Table 4.9 shows all variables except outsourcing and CAC are statistically significant for all the three statistical models; however, outsourcing and CAC are also significant for some models. Hence, the raised hypotheses in chapter 3 could be addressed using Table 4.9 and through the following step. Here is the hypotheses that check the coefficient  $\beta_j$ , (Kutner et al., 2005),

$$\begin{cases} H_0: \beta_j = 0 & \text{where } j=1, 2, \dots, P \\ H_A: \beta_j \neq 0 \end{cases}$$

In multiple linear regression, under the null hypothesis  $t_0 = \hat{\beta}_j / \widehat{se}(\hat{\beta}_j) \sim t(n-p-1)$ . The  $H_0$  is rejected if  $|t_0| > t(n-p-1, 1-\alpha/2)$ . This is a partial test because  $\hat{\beta}_j$  depends on all of the other predictors  $X_i, i \neq j$  that are in the model. Thus, this is a test of the contribution of  $X_j$  given the other predictors in the model. Hence, the hypotheses' results are shown in Table 4.10.

**Table 4.10. Hypotheses result**

Hypo.	Research Hypotheses	Hypotheses result
<b>H3.1</b>	The HPV variation for American, Japanese, and joint venture companies is significantly different	Research hypothesis was accepted. Japanese has the lowest HPV followed by joint venture plants
<b>H3.2</b>	HPV is correlated with the year of production	Research hypothesis was accepted.
<b>H3.3</b>	The car segment has a significant impact on the HPV	Research hypothesis was accepted. Segment C has the highest HPV followed by segment B
<b>H3.4</b>	HPV is correlated with CAC	Research hypothesis was accepted. Statistically significant at 10%
<b>H3.5</b>	HPV is correlated with the number of models	Research hypothesis was accepted.
<b>H3.6</b>	HPV is correlated with the vehicle variety	Research hypothesis was accepted.
<b>H3.7</b>	HPV is correlated with Platform strategy	Research hypothesis was accepted.
<b>H3.8</b>	HPV is correlated with annual production volume	Research hypothesis was accepted.
<b>H3.9</b>	HPV is correlated with flexibility	Research hypothesis was accepted.

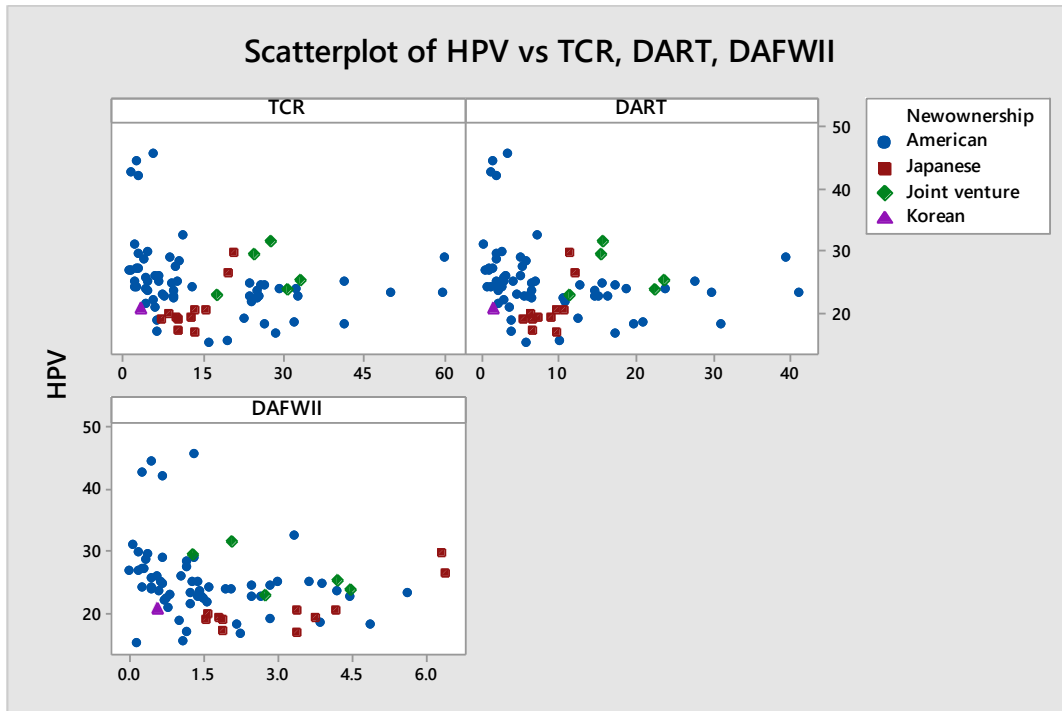
<b>Hypo.</b>	<b>Research Hypotheses</b>	<b>Hypotheses result</b>
<b>H3.10</b>	HPV is correlated with outsourcing	Research hypothesis was accepted.
<b>H3.11</b>	HPV is correlated with launching a new vehicle	Research hypothesis was accepted.
<b>H3.12</b>	HPV is correlated with annual available working days	Research hypothesis was accepted. (positive relation)
<b>H3.13</b>	HPV is correlated with hourly employees' percentage	Research hypothesis was accepted. (negative relation)

Table 4.10 indicates that all null hypotheses were statistically accepted.

Hence, all thirteen hypotheses (null or alternative hypothesis) were addressed. The Pearson correlation coefficients and their significance level are illustrated in Table AII (appendix). The H3.14 can be checked accordingly,

- H3.14: The ergonomics indices provided by OSHA has a relation with HPV

The scatter plot of the HPV versus total case rate (TCR), days away, restricted, and transfer (DART) case rate, and the days away from work (DAFWII) is illustrated for 77 plants in Figure 4.9. Figure 4.9 shows that there is a distinct difference between American, Japanese, and joint venture plants.



**Figure 4.9. Scatter plot of HPV versus TCR, DART, DAFWII**

The result of a Pearson correlation between HPV and TCR, DART, and DAFII are shown in Table 4.11. Table 4.11 depicts that there is a significant positive association between ergonomic indices and HPV for Japanese plants. It means that if the TCR, DART, and DAFWII are reduced the HPV will also improve. However, for the American plants it's vice versa and the Pearson correlation is statistically negative and much lower, which means by increasing the TCR, DART, and DAFWII, HPV will be reduced. There could be couple of reasons such as:

- The job cultural difference between the Japanese and American plants. In Japanese culture, employees usually are very loyal and won't be absent from the job for such low priority issues.
- Japanese plants could be more sensitive to the employees' ergonomics in comparison to the American; hence, the employees of Japanese plants may face a lesser risk during the work.
- There are four unusual American plants (all GM Bowling Green) that might affect the correlation value and resulting a negative value also.
- However, it's necessary to mention that the number of observations for Japanese plants are fewer in comparison with American plants (11 versus 61).

**Table 4.11. Pearson Correlation for the Ownership**

	HPV		
<b>Pearson Correlation</b>	American	Japanese	Joint Venture
<b>TCR</b>	<b>-0.32</b>	<b>0.828</b>	<b>0.137</b>
<b>P-Value</b>	0.012	0.002	0.827
<b>DART</b>	<b>-0.287</b>	<b>0.677</b>	<b>-0.184</b>
<b>P-Value</b>	0.025	0.022	0.767
<b>DAFWII</b>	<b>-0.277</b>	<b>0.829</b>	<b>-0.691</b>
<b>P-Value</b>	0.031	0.002	0.196
<b>Number of Plants</b>	61	11	5

#### **4.6.10. Model Selection Among the three Models**

The statistical results for the base model, base model with year as categorical, and base model with car model's types as categorical were presented in the previous section. In this section, the best statistical model will be chosen in terms of performance, simplicity, and interpretability.

Different criterion (performance measures, simplicity, and interpretability) for all three models are shown in Table 4.12. Comparing all the criterion indicates that the base model is the best statistical model among all three statistical models.

**Table 4.12. Discussed statistical model performance characteristic**

Regression methods	No. of var.	DF	R-sq	R-sq (adj)	AIC	Error DOF	Residual Standard Error (RSE)	P-value	Comment	Performance (adj R-sq, AIC, & RSE)	Simplicity	Interpretability
<b>Base model</b>	11	13	0.587	0.578	2,354	668	5.562	2.2E-16	Outsourcing and platform were not statistically significant	Medium	High	High
<b>Base model with year as categorical</b>	11	15	0.579	0.570	2,370	666	5.62	2.0E-16	Outsourcing and platform were not statistically significant	Low	Low	Medium
<b>Base model with model as categorical</b>	11	14	0.588	0.58	2,353	667	5.55	2.0E-16	Outsourcing, platform, and Types code.c were not statistically significant	High	Medium	Low

The final models for the two categorical variables, ownership and car segments, and launching a new vehicle as a dummy variable are shown in Table 4.6.

Here are the main conclusions that can be drawn from the final regression model:

- When the automakers are involved with a new product launch the HPV increased by 3.90 hours.
- The average HPV difference between car segments A and B is 2.90 hours without a new product launch, and 6.8 hours while launching a new product.
- The average HPV difference between car segments A and C is 9.99 hours without a new product launch, and 12.89 hours while launching a new product.
- Japanese companies are the most productive plants followed by joint venture. The HPV for Japanese and joint venture are lower than American plants with 2.94 and 2.46 hours on average, respectively.

#### **4.7. Conclusion**

In this chapter, all the defined and developed predictors and HPV other than year, segments, and ownerships were compared for American, Japanese, and joint venture. Interestingly, all regressors were statistically different for American, Japanese, and joint venture plants except annual available working days and launching a new product. Furthermore, all the hypotheses that were identified in chapter 3 were discussed and addressed. Finally, three statistical models were developed to estimate the HPV and the most appropriate one among those was chosen as the base model, further analysis will be done on the base model during the dissertation to enhance the prediction statistical model.

## 5 : Model Improvement

### 5.1. Introduction

In this chapter all the variables' transformations are checked for possibility of increasing the power of test and best variables transformation. Moreover, an approach about how the variables interaction will be taken into account is presented.

### 5.2. Transformations

Figure 4.3a and 4.4b show some potential non-linearity characteristic with respect to a few exploratory variables such as annual production volume (APV) and car assembly and capacity utilization (CAC). The nonlinearity characteristic of independent variables with respect to the dependent variable (HPV) can violate the assumption of linearity. There are some remedial measures to remove the effect of independent variables non-linearity (Kutner et al., 2005). Here are these methods,

1. Build a new regression model which is more appropriate.
2. Perform transformation the data such that new variables meet the linearity assumption of exploratory variables.

Both methods have their own advantages and disadvantages. The model which is provided from the first approach might be complex and difficult to interpret its parameters. However, the second approach (transformations) is more common and easier to handle. The transformation can be applied either on dependent or independent variables or both.

In general, transformations are used for various reasons such as (Box &Cox, 1964; Hoaglin et al., 1983; John &Draper, 1980; J. W. Tukey, 1960; J. W. Tukey, 1977)

1. Convenience: standardization of data
2. Reducing skewness
3. Equal spreads: Responding to the heteroscedasticity concern among the variables and moving towards homoscedasticity
4. Linear relationships: When linear regression is utilized, it's mandatory that the variables look linear with respect to the dependent variable

- Additive relationships: When additive models are used relationships are usually easier to analyze rather than multiplicative models.

In statistics, the power transform is a family of functions that are applied to create a monotonic transformation of data using power functions. This is a useful data transformation technique used to make the data more normal distribution-like, stabilize variance, improve the validity of measures of association such as the Pearson correlation between variables and for other data stabilization procedures. The main and common type of transformations was introduced by Tukey (1977) as Tukey’s ladder of transformation. A Tukey transformation can be found using Equation 5.1.

$$Y = \begin{cases} X^\lambda & \text{if } \lambda > 0 \\ \text{Log } X & \text{if } \lambda = 0 \\ -X^\lambda & \text{if } \lambda < 0 \end{cases} \quad (5.1)$$

A modified Tukey’s ladder of transformation is reproduced (when  $\lambda < 0$ ) and shown in Table 5.1.

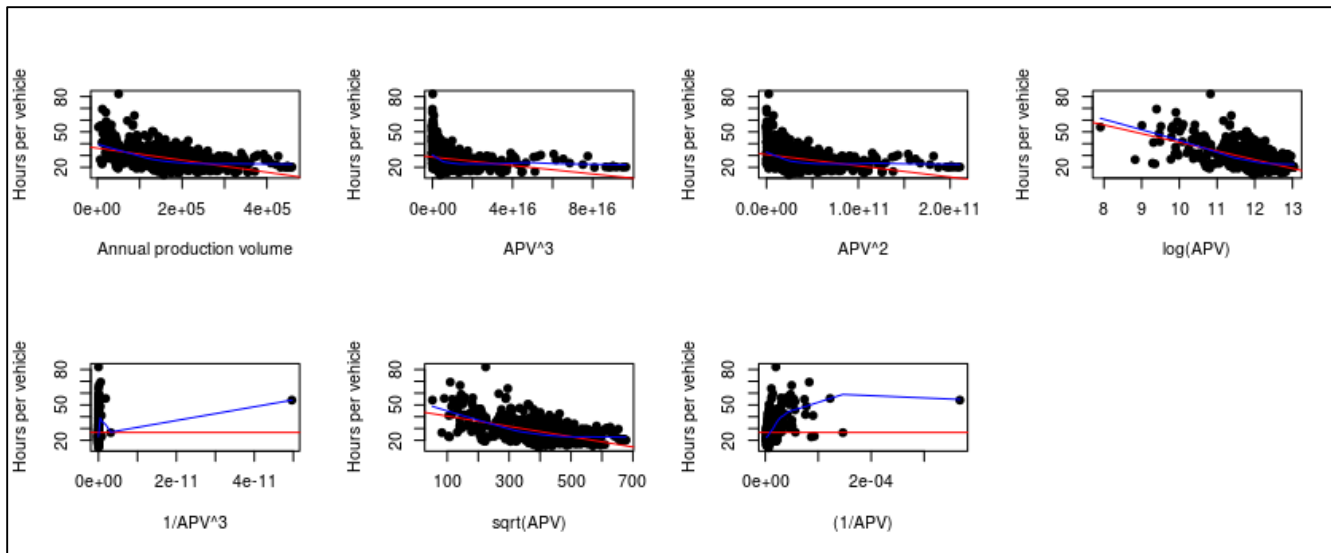
**Table 5.1. Modified Tukey’s ladder of transformation**

$\lambda$	-2	-1	-0.5	0	0.5	1	2
Y	$-1/X^2$	$-1/X$	$-1/X^{0.5}$	Log X	$X^{0.5}$	X	$X^2$

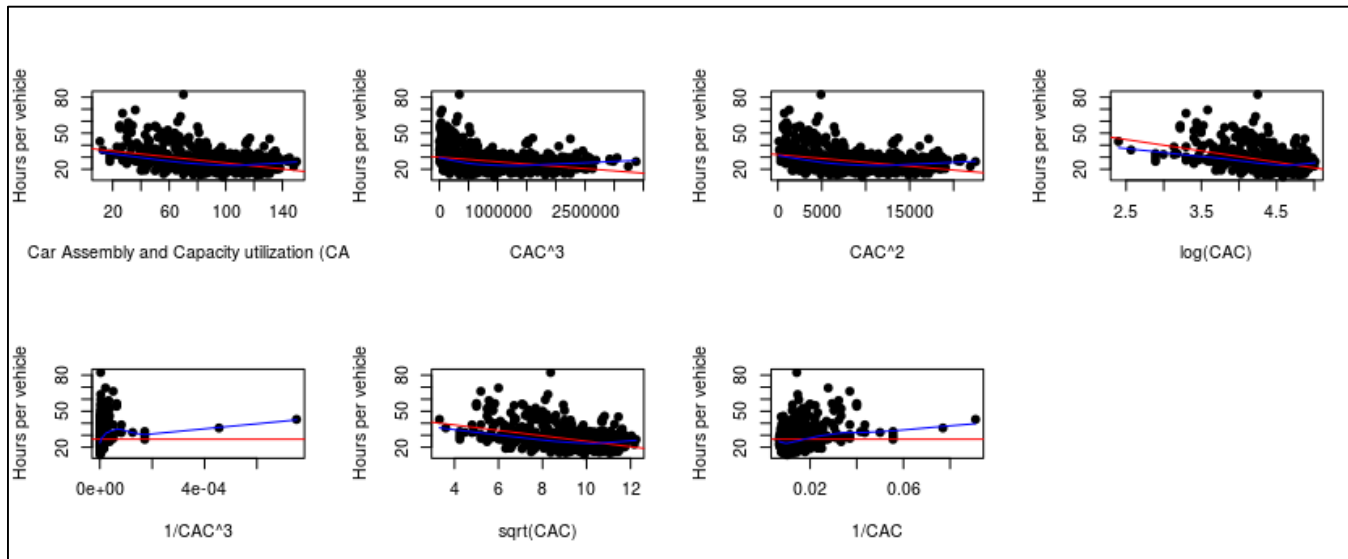
Different transformations (such as  $X^3$ ,  $X^2$ ,  $\text{Log}(X)$ ,  $X^{0.5}$ ,  $1/X$ , and  $1/X^3$ ) were applied for the APV and CAC to make them linear with respect to HPV. Figures 5.1 and 5.2 illustrate different transformation for both APV and CAC, respectively. In both situations, the Log normal transformation was more linear with respect to HPV than the other types of transformation. The blue line is LOESS Curve Fitting (Local Polynomial Regression) that is closer to the red line (true regression line) when Log transformation is used. Hence, it was decided to use Log transformation for AVP and CAC independent variables.

However, other variables (HPV, flexibility, platform strategy, percentage of hourly employee, and AWD) were also transformed to check whether or not the model performance will significantly improve. The results are illustrated in Table 5.2.





**Figure 5.1. Plotting different transformations on APV (the blue line is LOESS Curve Fitting (Local Polynomial Regression))**



**Figure 5.2. Plotting different transformations on CAC (the blue line is LOESS Curve Fitting (Local Polynomial Regression))**

**Table 5.2. Different Model Transformation comparison**

Type of Transformation for Different regression Models	Variable	AIC	R-sq	R-sq (adj)	Error D.O.F	Residual standard error	P-value	Comment
No transformation	11	2,354	0.5863	0.5783	668	5.562	<2.2 00E <sup>-16</sup>	Outsourcing was not significant
Ln (CAC and APV)	10	2,230	0.654	0.648	669	5.080	<2.00E <sup>-16</sup>	Ln (CAC), outsourcing, hourly/total were not significant
Ln (CAC, APV, and variety)	9	2,220	0.659	0.653	670	5.050	<2.00E <sup>-16</sup>	Ln(APV), hourly/total, Ln (platform strategy), outsourcing were not significant
Ln (CAC, APV, variety, outsourcing, and platform strategy)	9	2,220	0.660	0.653	669	5.040	<2.00E <sup>-16</sup>	Ln(APV), hourly/total, Ln(platform strategy), outsourcing were not significant
Ln (HPV and APV)	9	(2,468)	0.654	0.646	666	0.162	<2.00E <sup>-16</sup>	CAC, hourly/total, platform strategy, and outsourcing were not significant
Ln (CAC, APV, variety, outsourcing, and platform strategy) and Ln (HPV)	10	(2,490)	0.664	0.658	669	0.160	<2.00E <sup>-16</sup>	Ln (CAC), outsourcing, hourly/total not significant
Ln on all the variables	10	(2,488)	0.664	0.658	668	0.160	<2.00E <sup>-16</sup>	Ln (CAC), ln (platform strategy), and ln(hourly/total) were not significant

Table 5.2 indicates that the transformation of other variables didn't significantly improve the model performance (Adjusted R-squared and AIC). However, the adjusted R-squared and residual standard error were significantly improved from 0.5791 and 5.556 to 0.648 and 5.08. That means transformation of APV and CAC improve the model performance significantly; however, it's easier to work with two transformed variables. Hence, for the rest of the research it was decided to work with a simpler model that is also accurate enough and has only APV and CAC transformed variables.

### **5.3. Variables Interaction**

Interaction arises when the effect of one explanatory variables depends on specific level or value of other independent variables (Fitzmaurice, 2000).

Hence, it is worthy to identify those interactions that may likely influence the outcome. A priori knowledge is an effective way to decide about those; however, plotting the residuals for the additive regression model versus various interaction terms to determine which ones are influential in affecting the outcome variable could be another way (Kutner et al., 2005).

The detailed approach about considering the variables interaction will be discussed in section 6.5.

### **5.4. Model Validation**

In this step, the predictive capability of the selected model is checked by model validation procedures. Model validation can be done through estimating the test error using either a validation set approach or cross-validation approach. The set error or the cross-validation error for each obtained statistical model can be calculated, then the model which has the lowest estimated test error is selected. Since this method provides a direct estimate of the true model, it has an advantage relative to AIC, BIC, Cp, and adjusted R-squared. Moreover, fewer assumptions about a true underlying model are made comparing using AIC, BIC, Cp, and adjusted R-squared (James et al., 2013).

In the basic cross validation approach, k-fold CV, the entire data-set randomly is split to K equal sample size, called K "folds". A model is trained using K-1 of the folds as training data-set and the resulting model is validated on the remaining part of the data to calculate the performance

measure. The final reported model performance is the average of the values that was calculated in the loop. The major goal in the cross-validation might be to define how well the mean squared error (MSE) can be estimated. Another approach could be leave one out cross validation (LOOCV) that the data-set is split to the K-folds in which K is equal to the number of observations (K=n). This method may include exhaustive analysis.

The estimated observations  $\widehat{Y}_{ij}$  could be calculated through the HPV regression model. Then, the error term could be calculated as the difference between the actual response and estimated observations,  $Y_{ij} - \widehat{Y}_{ij}$  for each fold (in this study, 10). Consequently, the k-fold cross validation, CV, can be estimated as,

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^k MSE_i$$

Where,

$$MSE_i = \sum_{j=1}^{N_j} \frac{(Y_{ij} - \widehat{Y}_{ij})^2}{N_j} \quad (N_j \text{ is the number of observations of CV instance } j)$$

In general, cross-validation is performed to pursue a crucial goal, which is estimating of the test MSE.

Kohavi (1995) showed that using a cross validation with number of folds as ten outperform using different number of folds such as five, twenty, or the LOOCV method (Kohavi, 19995).

## 5.5. Conclusion

In this chapter the modified Tukey's ladder of transformations was checked for all the dependents and independent variables. Finally, it was decided to use ln transformations on APV and CAC only because of their impact on power of test incremental and simplicity. Also, an approach about how the two-way variables interaction will be utilized was presented.

## 6 : Different Statistical Analysis

### 6.1. Introduction

Linear least-squares regression can be very sensitive to unusual data. In this chapter several techniques are considered to down-weight the effect of the outliers on the obtained statistical models. These techniques include robust regression estimators, which attempt to down-weight or ignore unusual data such as M-estimators (Huber and Tukey's bisquare), MM-estimator, SMDM estimate, Least Trimmed Squares Robust (LTS) regression, and quantile-regression. Moreover, another common technique such as shrinking methods were used to build a parsimonious regression model (one that is simpler and easier to interpret). The common algorithms in shrinkage methods are ridge, lasso, and elastic-net regressions. The final model will be developed by a combination of robust and shrinkage methods to define the best possible statistical model. Finally, the outperforming residuals are analyzed to determine the existence of specific pursuant strategies that they may adhere to. Hence, RO4, RO5, and RO6 are addressed in this chapter.

### 6.2. Robust Regression

Robust regression can be used in any situation in which least squares regression can be utilized. Some outliers or high leverage data points could be found when a least squares regression is fitted. It was found that these data points were not data entry errors, neither were they from a different population than most of the data. So, there is no compelling reason to exclude those observations from the analysis. Since it was decided to not exclude these observations from the analysis and to treat all them in the regression model, robust regression might be a good remedy which attempts to down-weight or ignore unusual data (Fox & Weisberg, 2010; Statistical Consulting Group, 2015).

Harner (2015) mentioned that measuring the distance of an observation from the bulk of data could be an appropriate metric. The general class of squared distance function could be very useful. If  $M$  is a positive semi-definite matrix. Then, the squared distance between  $i^{\text{th}}$  observation,  $Y_i$ , and  $\bar{Y}^{\#}$  (some location estimate) is defined by  $D_i^2$ , Mahalanobis distance (Harner, 2015),

Where,

$$D_i^2 = (Y_i - \bar{Y}^{\#})^t M (Y_i - \bar{Y}^{\#})$$

There are different methods of robust regression such as M-estimates, MM-estimates, SMDM-estimate, least trimmed squares, and quantile-regression that will be reviewed in the next section.

### 6.2.1 M-Estimate

In an effort to provide a better estimation, robust regression procedures are utilized to fade out the influence of outliers. The class of M-estimator models includes all models that are derived to be maximum likelihood models (Alma, 2011). M-estimators are useful when the environment is noisy and possible outliers are probable (Kutner et al., 2005). Model adequacy checking using, Figures 4.5 to 4.7, indicates that some outliers are probable. M-estimator is one of the most common robust regression methods that utilize weighted least squares to down weight the outliers and reduce their influence. The M-estimators define a weight function with the following estimating equation (Fox & Weisberg, 2010; Statistical Consulting Group, 2015)

$$\sum_{i=1}^n W_i(Y - X_i^t \beta) X_i^t = 0$$

Where,

W = Weight matrix

The equation is solved utilizing **Iteratively Reweighted Least Squares (IRLS)**. The weights depend on the residuals and the residuals on the weights. For instance, the coefficient matrix at iteration j is:

$$B_j = [X^T W_{j-1} X]^{-1} X^T W_{j-1} Y$$

Where,

J = The matrix at a particular iteration

Hence, the unusual observations that have large residuals are given smaller weights. The weights that are revised in an iterative procedure yield new residuals until convergence is obtained. A summary of steps followed is (Kutner et al., 2005):

1. Choosing a weight function to weight the observations.
2. Obtaining starting weights for all observations.
3. Using the starting weights in weighted least squares and obtaining the residuals from the fitted regression function.
4. Using the residuals in step 3 to obtain revised weights.
5. Continuing the iterations until convergence is obtained.

The most widely used weight functions are Huber and Tukey’s bisquare weight functions. The objective functions and weight function for least-squares, Huber, and Tukey’s bisquare estimators are shown in Table 6.1.

**Table 6.1. Objective and weight functions for least-squares, Huber, and Tukey’s bisquare estimators (adapted from Fox and Weisberg (2010))**

Method	Objective function	Weight function
Least-squares	$\rho_{LS}(e) = e^2$	$W_{LS}(e) = 1$
Huber	$\rho_H(e) = \begin{cases} 1/2e^2 &  e  \leq k^* \\ K e  - 1/2e^2 &  e  > k \end{cases}$	$W_H(e) = \begin{cases} 1 &  e  \leq k^* \\ \frac{1.345}{ e } &  e  > k \end{cases}$
Tukey’s bisquare	$\rho_H(e) = \begin{cases} K^2/6[1 - [1 - (e/k)^2]^3] &  e  \leq k^* \\ K^2/6 &  e  > k \end{cases}$	$W_B(e) = \begin{cases} \left[1 - \left(\frac{e}{k}\right)^2\right]^2 &  e  \leq k^* \\ 0 &  e  > k \end{cases}$

\* The value k for the Huber and Tukey’s bisquare estimators is called a tuning constant

Smaller values of k produce more stability to outliers, but at the expense of lower efficiency when the errors are normally distributed. The tuning constant is generally picked in such a way as to give reasonably high efficiency in the normal case; specifically,  $k = 1.345\sigma$  for the Huber and  $k = 4.685\sigma$  for the Tukey’s bisquare (where  $\sigma$  is the standard deviation of the errors) produce 95-percent efficiency when the errors are distributed normally, and still offer protection against outliers (Fox & Weisberg, 2010).

In general, Tukey’s bisquare estimator is harsher than Huber and the observations in Tukey’s bisquare are usually down weighted more. We can see that the weight given to all observations (682) are significantly lower using Tukey’s bisquare weighting function (610) than Huber weighting function (642); however, the parameter estimates from these two different weighting methods may differ. When comparing the results of a regular OLS regression and a robust regression, if the results are very different, robust regression will most likely be used. Large differences suggest that the model parameters are being highly influenced by outliers.

Different functions have advantages and drawbacks. Huber weights can have difficulties with severe outliers, and Tukey’s bisquare weights can have difficulties in converging or may yield multiple solutions; however, the maximum number of iteration steps were increased to 50 to take care of the convergence issue. It can be seen from Table 6.2 that the results from the two

approaches are identical with the same coefficients sign and fairly close value. The main difference was for launching a new vehicle variable where the coefficients difference percentage is 13% (Huber coefficients value are considered as base). Moreover, Table 6.2 shows that outsourcing, platform strategy, and hourly/total variables are not statistically significant in either analysis; whereas, the others are statistically significant in both analyses.



**Table 6.2. Huber and Tukey's bisquare model comparisons**

	Huber result			Tukey's bisquare result			Coef. Diff. (%)
	Value	Std. Error	t value	Value	Std. Error	t value	
<b>(Intercept)</b>	1498.53	135.412	11.066	1342.108	135.008	9.941	0.10
<b>AWD</b>	0.046	0.007	6.483	0.044	0.007	6.187	0.04
<b>Variety.bodyandchassis</b>	0.615	0.088	6.993	0.517	0.088	5.893	0.16
<b>Flexibility</b>	-0.845	0.177	-4.77	-0.696	0.177	-3.944	0.18
<b>Model.types</b>	1.114	0.209	5.34	1.126	0.208	5.414	-0.01
<b>Year</b>	-0.705	0.068	-10.42	-0.631	0.067	-9.349	0.10
<b>ln(APV)</b>	-5.538	0.478	-11.577	-5.289	0.477	-11.089	0.04
<b>ln(APV)</b>	1.411	0.688	2.05	1.389	0.686	2.024	0.02
<b>Outsourcing</b>	-0.013	0.014	-0.937	-0.014	0.014	-0.993	-0.08
<b>Platform strategy</b>	0.425	0.684	0.622	0.825	0.682	1.21	-0.94
<b>Hourly/total</b>	-0.009	0.069	-0.129	0.016	0.069	0.229	2.78
<b>Newownership.Japanese</b>	-2.17	0.794	-2.733	-2.118	0.792	-2.676	0.02
<b>NewownershipJoint.venture</b>	-1.825	0.859	-2.124	-1.685	0.857	-1.967	0.08
<b>Vehicle.Launch</b>	3.034	0.434	6.989	2.58	0.433	5.961	0.15
<b>Segment.MD.B</b>	2.283	0.384	5.939	2.159	0.383	5.633	0.05
<b>Segment.MD.C</b>	8.015	0.779	10.295	8.198	0.776	10.561	-0.02
	Residual standard error: 4.1 on 666 degrees of freedom			Residual standard error: 3.96 on 666 degrees of freedom			

### 6.2.2 MM-Estimate

Some robust estimation methods use high breakdown points which is an efficient technique to define an outlier. The general idea of breakdown points is “the smallest proportion of the observations which can render an estimator meaningless” (Jensen et al., 2007). In other words, the breakdown point defines the amount of bad data points which may be available in our data-set. The good data are the majority of the data-set and the bad points are the minority data points.

Yohai (1987) developed MM estimator, multiple M-estimator, which is a special type of M-estimators. MM-estimation is a combination of efficient estimation, S estimation, and high breakdown value estimation. It was the first estimate with a high efficiency under normal error and high breakdown point (Alma, 2011). MM estimation procedure uses S estimation to estimate the regression parameter which minimizes the scale of the residual from M estimation and proceeds with M estimation accordingly (Susanti &Pratiwi, 2014).

MM estimation aims to provide estimates that have a high breakdown value and are more efficient. A common measure of the proportion of outliers is breakdown value that can be addressed before these observations affect the model (Susanti &Pratiwi, 2014). MM-estimator is the solution of

$$\sum_{i=1}^n \rho(u_i)X_{ij} = 0 \quad \text{or} \quad \sum_{i=1}^n \rho\left(\frac{Y_i - \sum_{j=0}^k X_{ij} \hat{\beta}_j}{S_{MM}}\right) X_{ij}$$

Where,

$S_{MM}$  = Standard deviation obtained from the residual of S estimation

$\rho$  = Tukey's biweight (bisquare) function

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^2} & -c \leq u_i \leq c \\ \frac{c^2}{6} & u_i < -c \text{ or } u_i > c \end{cases}$$

The MM estimator algorithm is as follows (Susanti &Pratiwi, 2014),

1. Estimate regression coefficients on the data using the OLS.
2. Test if the classical regression model assumptions are met.
3. Check the presence of outliers in the data.
4. Calculate residual value  $e_i = y_i - \hat{Y}_i$  of S estimate.
5. Calculate value of  $\hat{\sigma}_i = \hat{\sigma}_{sn}$ .

6. Calculate value  $u_i = \frac{e_i}{\hat{\sigma}_i}$

7. Calculate weighted value

$$W_i = \begin{cases} [1 - (\frac{u_i}{4.685})^2]^2, & |u_i| \leq 4.685 \\ 0 & |u_i| > 4.685 \end{cases}$$

8. Calculate  $\hat{\beta}_{MM}$  using WLS method with weighted  $W_i$ .

9. Repeat steps 5-8 to obtain a convergent value of  $\hat{\beta}_{MM}$ .

10. Test to determine whether independent variables have significant effect on the dependent variable.

The function `lmrob` in robust package in R is used for the MM-estimator in this study. This function computes a MM-type regression estimator as described in Yohai (1987) and Koller and Stahel (2011).

### 6.2.3 SMDM-Estimate

Koller and Stahel (2011) stated that there might be three main issues while using MM-estimates method as:

- The S-scale estimate might be biased
- Loss of efficiency of the estimated parameters
- The levels of tests may not be at the desired value

The scholars proposed two additional steps as a remedy and extend the standard MM-estimates. After MM-estimation step, the design-adaptive estimate will be calculated. Then, the regression parameters are reestimated and are used as initial estimate in MM-estimate. The argument setting of `lmrob.control` is provided to set alternative defaults as suggested in Koller and Stahel (2011) (using setting="KS2011", or its alternative setting="KS2014") (Koller &Stahel, 2011; Yohai, 1987).

It can be seen from Table 6.3 that the results from the two approaches are identical with the same coefficient signs and fairly close values. The main difference was for vehicle variety and flexibility variables that the coefficients difference percentage is 8% (MM coefficients value are base). Moreover, Table 6.3 depicts that CAC, outsourcing, platform strategy, and hourly/total variables are not statistically significant in either analysis; whereas, the others are statistically significant in both analyses.

**Table 6.3. MM and SMDM model comparisons**

	MM estimation				SMDM estimation				Coef. Diff. (%)
	Value	Std. Error	t value	Pr(> t )	Value	Std. Error	t value	Pr(> t )	
<b>(Intercept)</b>	1337.27	159.23	8.4	2.70E-16	1.41E+03	1.36E+02	10.31	< 2e-16	-0.05
<b>AWD</b>	0.0438	0.0081	5.4	9.20E-08	4.47E-02	7.18E-03	6.22	8.80E-10	-0.02
<b>Variety.bodyandchassis</b>	0.5141	0.1398	3.68	0.00025	5.57E-01	9.16E-02	6.08	2.00E-09	-0.08
<b>Flexibility</b>	-0.6922	0.2152	-3.22	0.00136	-7.50E-01	1.80E-01	-4.18	3.30E-05	-0.08
<b>Model.types</b>	1.1267	0.2136	5.28	1.80E-07	1.11E+00	2.08E-01	5.33	1.40E-07	0.01
<b>Year</b>	-0.6285	0.0788	-7.98	6.60E-15	-6.61E-01	6.81E-02	-9.71	< 2e-16	-0.05
<b>ln(APV)</b>	-5.2867	0.7925	-6.67	5.30E-11	-5.38E+00	4.94E-01	-10.89	< 2e-16	-0.02
<b>ln(APV)</b>	1.4007	0.9718	1.44	0.14996	1.37E+00	7.05E-01	1.95	0.0515	0.02
<b>Outsourcing</b>	-0.014	0.0131	-1.06	0.28729	-1.43E-02	1.42E-02	-1.01	0.313	-0.02
<b>Platform strategy</b>	0.8441	0.8445	1	0.31792	5.82E-01	6.84E-01	0.85	0.3948	0.31
<b>Hourly/total</b>	0.0149	0.0689	0.22	0.82849	1.03E-02	6.88E-02	0.15	0.8806	0.31
<b>Newownership.Japanese</b>	-2.1131	0.8206	-2.57	0.01024	-2.21E+00	7.93E-01	-2.79	0.0054	-0.05
<b>Newownership.Joint.venture</b>	-1.6789	0.7077	-2.37	0.01797	-1.78E+00	8.48E-01	-2.1	0.0364	-0.06
<b>Vehicle.Launch</b>	2.5685	0.5268	4.88	1.40E-06	2.76E+00	4.44E-01	6.2	9.70E-10	-0.07
<b>Segment.MD.B</b>	2.154	0.3613	5.96	4.00E-09	2.23E+00	3.81E-01	5.86	7.50E-09	-0.04
<b>Segment.MD.C</b>	8.1929	1.6693	4.91	1.20E-06	8.19E+00	8.04E-01	10.19	< 2e-16	0.00
	Residual standard error: 3.91 Multiple R-squared: 0.65 Adjusted R-squared: 0.643				Residual standard error: 4.35 Multiple R-squared: 0.65 Adjusted R-squared: 0.642				

#### 6.2.4 Least Trimmed Squares Robust (LTS) regression

Bounded influence methods effectively remove a large proportion of the cases and have a much higher breakdown point (as high as 50%). These methods can have trouble with small samples which is not our case (Jacoby, 2015).

One bounded-influence estimator is least-trimmed squares (LTS) regression. In this method, the squared residuals are ordered from smallest to largest:

$$(e^2)_{(1)}, (e^2)_{(2)}, \dots, (e^2)_{(n)}$$

The LTS estimator chooses the regression coefficients,  $\mathbf{b}$ , to minimize the sum of the smallest  $m$  of the squared residuals,

$$\text{LTS}(\mathbf{b}) = \text{Min} \left( \sum_{i=1}^m (e^2)_{(i)} \right)$$

Where,

$$m = [n/2] + [(p + 2)/2] \text{ (the brackets, } [ \ ], \text{ denote rounding down to the next smallest integer)}$$

$M$  is a little more than half of the observations. By using only 50% of the data that fits closest to the original OLS line, LTS completely ignores extreme outliers (the observations had the largest errors regardless if it's positive or negative). However, this method can misrepresent the trend in the data if it is characterized by different clusters of extreme cases or if the data set is relatively small. The mechanics of fitting the LTS estimator are a bit complicated while the LTS criterion is easily described. Moreover, bounded-influence estimators can produce unreasonable results in certain circumstances (Fox & Weisberg, 2010; Jacoby, 2015).

The result of LTS is shown in Table 6.4. Table 6.4 shows that CAC, outsourcing, and hourly/total variables are not statistically significant; whereas, the others are statistically significant in both analyses. However, Segment.MD.C was not significant at the level of 5% also.

**Table 6.4. Least-trimmed squares coefficients**

	Estimate	Std. Error	t value	Pr(> t )	
<b>(Intercept)</b>	1.07E+03	1.16E+02	9.25	< 2e-16	***
<b>AWD</b>	3.74E-02	6.18E-03	6.05	2.50E-09	***
<b>Variety.bodyandchassis</b>	5.15E-01	7.88E-02	6.54	1.30E-10	***
<b>Flexibility</b>	-6.49E-01	1.52E-01	-4.26	2.40E-05	***
<b>Model.types</b>	1.49E+00	1.77E-01	8.46	< 2e-16	***
<b>Year</b>	-5.00E-01	5.78E-02	-8.66	< 2e-16	***
<b>ln(APV)</b>	-4.15E+00	4.33E-01	-9.57	< 2e-16	***
<b>ln(APV)</b>	5.21E-01	6.01E-01	0.87	0.386	
<b>Outsourcing</b>	-1.28E-02	1.20E-02	-1.06	0.289	
<b>Platform strategy</b>	2.44E+00	5.91E-01	4.12	4.30E-05	***
<b>Hourly/total</b>	1.89E-03	5.87E-02	0.03	0.974	
<b>Newownership.Japanese</b>	1.81E+00	3.78E-01	4.79	2.10E-06	***
<b>Newownership.Joint.venture</b>	2.41E+00	3.20E-01	7.52	1.90E-13	***
<b>Vehicle.Launch</b>	4.90E+00	7.22E-01	6.79	2.70E-11	***
<b>Segment.MD.B</b>	-1.59E+00	6.66E-01	-2.39	0.017	*
<b>Segment.MD.C</b>	-9.74E-01	7.12E-01	-1.37	0.172	

Signif. codes: '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
Residual standard error: 3.58 on 615 degrees of freedom  
Multiple R-Squared: 0.612, Adjusted R-squared: 0.603  
F-statistic: 64.8 on 15 and 615 DF, p-value: <2e-16  
Note, for this calculation Alpha= 0.5

### 6.2.5 Quantile regression

The Ordinary Least Squares (OLS) models represent the relationship between one or more independent variables  $x$  and the conditional mean of a dependent variable  $y$ . However, quantile regression models the relationship between  $x$  and the conditional quantiles of  $y$  rather than just the conditional mean of  $y$ . Quantile regression is used to describe the distribution of the dependent variable at certain quantile.

A quantile regression gives a more comprehensive picture of the effect of the independent variables on the dependent variable. There are specific conditions to use quantile regression like continuity of dependent variable with no zeros or too many repeated values.

Quantile regression can be described by the following equation (Katchova, 2013):

$$Y_i = X_i^T \beta_q + e_i$$

Where,

$\beta_q$  = The vector of unknown parameters associated with the  $q^{\text{th}}$  quantile.

The OLS minimize  $\sum_{i=1}^n e_i^2$ , the sum of squares of the model prediction error  $e_i$ ; however, the quantile regression minimizes  $\sum_{i=1}^n q|e_i| + \sum_{i=1}^n (1 - q)|e_i|$ , a sum that gives the asymmetric penalties  $\sum_{i=1}^n q|e_i|$  for underprediction and  $\sum_{i=1}^n (1 - q)|e_i|$  for overprediction (where  $0 < q < 1$ ).

The  $q^{\text{th}}$  quantile regression estimator  $\widehat{\beta}_q$  minimizes the objective function over  $\beta_q$ . In contrast to OLS and maximum likelihood, the quantile regression computational implementation uses linear programming methods.  $B_q$  was used instead of  $\beta$  to make clear that different choices of  $q$  estimate different values of  $\beta$  (Katchova, 2013).

All variables coefficients VS for different quantiles, from 0 to 1.0, are shown in Figure 6.1. Figure 6.1 indicates that almost all the variables are needed and have the similar coefficients for various quantiles, other than year. In addition to coefficients, the prediction and confidence intervals for different coefficients for various quantiles are also illustrated in Figure 6.1.

### 6.2.6 Principal component regression (PCR)

PCR is a traditional multivariate method which its purpose is to estimate the response variable at the basis of some selected principle components (PCs) of the predictors (Filzmoser, 2001). The PCR mainly used for two reasons as removing the potential multicollinearity and reducing the dimensionality which both could be the interest of this research. In this research, principal component analysis (PCA) was used to find some linear combinations of the explanatory variables, without considering the categorical variables such ownership and car segment, that can be used to summarize the data without losing too much information in the process (Maitra & Yan, 2008).

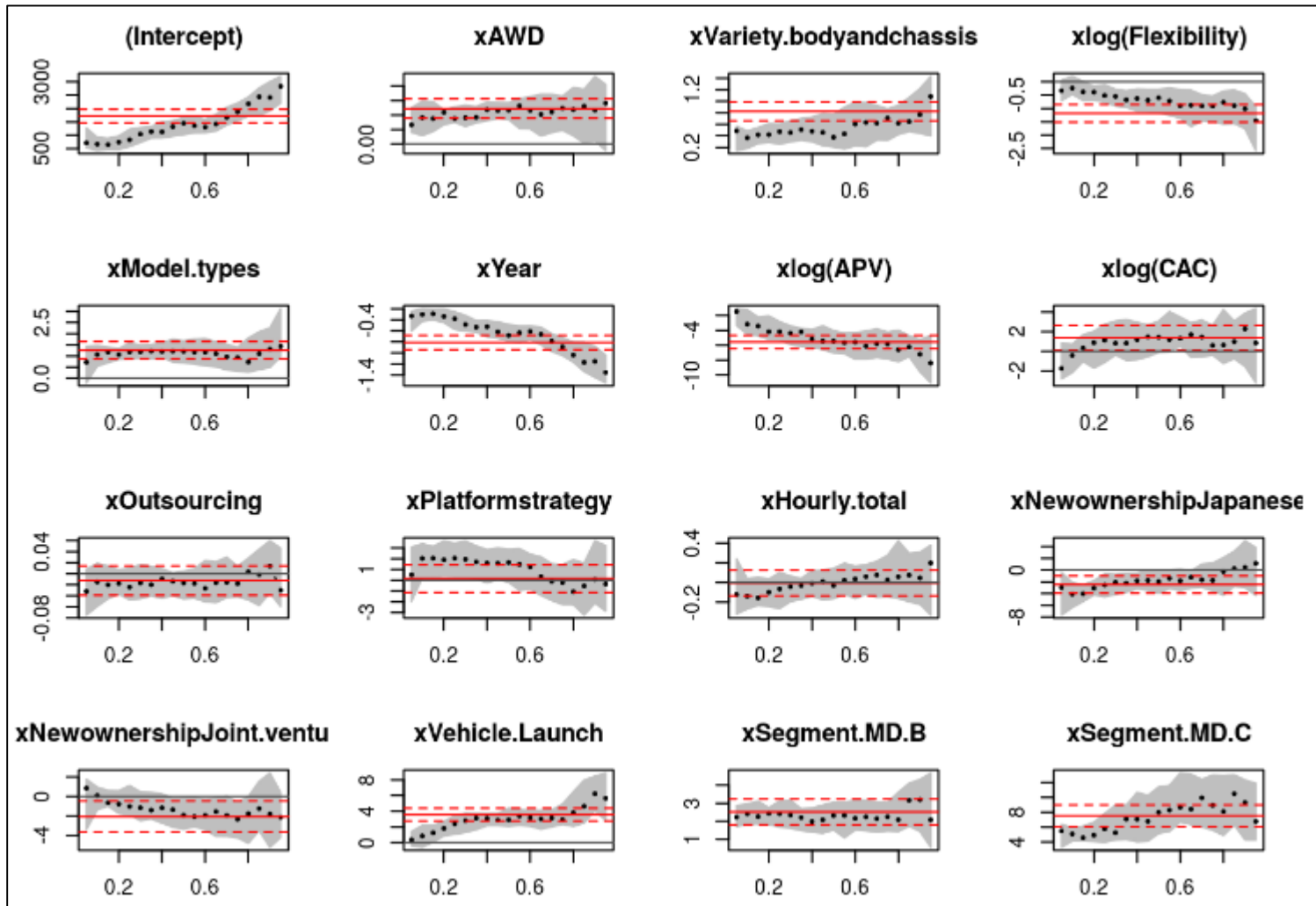


Figure 6.1. Coefficients of all variables VS different quantiles



### 6.3. Shrinkage Methods

The literature shows that shrinking the coefficient estimates towards zero may significantly reduce the variance. The three best-known methods to shrink the regression coefficients estimates towards zero are ridge, lasso, and elastic net regressions.

#### 6.3.1 Ridge Regression

The least squares fitting procedure estimates  $\beta_0, \beta_1, \dots, \beta_p$  utilizing the values that minimizes

$$RSS = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$$

Ridge regression is very similar to least squares, other than the fact that the coefficients are estimated using a different equation. The ridge regression coefficient estimates,  $\widehat{\beta}_\lambda^R$ , are the values that minimize (James et al., 2013).

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 = RSS + \lambda \sum_{j=1}^p \beta_j^2 \quad (6.1)$$

Where,

$$\lambda = \text{Tuning parameter and } \lambda \geq 0$$

Ridge regression is a tradeoff between minimizing the RSS (the coefficients estimate fits the data well) and the second term,  $\lambda \sum_{j=1}^p \beta_j^2$ , which is called the ridge shrinkage penalty. The shrinkage penalty will be small when  $\beta_1, \beta_2, \dots, \beta_p$  are close to zero; hence, it has the effect of shrinking the coefficient  $\beta_j$  towards zero. Since the estimated coefficients which are associated with each regressor variables need to be shrunk, the shrinkage penalty is applied to  $\beta_1, \beta_2, \dots, \beta_p$  not to the intercept  $\beta_0$ . The estimated intercept will be  $\widehat{\beta}_0 = \bar{Y} = \sum_{i=1}^n y_i / n$  when all the variables have been centered to have mean zero.

In statistical parlance, the ridge uses  $\iota_2$  penalty instead of  $\iota_1$  penalty. The  $\iota_2$  norm of a coefficient vector  $\beta$  is given by  $\|\beta\|_2 = \sqrt{\sum_{j=1}^p \beta_j^2}$ .

Ridge regression is applied after standardizing the predictors, using the following formula,

$$\widetilde{x}_{ij} = \frac{x_{ij}}{\sqrt{1/n \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}} \quad (6.2)$$

So all variables are all at the same scale. In (6.2), the denominator is the estimated standard deviation of the  $j$ th predictor. Hence, all of the standardized variables will have a standard

deviation of one. As a result, the final fit will not depend on the scale on which the predictors are measured. However, the ridge regression can return the unstandardized coefficients which I did that for my research.

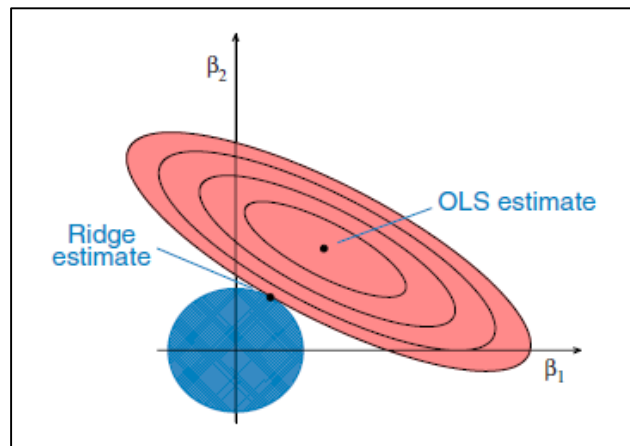
The turning parameter,  $\lambda$ , serves to control the relative impact of these two terms,  $RSS$  and  $\lambda \sum_{j=1}^p \beta_j^2$ , on the estimated coefficients. When  $\lambda=0$ , the penalty has no effect and the ridge regression produces exactly the same result as least squared regression does. However, as  $\lambda \rightarrow \infty$ , the impact of the shrinkage penalty will increase, and the ridge regression coefficients estimate must approach to zero to remedy this problem and ridge regression will give the null model in which all coefficients estimate other than  $\widehat{\beta}_0$  are equal to zero.

On the other hand, equation (6.1) is equivalent to the minimization of  $\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$

Subject to:

$$\sum_{j=1}^p \beta_j^2 \leq c \quad \text{when } c > 0$$

The geometric interpretation of ridge regression is shown in Figure 6.2.

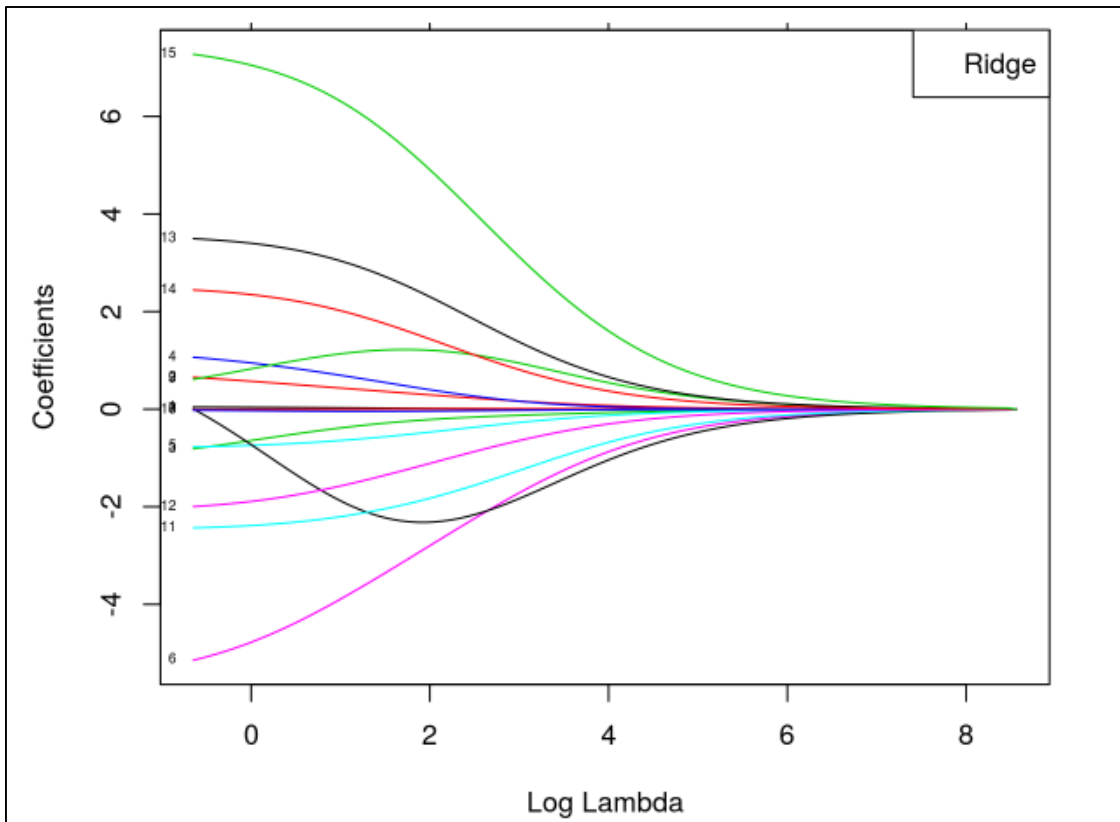


**Figure 6.2. Geometric interpretation of ridge regression (adapted from(PSU, 2016))**

The ellipses correspond to the contours of the residual sum of squares (RSS); the RSS is minimized at the ordinary least squares (OLS) and the inner ellipse has smaller RSS. For  $p=2$ , the constraints on ridge regression,  $\sum_{j=1}^p \beta_j^2 = \beta_1^2 + \beta_2^2 < c$ , corresponds to a circle. In the ridge regression, we are trying to simultaneously minimize the ellipse size and the circle. The point that the ellipse and

circle touch is the ridge estimate. Unlike least squares, which generates only one set of coefficient estimates, ridge regression will produce different sets of coefficients estimate,  $\hat{\beta}_\lambda^R = (X^T X + \lambda I)^{-1} X^T y$ , depending on the value of  $\lambda$ . The criterion for choosing  $\lambda$  can be obtained through cross-validation, by minimizing the cross-validated mean squared prediction error.

The ridge regression coefficients estimate can be plotted by calling glmnet package with alpha=0 in R software. The ridge regression coefficients estimate for the current data-set are displayed in Figure 6.3. Each curve corresponds to a variable for ridge coefficients estimate for our data-set. It shows the path of its coefficient against the  $\ell_2$ -norm of the whole coefficient vector as  $\lambda$  varies. Note that the x-axis is Log of  $\lambda$  instead of  $\lambda$ .



**Figure 6.3. The unstandardized ridge regression coefficients**

Figure 6.3 indicates that if  $\lambda$  is zero, the coefficients estimate are ordinary least squares estimates but as  $\lambda$  increases (moving to the right), the coefficients estimate will shrink and approach towards zero. For example, the pink solid line (#6) represents the ridge estimate for the  $\ln(\text{APV})$ , as  $\lambda$  is varied.

### 6.3.2 Lasso (least absolute shrinkage and selection operator) regression

Ridge regression has an obvious disadvantage, the ridge penalty term  $\lambda \sum_{j=1}^p \beta_j^2$  will shrink all the coefficients towards zero, but it doesn't force any coefficients to zero unless  $\lambda \rightarrow \infty$ . Increasing the value of  $\lambda$  will decrease the coefficients estimate magnitude but will not result in the exclusion of any variable. This might be a serious problem when the number of regressors are really high, which isn't the case for our study. The lasso can be a remedy for this disadvantage, the lasso coefficients,  $\hat{\beta}_\lambda^L$ , minimize the quantity

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| = RSS + \lambda \sum_{j=1}^p |\beta_j| \quad (6.3)$$

Comparing (6.1) to (6.3) indicates that ridge and lasso regressions have similar formulations, the only difference is the penalty terms, which in ridge regression is  $\lambda \sum_{j=1}^p \beta_j^2$  and in lasso regression is  $\lambda \sum_{j=1}^p |\beta_j|$ . In statistical parlance, the lasso uses  $\ell_1$  penalty instead of  $\ell_2$  penalty. The  $\ell_1$  norm of a coefficient vector  $\beta$  is given by  $\|\beta\|_1 = \sum |\beta_j|$ .

Similar to ridge regression, lasso regression shrinks the coefficients estimate toward zero when tuning parameter,  $\lambda$ , increases. However, in lasso the coefficients estimate potentially could be zero when the  $\lambda$  is large enough. Hence, similar to best subset selections and stepwise regression lasso can be used for variable selection. The lasso yields sparse model, models that involve subset of variables in which some variables could be zero.

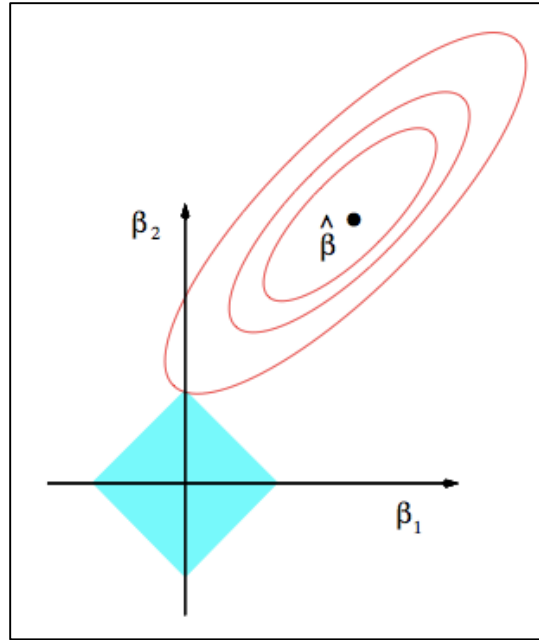
When  $\lambda=0$ , the lasso penalty has no effect and the lasso regression produces exactly the same result as least regression does. However, as  $\lambda \rightarrow \infty$ , the impact of the shrinkage penalty will increase, and the lasso regression coefficients estimate must approach zero. To tackle this situation, lasso regression gives the null model in which all coefficients estimate other than  $\widehat{\beta}_0$  are equal to zero.

Equation (6.3) is equivalent to minimization of  $\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$

Subject to:

$$\sum_{j=1}^p |\beta_j| \leq c \quad \text{when } c > 0$$

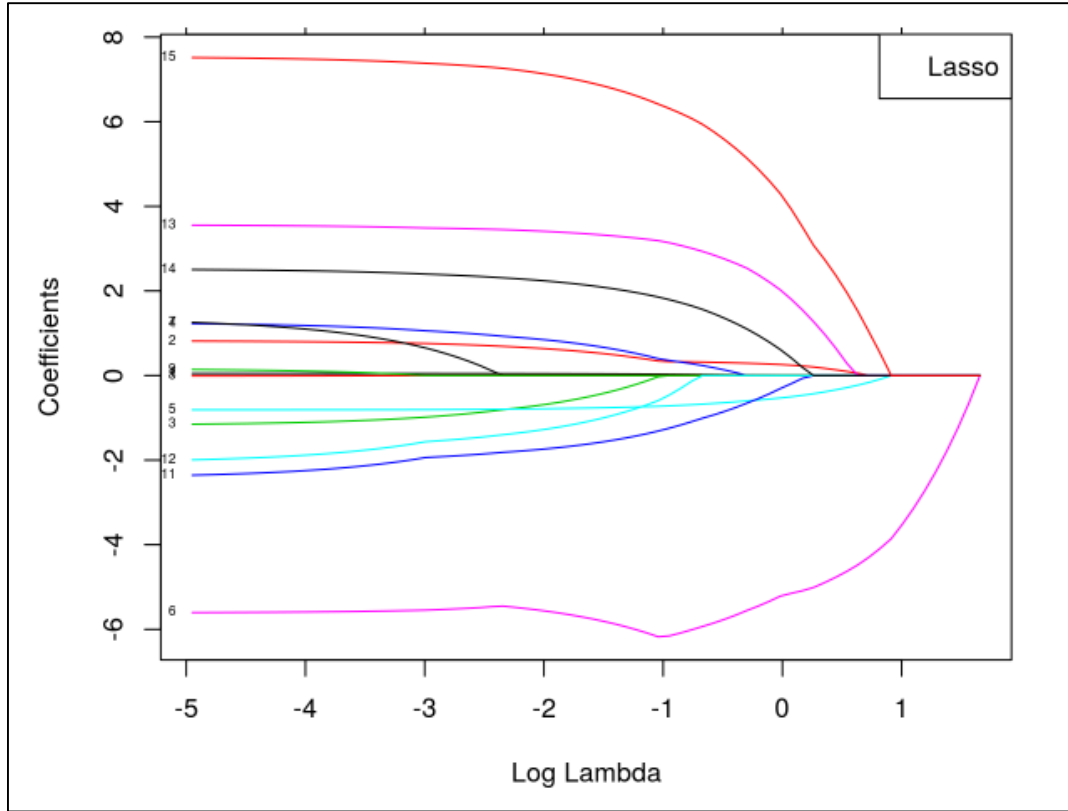
The geometric interpretation of lasso regression is shown in Figure 6.4.



**Figure 6.4. Geometric interpretation of lasso regression (adapted from (James et al., 2013))**

The ellipses correspond to the contours of the residual sum of squares (RSS); the RSS is minimized at the ordinary least squares (OLS), and the inner ellipse has smaller RSS. For  $p=2$ , the constraints on lasso regression,  $\sum_{j=1}^p |\beta_j| = |\beta_1| + |\beta_2| < c$ , corresponds to a diamond. In the lasso regression, we are trying to simultaneously minimize the ellipse size and the diamond. The point that the ellipse and diamond touch is the lasso estimate. Unlike ridge regression, the lasso constraint has corners at each of the axes, and the ellipse will often intersect the constraint region at an axis. When it happens, one of the coefficients will equal zero. Unlike least squares, which generate only one set of coefficients estimate, lasso regression will produce different sets of coefficients estimate,  $\hat{\beta}_\lambda^L$ , depending on the value of tuning parameter,  $\lambda$ . The criterion for choosing the  $\lambda$  can be defined through cross-validation, by minimizing the cross-validated mean squared prediction error.

The lasso regression coefficients estimate can be obtained by calling `glmnet` package with `alpha=1` in R software. The lasso regression coefficients estimate for our data-set are illustrated in Figure 6.5. Each curve corresponds to a variable for lasso coefficients estimate for our data-set. It shows the path of its coefficient against the  $\ell_1$ -norm of the whole coefficient vector as  $\lambda$  varies. Note that the x-axis is  $\text{Log of } \lambda$  instead of the  $\lambda$ .



**Figure 6.5. The unstandardized lasso regression coefficients**

Figure 6.5 indicates that if  $\lambda$  is zero, the coefficients estimate are ordinary least squared estimates but as  $\lambda$  increases (moving to the right), the coefficients estimate will shrink and approach zero. For example, the pink solid line (#6) represents the lasso estimate for the  $\ln(\text{APV})$ , as  $\lambda$  is varied.

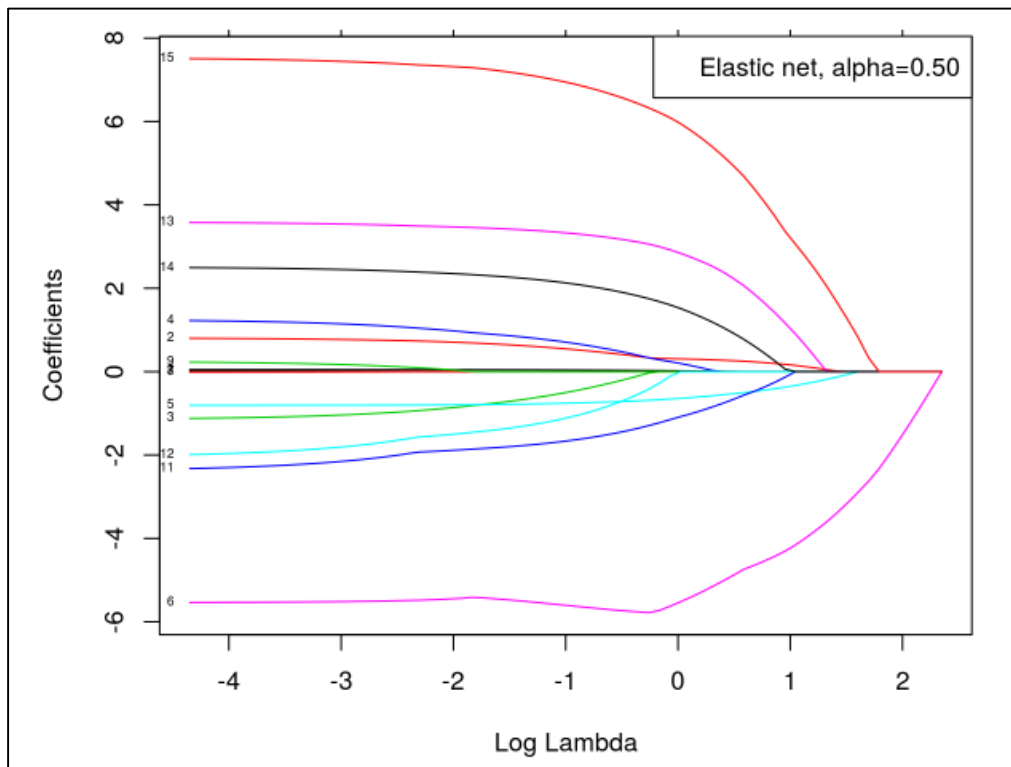
### 6.3.3 Elastic net regression

The ridge penalty shrinks the coefficients estimate of correlated predictors towards each other; however, lasso penalty tends to pick one of the coefficients estimate and discard the other variables. The elastic net penalty is the combination of ridge and lasso. It takes care of the stated disadvantages for both ridge and lasso (Zou & Hastie, 2005). The elastic-net regression is shown in equation 6.4.

$$\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p ((1 - \alpha)\beta_j^2 + \alpha|\beta_j|)$$

$$= \text{RSS} + \lambda \sum_{j=1}^p ((1 - \alpha)\beta_j^2 + \alpha|\beta_j|) \quad (6.4)$$

The term  $(1 - \alpha)\beta_j^2$  forces the correlated coefficients estimate to be averaged, while the term  $\alpha|\beta_j|$  encourages a sparse solution in the coefficients of the averaged features (Hastie et al., 2011). Elastic net regression is a hybrid approach that mixes both penalization of the  $t_1$  and  $t_2$  norms. The  $\alpha$ , hyper-parameter, varies between 0 and 1 and controls the  $t_1$  and  $t_2$  penalization (0 is ridge and 1 is lasso). The criterion for choosing  $\lambda$  is defined through cross-validation, by minimizing the cross-validated mean squared prediction error. However, in the elastic-net the tuning parameter,  $\lambda$ , depends on hyper-parameter,  $\alpha$ , which adds to the complexity of calculations. The elastic-net regression coefficients estimate can be obtained by calling glmnet package and defining the alpha value, alpha=0.5 for instance, in R software. As an example, the coefficients estimate for  $\alpha=0.5$  is shown in Figure 6.6. Each curve is elastic-net coefficients estimate for the study data-set. Note that the x-axis is Log of  $\lambda$  instead of the  $\lambda$ .



**Figure 6.6. The unstandardized elastic-net regression coefficients**

Figure 6.6 indicates that if  $\lambda$  is zero, the coefficients estimate are ordinary least squared estimates but as  $\lambda$  increases (moving to the right), the coefficients estimate will shrink and approach towards zero. For example, the pink solid line (#6) represents the lasso estimate for the  $\ln(\text{APV})$ , as  $\lambda$  is varied.

### 6.3.4 Selecting the Tuning Parameter

As already mentioned, the common criterion for choosing the tuning parameter,  $\lambda$ , in ridge, lasso and, elastic-net is using cross-validation, by minimizing the cross-validated mean squared prediction error. A grid of  $\lambda$  values are chosen and the cross validation error for each specific  $\lambda$  is calculated. Then,  $\lambda$  will be chosen in such a way that the cross-validation error is the smallest. Finally, the model refits the obtained tuning parameter which has the smallest value.

The function `cv.glmnet` from the R package `glmnet` does automatic cross-validation on a grid of  $\lambda$  values using for  $\ell_1$  and  $\ell_2$ -penalized depending on the regression methods, ridge, lasso, and elastic net. Cross-validation is an estimate of the expected generalization error for each  $\lambda$  which  $\lambda$  is sensibly chosen as the minimizer of this estimate. The `cv.glmnet` function returns two values of  $\lambda$ . The minimizer, `lambda.min`, and `lambda.1se`, which is a heuristic choice of  $\lambda$  producing a less complex model, for which the performance in terms of estimated expected generalization error is within one standard error of the minimum.

Figure 6.7 shows the mean squared error of the cross-validation VS different values of  $\lambda$  for the ridge and lasso regressions. The dashed vertical lines indicate the selected minimum value of  $\lambda$  and one standard error of the minimum value of  $\lambda$ .

### 6.3.5 Shrinkage Methods Assumptions

All the regularized regression techniques such as ridge, lasso, and elastic net regressions make a biased coefficients estimate and hoping to reduce the expected loss by exploiting the bias-variance trade off. The optimal value of the coefficients estimate will be selected by defining the tuning parameter through a cross-validation procedure.

Hoerl & Kennard (1970) proved that there always exists a value of tuning parameter,  $\lambda$ , that ridge/lasso regression coefficients estimate will lead to a smaller expected loss than ordinary least squares, OLS. This result usually doesn't hold any assumptions and is always true (Hoerl & Kennard, 1970). Hence, there is no need to do any further analysis on the shrinkage methods output and the coefficients could be considered as a proper estimates of the real coefficients.



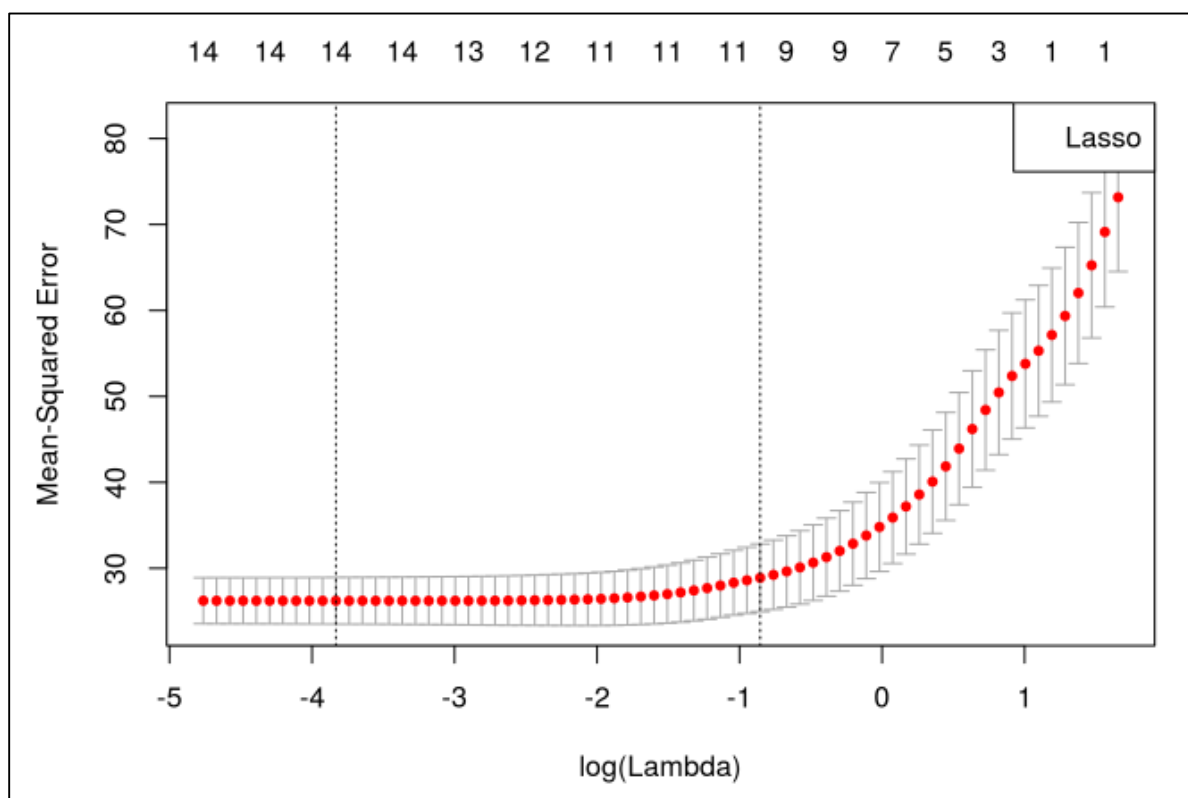
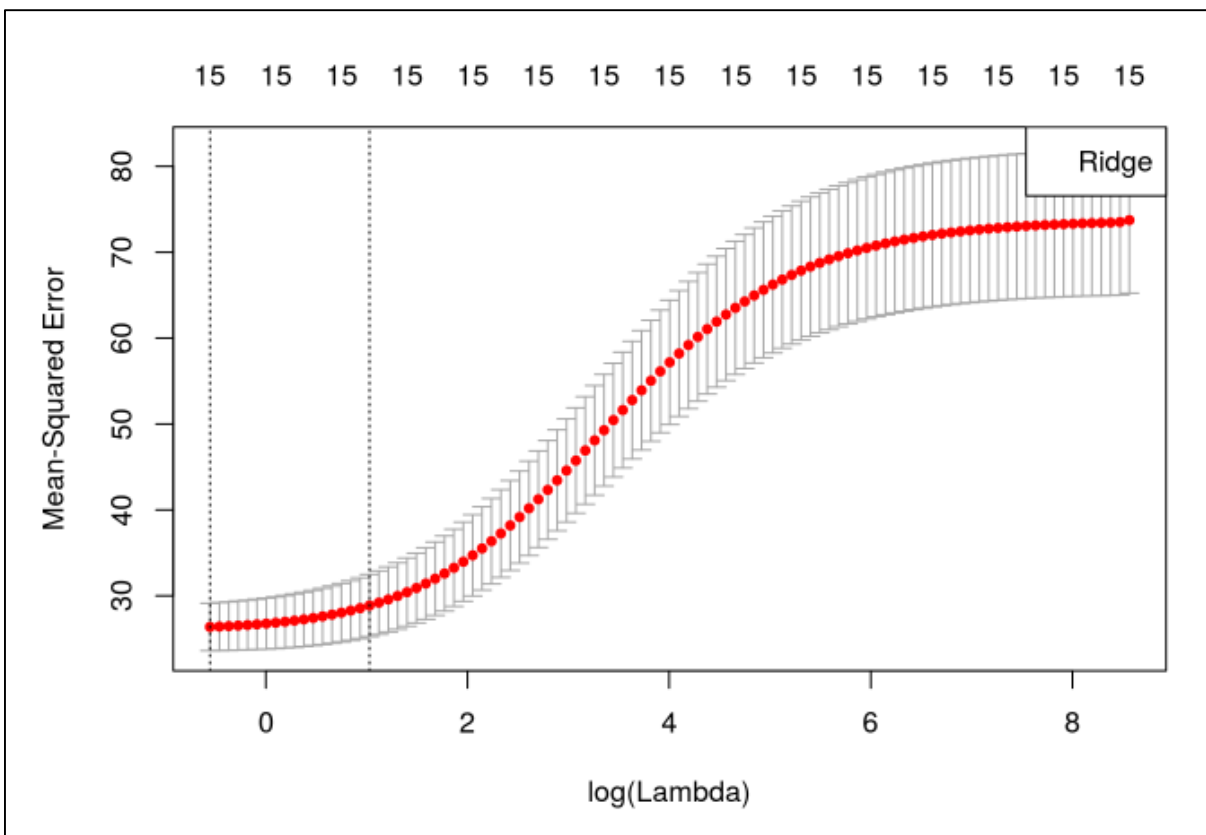


Figure 6.7. Mean-Squared error of ridge and lasso regressions for different values of  $\lambda$

## 6.4. Hybrid Methods

It was stated that there are some outliers in the data-set; hence, robust methods should be used to treat outliers and down-weight severe outliers. Moreover, shrinkage methods can be used to shrink the coefficients estimate and take care of any possible collinearity. Therefore, it was decided to use a combination of the robust and shrinkage methods. Consequently, each observation's weight (all plants) could be calculated and a matrix of weights formed in which the rows are the plants and the columns are different robust algorithms. Then the vector of weights are used in the shrinkage methods to reduce any possible collinearity. Utilizing the weights of each plant and considering the shrinkage methods, the equations (6.1), (6.3), and (6.4) may be rewritten as,

Ridge regression:

$$\min_{\beta} (Y - X\beta)^T W_{ij} (Y - X\beta) + \lambda \sum_{j=1}^p \beta_j^2 \quad (6.5)$$

Where,

- i= Observations weight, a nonnegative vector of length n, where n is the number of plants
- j= The weights can be calculated from the M-estimators (Huber and Tukey's Bisquare), MM-estimator, SMDM-estimator, and LTS regression weight

In short the  $W_{ij}$  matrix can be shown as,

$$\begin{array}{ccc} \text{Huber} & \dots & \text{LTS} \\ \left[ \begin{array}{ccc} W_{1,1} & \dots & W_{1,5} \\ \vdots & \ddots & \vdots \\ W_{682,1} & \dots & W_{682,5} \end{array} \right] \end{array}$$

Hence, the lasso and elastic net regressions, respectively, could be rewritten as:

$$\min_{\beta} (Y - X\beta)^T W_{ij} (Y - X\beta) + \lambda \sum_{j=1}^p |\beta_j| \quad (6.6)$$

$$\min_{\beta} (Y - X\beta)^T W_{ij} (Y - X\beta) + \lambda \sum_{j=1}^p ((1 - \alpha)\beta_j^2 + \alpha|\beta_j|) \quad (6.7)$$

The summary of all model estimations including the initial model (multiple linear regression without predictors transformation) and enhanced models: all robust regressions, M estimators, MM estimator, LTS regression, SMDM estimator and then ridge, lasso, and elastic net regressions, and the combination of both robust and shrinkage methods are shown in Table 6.5.

**Table 6.5. Comparison of the resulted different statistical models**

		Estimator	Adjusted R-squared	Residual standard error (RSD)	Weights sum	MSE based on 10 folds Cross Validation	Significant variables
Multiple linear regression	Multiple linear regression (without transformation)		0.578	5.56	668	5.71	All variables other than outsourcing and platform strategy
	Multiple linear regression (with transformation)		0.659	5	669	5.13	All variables other than outsourcing, platform strategy, and hourly employees' percentage
Robust regression	M- estimator (Huber)		0.651	4.10	642	5.21	All variables other than outsourcing, platform strategy, and hourly employees' percentage
	M- estimator (Tukey's bisquare)		0.636	3.96	606	5.26	All variables other than outsourcing, platform strategy, and hourly employees' percentage
	MM-estimator		0.643	3.91	604	5.27	All variables other than ln (CAC), outsourcing, platform strategy, and hourly employees' percentage
	SMDM-estimator		0.642	4.35	646	5.23	All variables other than outsourcing, platform strategy, and hourly employees' percentage
	LTS regression		0.603	3.58	615	-	All variables other than outsourcing, platform strategy, and hourly employees' percentage
	PCR		0.658	5.01	666	5.15	All new rotated variables, car segments, and ownership
Shrinkage Methods	Ridge regression		0.661	-	-	5.15	All variables
	Lasso regression		0.665	-	-	5.13	All variables other than hourly employees' percentage
	Elastic-net regression with $\alpha=0.5$		0.655	-	-	5.13	
Robust and Shrinkage Methods	Lasso with Huber weights		0.689	-	-	4.20	
	Lasso with Tukey's bisquare weights		0.692	-	-	3.83	
	Lasso with MM-estimator		0.692	-	-	3.81	All variables

Estimator	Adjusted R-squared	Residual standard error (RSD)	Weights sum	MSE based on 10 folds Cross Validation	Significant variables
Lasso with SMDM estimator	0.682	-	-	4.11	
Lasso with LTS weights	0.681	-	-	4.23	
Elastic with MM weight-alpha=0.25	0.692	-	-	3.82	
Elastic with MM weight-alpha=0.5	0.692	-	-	3.82	
Elastic with MM weight-alpha=0.75	0.692	-	-	3.82	

Here are the findings from Table 6.5.

- Robust regressions have lower residual standard error compared to multiple linear regression.
- Shrinkage methods have lower cross validation errors compared to multiple linear and robust regressions.
- The combination of robust and shrinkage methods has the lowest cross validation errors and highest adjusted R-squared when comparing them with the other methods.
- The best results were achieved by the combination of the lasso and MM-estimator.

In conclusion, the research objective 4, RO4, has been addressed and the best statistical model was achieved by the combination of the MM-estimator and lasso.

**6.5. Variables Interaction**

Here is the approach that was taken to consider the interaction of the variables in this research:

1. First of all, the base model that was obtained in section 5.2 (with log normal of APV and CAC) is chosen. Only the two-way interaction will be considered for ease of modeling and interpretation. There are  $\binom{13}{2} = 78$  two-way interactions which is a very huge number and a method must be used to reduce the number of interactions.

2. A stepwise regression with AIC was used in which the lower scope was of all the main effects, and the upper scope was of the main effects plus all priory two-way interactions. A model with 10 significant two-way interactions was resulted.
3. A model with 10 obtained two-way interactions will be used with lasso and MM estimator weights (based on results obtained from Table 6.5 which has the best performance). It was assumed that the new model has the best statistical performance among all developed statistical models with two-way interactions. However, a model with 10 two-way interactions is not easy to work with, and it was decided to reduce the number of two-way interactions.
4. Six two-way interactions were removed from the statistical model because either they were not significant or had a very small impact on the model.
5. The final statistical model has four two-way interactions of variables.

The statistical models with the two-way interactions are illustrated in Table 6.6.

**Table 6.6. Different statistical models comparison**

	Estimator	Adjusted R-squared	Residual standard error (RSD)	Weights sum	MSE based on 10 folds Cross Validation	Significant variables
<b>Robust and shrinkage methods</b>	Lasso with MM-estimator (10 two- way interaction)	0.714	-	-	3.74	All variables are significant
	Lasso with MM-estimator (4 two- way interaction)	0.711	-	-	3.73	All variables are significant

Table 6.6 indicates that by removing six two way interactions, our model performance (adjusted R-squared and cross validation error) did not deteriorate by a significant amount; hence, a parsimonious model has been obtained.

The final statistical model developed for different number of folds such as 5, 10, and leave one out cross validation is illustrated in Table 6.7. However, as already stated, the NOF with 10 will be used for the purpose of this research from now on.

**Table 6.7. The Final Model Coefficients**

Variables	NOF =5	NOF=10	NOF=682
<b>(Intercept)</b>	1,323.20	1320	1320
<b>AWD</b>	0.0393	0.0385	0.0384
<b>Varietybodyandchassis</b>	0.5111	0.513	0.513
<b>Flexibility</b>	-0.5845	-0.581	-0.58
<b>Model.types</b>	1.1348	1.15	1.15
<b>Year</b>	-0.6292	-0.63	-0.63
<b>ln(APV)</b>	-4.0332	-3.83	-3.79
<b>ln(CAC)</b>	1.0992	1.13	1.14
<b>Outsourcing</b>	-0.0089	-0.0087	-0.0087
<b>Platformstrategy</b>	1.01	1.04	1.05
<b>Hourly.total</b>	0.0138	0.0061	0.0043
<b>NewownershipJapanese</b>	-1.8059	-1.83	-1.84
<b>NewownershipJoint.venture</b>	-1.2779	-1.28	-1.28
<b>Vehicle.Launch</b>	-8.9346	-12.90	-13.8
<b>Segment.MD.B<sup>2</sup></b>	15.2131	18.70	19.5
<b>Segment.MD.C<sup>3</sup></b>	43.8966	47.30	48.1
<b>ln(APV):Segment.MD.B</b>	-1.0744	-1.36	-1.43
<b>ln(APV):Segment.MD.C</b>	-3.1162	-3.40	-3.47
<b>Platformstrategy:Vehicle.Launch</b>	-0.5927	-0.578	-0.575
<b>Flexibility:Vehicle.Launch</b>	-0.3882	-0.399	-0.402
<b>Hourly.total:Vehicle.Launch</b>	0.2146	0.260	0.271
Adjusted R-squared	0.711	0.711	0.711
Cross Validation	3.71	3.73	3.73

Tables 6.7 indicates that two factors directly reduce HPV and four factors increase it. Variables that directly improve productivity include APV and year of production. Year of production does not improve the HPV per se; rather, it is the other factors occurring in the same time period which result in improvements. However, the percentage of hourly employees, launching a new vehicle, flexibility and platform strategies are impacted by their two-way interactions. Since the two-way

<sup>2</sup> The car segment was defined from the outcome variable (HPV); however, it explained 5% to 8% of the adjusted R-squared, depending the statistical methods. Hence, it was decided to consider it as an independent variable in the statistical model.

<sup>3</sup> A statistical model without car segment has been developed also if someone doesn't want to use car segment as an exploratory variable (AIII, Appendix). Table AIII indicates that the adjusted R-squared and cross validation error were improved from 4.18 and 0.63 to 3.73 and .711, respectively.

interactions for some of the variables are significant, the finding for those variables aren't straightforward. For instance, besides vehicle launch, there are three significant two-way interactions (see Table 6.7) which must be taken into account while interpreting vehicle launch. Table 6.7 indicates that while launching a new vehicle common platform sharing, flexible manufacturing, and more salaried employees (lower hourly) strategies will help carmakers to overcome the side effect of launching new vehicles productivity penalization to some extent. For the segment A, while launching a new product the HPV is penalized by 2.63 hours and 2.87 hours for the mean and median of flexibility, Hourly.total, and platform strategy, respectively (Table 6.8).

**Table 6.8. The mean and median for the flexibility, hourly.total, and platformstrategy**

Segment A		
	Mean	Median
Flexibility	20	20.1
Hourly.total	90.3	91.8
Platformstrategy	0.61	0.5
Calculated - vehicle lunch - Mean/Median	2.63	2.87

Hence the developed HPV equation can be calculated as,

$$\begin{aligned}
 Exp(HPV) = & (1,320 + 0.0385AWD + 0.513Vehicle\ Variety - 0.581flexibility + 1.15Model - 0.63Year - 3.83\ln(APV) + 1.13\ln(CAC) \\
 & - 0.0087Outsourcing + 1.04platform.strategy + 0.0061Hourly.total - 12.90Vehicle.launch - 1.83Japanese - 1.28Joint.venture \\
 & + 18.7Segment.B + 47.30Segment.C - 1.36\log(APV):Segment.B - 3.40\log(APV):Segment.C - 0.578Platformstrategy:Vehicle.Launch \\
 & - 0.399Flexibility:Vehicle.Launch + 0.26Hourly.total:Vehicle.Launch)
 \end{aligned}
 \tag{6.8}$$

## 6.6. Analyzing the Relevant Outperforming Residuals

Research objective 6 (RO6) is reviewing the outperforming residuals in the final statistical model to determine the existence of specific pursuant strategies that they adhere to. To address this research objective, all the plants that their residuals are negative (outperforming residuals), based on the hybrid statistical model with two way interactions, have been extracted from the entire sample of plants. In general, there were 342 plants that outperformed the statistical model which are almost 50% of our data, as expected. To address RO6, the mean and median of the outperforming and not outperforming residuals were calculated and percentage changes for each variable computed and compared to determine any specific pursuant strategies. The results are shown in Tables 6.9 and 6.10. Also, the t-test for two samples were performed, at 5% statistical significance, for all the variables to test if there was any statistical difference between these two

samples (outperforming and not outperforming plants) in which the alternative hypotheses was that the difference in means was not equal to 0.



**Table 6.9. Outperforming and not outperforming residuals comparison**

Groups	True (N=342)		False (N=340)		Residuals comparison		t-test (two sample)	
	Mean	Median	Mean	Median	Mean	Median	P-value	95 percent confidence interval
<b>HPV</b>	22.9	22.4	30.4	27.6	0.328	0.232	<2e-16	(6.33, 8.65)
<b>Vehicle.Launch</b>	0.19	0	0.21	0	0.105	-	0.5	(-0.0414, 0.0789)
<b>AWD</b>	235	236	237	236	0.009	0.000	0.5	(-2.79, 5.39)
<b>Variety.bodyandchassis</b>	5.8	4	5.42	4	-0.066	0.000	0.2	(-0.973, 0.206)
<b>Flexibility</b>	5.97E+09	4.82E+08	2.43E+09	6.32E+08	-0.593	0.311	0.0003	(-5440000000, -1640000000)
<b>Model.types</b>	2.3	2	2.28	2	-0.009	0.000	0.8	(-0.191, 0.153)
<b>Year</b>	2003	2003	2003	2003	0.000	0.000	0.8	(-0.436, 0.335)
<b>APV</b>	183,000	188,000	189,000	196,000	0.033	0.043	0.4	(-8053, 18459)
<b>CAC</b>	87.4	92.5	87.5	93	0.001	0.005	0.9	(-3.86, 4.45)
<b>No.Platform</b>	1.19	1	1.29	1	0.084	0.000	0.02	(0.0155, 0.1808)
<b>No.Bodystyle</b>	3.19	2	2.76	2	-0.135	0.000	0.02	(-0.7808, -0.0758)
<b>No.Chassisconfiguration</b>	2.62	2	2.66	2	0.015	0.000	0.8	(-0.275, 0.353)
<b>Outsourcing</b>	63.3	63.6	61	63.6	-0.036	0.000	0.06	(-4.871, 0.126)
<b>Platformstrategy</b>	0.63	0.5	0.67	0.5	0.063	0.000	0.1	(-0.00972, 0.09680)
<b>Hourly.total</b>	91.2	92.1	91	92.2	-0.002	0.001	0.5	(-0.720, 0.328)

Here are the findings about the outperforming plants,

- They launched lower new products; however, there were not statistically significant differences between these two groups at 5% significance level.
- They were much more flexible when comparing to other plants, P-value was 0.0003.
- They were using a lower number of platforms and a higher number of body styles; however, there were statistically significant differences between these groups at 5% significance level. As a result of being more flexible and utilizing platform strategy, they were able to assemble a greater number of body styles.

Ownership and car segments have been considered separately in Table 6.10.

**Table 6.10. Outperforming and not outperforming residuals comparison for the ownership and car segment**

Ownership and car segment		True	False	Residuals comparison	Car Segment
Ownership	American	280	266	1.053	
	Japanese	44	53	0.830	
	Joint Venture	18	21	0.857	
car segment	Compact	41	34	1.206	B
	Full-Size Pickup	58	50	1.160	B
	Full-Size SUV	21	26	0.808	
	Large	14	4	3.500	B
	Large Van	13	3	4.333	B
	Luxury	9	11	0.818	
	Medium Duty	4	9	0.444	
	Midsize	56	39	1.436	A
	Midsize SUV	33	33	1.000	
	Minivan	11	42	0.262	
	Small Pickup	27	16	1.688	A
	Small SUV	22	5	4.400	B
	Sports Car	11	9	1.222	C
Sub Compact	23	58	0.397		

Here are the findings from Table 6.9:

- The percentage of American plants that outperformed the residuals were higher than Japanese and Joint Venture plants.
- Out of 14 car segments, eight car segments had the majority of the outperforming plants which were compact, full-size pickup, large, large-van, midsize, small pickup, small SUV, and sports car. Out of eight car segments five were in segment B and two in segment A.

### 6.7. Analyzing the Plants That Were not Productive

It was decided to study the plants that had the highest HPV during the study period to find out what has been happening to those. Hence, the 20 plants that had the highest HPV were extracted from the rest of the data. Among there 20 plants there were four plants during the study period that were Ford Cuautitlan Truck, Ford Cuautitlan, GM Janesville –Medium, and GM Lansing Craft Ctr. The trend of HPV for these four plants is shown in Figure 6.8.

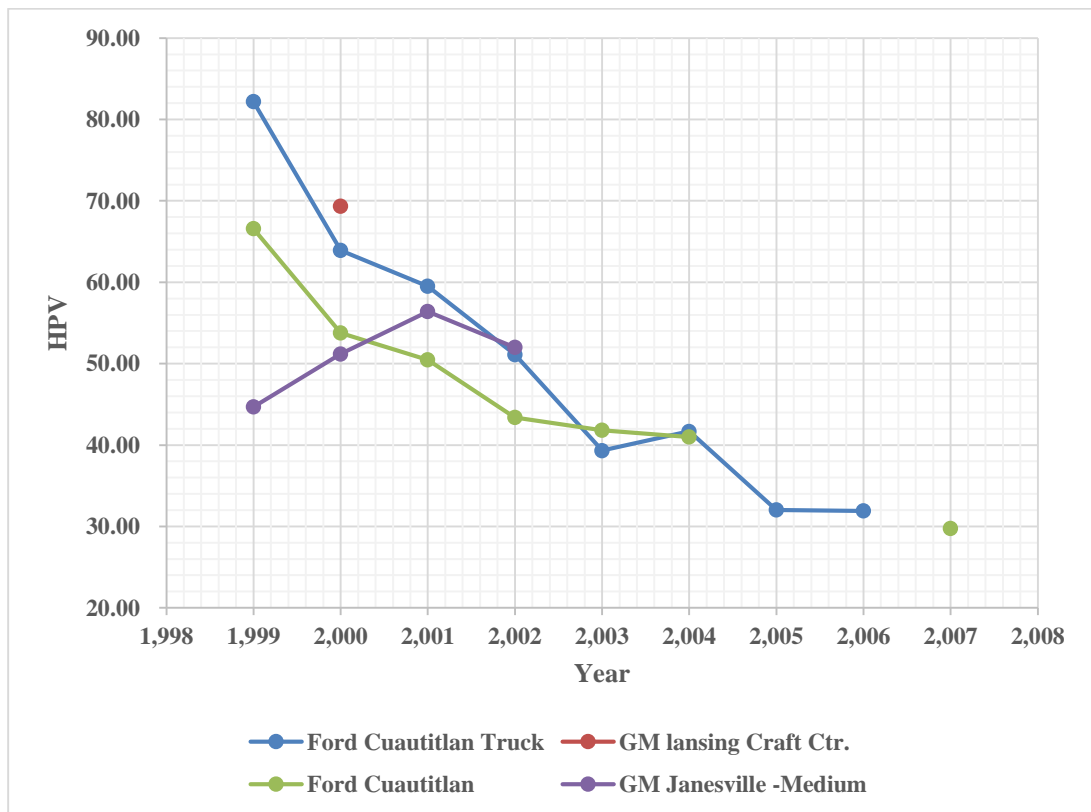


Figure 6.8. The trend of HPV for the plants that had the highest HPV over the study period

Figure 6.8 indicates that two plants tremendously reduced their HPV during the study period. For instance from 82 to 31.91 and 66.58 to 29.77 during the study period, they were Ford Cuautitlan Truck and Ford Cuautitlan, respectively. On the other hand, the plants who were not able to compete in this competition and were not able to reduce the HPV were forced to close down at some points. The Lansing Craft Center that started production in 1987 as the Reatta Craft Center closed in 2006. GM Janesville –Medium permanently closed in 2009.

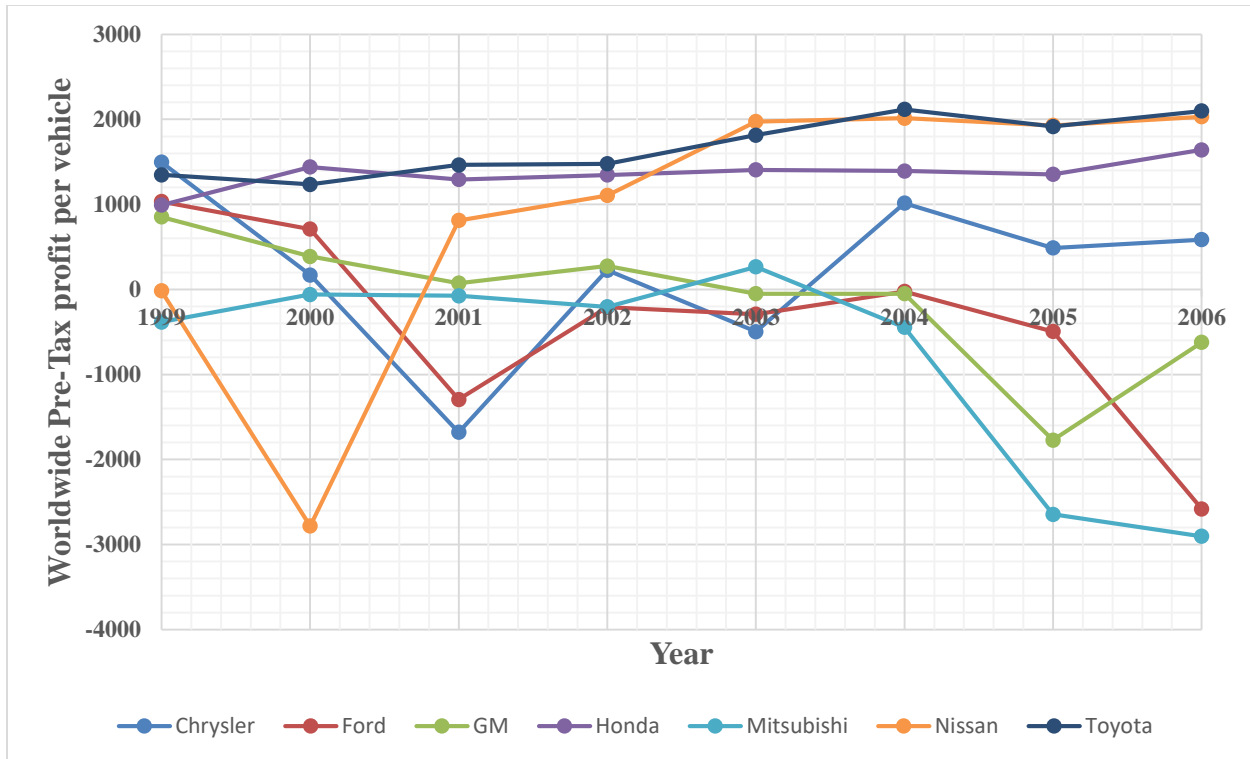
### 6.8. Overview of North American Car Manufacturers

Labor productivity might have a significant impact on the final cost of a car-manufacturer. Based on Harbour (2004), the labor productivity impact for Big Three is estimated and shown in Table 6.11.

**Table 6.11. Labor and benefit cost per vehicle for Big Three**

	<b>Chrysler</b>	<b>Ford</b>	<b>GM</b>	<b>Benchmark</b>
Labor Hours per Vehicle (Assembly, Stamping, Powertrain)	33.85	36.98	34.33	32.00
Weighted Labor Rate	37.88	40.32	41.25	38.00
Labor and benefit cost per vehicle	\$1,358	\$1,491	\$1,416	\$1,216

The worldwide profitability, pre-tax profit per vehicle (\$), is shown in Figure 6.9 for the available plants' brands during the study period.



**Figure 6.9. Worldwide profitability Pre-tax profit per vehicle for North American car manufactures**

Here are some findings from Figure 6.9:

- Honda and Toyota had the highest stable profit during the study period.
- Nissan had the highest worldwide pre-tax profit improvement per vehicle, from -\$2,782 in 2000 to \$2,030 in 2006. Moreover, Nissan was the most profitable brand among the six largest automakers in North America by having \$2,249 pre-tax profit per vehicle in the region.
- GM and Ford were profitable in 1999; however, they had significant loss in 2005 and 2006.
- Chrysler had a better position among the Big Three and quickly recovered after a huge loss in 2001.
- Almost during all the years from 1999 to 2006, Mitsubishi had loss and finally its sole plant in North America was closed in 2015.

There are several reasons that could be considered as the source of financial performance differences between American and Japanese plants such as (Harbour, 2002, 2003, 2004, 2005, 2006, 2007)

- Japanese automakers had a long term constructive relation with suppliers. They were giving them adequate profit by relentlessly analyzing every step in their shared design and production to take out waste and involved them in creating quality. Suppliers were looking for more work from Toyota and Honda- rather than GM, Ford, and Chrysler. In a survey of some 220 suppliers, it was found that GM, Ford, and Chrysler primarily focused on cost reduction and had little regard for supplier's intellectual property and proprietary information.
- GM and Ford were forced to escalate health care obligations with more than \$1.5 billion for the current and retired American employees in 2005 that penalize GM, Ford, and Chrysler for \$897, \$593, and \$370 per vehicle, respectively.
- Investment on the flexibility at the shop floor level could pay off. Ford invested \$375 million in one plant body shop to make it more flexible and assemble variety of body styles there, which the outcomes were promising.
- In 2005, 30.6 million vehicles were recalled mostly by GM, which suffered a huge loss for that reason.
- United Automobile Workers (UAW) could be considered as one of the biggest roadblocks to a plants' competitiveness especially for American plants.
- Absenteeism at some Ford plants has run above 10% on any given day, about twice as high as Nissan, Toyota, and Honda at their U.S. plants.
- Rising gasoline prices that help Japanese companies make a significant profit because of their low gas mileage despite American companies.
- Regulations at the national level.
- Shifting market share.

## **6.9. Conclusion**

In this chapter several robust regressions, shrinkage methods, and the combination of these methods were discussed. It was shown that the statistical model that was obtained by the combination of MM estimator and lasso regression, a hybrid method, has the best statistical performance. Accordingly, the outperforming plants were reviewed to discover any possible strategies that those adhere to, which the main two strategies were applying flexible manufacturing

and platform strategy. Hence, this statistical model was considered as a base and a final model, including the two-way interactions, was built upon this model and validated accordingly.

## 7 : Mixed Integer Programming

### 7.1. Introduction

Mathematical programming is an important area in applied mathematics and is widely used in academic and industrial areas. One of the most general types of finite-dimensional, single objective mathematical programs is mixed integer programming. In this chapter, mixed integer programming, containing both continuous and integer decision variables, and different approaches to solve such a problem are reviewed. These approaches include branch and bound and metaheuristics such as ant colony optimization algorithms. Hence, the research objectives 7 and 8, RO7 and RO8, developing a Mixed Integer Programming (MIP) model to specify target values for the productivity measure and a software program that utilizes the developed statistical model and is synergistically integrated with a Mixed Integer Programming will be discussed.

### 7.2. Mixed Integer Programming

The linear/nonlinear programming is used when all the decision variables are continuous, in which they could be allowed to be fractional. Often this is not a realistic assumption and the variables couldn't be continuous and might be an integer such as the number of models that an automaker could produce during a year. In this situation, the objective function and constraints could be defined as (S. Bradley et al., 1977),

$$\text{Max } \sum_{j=1}^n C_j x_j$$

Subject to:

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad (i=1, 2, \dots, m)$$

$$x_j \geq 0 \quad (j=1, 2, \dots, n)$$

$$x_j \text{ integer} \quad (\text{for some or all } j= 1, 2, \dots, n)$$

When all the decision variables are integers the problem is integer-programming. On the other hand, if some decision variables are not integers and some are, the problem is mixed integer programming; however, the problem could be linear or non-linear depending on the objective function or constraints.



In the literature, the most common deterministic approaches for both mixed integer linear and non-linear programming (MILP/MINLP) are branch and bound and metaheuristic (M. Schlueter, 2012).

### 7.3. Defining the Business Strategies to Find a Desired HPV Value

A desired target value for HPV can be obtained by using the developed regression model, equation (6.8), with each variable limitations in that car segment, and an optimization method such as mixed integer programming. For instance, the best achieved value of the HPV in car segment A was 16.37 hours (achieved by Nissan Smyrna, TN in 2003).

To achieve a target value for HPV, it is possible to use the branch and bound (BB) or metaheuristic optimization methods in which the obtained hybrid regression model is used as the objective function and the limitations on each of the independent variables is used as the constraints in the optimization model. For instance, the following mixed integer nonlinear programming (MINLP) formulation was used to achieve a target value of 10 hours for the HPV for car segment A, which was the benchmark of productivity during the study period at 16.37 hours. Z is the deviation from the target value for HPV, 10 hours in this example.

$$\begin{aligned} \text{Min}(Z) = & (1,320 + 0.0385\text{AWD} + 0.513\text{Vehicle Variety} - 0.581\text{flexibility} + 1.15\text{Model} - 0.63\text{Year} - 3.83\ln(\text{APV}) + 1.13\ln(\text{CAC}) \\ & - 0.0087\text{Outsourcing} + 1.04\text{platform.strategy} + 0.0061\text{Hourly.total} - 12.90\text{Vehicle.launch} - 1.83\text{Japanese} - 1.28\text{Joint.venture} \\ & + 18.7\text{Segment.B} + 47.30\text{Segment.C} - 1.36\log(\text{APV}): \text{Segment.B} - 3.40\log(\text{APV}): \text{Segment.C} - 0.578\text{Platformstrategy}: \text{Vehicle.Launch} \\ & - 0.399\text{Flexibility}: \text{Vehicle.Launch} + 0.26\text{Hourly.total}: \text{Vehicle.Launch} - 10)^2 \end{aligned} \quad (7.1)$$

Subject to:

$$\begin{aligned} 35 & \leq \text{AWD} \leq 336 && (\text{AWD is an integer}) \\ 2 & \leq \text{Vehicle Variety} \leq 22 && (\text{Vehicle variety is an integer}) \\ 13.8 & \leq \text{Flexibility} \leq 25.8 \\ 1 & \leq \text{Model.types} \leq 5 && (\text{Model types is an integer}) \\ 8.83 & \leq \ln(\text{APV}) \leq 13.03 \\ 2.56 & \leq \ln(\text{CAC}) \leq 5 \\ 1999 & \leq \text{Year} \leq 2007 && (\text{Year is an integer}) \\ 12.9 & \leq \text{Outsourcing} \leq 100 \\ 0.2 & \leq \text{Platform.strategy} \leq 1.5 \\ 74.1 & \leq \text{Hourly.total} \leq 96.7 \end{aligned}$$

Z is the expected value of the quadratic loss and incorporates both the variance of the estimator and its bias. This is also known as the Taguchi Loss Function. Z can be calculated as,

$$Z = (\bar{X}^2 - T)^2 + S^2$$

Where,

$\bar{X}$  = The sample mean

T = The specified target, in our research 13.5

$S^2$  = The sample variance

In a deterministic problem like this research, the sample variance will be zero at each iteration of the optimization, so the Z value should be very small as the mean hits the target. Several methods and algorithms are reviewed in the next sections in order to find an appropriate method for this study.

### 7.3.1. Branch and Bound Method

Land and Doig (1960) introduced the branch and bound (BB) method in 1960 for the first time. They considered an assumption in which the discrete decision variables of the MILP/MINLP were relaxable, which means the objective function and decision variables could also be evaluated for the continuous numbers; however, they actually must be discrete (Land & Doig, 1960; M. Schlueter, 2012). The BB can be applied for both MILP/MINLP problems. By recursive branching of the original optimization problem and calculating the decision variables, a decision tree over the discrete search space of the original problem can be generated. The BB method, per se, manages the tree and additional algorithms are required to solve the generated sub-problems. Taha (1982) discussed the detailed steps and procedure (Taha, 1982).

The following formulations, based on the branch and bound, were used to solve the problem with CPLEX software to get a value of 10 for HPV when there is a new product launch and AWD and CAC are 240 days and 100, respectively,

Minimize

$$\text{Obj: } u+v$$

Subject to

$$\begin{aligned} &1,320 + 0.0385AWD + 0.513Vehicle\ Variety - 0.581flexibility + 1.15Model - 0.63Year - 3.83\ln(APV) + 1.13\ln(CAC) \\ &- 0.0087Outsourcing + 1.04platform.strategy + 0.0061Hourly.total - 12.90Vehicle.launch - 1.83Japanese - 1.28Joint.venture \\ &+ 18.7Segment.B + 47.30Segment.C - 1.36\log(APV):Segment.B - 3.40\log(APV):Segment.C - 0.578Platformstrategy:Vehicle.Launch \\ &- 0.399Flexibility:Vehicle.Launch + 0.26Hourly.total:Vehicle.Launch + u - v = 10 \end{aligned}$$

Bounds

$35 \leq AWD \leq 336$  (AWD is an integer)  
 $2 \leq Vehicle\ Variety \leq 22$  (Vehicle variety is an integer)  
 $13.8 \leq Flexibility \leq 25.8$   
 $1 \leq Model.types \leq 5$  (Model types is an integer)  
 $8.83 \leq \ln(APV) \leq 13.03$   
 $2.56 \leq \ln(CAC) \leq 5$   
 $1999 \leq Year \leq 2007$  (Year is an integer)  
 $12.9 \leq Outsourcing \leq 100$   
 $0.2 \leq Platform.strategy \leq 1.5$   
 $74.1 \leq Hourly.total \leq 96.7$   
 $u\ and\ v \geq 0$

General

AWD  
 Vehicle\_Variety  
 Vehicle\_Launch  
 Model\_types  
 Year  
 Ownership  
 End

For the stated mixed integer programming, CPLEX output is shown in Table 7.1.

**Table 7.1. Desired factors values for HPV of 10H – CPLEX solver**

MIP - Integer optimal solution: Objective =0	
Solution time = 0.02sec. Iterations = 0 Nodes = 0	
Deterministic time = 0.03 ticks (1.68 ticks/sec)	
CPLEX> d sol var -	
Incumbent solution	
<b>Variable Name</b>	<b>Solution Value</b>
AWD	240.000000
Vehicle_Variety	14.000000
Vehicle_Launch	1.000000
Flexibility	24.260000
Model_types	5.000000
Year	2007.000000
ln(APV)	13.030000
ln(CAC)	2.560000
Outsourcing	25.920690
Platformstrategy	0.200000
Hourly.total	74.100000

Variable Name	Solution Value
Ownership	2.000000
All other variables in the range 1-14 are 0.	

However, there is an instinct restriction with CPLEX, LINDO, and GRUBI softwares which can only solve one specific case in which it is related to a specific car segment and ownership for the situation that if there is a new product launch. In short, it means they are not able to handle the general formulation of the problem that includes “If statements” for different conditions.

### 7.3.2. Stochastic Metaheuristics

In the literature, several attempts to apply stochastic metaheuristics on MILP/MINLP have been proposed, such as: Genetic Algorithm (Cheung et al., 1997; Munawar et al., 2011), Particle Swarm Optimization (PSO) (Yiqing et al., 2007), Artificial Bee Colony Optimization (ABCO) (Bhattacharjee et al., 2011), and Ant Colony Optimization (ACO) (Fan et al., 2003). Deterministic or local search algorithms are often combined with stochastic metaheuristics to improve the local behavior convergence. These combined algorithms are referred to as hybrid algorithms.

### 7.3.3. Ant Colony Optimization (ACO)

ACO (Ant Colony Optimization) has been applied in a large number of path planning research due to its strength in finding a route. This method is an inspired optimization algorithm which belongs to the class of evolutionary methods, where a population of agents share some information about their path to achieve some goals. The ants randomly start to explore the area around their nest to find the food. If an ant succeeds in finding a food source, it will return to the nest, laying down a chemical pheromone trail to mark its path for the other ants. This pheromone trail attracts other ants in the hope of finding food again (M. Schlueter, 2012).

The basic idea of ACO is similar to the nature in which some artificial ants randomly start to search the feasible space in the hope of finding the optimal solution; however, the artificial ants use some pheromone to make their pace traceable. Schlüter et al. (2009) extend the application of this concept in the mixed integer search domain (Schlüter et al., 2009).

### 7.3.4. Mixed Integer Distributed Ant Colony Optimization (MIDACO)

Mixed integer distributed ant colony optimization (MIDACO) in combination with the oracle penalty function method that was developed by Schlüter (2014) was used for this study. MIDACO is a nonlinear programming solver that can robustly solve problems with critical function properties like high non-convexity, non-differentiability, flat spots, and even stochastic noise (M. Schlueter, 2014).

The factors' range for segment A, their optimal values and mean squared error (Z) are shown in Table 7.2. In our study the Z was 3.5852268e-022.

Any new constraints based on the available plants' resources and strategies can be added. This approach can be repeated for segments B and C to find the desired setting for any other hypothetical situation. Hence, automakers may improve their HPV for a variety of different factors and limitations by using excellent tools such as hybrid regression in conjunction with optimization techniques.

Special attention should be paid to each manufacturing plant's constraints while using the HPV equations; for instance, it would make little sense for the plant to reduce its operating hours (AWD) to improve HPV.

**Table 7.2. Desired factors values for HPV of 10H – MIDACO Algorithm**

Variable	Values range based on the available data during the study period**	Obtained Setting***
HPV	(13.57, 38.8)	10
AWD*	(35, 301)	240
Vehicle Variety	(1, 14)	1
Flexibility	(13.8, 24.26)	24.2217
Model types	(1, 5)	1
LN(APV)	(8.83, 13.03)	13.029
LN(CAC)	(2.56, 5)	3.1445
Year	(1999, 2007)	2007
Outsourcing	(12.9, 100)	55.05
Platformstrategy	(0.2, 1.5)	1.46
Hourly.total	(74.1, 96.7)	83.12
Ln(APV):Segment	-	-

Variable		Values range based on the available data during the study period**	Obtained Setting***
Platformstrategy:Vehicle.Launch		-	-
Flexibility:Vehicle.Launch		-	-
Hourly.total:Vehicle.Launch		-	-
Vehicle Launch		0, 1	1
Ownership		(American, Japanese, Joint venture)	Japanese
Segment		A	A
Response	Fit	Mean Squared Error (Z)	
HPV	10	3.5852268e-022	

\*It was assumed that a plant was operating for 240 days

\*\*The variables range for segments A, B, and C are in Table AIV in the Appendix.

\*\*\*The optimization solver was run for the 100 iterations

A schematic view of the software program with “If Statements” is shown in Table 7.3. The developed model can be used for all ownerships and car segments regardless of launching a new product. The developed software program has the advantage of handling the “If Statements” that other programming software such as CPLEX, LINDO, and GRUBI are not able to solve the general format (six coefficients have the if statements, see Table 7.3).

The benchmark of productivity without any product launch was Chrysler Toledo with 13.57 hours which was achieved in 2007. However, a similar mixed integer nonlinear programming (MINLP) formulation was used to achieve a target value of 10 hours for the HPV for car segment A. The results are shown in the Appendix (Table AV).

**Table 7.3. The optimization model adapted for all car segments and ownerships**

<b>Optimization model</b>					
	<b>Coefficients</b>	<b>Input Control</b>	<b>LCL</b>	<b>UCL</b>	<b>Integer</b>
<b>Intercept</b>	1320				
<b>AWD</b>	0.0385	240	240	240	Y
<b>Vehicle_Variety</b>	0.513	1	1	14	Y
<b>Vehicle_Launch</b>	-12.9	1	1	1	Y
<b>Flexibility*</b>	-0.98	24.25961814	13.8	24.26	
<b>Model_types</b>	1.15	1	1	5	Y
<b>Year</b>	-0.63	2003	1999	2007	Y
<b>Segment*</b>	0	1	1	1	Y
<b>ln(APV)*</b>	-3.83	12.38193266	8.83	13.03	
<b>ln(CAC)</b>	1.13	4.071962427	2.56	5	
<b>Outsourcing</b>	-0.0087	55.04785686	12.9	100	
<b>Platformstrategy*</b>	0.462	1.458417071	0.2	1.5	
<b>Hourly.total*</b>	0.2661	83.11927093	74.1	96.7	
<b>Ownership*</b>	-1.83	2	2	2	Y
<b>Ln(APV):Segment</b>	0	1	1	1	
<b>Platformstrategy:Vehicle.Launch</b>	-0.578	1	1.458417071	1.458417	
<b>Flexibility:Vehicle.Launch</b>	-0.399	-0.5845	24.25961814	24.25962	
<b>Hourly.total:Vehicle.Launch</b>	0.26	0.0138	83.11927093	83.11927	
<b>HPV</b>	10				

Note: \* Have the embedded if statements (six of the coefficients)

<b>American</b>	1
<b>Japanese</b>	2
<b>Joint ventire</b>	3

<b>Segment A</b>	1
<b>Segment B</b>	2
<b>Segment C</b>	3

#### 7.4. Conclusion

In this chapter, mixed integer programming concept and its solution procedure were reviewed. It was shown that MIDACO has the advantage when comparing the CPLEX, LINDO, and GRUBI solvers. Hence, MIDACO was used to show how a plant manager can achieve a target value for HPV according to available resources and constraints.

## **8 : Conclusion, Contributions, Limitations, and Future Research**

### **8.1. Conclusion**

The most important process to improve a car manufacturer's productivity is the production process itself. HPV is a widely practiced and recognized measure that automakers use to gauge labor productivity; however, it seems that there is a limited understanding of the set of factors that affect HPV at both the conceptual and technical levels in the automotive industry. Hence, the focus of this research was to define the set of strategies that improve productivity in the automotive industry for the production process.

In this study, the impact of 13 variables on HPV was considered and the combination of MM-estimator and lasso regression was used to develop a robust regression model to show the effect of these factors on HPV. The factors were selected based on a literature survey and practical considerations.

It was shown that product variety was often considered a primary factor in decreasing productivity; a one-unit increase in product variety results in an HPV increase of 0.513 hours. Adoption of complementary activities to product variety such as increasing the production volume, platform sharing strategy, and flexible manufacturing were shown to reduce the HPV. However, this reduction may come at a cost in terms of adverse effects on other variables such as flexibility and platform sharing. While launching a new vehicle, using platform strategy, flexible manufacturing, and reducing the percentage of hourly employees help to improve productivity and reduce HPV.

Flexibility and platform sharing strategies are two of the most effective factors in this regard. They could be considered technology, worker skills (implicit and explicit knowledge), the organizational capabilities for rapid changeover from one product to another, and assembling more variety of products based on fewer platforms. They enable plants to respond more effectively to a wide variety of changes in their competitive environment.

It must be mentioned that the overall customer satisfaction, facility performance level, and organization profitability can be improved simultaneously. During the study period, Nissan was the benchmark in this respect. Despite its variety of car models throughout the years, the HPV value was held at a reasonable level by moving toward flexibility, increasing annual production volume, and implementing a platform sharing strategy. However, it was the most profitable North



American car manufacturer in 2005 and increased its pretax profit per vehicle by \$4,812 during five years.

The methodology developed in this research is applicable to all types of vehicles and can be used by any auto manufacturer in the world regardless of their plant location. Therefore, any automaker may use the results of the current study to explore economical and practical ways to improve productivity. Automotive plants may note that our regression equation is data specific and our methodology can be used as a guide to enhance productivity. Finally, to the best of our knowledge, this study provides the first piece of empirical evidence on how productivity can be improved in the automotive manufacturing sector.

## **8.2. Contributions**

There are several contributions that are related for each research objectives as,

- RO3: Define and develop some new variables based on the available data restriction and practical consideration to measure productivity (HPV). These variables didn't exist or utilized in the literature survey. Here is the list of the variables:
  - Car segmentation
  - Flexible manufacturing
  - Platform sharing
  - Manufacturing Ownership
  - Outsourcing
  - Hourly employees percentage
  - Number of models
  - Car Assembly and Capacity (CAC) utilization
  - Annual available working days
- RO4: Develop a hybrid statistical method to estimate the productivity measure
  - Through a very sophisticated and rational procedure, it was shown that how the pursued steps in this research improve the performance of the developed statistical model (from 0.58 to 0.71 and 5.71 to 3.73 for the adjusted R-squared and mean squared cross validation, respectively). These steps are including but not limited to developing the base model with several transformations, using robust and shrinkage methods, and developing a hybrid method (a combination of both robust and shrinkage methods).
- RO6: Outperforming residuals characteristics

- Another novelty in this research was considering the characteristic of the outperforming residuals (plants) to derive any particular business strategies that these plants are adhere to. The results showed that those plants were following specific strategies such as flexibility and platform sharing
- RO7 & RO8: Combining applied statistical models and optimization techniques
  - A software program was developed that was able to solve a mixed integer nonlinear programming to get a desirable value for the HPV. The developed statistical model was considered as an objective function to find the optimal settings that a plant manager can adhere to achieve a desired target value for HPV, based on its plant constraints and resources. It was shown that the developed statistical model, a robust and advanced approach, can overcome the plants that were the benchmark of productivity during the study period (Nissan Smyrna, TN in 2003 and Chrysler Toledo, OH in 2007).

### **8.3. Limitations**

In most empirical studies, researchers will encounter limitations and data deficiencies, a few of which should be mentioned in relation to this study. Conditions present in this study that should be accounted for include plants that stopped the production of a specific product, which included 38 cases. Furthermore, data available for manufacturers who produced in a variety of segments have been considered for their highest volume segment. Since this study was conducted for the available data from 1999-2007, any extrapolation for the current situation isn't advisable and the research must have considered in the scope of available data.

### **8.4. Future Research**

Possibilities for future work include:

- Identifying appropriate and feasible measures for new predictors such as product complexity, logistical advancement and processes, level of plant's automation, the country the plant was located such as USA, Canada, and Mexico, the type of union relationship such as union or not union, and lean manufacturing also can be considered to be included in the HPV model.

- The analyses can be extended to European automakers and North American and European companies can be compared to find any significant differences in HPV and relevant strategies to improve HPV.
- Prioritizing the body, paint, and trim lines in terms of their roles in productivity enhancement.
- Defining the strategies and techniques to improve energy efficiency for North American car manufacturers including the following steps:
  - Developing an advanced and robust statistical model to estimate the energy performance indicator.
  - Defining the strategies to improve energy performance indicators through optimization methods such as MIDACO.
- Defining the strategies and techniques to simultaneously improve productivity and energy efficiency for North American car manufacturers including the following steps:
  - Developing a multivariate statistical model to estimate productivity and energy efficiency.
  - Defining the strategies to improve productivity and energy efficiency by solving a multi objective function that was developed in the previous step.

## Literature Cited

- Al-Darrab, I. A. (2000). Relationships between productivity, efficiency, utilization, and quality. . *Work Study*, 49(3), 97-104.
- Al-Tahat, M. D., & Jalham, I. S. (2015). A structural equation model and a statistical investigation of lean-based quality and productivity improvement. *Journal of Intelligent Manufacturing*, 26(3), 571-583.
- Alden, J. M., Burns, L. D., Costy, T., Hutton, R. D., Jackson, C. A., Kim, D. S., & Veen, D. J. V. (2006). *General Motors increases its production throughput*. Paper presented at the Informs.
- Alma, Ö. G. (2011). Comparison of robust regression methods in linear regression. *Int. J. Contemp. Math. Sciences*, 6(9), 409-421.
- Almström, P., & Kinnander, A. (2011). The Productivity Potential Assessment method: assessing and benchmarking the improvement potential in manufacturing systems at shop-floor level. *International Journal of Productivity and Performance Management*, 60(7), 758-770.
- Aspen, U., Bråthen, A., Cassel, P., Ericsson, P., & Marelius, M. (1991). *Produktivitetens utveckling inom svenskt näringsliv—en studie baserad på nationalräkenskaperna*. Stockholm.: Allmänna Förlaget.
- Baldwin, C. Y., & Clark, K. B. (2009). *Managing in the Modular Age: Architectures, Networks, and Organizations*: Wiley.
- Baraldi, E. C., & Paulo, C. (2011). Ergonomic planned supply in an automotive assembly line. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 21(1), 104-119.
- Battini, D., Faccio, M., Persona, A., & Sgarbossa, F. (2011). New methodological framework to improve productivity and ergonomics in assembly system design. *International Journal of Industrial Ergonomics*, 41(1), 30-42.
- Bayo-Moriones, A., Bello-Pintado, A., & Merino-Díaz de Cerio, J. . (2010). 5S use in manufacturing plants: contextual factors and impact on operating performance. *International Journal of Quality & Reliability Management*, 27(2), 217-230.
- Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131-160.
- Bernolak, I. (1997). Effective measurement and successful elements of company productivity: The basis of competitiveness and world prosperity. *International Journal of Production Economics*, 52(1), 203-213.
- Bhattacharjee, P., P. , Rakshit, I. , Goswami, A. , Konar, A., & Nagar, K. . (2011). *Multi-robot path-planning using artificial bee colony optimization algorithm*. Paper presented at the Nature and Biologically Inspired Computing (NaBIC).
- Biesebroeck, J. V. (2007). Complementarities in automobile production. *Journal of Applied Econometrics*, 22(7), 1315-1345.

- Bogue, R. (2013). Robotic vision boosts automotive industry quality and productivity. *Industrial Robot: An International Journal*, 40(5), 415-419.
- Box, G. E., & Cox, D. R. . (1964). An analysis of transformations. . *Journal of the Royal Statistical Society. Series B (Methodological)*,, 211-252.
- Bradley, D., Bruns, M., Fleming, A., Ling, J., Margolin, L., Roman, F., & Flury, A. (2005). Automotive Industry Analysis.
- Bradley, S., Hax, A., & Magnanti, T. . (1977). *Applied mathematical programming*.
- Broberg, O. , Andersen, V., & Seim, R. (2011). Participatory ergonomics in design processes: The role of boundary objects. *Applied ergonomics*, 42(3), 464-472.
- Brondo, K., & Baba, M. (2010). Last In, First Out: A Case Study of Lean Manufacturing in North America's Automobile Industry. *Human Organization*, 69(3), 263-274.
- Brynjolfsson, E., & Yang, S. . (1996). Information technology and productivity: a review of the literature. *Advances in computers*, 43, 179-214.
- Callen, J. L., Morel, M., & Fader, C. (2005). Productivity measurement and the relationship between plant performance and JIT intensity. *Contemporary Accounting Research*, 22(2), 271-309.
- Cheung, B. S., Langevin, A., & Delmaire, H. . (1997). Coupling genetic algorithm with a grid search method to solve mixed integer nonlinear programming problems. . *Computers & Mathematics with Applications*,, 34(12), 13-23.
- Chew, W. B. (1988). No-nonsense guide to measuring productivity. *Harvard Business Review*, 66(1), 110- 118.
- Cholasuke, C., Bhardwa, R. , & Antony, J. (2004). The status of maintenance management in UK manufacturing organizations: results from a pilot survey. *Journal of Quality in Maintenance Engineering*, 10(1), 5-15.
- Cobb, C. W., & Douglas, P. H. . (1928). A theory of production. *The American Economic Review*, 18(1), 139-165.
- Direction, S. . (2005). Henry Ford would just love Toyota: Car makers' strategy for market share. *Strategic Direction*, 21(4), 25 - 27.
- Doran, D. (2003). Supply chain implications of modularization. *International Journal of Operations & Production Management*,, 23(3), 316-326.
- Dorf, R. C. (1998). *The technology management handbook*: CRC Press.
- Duguay, C. R., Landry, S., & Pasin, F. (1997). From mass production to flexible/agile production. *International Journal of Operations & Production Management*, 17(12), 1183-1195.
- El-Khalil, R. (2014). Managing and improving robot spot welding efficiency: a benchmarking study. *Benchmarking: An International Journal*, 21(3), 344-363.
- El-Khalil, R. . (2015). Simulation analysis for managing and improving productivity: A case study of an automotive company. . *Journal of Manufacturing Technology Management*, 26(1), 36-56.

- Fan, X. X., Luo, S., Yi, S., Yang, H., & Proceedings., Zhang "" . (2003). *Optimal path planning for mobile robots based on intensified ant colony optimization algorithm*. Paper presented at the 2003 IEEE International Conference on.
- Faraway, J. J. . (2005). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*: CRC press.
- Filzmoser, P. (2001). Robust principal component regression (pp. 132-137): Belarusian State University, Minsk.
- Fisher, T. J. (1990). Business productivity measurement using standard cost accounting information. *International Journal of Operations & Production Management*, 10(8), 61-69.
- Fitzmaurice, G. . (2000). The meaning and interpretation of interaction. *Nutrition*, 16(4), 313-314.
- Fortune. (2014). GLOBAL 500 2014. Retrieved 23rd September 2014, from <http://fortune.com/global500/>
- Fox, J., & Weisberg, S. (2010). *An R companion to applied regression*. : Sage.
- Golhar, D. Y., & Deshpande, S. P. (1999). Productivity comparisons between Canadian and US TQM firms: An empirical investigation. . *International Journal of Quality & Reliability Management*, 16(7), 714-722.
- Gonzalez-Benito, J., & Spring, M. (2000). JIT purchasing in the Spanish auto components industry - Implementation patterns and perceived benefits. *International Journal of Operations & Production Management*, 20(9), 1038-1061. doi: 10.1108/01443570010339145
- Gopal, A., Goyal, M., Netessine, S., & Reindorp, M. (2013). The impact of new product introduction on plant productivity in the North American automotive industry. *Management Science*, 59(10), 2217-2236.
- Gunasekaran, A., & Cecille, P. (1998). Implementation of productivity improvement strategies in a small company. *Technovation*, 18(5), 311-320.
- Gupta, S., & Prasad, A. (2011). Productivity-based hybrid model: learning from the Indo-Japanese and Indian auto sector. *Journal of Advances in Management Research*, 8(1), 158-171.
- Harbour. (2002). The Harbour Report™ North America 2002. Detroit, Michigan.
- Harbour. (2003). The Harbour Report™ North America 2003. Detroit, Michigan.
- Harbour. (2004). The Harbour Report™ North America 2004. Detroit, Michigan.
- Harbour. (2005). The Harbour Report™ North America 2005. Detroit, Michigan.
- Harbour. (2006). The Harbour Report™ North America 2006. Detroit, Michigan.
- Harbour. (2007). The Harbour Report™ North America 2007. Detroit, Michigan.
- Harner, J. (2015). *Modeling Multivariate Data* West Virginia University.
- Hastie, Trevor J., Tibshirani, R. J., Friedman, J. H., & Chicago. (2011). *The elements of statistical learning: data mining, inference, and prediction*. .

- Herron, C., & Hicks, C. (2008). The transfer of selected lean manufacturing techniques from Japanese automotive manufacturing into general manufacturing (UK) through change agents. *Robotics and Computer-Integrated Manufacturing*, 24(4), 524-531. doi: 10.1016/j.rcim.2007.07.014
- Hill, K., Menk, D., & Cooper, A. (2015). Contribution of the automotive industry to the economies of all fifty states and the United States. *Center for Automotive Research*, 1-42.
- Hill, K., Menk, D., Cregger, J., & Schultz, M. (2015). Auto jobs and economics. Retrieved February 19 2015, from <http://www.autoalliance.org/files/dmfile/Tax-Revenue-Generated-by-Auto-Sector-2013.pdf>
- Hill, T. . (1993). *Manufacturing strategy: the strategic management of the manufacturing function*.
- Hirsh, E., Kakkar, A., Singh, A., & Wilk, R. (2015). 2015 Auto Industry Trends. Retrieved February 16 2016, from <http://www.strategyand.pwc.com/perspectives/2015-auto-trends>
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis (Vol. 3)*: New York: Wiley.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- IEA. (2016). Definition and Domains of Ergonomics. Retrieved March 29 2016, from <http://www.iea.cc/whats/index.html>
- Indiana, U. (2016). Types of Variables. Retrieved March 29 2016, from [http://www.indiana.edu/~educy520/sec5982/week\\_2/variable\\_types.pdf](http://www.indiana.edu/~educy520/sec5982/week_2/variable_types.pdf)
- Jacoby, B. . (2015). Lecture 12: Robust regression: Michigan State University.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. . (2013). *An introduction to statistical learning (Vol. 112)*: New York: springer.
- Jensen, W. A., Birch, J. B., & Woodall, W. H. (2007). High breakdown estimation methods for phase I multivariate control charts. . *Quality and Reliability Engineering International*, 23(5), 615-629.
- Jiang, B., Frazier, G. V., & Prater, E. L. (2006). Outsourcing effects on firms' operational performance - An empirical study. *International Journal of Operations & Production Management*, 26(12), 1280-1300. doi: 10.1108/01443570410710551
- John, J. A., & Draper, N. R. . (1980). An alternative family of transformations. *Applied Statistics*, 190-197.
- Kaplan, Robert S., & Cooper, Robin. (1998). Cost & effect: using integrated cost systems to drive profitability and performance. . In Harvard Business Press (Ed.).
- Katchova, A. (2013). Econometrics Academy.
- Kini, R. B. . (2002). IT in manufacturing for performance: the challenge for Thai manufacturers. *Information management & computer security*, 10(1), 41-48.
- Kochan, A. (2005). BMW uses even more robots for both flexibility and quality. *Industrial Robot-an International Journal*, 32(4), 318-320. doi: 10.1108/01439910510600173

- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, 14(2), 1137-1145.
- Koller, M., & Stahel, W. A. . (2011). Sharpening wald-type inference in robust regression for small samples. . *Computational Statistics & Data Analysis*, 5(8), 2504-2515.
- Koss, E., & Lewis, D. A. (1993). Productivity or efficiency—Measuring what we really want. *National Productivity Review*, 12(2), 273-284.
- Kuo, C. . (2014). Effect of Enterprise Resource Planning Information System on Business Performance: An Empirical Case of Taiwan. *Journal of Applied Finance and Banking*, 4(2), 1-19.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W., & Chicago. (2005). *Applied linear statistical models* (Vol. 4): Chicago.
- Land, A. H., & Doig, A. G. . (1960). An automatic method of solving discrete programming problems. . *Econometrica: Journal of the Econometric Society*, 497-520.
- MacDuffie, J. P. . (1996). Automotive White-Collar: Broken Ladders: Managerial Careers in the New Economy: *Managerial Careers in the New Economy* (pp. 81).
- MacDuffie, J. P., Sethuraman, K., Fisher, M. L. (1996). Product variety and manufacturing performance: Evidence from the International Automotive Assembly Plant Study. *Management Science*, 42(3), 350-369. doi: 10.1287/mnsc.42.3.350
- Maitra, S., & Yan, J. . (2008). Principle component analysis and partial least squares: Two dimension reduction techniques for regression (Vol. 79): Casualty Actuarial Society.
- May, A., & Carter, C. (2001). A case study of virtual team working in the European automotive industry. *International Journal of Industrial Ergonomics*, 27(3), 171-186. doi: 10.1016/s0169-8141(00)00048-2
- McKinsey. (2013). The road to 2020 and beyond: What's driving the global automotive industry? (pp. 1-24).
- Misterek, S. D. , Dooley, K. J., & Anderson, J. C. (1992). Productivity as a performance measure. *International Journal of Operations & Production Management*, 12(1), 29-45.
- Montgomery, D., C., (2012). *Design and analysis of experiments* (8th ed.): John Wiley & Sons.
- Moreno, A., & Terwiesch, C. (2015). Pricing and production flexibility: an empirical analysis of the us automotive industry. *Manufacturing & Service Operations Management*, 17(4), 428-444.
- Moseng, B., & Rolstadås, A. . (2001). *Success factors in the productivity process*. Paper presented at the In 10th World Productivity Congress.
- Munawar, A., Wahib, M., Munetomo, M., & Akama, K. In 2011 IEEE Congress of Evolutionary Computation (CEC). (2011). *Advanced genetic algorithm to solve MINLP problems over GPU*. Paper presented at the IEEE.
- OICA. (2015). World Motor Vehicle Sales. Retrieved June 5th 2015, from <http://www.oica.net/wp-content/uploads//total-sales-2014-2.pdf>
- OSHA. (2011). Establishment Specific Injury & Illness Data. Retrieved March 2ed 2016, from [https://www.osha.gov/pls/odi/establishment\\_search.html](https://www.osha.gov/pls/odi/establishment_search.html)



- Porter, M. E. (1985). *Competitive advantage of nations: creating and sustaining superior performance*: Simon and Schuster.
- PSU. (2016). Applied Data Mining and Statistical Learning. Accessed 10/18/2016, from <https://onlinecourses.science.psu.edu/stat857/node/155>
- Pucik, V. . (1984). White collar human resource management: A comparison of the US and Japanese automobile industries. *Columbia Journal of World Business*.
- Quazi, H. A., & Jacobs, R. L. (2004). Impact of ISO 9000 certification on training and development activities an exploratory study. *International journal of quality & reliability management*, 21(5), 497-512.
- Rathilall, R., & Singh, S. (2011). Improving quality and productivity at an automotive component manufacturing organisation in Durban-South Africa. *African Journal of Business Management*, 5(22), 8854-8874.
- Sakai, H., & Amasaka, K. (2007). Development of a robot control method for curved seal extrusion for high productivity in an advanced Toyota production system. *International Journal of Computer Integrated Manufacturing*, 20(5), 486-496. doi: 10.1080/09511920601160262
- Schlueter, M. (2014). MIDACO software performance on interplanetary trajectory benchmarks. *Advances in Space Research*, 54(4), 744-754.
- Schlueter, M. . (2012). *Nonlinear mixed integer based optimization technique for space applications*. (Doctoral dissertation), University of Birmingham.
- Schlüter, M., Egea, J. A., Antelo, L. T., Alonso, A. A., & Banga, J. R. . (2009). An extended ant colony optimization algorithm for integrated process and control system design. *Industrial & Engineering Chemistry Research*, 48(14), 6723-6738.
- Sehgal, B., & Gorai, P. (2012). Platform Strategy will Shape Future of OEMs: Flexibility to Drive Growth. from <http://sandhill.com/exec-briefing/platform-strategy-will-shape-future-of-oems-flexibility-to-drive-growth/>- Accessed July 12, 2015
- Sethi, A. K., & Sethi, S. P. (1990). Flexibility in manufacturing: a survey. *International journal of flexible manufacturing systems*, 2(4), 289-328.
- Shunk, C. (2011). Which automaker is the most productive? Retrieved June 12, 2015, from <http://www.autoblog.com/2011/04/06/which-automaker-is-the-most-productive/>
- Sink, D. S., & Tuttle, T. C. (1989). Planning and measurement in your organization of the future. *Industrial Engineering And Management*.
- Soković, M. , Pavletić, D., & Krulčić, E. (2006). Six Sigma process improvements in automotive parts production. *Journal of Achievements in Materials and Manufacturing Engineering*, 19(1), 96-102.
- Statistical Consulting Group, UCLA. (2015). Introduction to SAS.
- Steiner, P. (1987). Le projet physiocratique: théorie de la propriété et lien social. *Revue économique*, 1111-1128.

- Sumanth, D. J. (1984). *Productivity engineering and management: Productivity measurement, evaluation, planning, and improvement in manufacturing and service organizations.*: McGraw-Hill College.
- Sundin, A., Christmansson, M., & Larsson, M. (2004). A different perspective in participatory ergonomics in product development improves assembly work in the automotive industry. *International Journal of Industrial Ergonomics*, 33(1), 1-14. doi: 10.1016/j.ergon.2003.06.001
- Susanti, Y., & Pratiwi, H. (2014). M Estimation, S Estimation, And Mm Estimation In Robust Regression. . *International Journal of Pure and Applied Mathematics.*, 91(3), 349-360.
- Taha, H. A. (For VTU). (1982). *Operations Research: An Introduction* Pearson Education India.
- Taj, S. , & Berro, L. . (2006). Application of constrained management and lean manufacturing in developing best practices for productivity improvement in an auto-assembly plant. *International Journal of Productivity and Performance Management*, 55(3/4), 332-345.
- Tangen, S. . (2005). Demystifying productivity and performance. *International Journal of Productivity and performance management*, 54(1), 34-46.
- Tay, H. K. (2003). Achieving competitive differentiation: the challenge for automakers. *Strategy & Leadership*, 31(4), 23-30.
- Thun, J. H., Lehr, C. B., & Bierwirth, M. (2011). Feel free to feel comfortable—an empirical analysis of ergonomics in the German automotive industry. *International Journal of Production Economics*, 133(2), 551-561.
- Thurrow, L.C. , Christopher, W.F. , & Thor, C.G. (1993). *Handbook for Productivity Measurement and Improvement*. Portland.
- Trimble, R., Copeland, K., MacIntyre, J., & Smith, P. (2013). The Use and Impact of Manufacturing Productivity Improvement Tools and Methodologies within the Automotive Component Industry. *Engineering Management Research*, 2(1), 21-33.
- Tukey, J. W. (1960). The practical relationship between the common transformations of percentages of counts and of amounts. . In *Statistical Techniques Research Group Technical Report* (Ed.).
- Tukey, J. W. . (1977). *Exploratory data analysis*.
- Uludag, E., & Weisenstein, S. (2005). How To Reduce Hours Per Vehicle. *Automotive Design & Production*, 117(11), 12-15.
- Van Biesebroeck, J. (2003). Productivity dynamics with technology choice: An application to automobile assembly. *The Review of Economic Studies*, 70(1), 167-198.
- Van Biesebroeck, J. . (2007). The cost of flexibility. *Assembly Automation*, 27(1), 55-64.
- Voordijk, H., Meijboom, B. , & De Haan, J. (2006). Modularity in supply chains: a multiple case study in the construction industry. *International Journal of Operations & Production Management*, 26(6), 600-618.
- Wayman, O. (2006). *Creating Value by Improving Marketing Return on Investment*. Retrieved 24 September 2014, from [http://www.oliverwyman.com/content/dam/oliverwyman/global/en/files/archive/2006/MktSales06-MktReturn\\_on\\_Investment.pdf](http://www.oliverwyman.com/content/dam/oliverwyman/global/en/files/archive/2006/MktSales06-MktReturn_on_Investment.pdf)

- Weyer, Matthias. (2011). Hours-per-vehicle controlling – the renaissance of staff productivity. *International Journal of Production Research*, 49(11), 3271-3284.
- Yiqing, L., Xigang, Y., & Yongjian, L. (2007). An improved PSO algorithm for solving non-convex NLP/MINLP problems with equality constraints. . *Computers & chemical engineering*, 31(3), 153-162.
- Yohai, V. J. . (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 642-656.
- Zahoor Sarwar, S., Ishaque, A., Ehsan, N., Saeed Pirzada, D., & Moeen Nasir, Z. . (2012). Identifying productivity blemishes in Pakistan automotive industry: a case study. *International Journal of Productivity and Performance Management*, 6(12), 173-193.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. . *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.

## Appendix

**Table AI - The range of all variables based on the available data during the study period**

<b>Variable</b>	<b>N</b>	<b>Min</b>	<b>Max</b>	<b>Range</b>	<b>Median</b>	<b>Mean</b>	<b>Se.Mean</b>	<b>Std.Dev</b>
<b>Vehicle Launch</b>	682	-	1	1	-	0	0	0
<b>Model types</b>	682	1	7	6	2	2	0	1
<b>HPV</b>	682	14	82	69	24	27	0	9
<b>AWD</b>	682	35	336	301	236	236	1	27
<b>Variety body and chassis</b>	682	1	25	24	4	6	0	4
<b>Ln(Flexibility)</b>	682	11.82	25.80	25.80	20.11	22.16	20.01	23.26
<b>APV</b>	682	2,720	459,000	456,000	192,000	186,000	3,370	88,100
<b>CAC</b>	682	11	150	139	93	88	1	28
<b>Outsourcing</b>	682	13	100	87	64	62	1	17
<b>Platform strategy</b>	682	0	3	3	1	1	0	0
<b>Hourly total</b>	682	74	99	25	92	91	0	3

# Histogram of HPV, Vehicle Laun, AWD, Variety(body, Ln(Flexibili, ... Normal

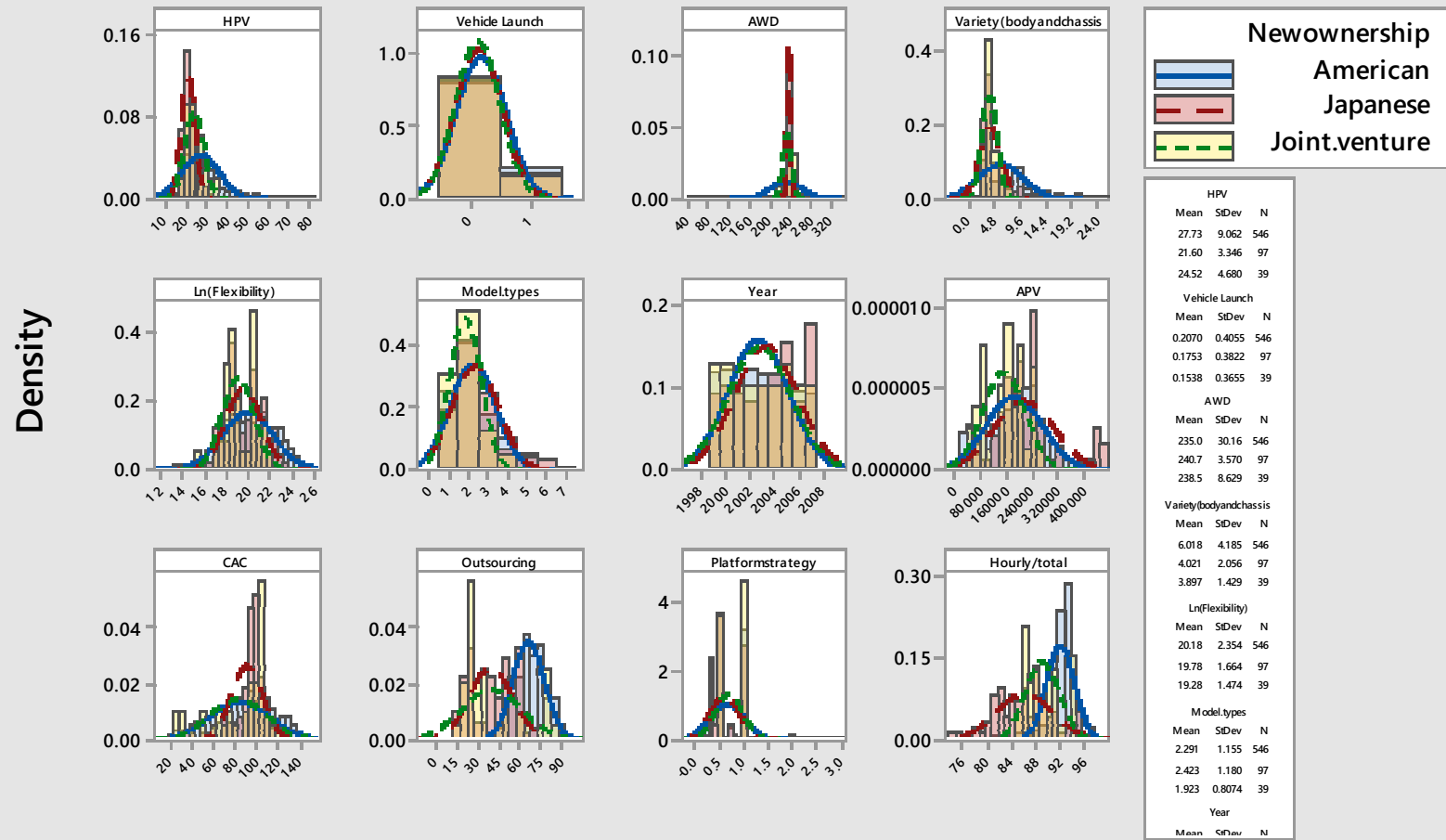


Figure AI- Density function for all variables versus different ownerships

**Table AII – Pearson correlation coefficients and their significance level**

Corr. Coeff.	Vehicle Launch	Model types	Year	HPV	AWD	Variety body and chassis	Flexibility	APV	CAC	Outsourcing	Platform strategy	Hourly Total
Vehicle Launch	1	0.14	-0.03	0.24	0.05	-0.01	-0.04	-0.07	-0.08	0.14	0.04	0.01
Model types	0.14	1	0.1	0.02	0.05	0.02	-0.09	0.1	0	0.15	0.05	0.06
Year	-0.03	0.1	1	-0.3	-0.08	-0.07	-0.08	-0.03	-0.03	0.02	-0.07	-0.01
HPV	0.24	0.02	-0.3	1	-0.04	0.25	0.12	-0.53	-0.4	0.11	-0.01	-0.03
AWD	0.05	0.05	-0.08	-0.04	1	0.15	0.04	0.33	0.35	-0.09	-0.06	-0.02
Variety body and chassis	-0.01	0.02	-0.07	0.25	0.15	1	0.73	-0.02	0.15	-0.05	-0.71	0.15
Flexibility	-0.04	-0.09	-0.08	0.12	0.04	0.73	1	-0.05	0.1	-0.06	-0.47	0.1
APV	-0.07	0.1	-0.03	-0.53	0.33	-0.02	-0.05	1	0.75	-0.03	-0.07	0.18
CAC	-0.08	0	-0.03	-0.4	0.35	0.15	0.1	0.75	1	-0.07	-0.21	0.13
Outsourcing	0.14	0.15	0.02	0.11	-0.09	-0.05	-0.06	-0.03	-0.07	1	0.02	0.36
Platform strategy	0.04	0.05	-0.07	-0.01	-0.06	-0.71	-0.47	-0.07	-0.21	0.02	1	-0.17
Hourly total	0.01	0.06	-0.01	-0.03	-0.02	0.15	0.1	0.18	0.13	0.36	-0.17	1

P	Vehicle Launch	Model types	Year	HPV	AWD	Variety body and chassis	Flexibility	APV	CAC	Outsourcing	Platform strategy	Hourly total
Vehicle Launch	-	0.0004	0.4453	0	0.226	0.7277	0.3139	0.077	0.05	0.0003	0.26	0.7892
Model types	0.0004	-	0.0118	0.6662	0.1999	0.6478	0.0189	0.006	0.968	0.0001	0.2263	0.1001
Year	0.4453	0.0118	-	0	0.0452	0.0877	0.0282	0.414	0.42	0.5931	0.0519	0.843
HPV	0	0.6662	0	-	0.3457	0	0.0021	0	0	0.0028	0.771	0.3819
AWD	0.226	0.1999	0.0452	0.3457	-	0.0001	0.3456	0	0	0.024	0.112	0.5207
Variety body and chassis	0.7277	0.6478	0.0877	0	0.0001	-	0	0.628	1E-04	0.2268	0	0
Flexibility	0.3139	0.0189	0.0282	0.0021	0.3456	0	-	0.208	0.012	0.1461	0	0.0131
APV	0.0774	0.0064	0.4138	0	0	0.6284	0.2083	-	0	0.4126	0.0693	0
CAC	0.0502	0.9675	0.4197	0	0	0.0001	0.0118	0	-	0.0577	0	0.0008
Outsourcing	0.0003	0.0001	0.5931	0.0028	0.024	0.2268	0.1461	0.413	0.058	-	0.5968	0
Platform strategy	0.26	0.2263	0.0519	0.771	0.112	0	0	0.069	0	0.5968	-	0
Hourly total	0.7892	0.1001	0.843	0.3819	0.5207	0	0.0131	0	8E-04	0	0	-

**Table AIII- The developed statistical model without the car segment, with 10 folds cross validation**

Variables	NOF=10
<b>(Intercept)</b>	1.41E+03
<b>AWD</b>	4.81E-02
<b>Variety.bodyandchassis</b>	4.87E-01
<b>Flexibility</b>	-5.66E-01
<b>Model.types</b>	1.32E+00
<b>Year</b>	-6.56E-01
<b>ln(APV)</b>	-6.69E+00
<b>ln(CAC)</b>	1.52E+00
<b>Outsourcing</b>	1.95E-03
<b>Platformstrategy</b>	1.98E+00
<b>Hourly.total</b>	5.74E-04
<b>NewownershipJapanese</b>	-2.60E+00
<b>NewownershipJoint.venture</b>	-2.21E+00
<b>Vehicle.Launch</b>	.
<b>Platformstrategy:Vehicle.Launch</b>	5.31E-01
<b>ln(Flexibility):Vehicle.Launch</b>	-3.20E-01
<b>Hourly.total:Vehicle.Launch</b>	9.23E-02
<b>Adjusted R-squared</b>	0.63
<b>Cross Validation</b>	4.18

**Table AIV – Ranges for HPV and the other independent variables**

Variable	Values range		
<b>HPV</b>	(13.57, 38.8)	(17.5, 82.2)	(21.5, 69.3)
<b>AWD</b>	(35, 301)	(91, 336)	(78, 248)
<b>Vehicle Variety</b>	(1, 14)	(1, 25)	(2, 12)
<b>Flexibility</b>	(13.8, 24.26)	(13.33, 25.8)	(11.82, 22.56)
<b>Model types</b>	(1, 5)	(1, 7)	(0.008, 0.444)
<b>LN(APV)</b>	(8.83, 13.03)	(9.31, 12.96)	(7.91, 12.39)
<b>CAC</b>	(13, 148)	(11, 150)	(20, 114)
<b>LN(CAC)</b>	(2.56, 5.00)	(2.4, 5.01)	(3.0, 4.74)
<b>year</b>	(1999, 2007)	(1999, 2007)	(1999, 2007)
<b>Outsourcing</b>	(12.9, 100)	(21.6, 96.2)	(50, 87.8)
<b>Platform strategy</b>	(0.2, 1.5)	(0.14, 2.0)	(0.25, 3.0)
<b>Hourly total</b>	(74.1, 96.7)	(80.9, 99.4)	(78.6, 95.5)
<b>Vehicle Launch</b>	0, 1	0, 1	0, 1
<b>Ownership</b>	American, Japanese Joint venture	American, Japanese, Joint venture	American, Japanese, Joint venture
<b>Segment</b>	A	B	C



**Table AV - Desired factors values for HPV of 10H (when there was no product launch) – MIDACO Algorithm**

Variable	Values range based on the available data during the study period **	Obtained Setting ***
HPV	(13.57, 38.8)	10
AWD*	(35, 301)	240
Vehicle Variety	(1, 14)	3
Flexibility	(13.8, 24.26)	23.4768
Model types	(1, 5)	2
LN(APV)	(8.83, 13.03)	13.0004
LN(CAC)	(2.56, 5)	3.1445
Year	(1999, 2007)	2007
Outsourcing	(12.9, 100)	96.6392
Platformstrategy	(0.2, 1.5)	1.4999
Hourly.total	(74.1, 96.7)	80.3506
Ln(APV):Segment	-	-
Platformstrategy:Vehicle.Launch	-	-
Flexibility:Vehicle.Launch	-	-
Hourly.total:Vehicle.Launch	-	-
Vehicle Launch	0, 1	1
Ownership	(American, Japanese, Joint venture)	American
Segment	A	A
Response	Fit	Mean Squared Error (Z)
HPV	10	1.8562103e-021

\*It was assumed that a plant was operating for 240 days

\*\*The variables range for segments A, B, and C are in Table AIV in the Appendix.

\*\*\* The optimization solver was run for the 100 iterations in 139.847 seconds