Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

Car demand modelling in the Czech Republic

Author: **Adam Rückl**

Supervisor: **Ing. Barbara Pertold-Gebická Ph.D.**

Academic Year: **2013/2014**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 15, 2014

Signature

Acknowledgments

I would like to express my gratitude to Ing. Barbara Pertold-Gebická Ph.D. for valuable comments, ideas and suggestions during the supervision of my thesis. My acknowledgement also belongs to my former supervisor Ing. Zdeněk Hrubý CSc., who unfortunately is no longer with us.

Finally, I thank my family and beloved for supporting me all the time.

Bibliographic record

Rückl A. (2014). *Car demand modelling in the Czech Republic*. Bachelor thesis. Charles University in Prague.

Character count: 87152

Abstract

The main purpose of this thesis is to construct a model of passenger cars demand and estimate it using data on new car registrations from the Czech Republic. The constructed model takes into account factors affecting not only the demand, but also the supply, as both of these variables form a simultaneous equations system. In order to estimate the model consistently, the econometric theory of simultaneous equations model is presented. At first the basic notations are introduced, the inconsistency of ordinary least squares estimates is proved and consistent methods of estimation are described, notably the two stage least squares and the indirect least squares methods.
Observation of the data suggests that new passenger cars registrations in the Czech Republic were influenced by a specific taxation policy, which up to April 2009 did not allow the value added tax deduction in case of passenger cars. A large proportion of passenger cars were therefore registered as light utility vehicles. This fact has to be taken into account when studying the passenger cars demand. Results of the estimation by the two stage least squares method showed that the demand for new passenger in the Czech Republic is elastic in price and income. Significant sensitivity of the demand was observed also with respect to used cars price, which act as substitutes to new cars. In case of the fuel price, no relevant effect on new cars sold was detected. The highest impact is attributed to seasonality, which probably acts as a demonstration of consumers´ subjective preferences.

JEL Classification	C01, C30, D11, D12, L62, R41
Keywords	demand modelling, passenger cars, regression analysis, simultaneous equations model, two stage least squares method

Author's e-mail	ruckladam@volny.cz
Supervisor's e-mail	gebicka@fsv.cuni.cz

Abstrakt

Cílem této práce je vytvořit a s použitím dat registrací nových vozidel v České republice odhadnout model poptávky po osobních automobilech. Sestavený model bere zvlášť v potaz faktory ovlivňující nejen poptávku, ale i nabídku, přičemž obě tyto veličiny tvoří systém simultánních rovnic. Za účelem správného odhadu je prezentována ekonometrická teorie modelu simultánních rovnic. Nejprve jsou zavedeny základní pojmy, je dokázána nekonzistence metody nejmenší čtverců a představeny vhodné metody odhadů, zejména pak dvoustupňová metoda nejmenších čtverců a nepřímá metoda nejmenších čtverců. Při následné empirické analýze bylo zjištěno, že registrace nových osobních automobilů v České republice jsou ovlivněny specifickou daňovou politikou, která do dubna 2009 neumožňovala odpočet daně z přidané hodnoty u osobních automobilů. Velká část osobních automobilů byla tedy registrována jako lehké užitkové vozy a tento fakt je při studiu poptávky po automobilech nutné vzít v potaz. Výsledky regrese pomocí dvoustupňové metody nejmenších čtverců ukázaly, že poptávka po nových automobilech v České Republice je cenově a důchodově elastická, značná citlivost poptávky byla zjištěna i na cenu ojetých automobilů jakožto jejich substitutu. Naopak u ceny paliva nebyl prokázán vliv na počet kupovaných automobilů. Největší dopad je připisován sezónnosti, která je pravděpodobně výsledkem subjektivních preferencí spotřebitelů.

Klasifikace	C01, C30, D11, D12, L62, R41
Klíčová slova	modelování poptávky, osobní automobily, regresní analýza, model simultánních rovnic, dvoustupňová metoda nejmenších čtverců

E-mail autora	ruckladam@volny.cz
E-mail vedoucího práce	gebicka@fsv.cuni.cz

Contents

List of Tables

List of Figures

Acronyms

SEM	Simultaneous equations model
OLS	Ordinary Least Squares
2SLS	Two Stage Least Squares
CPI	Consumer price index
PPI	Producer price index
CSO	Czech Statistical Office
CNB	Czech National Bank
CIA	Car Importers Association
ACEA	European Automobile Manufacturers Association
CVR	Central Vehicle Register
AIA	Automotive Industry Association
LUV	Light utility vehicles
VAT	Value added tax
AR(1)	Autoregressive process of order one

Bachelor Thesis Proposal

Author	Adam Rückl
Supervisor	Ing. Barbara Pertold-Gebická Ph.D.
Proposed topic	Car demand modelling in the Czech Republic

Topic characteristics

During the twentieth century, the passenger car market has developed from a small, luxury oriented business to one of the most important branches in the economics. Car manufacturers produce millions of cars every year offering a great number of working posts and providing income for governments by taxes. Changes in the car demand can deeply influence the whole economy of the country as we have seen recently during the world economic crisis. That definitely holds also for the Czech Republic, a quite small, car-production oriented country. It would be very convenient to determine what influences the new cars demand. With this information, we could face negative demand deviations more efficiently and minimize the negative effects.

In my thesis, I study various models estimating the car demand. In the next step, factors influencing the supply and demand are presented and the methodology of estimation are derived. A suitable dataset for the Czech Republic environment is analysed and finally, I estimate the model by an appropriate econometric method.

Core bibliography

Carlson, R. (1978). Seemingly Unrelated Regression and the Demand for Automobiles of Different Sizes, 1965-75. *The Journal of Business*, vol. 51, issue 2, pp. 243-62.

Jansson, J. (1989). Car Demand Modelling and Forecasting. *Journal of Transport Economics and Policy*, vol. 23, No. 2, pp. 125-140

Mikas, T. (2011). *Faktory ovlivňující poptávku po osobních automobilech v ČR*. Diploma thesis. Pardubice: University of Pardubice, Jan Perner Transport Faculty.

Samuelson, P. (2005). *Economics*. 18th ed. Boston: McGraw-Hill/Irwin.

Thomassen, O. (2007). *A Pure Characteristics Model of Demand for Cars*, University of Oxford.

Witt, S. & Johnson, R. (1986). An econometric model of new-car demand in the UK. *Managerial and Decision Economics*, vol. 7, issue 1, pp. 19-23.

Wooldridge, J. (2009). *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, OH: South Western, Cengage Learning.

Author Supervisor

1 Introduction

During the 20th century passenger cars became one of the most common means of transport. High level of flexibility and comfort quickly overcame negative aspects and most of people nowadays prefer passenger cars to their available alternatives. With a massive extension of cars all over the world, especially in more developed countries, completely new markets have emerged. Thanks to the fact that a car has always been a complex and expensive product, its production, distribution and servicing involve great amount of labour and capital. In the Czech Republic, rather small, industrial and export oriented European country, the automotive industry plays a leading role in its economy for two main reasons. At first, expenditures on new cars represent substantial part of aggregate expenditures and therefore the condition of the car market may significantly influence the whole economy. The second reason is that automotive industry companies, which either produce cars or its parts substantially contribute to the Czech employment and industrial production. According to Automotive Industry Association (2014) about 110 000 people of various specializations are directly employed in Czech automotive companies and in 2012 the Czech automotive industry accounted for approximately 20 % of the whole Czech industrial production. In terms of car production, the Czech Republic belongs to the world leaders. With 1 128 473 cars produced in 2013, it achieved the 13th place in the world by the produced volume and when comparing cars per capita production, with 107 cars per 1000 inhabitants the Czech Republic stands at the top, competing with its neighbour Slovakia. Just from these basic indicators the economic importance of automotive industry for the Czech Republic is obvious. The special position is also stressed by the interests of governments in their interventions, especially during the period of unfavourable economic conditions.

With respect to these facts, it is important to identify factors which determine the number of new cars sold within the country and analyse the elasticity of demand with repsect to these factors. The obtained information could serve to car producers and to the government to interpret the market situation, to predict the future development, and design policies supporting the automotive market. Although quite a lot of research was done on this topic, these works are usually concentrating on one particular country reflecting its unique features. The author of this work is of the opinion that because of different habits and economy situation in each country these models cannot be simply transferred to the Czech environment and a unique model for the Czech Republic has to be found. Unfortunately, the Czech Republic stood apart of the researchers interest and only one actual work on this topic was produced. It is a diploma thesis of Tomáš Mikas (2011). But although Mikas (2011) uses a solid methodology, he does not achieve any satisfactory results.

Therefore the objective of this thesis is to propose and estimate a model which explains the new passenger car demand in the Czech Republic by means of economic indicators.

The first section provides the literature review of published papers which are focused on the car demand estimation. The most important ideas and results of each article are stressed. The second section introduces the theoretical background of the model used for estimation. The expected effect of chosen variables according to the microeconomics theory and empirical findings is analysed. The methodology of the new car demand estimation using the Simultaneous Equations Model is derived in the next, third section. The fourth section is dedicated to the data description and analysis, introducing some unique concepts typical to the Czech market environment. The models of the new cars demand in the Czech Republic are estimated. A final section summarizes the research and points out the most important implications.

2 Literature review

2.1 Review

Since the estimation of car demand has attracted researchers´ interest for decades, there are many papers written on this topic. One of the first modern researchers who dedicated their interest to car demand analysis was Daniel B. Suits (1958). In 1958 he published a paper studying the demand for new cars in the United States observing years 1929-1956. Suits estimated a model where the real disposable income and the stock of cars played a major role.

A much more complex approach presented Rodney L. Carlson (1978) approximately 20 years later. In his work called Seemingly Unrelated Regression and the Demand for Automobiles of Different Sizes he developed a theory for estimating new cars demand with a disaggregate approach. He observed car sales of different sizes in the United States over years 1965-1975 and estimated a model for each car class. Since in case of estimating a set of regression equations whose disturbances are correlated the OLS estimator is not efficient, Carlson estimated the equations by Zellner´s two-stage Aitken method based on the generalized least squares and compared the outcome with the OLS estimates. The results of the model showed that the demand for cars of lower classes (subcompact) is price-inelastic. On the other hand, cars of higher classes, which are usually more expensive tend to be price-elastic. The highest elasticity was observed in case of the full size class, where the price increased by 1 % causes the decrease in demand by 2.07 %. Interesting is that the demand for luxury cars is again price-inelastic, an increase of the price by 1 % lowers the demand only by 0.347 %. That is less than half that the demand for subcompact cars would. Increase in the gasoline price causes the demand for subcompact and compact class to be higher and lowers the demand for cars of higher classes. According to Carlson, this can be explained in the way that people tend to buy smaller cars with lower fuel consumption in order to diminish their expenditures. As expected, the increased purchasing power increases the demand for cars. The size of this impact is highest in case of the subcompact class, income increased by 1 % results in increase of the demand by 3.259 %. The income elasticity decreases with larger and more expensive cars, for example luxury cars respond to 1% positive change in income by increase of the demand only by 0.839 %. According to the Carlson´s research, passenger cars behave like normal goods and an increase of their prices causes the car demand to be lower.

Stephen. F. Witt and Raymond Johnson (1986) published a research paper An Econometric Model of New-car Demand in the UK in which they were estimating new car registrations using the data for years 1961-1981. Using the method of OLS they estimated the model where the logarithmic form of new-car registrations per capita served as a dependent variable representing the demand. In the Witt and Johnson´s model, the real price of motor fuel turned out to have no important impact and it was excluded from the model. They did not use the disaggregate approach to estimate effects for different car sizes, so the effect of small vehicles preference caused by the price increase could not be captured. Demand for new cars was estimated to be price-inelastic, when 1% increase in price results in a 0.341% decrease in the car demand per capita. High impact on the dependent variable was recorded in case of the real personal disposable income. Its increase by 1 % results in increase of the demand by 1.947 %.

An interesting idea of the model is including an interest rate and a minimum percentage deposit as conditions for financing of a new car. 1% increase in the cost of borrowing causes a decrease in the demand by 0.436 %. Also the minimum percentage deposit has impact on the demand, 1% increase is followed by a decrease of the demand by 0.105 %. The regression has also shown negative influence of the oil crisis and a strong dependence of new registrations on the number of registrations in the previous year, since the coefficient for lnN_{t-1} was estimated to be 0.404. Authors explain this as a habit persistence in the buyers´ behaviour when the most of new registrations in UK in 70´s were found to be replacement purchases.

In the last three decades, researchers adopted more comprehensive approaches of studying the demand, desiring to reveal a complicated process of consumer decision making. Jansson (1988) build a demand model based on the idea of car ownership entry- and exit propensity. McCarthy (1996) focused on studying market price and income elasticity, using a logit model with a wide range of variables representing a vehicle style, qualities and reputation. A paper estimating the car demand in Norway using advanced methods was presented by Thomassen (2007). He incorporated an observed effect of vehicle characteristics preferences to estimate demand models for various types of vehicles.

Although the car demand estimation became a rather complicated issue in case of certain papers, simple models capturing the effect of economic key performance indicators did not lose popularity. For example, a recent work focused on estimating a simple car demand model was published by Hui-Yen Lee and Hsin-Hong Kang (2008). They studied the car demand in an emerging market of Brazil over the years 2001 to 2006. They used the method of Ordinary Least Squares to estimate their model. Lee and Kang observed strong income elasticity of the car demand in Brazil – if the GDP (proxy for income) increases by 1 %, the demand will increase by 1.821668 %. The cars are supposed to be luxury goods in Brazil. Another finding is that a 1% increase in the lending rate variable results in 0.319251% decrease in the quantity demanded. The price of fuels did not reveal any important impact, the change of the demand caused by a 1% increase of the fuel price is positive, 0.006477 %.

Probably the most actual work modelling the car demand in the Czech Republic is the diploma thesis of Tomáš Mikas from 2011. In this work, he analysed the passenger cars demand in the Czech Republic for years 2007 – 2011. Although Mikas used quite a solid methodology including the time demeaning, detrending and controlling for serial correlation, he did not reach any significant multiple regression linear model. The only result he stated was the simple linear regression model where the number of new car registrations served as a dependent variable and the average monthly wage, time trend and seasonal effects were the explanatory variables.

We found a mistake in the data collection which could cause Mikas´s results to be biased and could be also responsible for not reaching any satisfactory outcome. The problem is associated with a Czech car market unique specification. Before April 2009, there was no possibility to subtract the value added tax (VAT) for passenger cars (category M1) used for commercial purpose. Nevertheless, one could transform his car into a Light Utility Vehicle (category N1) and subtract the VAT from the price. The transformation was very easily provided by installing a counter dividing the trunk and the room for passengers. This led to a curious situation when number of passenger cars including also very luxury cars were sold as LUV. In April 2009 the Czech law 235/2004 Sb. treating the VAT was revised and allowed also VAT deduction for M1 category. Because Mikas studied only M1 category and the breakpoint was approximately in the middle of the period he studied, the data could be affected by this policy change. The correctness of the hypothesis and the way how to avoid this problem is studied in the section dedicated to methodology (Portal of the Public Administration, 2014).

2.2 Summary

Although the results of discussed studies cannot be directly compared because each study differs by the econometric model which was used, country which was observed and the time period the study was performed, the general outcome can be summarised as follows.

It was observed that new passenger cars are normal goods, the quantity demanded for cars is usually positively correlated with a consumer´s income or with another variable which can represent the wealth of population, for example a GDP. In terms of the income elasticity, new cars are luxury goods, e.g. they are income elastic. The only category that was found to behave like income inelastic goods were the luxury cars, whose buyers represent a specific group of population.

In works where the price of new passenger cars was included in the model the quantity demanded for cars decreased with their price, therefore, we can say that cars are ordinary goods.

In the studies which estimated an aggregate demand, i.e. all classes of cars, the fuel price seemed to be uncorrelated with the quantity demanded, in some cases the correlation was even positive. This was probably caused by the fact that the fuel price did not increase enough to overwhelm another effects on the demand so far. A possible explanation brings a disaggregate approach performed by Carlson (1978) or Thomassen (2007). According to Carlson (1978), when the price of the fuel increases, people tend to buy smaller cars with lower fuel consumption to diminish their operational costs. But the number of cars sold is relatively unchanged and therefore this effect cannot be captured by the aggregate models.

Although a lot of advanced and comprehensive methods of modelling were applied so far, none of the above mentioned studies takes into account the supply of cars as a force participating on the traded quantity. Therefore, one of the objectives of this work is to present a new cars demand model involving both supply and demand to fill this gap in the field of research.

3 Theoretical model

The aim of the chapter is to determine factors which are involved in explaining the supply and demand of new passenger cars in the Czech Republic and could be used as explanatory variables in the econometric model. At first, determinants of the supply and demand are presented separately using the economics theory. The choice of variables is inspired also by previous studies, especially by Carlson (1978), Witt and Johnson (1986) and Mikas (2011). Then, an equilibrium as a result of the supply and demand interaction is studied.

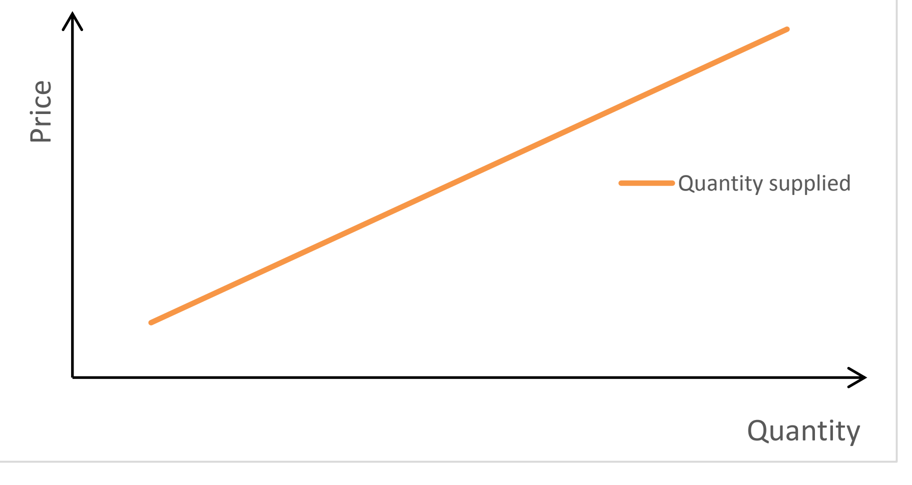
3.1 New cars supply determinants

The new cars price, producer´s costs and the exchange rate have been taken into account as variables significantly affecting the new cars supply.

3.1.1 New cars price

Price is probably the most important factor which determines the quantity supplied. If we keep all other factors than price and quantity fixed, we obtain a relationship between the price and the quantity traded. Under the basic microeconomic assumptions about the behaviour of economic agents, the demand is usually represented by an upward sloping curve in the quantity-price space. Therefore *ceteris paribus* the higher is the price, the larger amount of goods are the firms willing to sell. This fact is called *the law of increasing supply* (Samuelson, 2005).

Figure 3.1: Supply curve



Source: author's computations

3.1.2 Producer´s costs

According to the microeconomics theory, each firm in the market maximizes its profit Π. For $TR(q)$ being a total revenue function of quantity produced (q) and $TC(q)$ being a total cost function of q, the profit function as a function of q is given as:

$$\Pi(q) = TR(q) - TC(q) \tag{3.1}$$

Taking the partial derivative with respect to q we obtain:

$$\frac{\partial \Pi(q)}{q} = MR(q) - MC(q) \tag{3.2}$$

Where MR(q) is a marginal revenue function and MC(q) represents a marginal cost function. It is usually assumed that the marginal cost function is increasing with respect to q and the marginal revenue function is nondecreasing with respect to q.

If we assume that the profit function is concave and at least for some q nonnegative, the first order condition for maximizing the profit is:

$$\frac{\partial \Pi(q)}{q} = 0 \tag{3.3}$$

And therefore:

$$MR(q) = MC(q) \tag{3.4}$$

Now, if the producer costs per a unit produced increase, *ceteris paribus* the marginal costs increase as well (the $MC(q)$ curve shifts up) whereas the marginal revenue stays unchanged. Because the marginal cost is assumed to be increasing with q, in this situation, the producer lowers the quantity produced to preserve the profit-optimal relation $MR(q) = MC(q)$. The supply decreases (the supply curve in the quantity-price space shifts down). In an opposite case, when the producer´s costs decrease, using the same logic the supply increases and the supply curve shifts up (Koubek, 2014).

3.1.3 Exchange rate

Foreign producers usually sell the imported goods for a domestic currency and then exchange their revenues in the foreign currency. Because the competitive market prices are up to some extent rigid in the short run, foreign producers cannot change the price appropriately when the exchange rate of the domestic currency increases, in other words when the domestic currency depreciates. That would lead to abandoning the market position to domestic producers or another producers from countries which are not influenced by the exchange rate change. Therefore, instead of increasing the prices they sell goods for the same or only slightly higher domestic price and obtain less for their goods. The effect is similar as if the price decreased. The willingness of foreign producers to sell cars in such a market is lower, the supply decreases and the supply curve in the quantity-price space shifts down. Naturally, this relation works also in an opposite direction. If the exchange rate decreases, the supply curve shifts up.

Since approximately 60 % of the new passenger cars sold within the Czech Republic are imported from foreign economies, the exchange rate is expected to be an important variable in the supply and demand model.

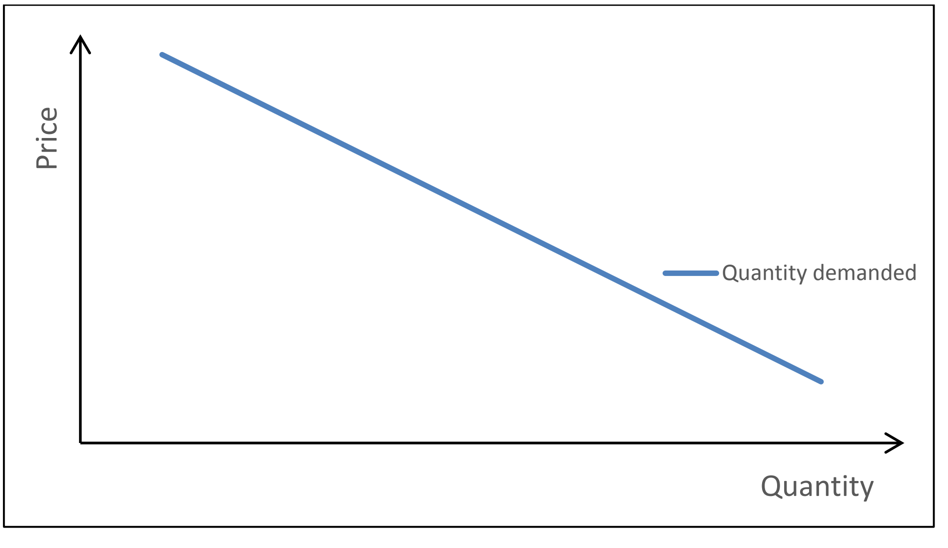
3.2 New cars demand determinants

For variables which could have a significant effect on the new cars demand, the new cars price, used cars price, consumer´s income, fuel price, population, market saturation and "consumer taste" have been chosen.

3.2.1 New cars price

Similarly as in case of the supply analysis, keeping all other factors than price and quantity fixed, we obtain a relationship between the price and the quantity traded. Under the basic microeconomic assumptions about the behaviour of economic agents, the demand is usually represented by a downward sloping curve in the quantity-price space. This means that *ceteris paribus* the higher is the price, the smaller amount of goods is the consumer demanding. It is called the *law of decreasing demand* (Samuelson, 2005).

Figure 3.2: Demand curve



Source: author's computations

According to the way how the demand Q^Dreacts to change in price p the microeconomics theory distinguishes two general types of goods.

If $\frac{\Delta Q^D}{\Delta p} < 0$, i.e. if the demand curve is downward sloping, we talk about an *ordinary good.*

In case that $\frac{\Delta Q^D}{\Delta p} > 0$, i.e. if the demand curve is upward sloping, we talk about a *Giffen good.*

The intensity the quantity demanded responds to a change in price with is measured by the price elasticity of demand.

> Price elasticity of demand:
> *"Price elasticity of demand e_p^i for i^{th} good is defined as a percentage change of demand $Q^D(p)$ caused by price change by one percent."* (Rees, 2004)
>
> $$e_p^i = \frac{\frac{\Delta Q^D(p)}{\Delta p}}{\frac{Q^D(p)}{p}}$$

If $e_p^i < -1$, i.e. the percentage change of demand is higher than the percentage change of price, we say that the good is price elastic.

If $-1 < e_p^i < 0$, i.e. the percentage change of demand is lower than the percentage change of price, we say that the good is price inelastic.

Cars are expected to be ordinary goods and therefore, the quantity demanded should decrease with price. Since consumers have a possibility to choose a very close substitute goods to new cars, which are used cars or public transport, the demand for new cars should tend to be price elastic.

3.2.2 Consumer Income

Also the consumer's income affects their demand for goods. According to the microeconomic theory, a consumer makes his decision about the demanded quantity within the budget constraint which is given by his disposable income.

Two types of goods according to the demand Q^D reaction to changes in consumer´s income M are distinguished:

If $\frac{\Delta Q^D}{\Delta M} < 0$, i.e. the demand for good decreases with increasing income, we say it is the *inferior good.*

If $\frac{\Delta Q^D}{\Delta M} > 0$, i.e. the demand for good increases with income, we say it is the *normal good*.

Another classification of goods is based on the income elasticity of demand.

Income elasticity of demand:

"Income elasticity of demand e_M^i is a percentage change of a demand for i^{th} good $Q^D(p, M)$ caused by a change of consumer income M by one percent." (Rees, 2004)

$$e_M^i = \frac{\frac{\Delta Q^D(p,M)}{\Delta M}}{\frac{Q^D(p,M)}{M}}$$

In case that $e_M^i > 1$, i.e. the percentage change of demand is higher than the percentage change of income, we talk about a luxury good.

In case that $0 < e_M^i < 1$, the percentage change of demand is smaller than the percentage change of income, we talk about a necessity good.

In terms of the microeconomic theory, cars are also definitely normal goods, therefore, it is expected that an increase of the consumer´s income will have a positive effect on the car demand. Using a real life experience cars are still today rather luxury goods with a cheaper and quickly available complement of public transport. That means the demand should be income-elastic, e.g. 1% increase in consumer income will cause an increase of the demand by more than 1 %.

3.2.3 Used cars price

A car is a long term consumption good. Therefore, there also exists a market with used cars. In general, a consumer who is decided to buy a car is facing two options – to buy a new car or a used one. Therefore, used cars can be treated as substitutes to new cars. The relationship between the demand for new cars and the price of their substitute, used cars, can be expressed in terms of cross price elasticity of demand.

Cross-price elasticity of demand:

"Cross-price elasticity of demand $e_{p_i}^{j}$ *is a percentage change of a demand for* j^{th} *good* $Q_j^D(p, M)$ *caused by a change of* i^{th} *good price* p_i *by one percent.*

$$e_{p_i}^{j} = \frac{\frac{\Delta Q_{j\ j}^{D}(p, M)}{\Delta p_i}}{\frac{Q_j^D(p, M)}{p_i}}$$

For $e_{p_i}^{j} > 0$ *the* j^{th} *good is called to be a gross substitute of the* i^{th} *good. For* $e_{p_i}^{j} < 0$ *the* j^{th} *good is called to be a gross complement of the* i^{th} *good." (Rees, 2004)*

Because used cars are expected to be a complementary good to new cars, used cars price should have a positive effect on new cars demand. It is important to be very careful when interpreting the causality of new and used cars prices. Up to a large extent, these variables are being mutually influenced by each other, but they cause contrary effects on the demand of new cars and neither of them cannot be omitted.

3.2.4 Fuel price

According to the above mentioned theory of supplementary and complementary goods, fuel is probably the most important complement of cars. It is an irreplaceable part of a car usage because cars consume fuel in order to be operational. Price of fuel used for cars is, therefore, expected to influence new car demand and with rising fuel price the car demand should be lower. According to Carlson (1978), effects of fuel price change are more evident when studying the demand for particular car classes. For aggregated data, to force a significant portion of car owners to unconditionally give up the car possession, the fuel price increase would have to be probably much higher than it was ever observed. This assumption confirm also results of Lee and Kang (2008) and Witt and Johnson (1986).

3.2.5 Market saturation

Within the Czech market, the aggregate demand is limited not only by the economic performance of the country but also by the population. Using a simple logic the more people are located in the marked the higher the aggregate demand is. As the consumer utility is a concave function, the marginal effect from buying more goods (cars) is diminishing. In other words, the more cars people already have the less they demand. Therefore, capturing the population effect and the market saturation is important for estimating and predicting the aggregate demand. Naturally, this theory assumes that each car is of the same quality, but in the real world people might have intentions to change their car for a newer, more luxurious one, etc.

3.2.6 Consumer taste

Mikas (2011) in his work mentioned a factor of "a consumer taste". Car brand, design, specific preferences or even time of the year may explain a lot of consumer behaviour. We are aware of the fact that it is very difficult to specify the consumer taste through the objective indicators. Since the scope of this work is to estimate the demand using the economic factors, the factors of consumer preferences will not be subject of the analysis. The only exception are probably the seasonal effects, which are well observable and as preliminary studies proved, the car demand shows a high variance for particular months.

3.3 Supply and demand interaction

According to the microeconomic theory, the number of new cars sold within the market is a result of meeting aggregate demand with aggregate supply. The demand is represented by consumers wishing to buy cars and the supply is represented by firms which intend to sell cars. These are the two groups (powers) which interact on the particular market. Together they set the quantity and price at which the cars are traded. (Samuelson, 2005) Let us denote the quantity demanded as Q^D and quantity supplied as Q^S. If we relate these quantities linearly to their respective explanatory factors, then

$$Q^D = a_0 + a_1 p + \sum_{i=1}^{n} a_{i+1} D_i$$
$$Q^S = b_0 + b_1 p + \sum_{j=1}^{m} b_{j+1} S_j \tag{3.5}$$

where a_0 and b_0 are the intercepts, p is the price and D_i, S_j are other explanatory factors while a_{i+1},b_{j+1} represent their corresponding coefficients.

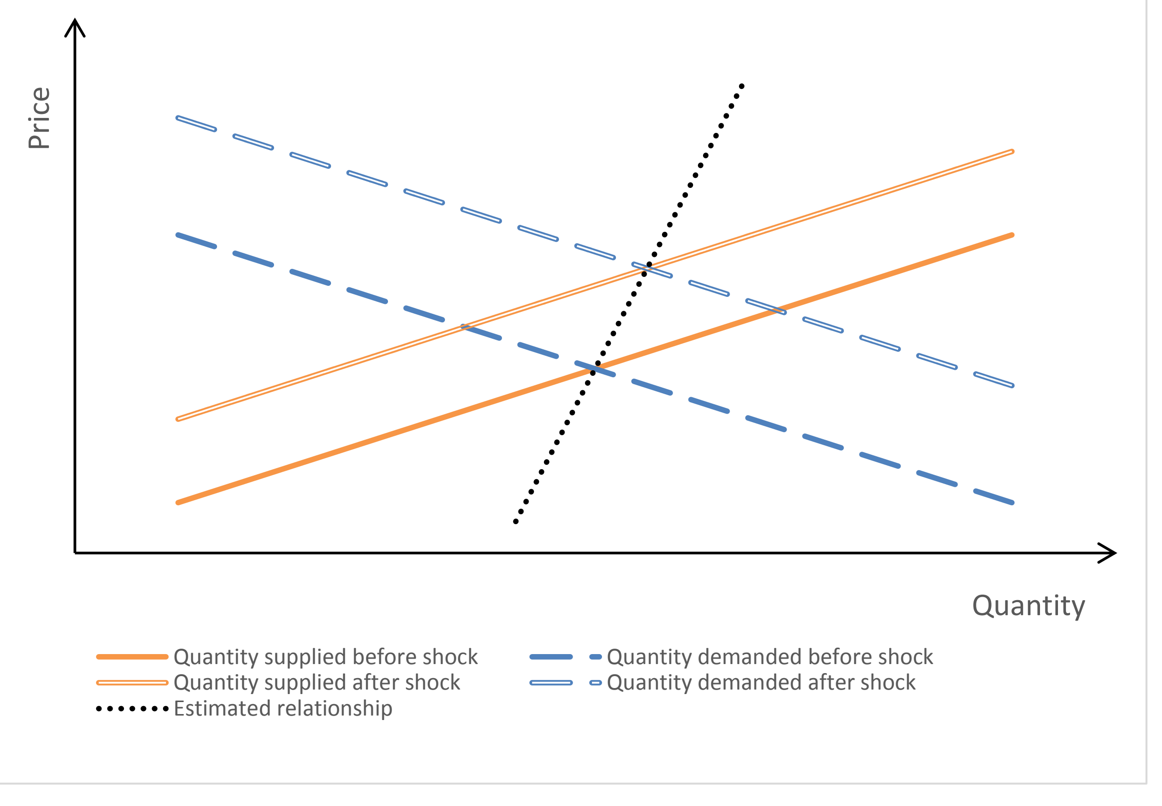
We observe the traded quantity for which the equilibrium condition holds:

$$Q^D = Q^S \tag{3.6}$$

The main problem we face when trying to estimate the demand curve or supply curve separately is that the demand and supply are determined simultaneously by many different factors and the observed quantity is a result of interaction of all these forces. That means that we do not observe the isolated demand curve or the supply curve but the points determined by intersections of these two curves as they shift up and down. Therefore, the relationship between the demand for cars and the traded quantity is implicit (Wooldridge, 2009).

For example, let us imagine a case when a traded quantity of some good is observed in two time periods in order to estimate the demand. In the first one, we obtain the traded quantity as a market equilibrium, i.e. a point, where the demand meets supply. Now suppose that in the second period the demand is affected by a positive shock and the supply is affected negatively. Both supply and demand curve shift up and the new equilibrium is observed. But if we study only exogenous factors affecting the demand, we do not know anything about how the supply curve shifted. Therefore, we cannot identify the demand equation and the estimated relationship will be biased. The situation is illustrated in the graph.

Figure 3.3: Supply and demand interaction



Source: author's computations

To estimate the effect of price on Q^D efficiently, the interaction between the supply and demand has to be taken into account and a simultaneous equations model (SEM) has to be applied (Wooldridge, 2009).

4 Econometric approach

In this chapter, the theory of simultaneous equations models estimation is being discussed. The simultaneous equations model is represented by a finite set of equations whose common solutions are searched. Using the supply and demand equations from the previous chapter, a model of two simultaneous linear equations which would represent the quantities demanded and supplied can be constructed as follows:

$$Q^D = a_0 + a_1 p + \sum_{i=1}^{n} a_{i+1} D_i + u_1$$
$$Q^S = b_0 + b_1 p + \sum_{j=1}^{m} b_{j+1} S_j + u_2 \tag{4.1}$$

Where Q^S Q^D represent the quantities supplied and demanded respectively, a_0 and b_0 are the intercepts, p is the price and D_i, S_j are other explanatory factors while a_{i+1},b_{j+1} represent their corresponding coefficients. The u_1, u_2 are the random error terms representing other unobserved factors that affect the demand and supply, respectively.

4.1 Structural and reduced form equations

For simplicity, let us now assume there is only one additional explanatory factor in each equation, i.e. $n = m = 1$. Let us call these factors w and z. Now if the equilibrium condition holds, then $Q^D = Q^S$ and we can denote the quantity simply as $q = Q^D = Q^S$. Using a simple computation we derive our model:

$$q = \alpha_0 + \alpha_1 p + \alpha_2 w + u_1 \tag{4.2}$$

$$p = \beta_0 + \beta_1 q + \beta_2 z + u_2 \tag{4.3}$$

where for $k = 1,2,3$: $\alpha_k = a_k$; $\beta_0 = -\frac{b_0}{b_1}$, $\beta_1 = -\frac{1}{b_1}$ and $\beta_2 = -\frac{b_2}{b_1}$. The u_1, u_2 are the random error terms.

We can now directly see that the price and the quantity are determined simultaneously by these equations - we can say that they are endogenous in this model.

Endogenous and exogenous variables:
"Endogenous variables are variables whose values are determined by the interaction of the relationships in the model. Exogenous ones are those whose values are determined externally." *(Dougherty, 2011)*

Therefore, apart from two endogenous variables q and p, our system has also two exogenous variables z and w. The main point of this distinction is that exogenous variables are assumed to be uncorrelated with error terms u_1, u_2 and together with these error terms they determine the endogenous variables. The above stated equations are called structural form equations.

Reduced form equations:
"*The mathematical relationships expressing the endogenous variables in terms of the exogenous variables and disturbance terms are known as the reduced form equations.*" *(Dougherty, 2011)*

The reduced form equations for q and p can be obtained by a substitution. At first for q:

$$q = \alpha_0 + \alpha_1 p + \alpha_2 w + u_1$$

$$= \alpha_0 + \alpha_1(\beta_0 + \beta_1 q + \beta_2 z + u_2) + \alpha_2 w + u_1$$

or

$$q(1 - \alpha_1\beta_1) = \alpha_0 + \alpha_1\beta_0 + \alpha_1\beta_2 z + \alpha_2 w + u_1 + \alpha_1 u_2$$

With assumption that

$$\alpha_1\beta_1 \neq 1 \qquad (4.4)$$

$$q = \frac{\alpha_0 + \alpha_1\beta_0 + \alpha_1\beta_2 z + \alpha_2 w + u_1 + \alpha_1 u_2}{(1 - \alpha_1\beta_1)} \qquad (4.5)$$

According to Wooldridge (2009), this can be rewritten as

$$q = \pi_{10} + \pi_{11}z + \pi_{12}w + v_1 \tag{4.6}$$

where $\pi_{10} = \frac{\alpha_0+\alpha_1\beta_0}{(1-\alpha_1\beta_1)}$, $\pi_{11} = \frac{\alpha_1\beta_2}{(1-\alpha_1\beta_1)}$ and $\pi_{11} = \frac{\alpha_2}{(1-\alpha_1\beta_1)}$ are the reduced form parameters

and $v_1 = \frac{u_1+\alpha_1 u_2}{(1-\alpha_1\beta_1)}$ is the reduced form error.

The same computation for p with assumption that $\alpha_1\beta_1 \neq 1$ leads to

$$p = \frac{\beta_0 + \alpha_0\beta_1 + \alpha_2\beta_1 w + \beta_2 z + u_2 + \beta_1 u_1}{(1-\alpha_1\beta_1)} \tag{4.7}$$

Respectively

$$p = \pi_{20} + \pi_{21}z + \pi_{22}w + v_2 \tag{4.8}$$

where $\pi_{20} = \frac{\beta_0+\alpha_0\beta_1}{(1-\alpha_1\beta_1)}$, $\pi_{21} = \frac{\alpha_2\beta_1}{(1-\alpha_1\beta_1)}$, $\pi_{22} = \frac{\beta_2}{(1-\alpha_1\beta_1)}$ and $v_2 = \frac{u_2+\beta_1 u_1}{(1-\alpha_1\beta_1)}$.

4.2 Simultaneous equations bias

In most of the cases, estimating SEM models by OLS does not fulfil the 3rd Gauss-Markov assumption $E(u|x) = 0$, because the error term is not uncorrelated with explanatory variables (p and q in our case). This generally leads to biased and inconsistent estimates. Dougherty (2011) shows it using the simple model from the previous chapter.

Let us recall the reduced form expressing p (4.8). Assuming that z and w are exogenous and therefore uncorrelated with u_1 and u_2, we are able to consistently estimate the reduced form parameters π_{20}, π_{21} and π_{22}, because the reduced form error v_1 is only a linear function of u_1 and u_2 and hence it is also uncorrelated with z and w. The problem arises when we try to estimate coefficients of the structural equation, α_i

or β_i, respectively, where $i = 0,1,2$. From the following computation we can see that both q and p depend on u_1, u_2.

$$p = \frac{\beta_0 + \alpha_0\beta_1}{(1-\alpha_1\beta_1)} + \frac{\alpha_2\beta_1}{(1-\alpha_1\beta_1)}w + \frac{\beta_2}{(1-\alpha_1\beta_1)}z + v_1 \tag{4.9}$$

Where $v_1 = \frac{u_1+\alpha_1 u_2}{(1-\alpha_1\beta_1)}$

Since z, w are exogenous

$$E(z|u_1,u_2) = z \tag{4.10}$$
$$E(w|u_1,u_2) = w \tag{4.11}$$

Together with assumption that

$$E(u_1|p,w) = 0 \tag{4.12}$$
$$E(u_2|q,z) = 0 \tag{4.13}$$

Taking expectation of p results in

$$E(p) = \frac{\beta_0 + \alpha_0\beta_1}{(1-\alpha_1\beta_1)} + \frac{\alpha_2\beta_1}{(1-\alpha_1\beta_1)}w + \frac{\beta_2}{(1-\alpha_1\beta_1)}z \tag{4.14}$$

Hence

$$p - E(p) = v_1$$

And also

$$u_1 - E(u_1) = u_1$$

Therefore

$$cov(p,u_1) = E[p-E(p)][u_1 - E(u_1)] =$$

$$cov(p,u_1) = E\left[\frac{u_1+\alpha_1 u_2}{(1-\alpha_1\beta_1)}\right][u_1] = E\left[\frac{(u_1)^2 + \alpha_1 u_2 u_1}{1-\alpha_1\beta_1}\right] \tag{4.15}$$

If we denote the variance of u_1 as $\sigma_{u_1}^2$ and assume that disturbances u_1 and u_2 are uncorrelated, we can rewrite the covariance as

$$cov(p,u_1) = \frac{\sigma_{u_1}^2}{1-\alpha_1\beta_1} \tag{4.16}$$

With assumption that $\sigma_{u_1}^2$ is a positive variance of u_1 and that $\alpha_1\beta_1 \neq 1$, we have proved $cov(p,u_1) \neq 0$. In the same way it could be shown that $cov(q,u_2) \neq 0$.

Fact that p depends on u_1 causes breaking the assumption of a regressor distributed independently of the disturbance term. $Cov(p,u_1) \neq 0$ is the reason why estimating

$$q = \alpha_0 + \alpha_1 p + \alpha_2 w + u_1$$

by OLS would lead to biased and inconsistent estimates of α_i. We say that OLS suffers from "simultaneity bias."

In order to derive the sign of the bias, we can divide the estimate of α_1 into the true value and the error term:

$$\hat{\alpha}_1^{OLS} = \alpha_1 + \frac{cov(p, u_1)}{var(p)} = \alpha_1 + \frac{1}{1 - \alpha_1 \beta_1} \cdot \frac{\sigma_{u_1}^2}{\sigma_p^2} \tag{4.17}$$

Taking expectation we obtain

$$E(\hat{\alpha}_1^{OLS}) = E\left(\alpha_1 + \frac{cov(p, u_1)}{var(p)}\right) = \alpha_1 + \frac{1}{1 - \alpha_1 \beta_1} \cdot E\left(\frac{\sigma_{u_1}^2}{\sigma_p^2}\right) \tag{4.18}$$

Since

$$\sigma_{u_1}^2 > 0$$

$$\sigma_p^2 > 0$$

the bias of $\hat{\alpha}_1^{OLS}$ is determined by the sign of $1 - \alpha_1 \beta_1$. Recalling our model, under the microeconomics assumptions about supply and demand both α_1 and β_1 are expected to be negative. Therefore, $\alpha_1 \beta_1 > 0$ which implies that $1 - \alpha_1 \beta_1 < 1$. The resultant sign of the bias depends on whether $\alpha_1 \beta_1 > 1$ or $\alpha_1 \beta_1 < 1$, in other words, whether the supply and demand are elastic. From the consumer´s point of view cars are expected to behave like luxury goods and therefore assuming$|\alpha_1| > 1$ should be right. If we expect that at least in the long run the supply is elastic, i.e. $|\beta_1| > 1$, using the method of OLS would result in underestimated coefficient α_1.

4.3 Instrumental variables estimation

As we discussed in previous chapter, in case of the SEM the OLS estimates suffer from bias, because there is present an endogeneity issue:

$$Cov(p, u_1) \neq 0 \tag{4.19}$$

$$Cov(q, u_2) \neq 0 \tag{4.20}$$

However, from the reduced form equations we can see that also $Cov(p, z) \neq 0$. And since z is exogenous and it is not involved in the structural equation on its own right, it is possible to use z as a proper instrument for p. Using the method of instrumental variables, we are able to obtain estimates of α_0, α_1 (Dougherty, 2011).

In case of simple regression (equation $q = \alpha_0 + \alpha_1 p + u_1$), the IV estimator of α_1 using z as instrument is given by

$$\alpha_1^{IV} = \frac{Cov(z, q)}{Cov(z, p)} \tag{4.21}$$

Using the substitution for x, it can be shown that α_1^{IV} is a consistent estimator of α_1:

$$\alpha_1^{IV} = \frac{Cov(z, [\alpha_0 + \alpha_1 y + u_1])}{Cov(z, p)} = \frac{Cov(z, \alpha_0) + Cov(z, \alpha_1) + Cov(z, u_1)}{Cov(z, p)}$$

Using the fact that $Cov(z, \alpha_0) = 0$ and $Cov(z, \alpha_1 p) = \alpha_1 Cov(z, p)$, we get

$$\alpha_1^{IV} = \alpha_1 + \frac{Cov(z, u_1)}{Cov(z, p)} \tag{4.22}$$

Since z is exogenous, it is distributed independently of u_1 and $plim\ Cov(z, u_1) = 0$. With the fact that p is determined by z and therefore $plim\ Cov(z, p) \neq 0$ we have

$$plim\ \alpha_1^{IV} = \alpha_1 \tag{4.23}$$

With instrumental variables estimator, we are able to obtain consistent estimates of SEM. In contrast with OLS the IV estimates are distributed around the true values, but given the nature of IV the standard errors tend to be higher than in case of OLS (Dougherty, 2011).

4.4 Identification of structural equations

Before we proceed to specific methods of SEM estimation it is necessary to determine at which conditions the estimates of the parameters of a structural equation can be obtained. In case that we are able to obtain the estimates, we say that the particular equation is identified. Otherwise, we say that the equation is underidentified. Identified equation can also be "just identified" or "overidentified". Let us show these situations separately for two-equation models. The following examples are inspired by Gujarati (2004).

4.4.1 Underidentification

Let us suppose the model

$$q = \alpha_0 + \alpha_1 p + u_1 \quad (4.24)$$

$$p = \beta_0 + \beta_1 q + u_2$$

The reduced form equations for q and p are

$$q = \pi_1 + v_1 \quad (4.25)$$

$$p = \pi_2 + v_2$$

Where

$$\pi_1 = \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1}$$

$$\pi_2 = \frac{\beta_0 + \alpha_0\beta_1}{1 - \alpha_1\beta_1} \quad (4.26)$$

Since the reduced form errors v_1, v_2 are only a linear combination of the structural errors u_1, u_2 and therefore, the reduced form coefficients π_1, π_2 may be estimated consistently (no endogneity is involved), we might try to obtain the structural coefficients $\alpha_0, \alpha_1, \beta_0, \beta_1$ as a solution of a system of equations. Unfortunately, in our case, we have a system of only 2 equations and four parameters. According to linear algebra, to obtain n unknow parameters we need a system of at least n equations. Hence, it is not possible to obtain all four parameters.

Alternative explanation is that there is no exogenous variable in the system of equations. Without exogenous regressor, we are not able to cut the circularity between the q and the p, because a variable which could be used as an instrument for q or p respectively is not present.

We say, that both of the equations are underidentified and they cannot be estimated. This fact makes also the whole model underidentified.

4.4.2 Exact identification

Now let us add some exogenous variable into both of the models

We have

$$q = \alpha_0 + \alpha_1 p + \alpha_2 z + u_1 \tag{4.27}$$

$$p = \beta_0 + \beta_1 q + \beta_2 s + u_2$$

where z and s are exogenous.

We use the same logic as in the previous case of underidentified equations. Reduced form equations obtained by substitution are as follows

$$q = \pi_{10} + \pi_{11} s + \pi_{12} z + v_1 \tag{4.28}$$

$$p = \pi_{20} + \pi_{21} z + \pi_{22} s + v_2$$

where

$$\pi_{10} = \frac{\alpha_0 + \alpha_1 \beta_0}{1 - \alpha_1 \beta_1}$$

$$\pi_{11} = \frac{\alpha_1 \beta_2}{1 - \alpha_1 \beta_1}$$

$$\pi_{12} = \frac{\alpha_2}{1 - \alpha_1 \beta_1}$$

$$\pi_{20} = \frac{\beta_0 + \alpha_0 \beta_1}{1 - \alpha_1 \beta_1} \tag{4.29}$$

$$\pi_{21} = \frac{\alpha_2 \beta_1}{1 - \alpha_1 \beta_1}$$

$$\pi_{22} = \frac{\beta_2}{1 - \alpha_1 \beta_1}$$

And the reduced form errors v_1, v_2 are again linear combinations of structural errors u_1, u_2.

In our model, we have six structural coefficients $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2$. These coefficients can be obtained from a system of at least six equations. As we can see, we are able to estimate exactly six reduced form coefficients $\pi_{10}, \pi_{11}, \pi_{12}, \pi_{20}, \pi_{21}, \pi_{22}$ – together they form a system of equations which allows us to obtain all the structural coefficients. The parameters of both of the simultaneous equations can be identified as well as the whole model.

To proceed an estimation using the instrumental variables, it is important that exactly one endogenous and one exogenous variable are present in each equation. This allows us to use z as an instrument for q and s an instrument for p. The issue of uncertain causality is solved and we are able to estimate consistently both of the equations.

We say that the equations are just- or exactly identified. Such equations can be estimated.

Interesting point is that it is the exogenous variable included in the first equation which makes the second equation identified. If we had only one exogenous variable in the model, let us say z and the equations would look like

$$q = \alpha_0 + \alpha_1 p + \alpha_2 z + u_1 \qquad (4.30)$$

$$p = \beta_0 + \beta_1 q + u_2$$

we would be able to use z as an instrument for q and estimate the second equation. But there would be no variable left to identify and estimate the first equation. In this case, the second equation would be identified, but the first one would stay unidentified.

4.4.3 Overidentification

In the last example, we add more exogenous variables, so that there is one endogenous and two exogenous variables in each equation.

$$q = \alpha_0 + \alpha_1 p + \alpha_2 z + \alpha_3 w + u_1 \qquad (4.31)$$

$$p = \beta_0 + \beta_1 q + \beta_2 s + \beta_3 t + u_2$$

Using substitution again to derive reduced form

$$q = \pi_{10} + \pi_{11} s + \pi_{12} t + \pi_{13} z + \pi_{14} w + v_1 \qquad (4.32)$$

$$p = \pi_{20} + \pi_{21} s + \pi_{22} t + \pi_{23} z + \pi_{24} w + v_2$$

where

$$\pi_{10} = \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1}$$

$$\pi_{11} = \frac{\alpha_1\beta_2}{1 - \alpha_1\beta_1}$$

$$\pi_{12} = \frac{\alpha_1\beta_3}{1 - \alpha_1\beta_1}$$

$$\pi_{13} = \frac{\alpha_2}{1 - \alpha_1\beta_1}$$

$$\pi_{14} = \frac{\alpha_3}{1 - \alpha_1\beta_1}$$

$$\pi_{20} = \frac{\beta_2}{1 - \alpha_1\beta_1} \quad (4.33)$$

$$\pi_{21} = \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1}$$

$$\pi_{22} = \frac{\alpha_1\beta_2}{1 - \alpha_1\beta_1}$$

$$\pi_{23} = \frac{\beta_2}{1 - \alpha_1\beta_1}$$

$$\pi_{24} = \frac{\beta_3}{1 - \alpha_1\beta_1}$$

Reduced form errors v_1, v_2 are linear combinations of structural errors u_1, u_2.

Now we have 10 reduced form coefficients and only 8 structural form coefficients - $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3$. That means the number of equations is higher than the number of unknowns. Just from solving the system of equations, there is no guarantee we will be able to obtain unique estimates of all coefficients. Some of them might give two different results.

We say that the equations are overspecified since they contain more than necessary information to be identified and estimated.

If we wanted to estimate structural equations by the method of instrumental variables, we would face a dilemma – which exogenous variable is the proper instrument? For example, as an instrument for q, there could be used z as well as w. Both options allow us to obtain consistent estimates, but usually we want to minimize the variance of estimates. Therefore, an instrument with higher correlation with q would be preffered. In the next chapter, a method of 2SLS that allows us to estimate overidentified equations efficiently will be discussed.

4.4.4 General rules

Rules which allow us to find out whether an equation may be estimated, e.g. it is identified are given as two conditions. The first one is called the Order condition and it is a necessary condition for the second one, Rank condition. Rank condition is necessary and sufficient condition for equation to be identified (Gujarati, 2004).

Before the rules are stated, let us introduce the notation which is used in Gujarati (2004):

M = number of endogenous variables in the model

m = number of endogenous variables in a given equation

K = number of predetermined variables in the model including the intercept

k = number of predetermined variables in a given equation

Order condition:

“In a model of M simultaneous equations, in order for an equation to be identified, the number of predetermined variables excluded from the equation must not be less than the number of endogenous variables included in that equation less 1, that is,

$$K - k \geq m - 1$$

If $K - k = m - 1$, the equation is just identified, but if $K - k > m - 1$, it is overidentified.“ (Gujarati, 2004)

Rank condition:

„In a model containing M equations in M endogenous variables, an equation is identified if and only if at least one nonzero determinant of order $(M - 1)(M - 1)$ can be constructed from the coefficients of the variables (both endogenous and predetermined) excluded from that particular equation but included in the other equations of the model.“ (Gujarati, 2004)

4.5 Methods of estimation

4.5.1 Two stage least squares

We have already proved that structural equations parameters estimated by OLS suffer from a simultaneity bias. There was also shown that the solution of this issue is the instrumental variables estimation which gives consistent estimates. However, in case that the equation is overidentified, we face a problem which variable should we choose as an instrument. The method of Two-Stage Least Squares allows us to obtain consistent estimates without need to choose only one variable as an instrument. Instead of that, the method uses a new variable which is a linear function of all potential instruments. Dougherty (2011) mentions two main reasons why the 2SLS is preferred. It is expected that with properly chosen coefficients a linear combination of instruments will be more efficient than individual instruments. Moreover, we avoid a problem of several conflicting estimates. We will show the method in practice using the procedure inspired by Dougherty (2011).

For our purpose, let us assume the model

$$q = \alpha_0 + \alpha_1 p + u_1 \tag{4.34}$$

$$p = \beta_0 + \beta_1 q + \beta_2 z + \beta_3 w + u_2$$

Variables z, w are assumed to be exogenous and we are interested in estimating the first equation. It is easy to realize that our model is overidentified since the second equation contains two exogenous variables which are excluded from the first one. As the name of the method implies, the estimation consists of two stages.

At first, we construct a new variable, let say V, as a linear function of all potential instruments. In our case, the assumptions to be a proper instrument fulfil the variables z and w. Then we have

$$V = h_0 + h_1 z + h_2 w \tag{4.35}$$

Coefficients of the linear combination can be obtained easily by OLS. We regress q on z and w, then we predict the fitted values of q. These fitted values we will use as the demanded variable V.

$$V = \hat{q} = h_0 + h_1 z + h_2 w \tag{4.36}$$

Under the premise that the model is correctly specified and the Gauss Markov assumptions are fulfilled, OLS is BLUE – best linear unbiased estimator. It means that the h coefficients are chosen in the way to minimize the sum of the square residuals (SSR), maximize the R-squared and maximize the correlation between the fitted and the true values of q.

The second stage is already known to us, as it is the instrumental variable estimation of β_1, with V being used as an instrument.

$$\beta_1^{2SLS} = \frac{Cov(V,q)}{Cov(V,p)} = \frac{Cov(\hat{p},q)}{Cov(\hat{p},p)} \tag{4.37}$$

4.5.2 Indirect Least Squares

In case of exactly identified simultaneous equations, there is another possible method of estimation. We call it the method of Indirect least squares and it is based on estimation of reduced form coefficients by OLS and then obtaining structural coefficients from system of equations. Recalling the model from previous chapter (4.27), which can be rewritten as

$$q = \pi_{10} + \pi_{11}s + \pi_{12}z + v_1 \tag{4.38}$$

$$p = \pi_{20} + \pi_{21}z + \pi_{22}s + v_2$$

where

$$\pi_{10} = \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1}$$

$$\pi_{11} = \frac{\alpha_1\beta_2}{1 - \alpha_1\beta_1}$$

$$\pi_{12} = \frac{\alpha_2}{1 - \alpha_1\beta_1}$$

$$\pi_{20} = \frac{\beta_0 + \alpha_0\beta_1}{1 - \alpha_1\beta_1} \tag{4.39}$$

$$\pi_{21} = \frac{\alpha_2\beta_1}{1 - \alpha_1\beta_1}$$

$$\pi_{22} = \frac{\beta_2}{1 - \alpha_1\beta_1}$$

The reduced form coefficients $\pi_{10}, \pi_{11}, \pi_{12}, \pi_{20}, \pi_{21}, \pi_{22}$ can be estimated by OLS consistently. Therefore, we just solve the system of 6 equations and compute structural coefficients $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2$.

Dougherty (2011) stresses that the application ILS is limited since in case of overidentification, the computation of structural coefficients might result in several different estimates of one parameter, therefore, it is more convenient to prefer the 2SLS.

„ILS has no advantages over IV and has the disadvantage of requiring more computation. If an equation is underidentified for IV, it is underidentified for ILS; if it is exactly identified, IV and ILS yield identical estimates; if it is overidentified, ILS yields conflicting estimates, a problem that is resolved with IV by using 2SLS.“ (Dougherty, 2011)

4.6 Preliminary Model

The aim of this thesis is to estimate coefficients of the demand equation in the new cars supply-demand model. Based on the explanatory factors derived in the previous chapter the expected form of the model is following:

$$q = \alpha_0 + \alpha_1 p_n + \alpha_2 p_u + \alpha_3 p_f + \alpha_4 i + \alpha_5\, m + \alpha_6 s\; + u_1 \qquad (4.40)$$

$$q = \beta_0 + \beta_1 p_n + \beta_2 c + \beta_3\, r + u_2$$

where the first equation represents the demand, the second equation represents the supply and where q is the quantity of new cars sold, p_n is the new cars price, p_u is the used cars price, p_f is the fuel price, i is the consumer income, s is the market saturation, t is the seasonality representing the consumer taste, c represents the producer costs, r is the exchange rate, u_1 and u_2 are the random disturbances and α_i and β_j represent the appropriate regression coefficients. Only the variables q and p_n are assumed to be endogenous.

According to the Rank condition, the model is identified and the demand equation is overidentified. In other words, there is more than one possible instrument for the new cars price, hence, the preferred method is the 2SLS. The hypothesis is that the coefficients of the demand equation estimated by OLS will be underestimated in comparison with those estimated by 2SLS.

5 Empirical analysis

This section is dedicated to the data collection and their analysis. In the next step the new cars demand is estimated and the results are discussed. STATA/SE 12 and MS Office Excel 2013 software was used for computations.

5.1 Data set

The following chapter introduces the data which are used to represent the variables in the supply and demand model. To each kind of data, their development is briefly described and the summary statistics is included.

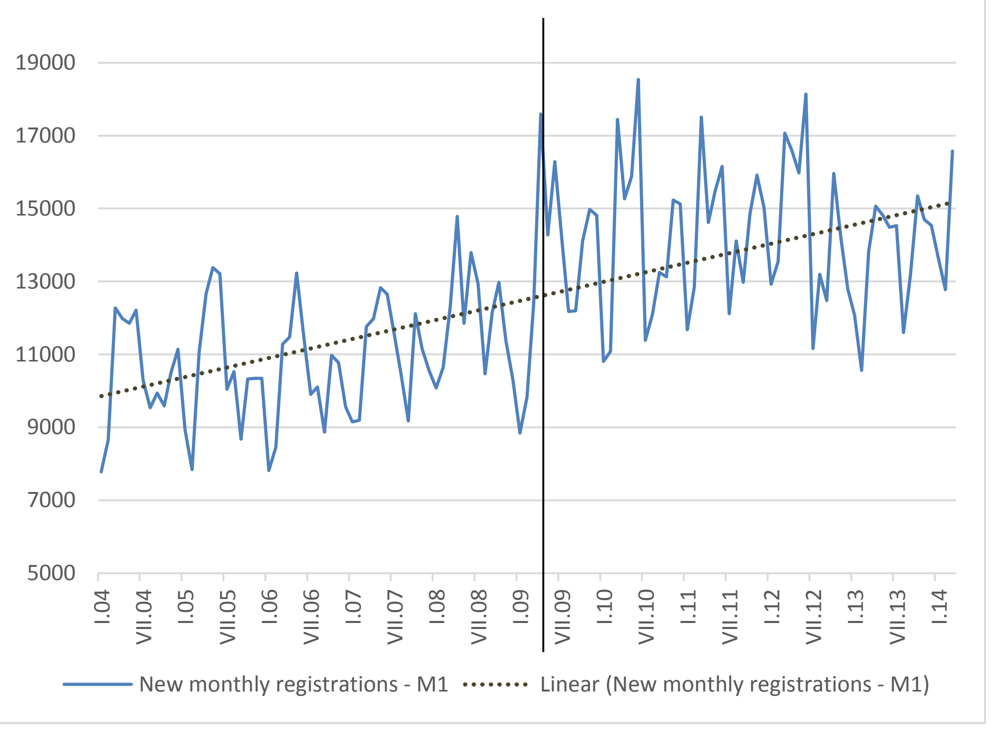
Most of the data samples are provided by the Czech Car Importers Association (CIA), the Czech statistical office (CSO) and the Czech National Bank (CNB). The goal was to obtain the data sample for estimation as large as possible, the main limitation is caused by the dependent variable which is available only from 1/2004. To provide the highest possible number of observations the time series data are examined monthly.

5.1.1 Number of new cars registrations

The data which probably best represent the traded quantity of new cars as a dependent variable in the model are the new car registrations. The CIA publishes the data acquired from the Central Vehicle Register (CVR) administered by the Czech Ministry of Transport. The available statistics captures the monthly number of new cars registered to the CVR over the period from January 2004 to March 2014. For purposes of the passenger car demand estimation, the two categories are important. The first category are the passenger cars which are classified as M1 or OA. The second one are the Light Utility Vehicles classified as N1 or LUV.

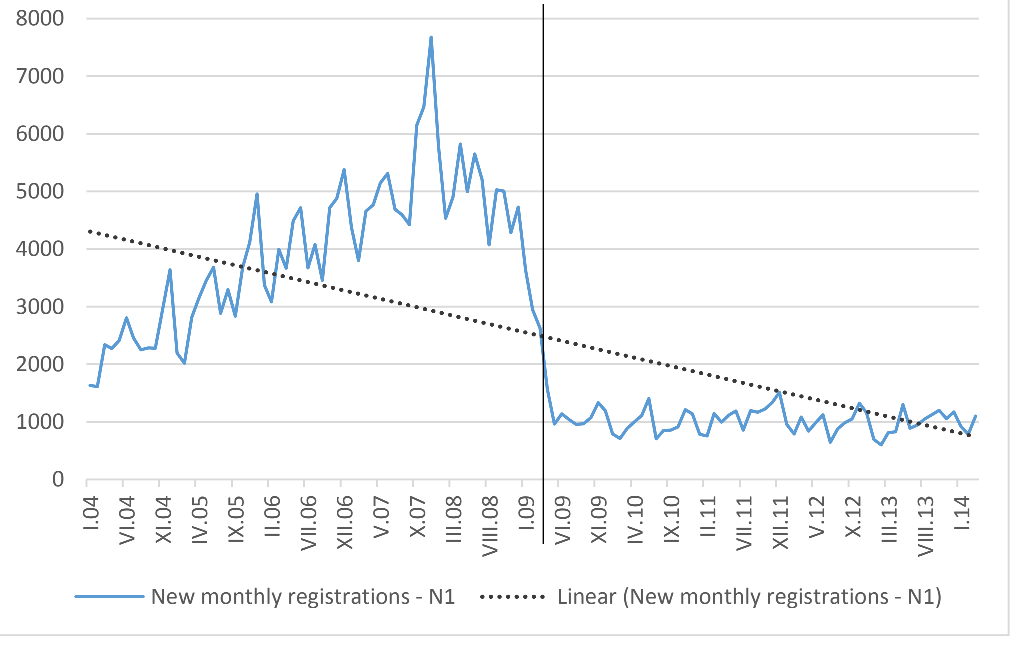
Graphs shown below represent the development of the M1 and N1 registrations in the Czech Republic from 1/2004 to 3/2014 with a linear time trend included.

Figure 5.1: New monthly registrations M1 - passenger cars. Vertical line represents a change in the taxation policy in 4/2009



Source: CIA (2014), author's computations

Figure 5.2: New monthly registrations N1 – light utility vehicles. Vertical line represents a change in the taxation policy in 4/2009



Source: CIA (2014), author's computations

Before any decision about the usage of the data is made, it is necessary to analyse how the M1 and N1 categories interact with each other, since the vehicles in these groups often have very similar characteristics.

Light Utility Vehicles Phenomenon

A quite interesting fact related to the Czech car market is an extreme popularity of "Light Utility Vehicles". During the years 2004 – 2009 they accounted for up to 70 % of passenger cars sold in a particular month. Although these vehicles are registered separately from passenger cars, they usually do not differ from them very much. The reason is that the LUV´s may be produced by a simple transformation of ordinary passenger cars, mostly by installing a counter dividing the trunk and the room for passengers. But what caused such a high share of these vehicles in new registrations? The answer is quite straightforward.

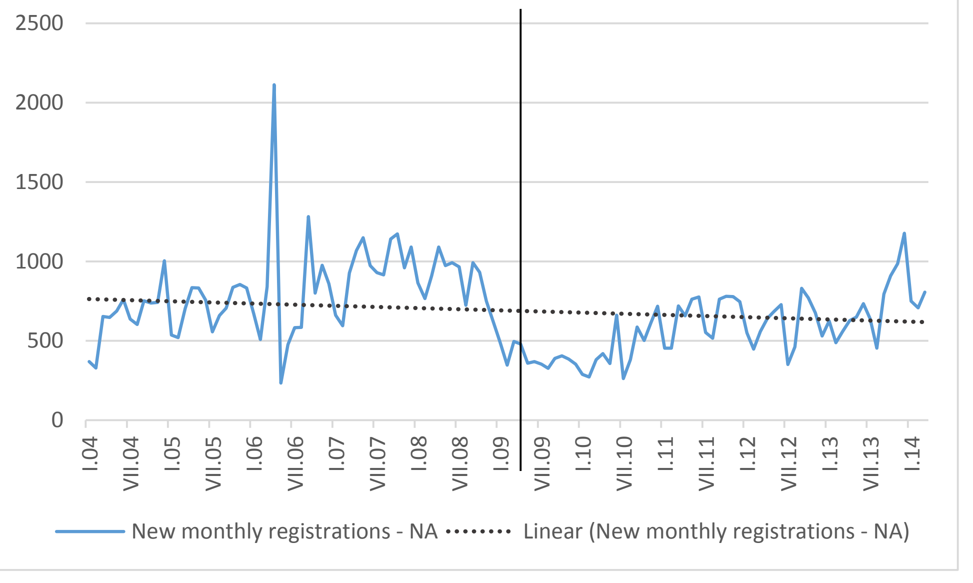
Before the April 2009, there was no possibility to subtract VAT for passenger cars (category M1) used for commercial purpose. Nevertheless, one could transform his car into Light Utility Vehicle (category N1) and subtract the VAT from the price. The transformation was very easily provided by installing a counter dividing the trunk and the room for passengers. This lead to a curious situation when number of passenger cars including also very luxury cars were sold as LUV. The data provided by SDA are quite detail, so one can easily find out that it was not unusual that in some months half of the cars manufactured by Porsche, which is focused on luxury sports cars, were sold as LUV. In April 2009 the Czech law 235/2004 Sb. treating the VAT was revised and allowed also VAT deduction for M1 category.

Our hypothesis is that the change caused huge downturn of N1 sales and rise of M1, because consumers were forced to buy N1 vehicles only to subtract the VAT, not because they would really demand the utility vehicles.

The CIA provides monthly data of new car registrations separately for M1 and N1 categories. Using the data from January 2004 to February 2014 it is easy to identify the downturn of N1 registrations concurrently with implementing the new policy and increase of M1 registrations.

An issue is a fact that approximately at the same time the new policy was introduced, a global financial crisis arrived. To find out whether the downturn was caused by the change of the policy or by the global economic crisis, we can use the difference in difference estimation. Monthly registrations in the NA category (trucks) should not be affected by the change in policy, therefore, it can be used as control group. The treatment group are the registrations in the N1 category and M1 respectively. The graph below shows the NA category registrations in the Czech Republic from 1/2004 to 3/2014 with a linear time trend included.

Figure 5.3: New monthly registrations NA – trucks. Vertical line represents a change in the taxation policy in 4/2009



Source: CIA (2014), author's computations

Estimating the following models by OLS should deliver us the right answer.

$$\mathrm{PN}_t = \alpha_1 + \alpha_2\, NewPolicy_t + \alpha_3 N1_t + \alpha_4\, N1_t * NewPolicy_t + \alpha_5\, t + u_{1t}$$

$$\mathrm{PM}_t = \beta_1 + \beta_2\, NewPolicy_t + \beta_3 M1_t + \beta_4\, M1_t * NewPolicy_t + \beta_5\, t + u_{1t}$$

where

- PN_t is a pooled dataset of the NA and the N1 categories monthly registrations in time t
- PM_t is a pooled dataset of the NA and the M1 categories monthly registrations in time t
- $\mathrm{M1cat}_t$ is a dummy for passenger cars (M1 category), $M1_t = \begin{cases} 0 \text{ if M1} \\ 1 \text{ otherwise} \end{cases}$

- $N1cat_t$ is a dummy for LUV (N1 category), $N1_t = \begin{cases} 0 \text{ if N1} \\ 1 \text{ otherwise} \end{cases}$
- $NewPolicy_t$ is a dummy for policy change, $NewPolicy_t = \begin{cases} 0 \text{ for } t < 64 \\ 1 \text{ for } t \geq 64 \end{cases}$
- t stands for a monthly time variable, t=1,2,…,123

Table 5.1: Difference in difference OLS estimation – LUV (N1) + trucks (NA)

Variable	Coefficient	St. Error	*t* – statistics	*P* - value
const.	293.554	105.105	2.79	0.006
NewPolicy	-1171.318	175.743	-6.66	0.000
N1	3154.254	109.855	28.71	0.000
N1*NewPolicy	-2714.504	157.289	-17.26	0.000
t	15.622	2.213	7.06	0.000

Source: author's computations

Table 5.2: Difference in difference OLS estimation – Passenger cars (M1) + trucks (NA)

Variable	Coefficient	St. Error	*t* – statistics	*P* - value
const.	560.154	213.394	2.62	0.009
NewPolicy	-658.946	356.811	-1.85	0.066
M1	10070.17	223.040	45.15	0.000
M1*NewPolicy	3594.259	319.344	11.26	0.000
t	7.291	4.492	1.62	0.106

Source: author's computations

Results of the estimation can be interpreted in the way that the change of the policy is responsible for a decrease of N1 category by 2 714.5 registrations on average. The estimated downturn of both groups by 1 171.3 can be explained as a negative impact of the financial crisis, which arrived approximately at the same time as the policy was introduced.

In case of the second model, results correspond with the first estimation. After changing the VAT policy the M1 registrations average increase was estimated to be 3594.3. The overall decrease by -658.95 can again be explained as a result of the crisis.

Having analysed obtained results, we can say that the policy caused a shift in registrations from N1 category to M1. That confirms our hypothesis. If we take into account the fact that the financial crisis stroke at the same time as the policy was introduced, we can explain the overall decrease in sales rather to be a result of the crisis than introducing the policy. This statement can be supported by a logical assumption, that allowing the companies to subtract VAT in case of both categories is a positive step which should not cause lower registrations.

As the hypothesis was right, for estimating the car demand over years 2004 and 2014 it is necessary to use registrations aggregated for categories M1 and N1 since capturing the exact effect of the policy change on only one group would be very difficult.

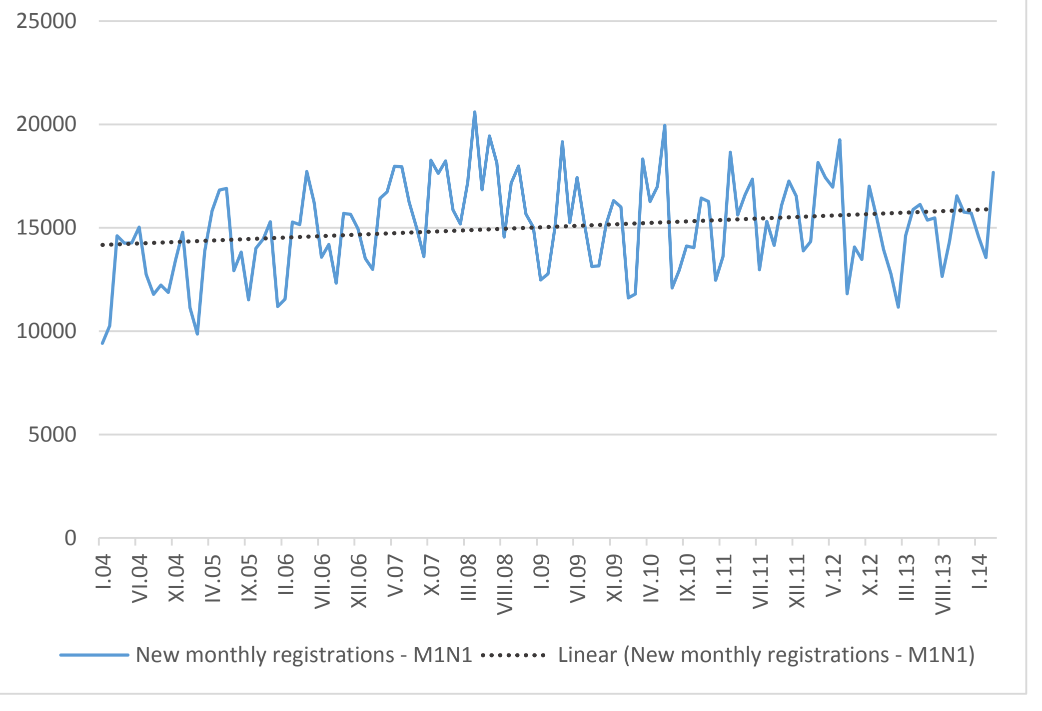
This research also implies one important consequence. Because Mikas (2011) in his work studied only the M1 category and the policy breakpoint was approximately in the middle of the period he studied, the data wrongly showed an increasing time trend and the whole estimation was biased and incorrect. His estimates therefore should not be used for a reliable prediction of the new car demand.

Since only the aggregated data for M1 and N1 categories make sense as a variable representing the new car demand over the years from 2004 to 2014, let us study this merged data. The best results would be probably reached by using a disaggregate approach introduced by Carlson (1978), but unfortunately, the N1 category is not sorted by the same classes as in the M1 category and it would be very difficult and inaccurate to distinguish registrations in particular classes. Hence, the data will be estimated as aggregated registrations. This will probably affect the fuel price impact, because Carlson (1978) observed a shift between car classes when the fuel price changed. Using the net registrations, e.g. new registrations less decommissioned cars, would probably be the most convenient to estimate the actual willingness of consumers to buy a car. Mikas (2011) could not include the number of cars decommissioned from the CVR, because the data showed to be inaccurate and did not represent the reality. A lot of still registered vehicles do not exist already and they are being decommissioned from the register rather randomly. After a consultation with the SDA representative, which confirmed Mikas´s statement, the author has also decided to focus only on the new cars registrations.

Using the new cars registrations as an indicator of the quantity traded delivers a disadvantage in terms that the Central vehicle register does not capture the exact reality in the car market. During the last years car importers started to perform the so-called reexport of new cars. The reexported cars are registered as new cars in the CVR, but immediately after the registration these cars are exported. Supposingly, the main reason for this action are low prices of new cars in the Czech Republic and low sales of car importers. Therefore, they improve the revenues by fictive sales and sell cars unofficially abroad. It is estimated that 7.5 % of new cars registered in 2012 and 9.2 % in 2013 were subject to the reexport. Unfortunately, these estimates are inaccurate and unofficial, the CIA refused to provide any information about the problem. Because there is still no better source of information about new cars sold than the CVR, this bias has to be tolerated.

The graph below represents the aggregated new car registrations of M1 and N1 categories in the Czech Republic from 1/2004 to 3/2014 including the linear time trend.

Figure 5.4: New monthly registrations M1N1 - passenger cars and LUV



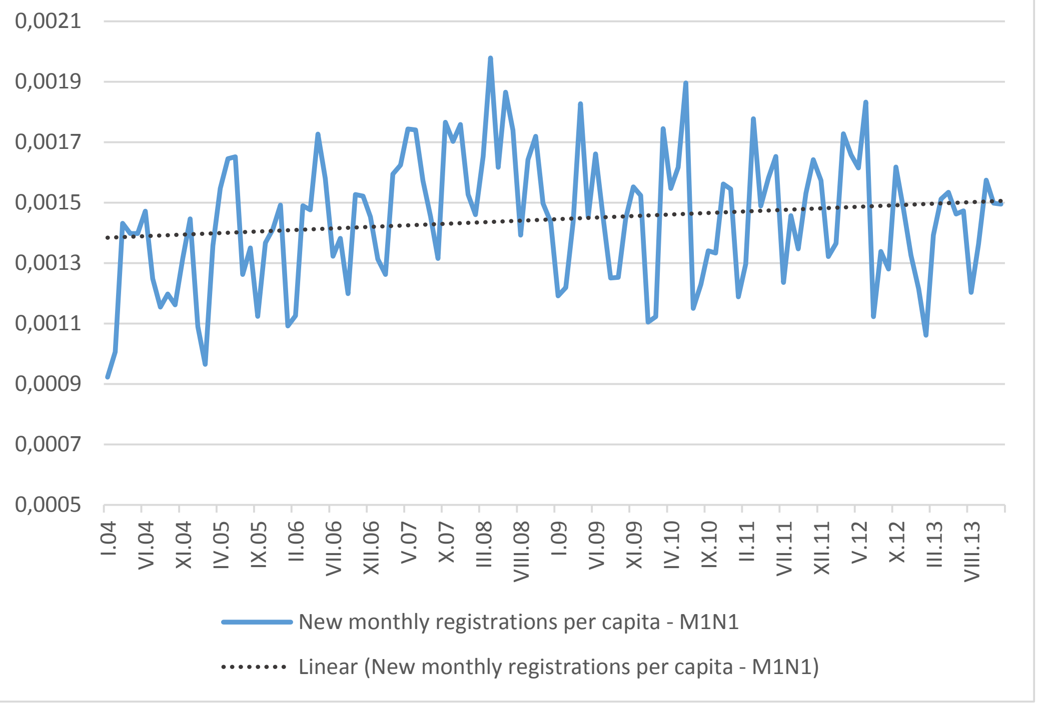
Source: CIA (2014), author's computations

From 2004 to 2008, the M1 and N1 vehicles registrations were steadily increasing. In 2008, the registrations reached the peak and then dropped immediately with the economic crisis. Since 2009 the registrations numbers stabilized, but they are still waiting for recovery to pre-crisis performance. A slightly increasing time trend is present and the F-test confirmed a strong seasonality for particular months, therefore, the seasonality will have to be treated in the regression by adding the dummies for particular months.

5.1.2 Population

The monthly statistics provided by the CSO offers quite a detailed overview about the population development. If we compare the mid-period population for each month, e.g. the average of two consecutive months, we find out that over last ten years the number of inhabitants of the Czech Republic gradually increased from 10 209 003 in 1/2004 to 10 513 164 in 12/2013. To reflect this change in terms of new cars registrations, the number of new registrations per capita was computed for each month. The following graph represents new cars registrations for M1 and N1 categories per capita.

Figure 5.5: New monthly registrations per capita M1N1- passenger cars and LUV



Source: CIA (2014), author's computations

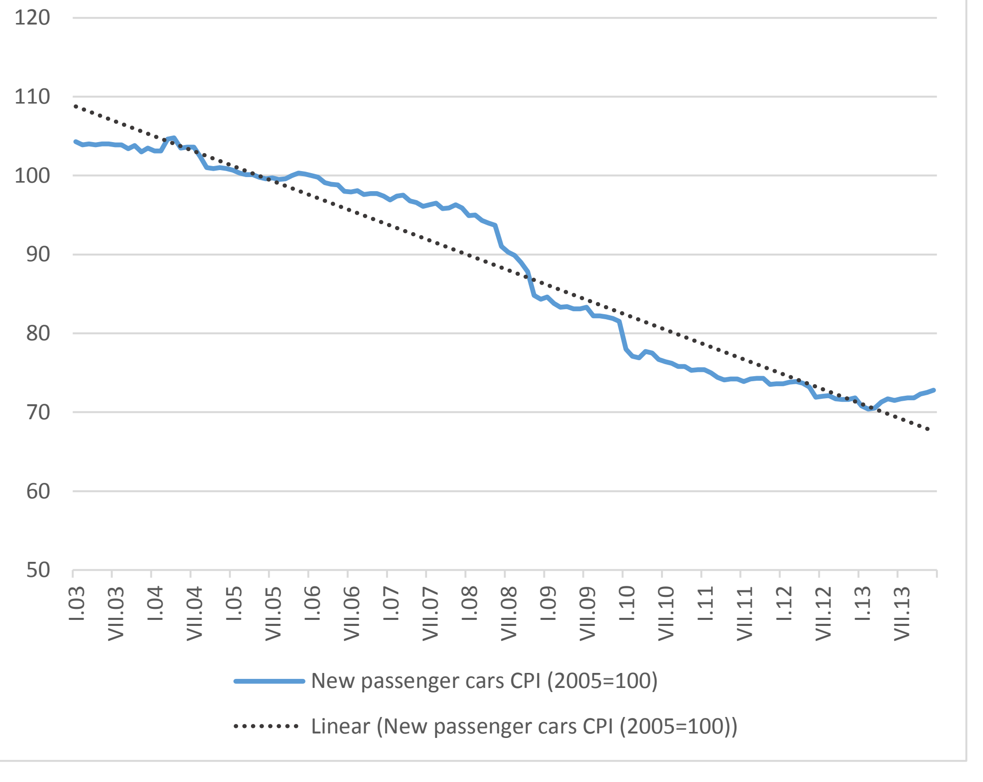
5.1.3 Number of registered passenger cars

AIA provides a basic information about the development of registrations in the CVR. As of 31.12.2013 there were 4 787 849 registered passenger cars in the Czech Republic, e.g. thus there were 455 registered passenger cars on 1000 inhabitants. This ratio is therefore lower than in developed countries of the Western Europe or in US but higher than in most of former Eastern Bloc countries (ACEA, 2014). AIA stresses the insufficient car renewal ratio around 3.5 % per year and the high average age of motor vehicles, which in case of passenger cars reached 14.2 years in 2013. Unfortunately, because of the same reasons as in case of the number of cars decommissioned the data were found to be inaccurate and cannot be taken into account as an explanatory variable. However, although the ability of a market to absorb new cars is limited and at some level of saturation the elasticity of consumer demand decreases, with respect to the average age of registered vehicles and the market saturation in developed countries, it can be assumed that the Czech car market did not reach the level of saturation which would require a special treatment.

5.1.4 New passenger cars price

A monthly basic CPI of new cars in the Czech Republic related to year 2005 (average of 2005=100) was obtained from the CSO. The index is calculated from consumer prices of the most common representatives of the Czech car market. The graph below shows the Czech new passenger cars CPI development from 1/2003 to 12/2013.

Figure 5.6: New passenger cars CPI (2005=100)



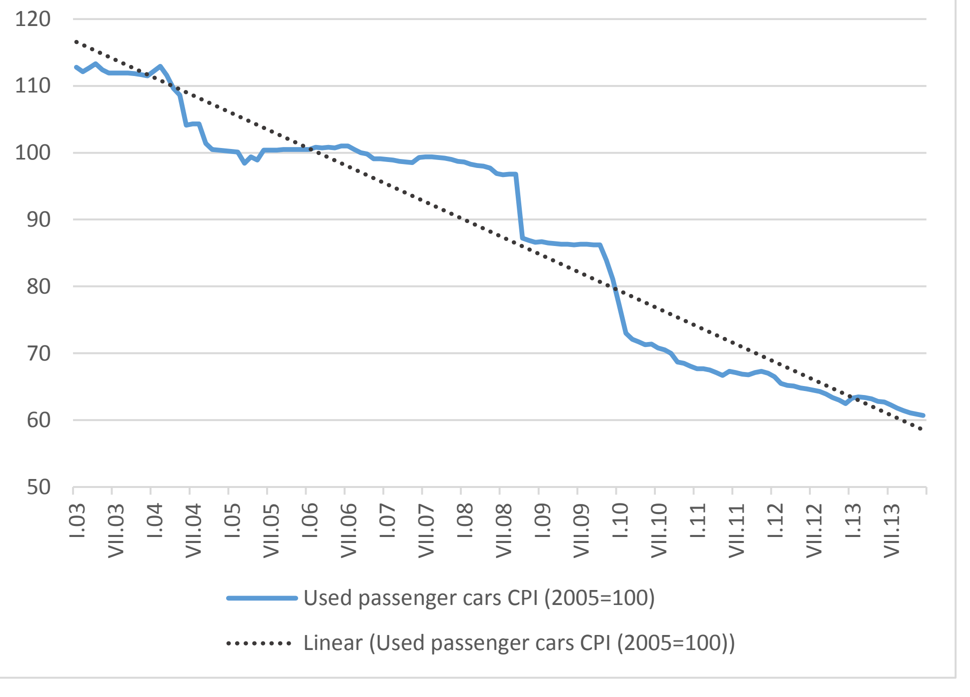
Source: CSO (2014), author's computations

As we can see, the new cars prices in the Czech Republic have a decreasing tendency falling on average by -3.718 % a month. The highest prices were recognized in April 2004 when the basic index reached 104.8. The lowest prices February 2013, the price index was only 70.4. In terms of the year over year change, the dramatic downturn appeared in 2008 and 2009, probably as a consequence of the economic crisis. The car market in the Czech Republic of last few years is known for its competitive environment and customers´ high price sensitivity pushing the car prices down. In the beginning of 2014, which is not in scope of the estimation because some data are not available, the prices probably reacted to CZK exchange rate depreciation and started to increase. The future development will show whether this trend will preserve in the long run.

5.1.5 Used passenger cars price

Similarly as in case of the new cars price, the CSO publishes the monthly basic CPI of used cars in the Czech Republic related to year 2005 (average of 2005=100). The index is calculated from consumer prices of the most common representatives of the Czech car market. The graph below shows the data development from 1/2003 to 2/2014.

Figure 5.7: Used passenger cars CPI (2005=100)



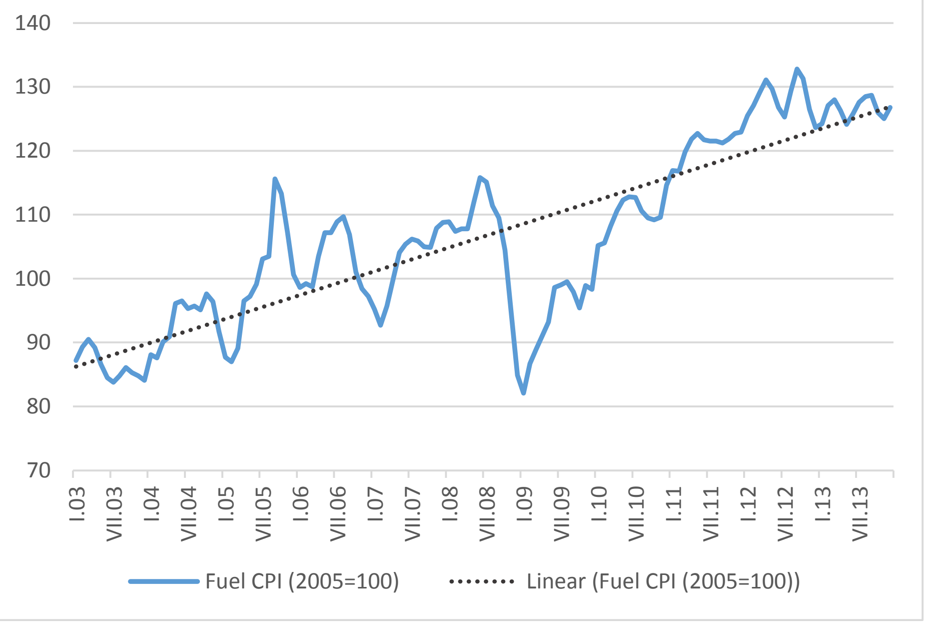
Source: CSO (2014), author's computations

Over the last 10 years the used cars prices in the Czech Republic followed basically the same development as the new cars prices did. The price index reached its maximum 113.3 in April 2003, than decreased till present on average by 6.23 % with three significant drops in 2005, 2008 and 2010. The lowest index 60.5 was recorded in the last observation.

5.1.6 Fuel price

To represent the fuel price the monthly basic CPI of fuel in the Czech Republic related to year 2005 (average of 2005=100) was chosen. It is published by the Czech statistical office. The index is calculated from consumer prices of chosen car fuels (gasoline, oil, LPG) with weights representing the share of the fuel type on total fuel traded. The graph below shows the data development from 1/2003 to 2/2014.

Figure 5.8: Fuel CPI (2005=100)



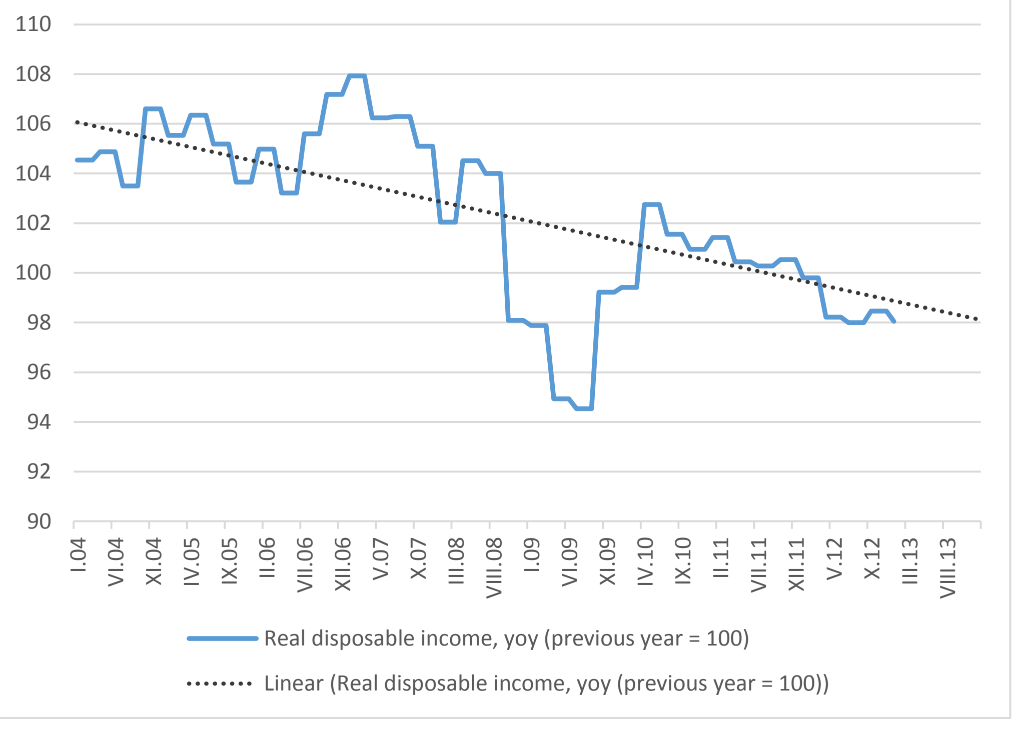
Source: CSO (2014), author's computations

Unlike the new and used cars prices, the fuel price index has been significantly increasing over the studied period. The only decrease was recognized during the year 2008 when the price hit the bottom of 82.1. Nevertheless, it quickly recovered to an increasing trend.

5.1.7 Real gross disposable income

As a proxy variable representing the consumer´s income the real gross domestic income proved to be a good choice. The year over year index for the Czech Republic is provided by the Czech statistical office quarterly, representing the percentage change with respect to the previous year. The year over year form is convenient since it avoids the possible seasonality and it could also well represent the consumer´s expectations and willingness to spend money, since the GDI change is a frequently cited economic indicator. In order not to lose a high number of observations, the quarterly data have been transformed to the monthly form by assigning the quarter output to appropriate months. The monthly data development from 1/2004 to 12/2013 is summarized by the graph below.

Figure 5.9: Real gross domestic income (previous year = 100)



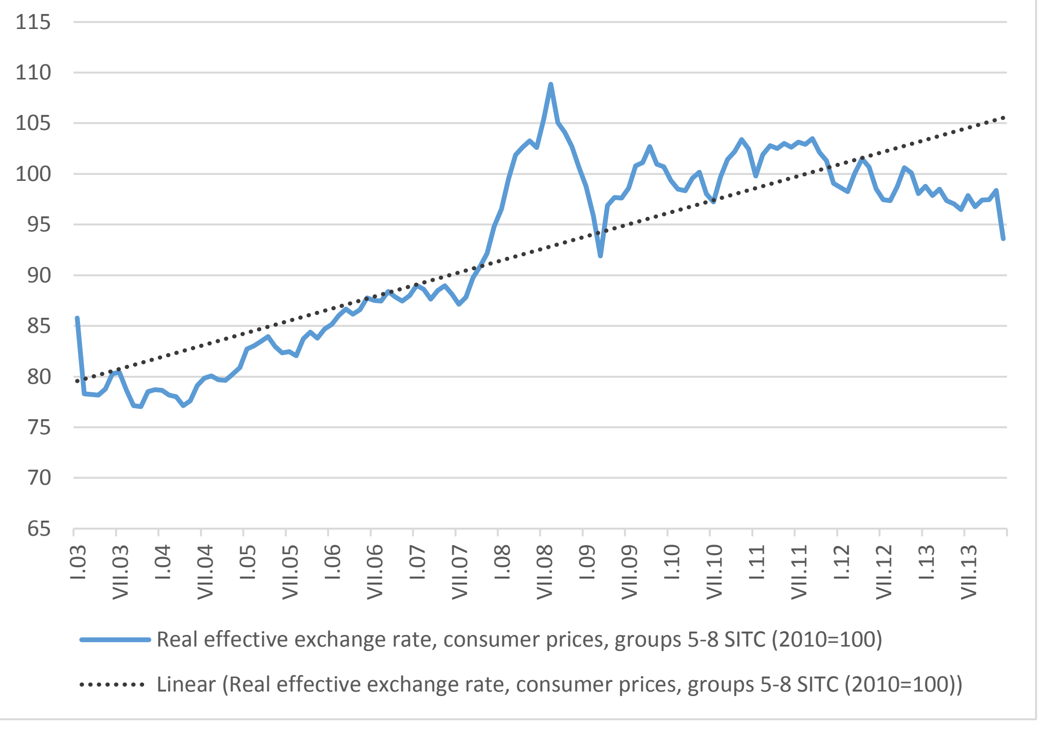
Source: CSO (2014), author's computations

Up to the seasonal fluctuations (which are not reflected by the year over year change), the real GDP in the Czech Republic was increasing over the whole studied period. The average rate of growth was 3.31 %. The highest recorded change was an increase by 7.933 % in the first quarter of 2007 and the lowest one, when the real GDP rose only by 0.2025 %, was in the first quarter of 2012. Since 2008 there is a trend of diminishing increase of the GDP, the years 2012 and 2013 were in the name of only very low economy growth.

5.1.8 Real effective exchange rate of CZK

For purposes of the estimation, probably the best variable representing the exchange rate effect on the imported cars supply is the real effective exchange rate of CZK. It is a weighted average of real exchange rates of countries the Czech Republic trades with. The weights are chosen according to the trade turnover between the Czech Republic and the particular foreign country. It can interpreted as follows. *"If the real exchange* *rate is high, foreign goods are relatively cheap, and domestic goods are relatively expensive. If the real exchange rate is low, foreign goods are relatively expensive, and domestic goods are relatively cheap*." (Mankiw, Macroeconomics 5th edition). The Czech National Bank (CNB) publishes the real effective exchange rate monthly index in consumer prices, 2010 = 100. Weights are chosen according to the foreign trade turnover in groups 5-8 of Standard International Trade Classification (SITC), 4th revision. The graph below shows the monthly data development from 1/2003 to 12/2013.

Figure 5.10: Real effective exchange rate, consumer prices, 5-8 SITC (2010=100)



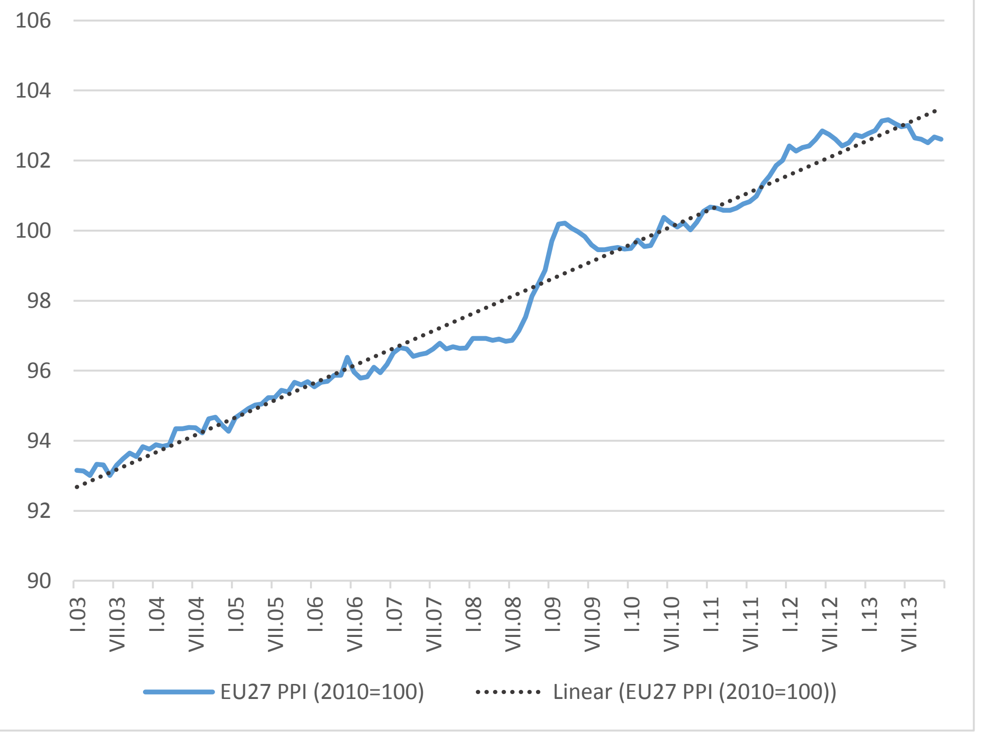
Source: CNB (2014), author's computations

Over the studied period, the real effective exchange rate reached the lowest value in when it accounted only for of the 2010 average. It increased incrementally, then the index recorded a sharp rise to the maximum of 110.61 and immediately dropped down. Next years, the index increased again whereas in the last two years a decreasing tendency occurred again.

5.1.9 Producer costs

Unfortunately, a problem with collecting the producer costs data occurred. There was not found any index which would directly represent costs of motor vehicles producers. The producer price index reflects prices the firm obtains for produced goods. Therefore, it can be a good proxy for the producer´s costs from the side of his subcontractors but it is also highly correlated with the consumer price. Hence, the ability of producer price index to serve as an exogenous supply shifter is limited. Two production price indices were obtained from Eurostat. The first one represents the monthly basic index of producer prices for manufacturers of motor vehicles, trailers, semi-trailers and of other transport equipment summarized for 27 European countries, related to the average of 2010 = 100. Since the most of cars imported into the Czech Republic originates from the EU, the index could represent the costs of these cars´ producers. The data development from 1/2003 to 12/2013 and the year to year percentage change are shown in the graphs below.

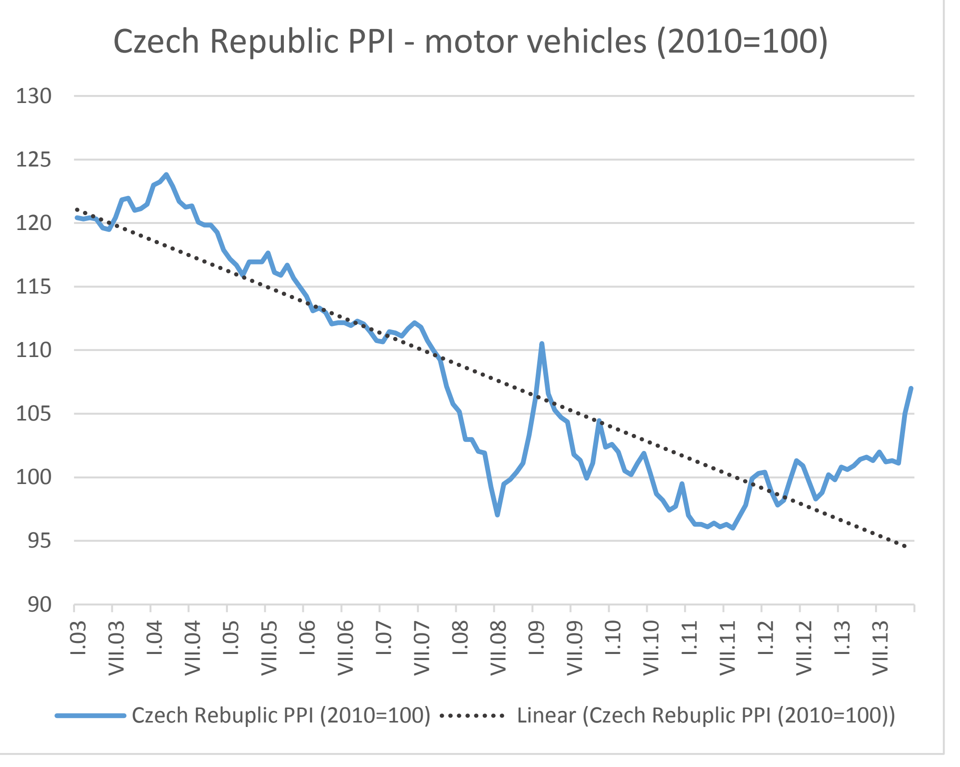
Figure 5.11: EU27 PPI, motor vehicles manufacturers (2010=100)



Source: Eurostat (2014), author's computations

The second index is a basic producer price index for Czech manufacturers of motor vehicles, trailers, semi-trailers and of other transport equipment related to average of 2010=100. In this case an issue is the fact that only approximately 60 % of new cars sold in the Czech Republic were imported. The index, therefore, reflects only one third of the production sold, but it fulfils the assumptions to serve as a proper instrument since it is positively correlated with the new cars CPI. Therefore, it is preferred over the EU27 PPI.

Figure 5.12: CZ PPI, motor vehicles manufacturers (2010=100)



Source: Eurostat (2014), author's computations

5.1.10 Summary

Table 5.3: Data summary statistics

Variable	Observations	Mean	Std. Dev	Min	Max
M1N1 reg. pc	123	15035.07	2252.051	9417	20603
New cars CPI	134	87.963	12.407	70.4	104.8
Used cars CPI	134	87.138	17.719	60.5	113.3
Fuel CPI	134	106.85	14.168	82.1	132.8
Real GDI	120	3.310	2.225	.203	7.933
Real eff. ex. rate	134	92.535	8.894	77.04	108.86
CZ PPI	134	107.707	8.558	96	123.81
EU 27 PPI	134	96.564	9.543	80.5	109.99

Source: author's computations

5.2 Time series analysis

In order to make sure that all the data follow a stochastic, weakly dependent process and that the Gauss Markov assumptions for linear regressions are fulfilled, the correlation analysis, test for a unit root and test for a linear time trend and seasonality were performed.

5.2.1 Unit root test

The augmented Dickey-Fuller test was applied to the dependent and all the explanatory variables. Unfortunately, for all the explanatory variables the test did not provide enough evidence to reject the null hypothesis of unit root present. The problem occurred probably because of a persistent nature of the index data. Since an attempt to find another data or suitable proxy failed, the problem had to be treated by a transformation of the persistent data. Although the differencing would solve the problem of high persistence the interpretation of the resulting data would not make much sense. Using the year on year (yoy) data showed to be a better way how to settle the issue. The year on year data were computed for each variable (except the gross domestic real income which was already obtained as year on year index) using the following formula:

$$\text{for } t \geq 12\text{:} \ x_t^{yoy} = 100 * \frac{(x_t - x_{t-12})}{x_{t-12}} \tag{5.1}$$

Unfortunately the Dickey-Fuller test of the new and used cars CPI and the domestic income did not provide satisfactory results even after the transformation. However the test for an autoregressive process of order one (AR(1)) showed that the persistence of transformed data is lower and it can be assumed they do not follow a unit root process and therefore, they are weakly dependent.

5.2.1 Correlation analysis

Computing the pairwise correlation coefficients offers an initial information about the mutual relationships between the explanatory variables. The data already transformed because of the unit root were used. The correlation matrix is shown below.

Table 5.4: Correlation matrix

Variable	1	2	3	4	5	6	7
1 New cars CPI	1.000						
2 Used cars CPI	0.708	1.000					
3 Fuel CPI	0.301	-0.04	1.000				
4 Real GDI	0.504	0.363	0.234	1.000			
5 PPI CZ	0.128	0.049	-0.390	-0.426	1.000		
6 PPI EU27	-0.426	-0.139	-0.541	-0.484	0.453	1.000	
7 Real eff. ex. rate	0.046	0.201	0.344	0.468	-0.919	-0.440	1.000

Source: author's computations

The analysis showed that the new cars and the used cars prices are highly positively correlated, but this is an expected result since new and used cars are close substitutes. Important is the fact that in any case the correlation is not near one and therefore the issue of multicollinearity does not arise. Moreover, according to Johnston (1997) the method of 2SLS yields valid estimates also in case of nonstationary and cointegrated time series. Another finding of the correlation analysis is that the new cars price and the European PPI are negatively correlated. This is at variance with expectations and it can be probably explained by a very low Czech price level of new cars in comparison to other countries in the region. Producers sell cars cheap in order to keep their market shares although the prices and costs in Europe increased. According to this fact using the available data the European PPI effect on the new cars price is unobservable and it cannot serve as a proper instrument. Even in the yoy form the Czech PPI and the real effective exchange rate show a high negative correlation, approximately -0.92. The further analysis confirmed that the data development of these two variables is basically the same, but only inverse. Using both variables as instruments for the new cars CPI resulted in a spurious regression suffering by a multicollinearity, therefore, only one of the variables might be used at one time. Better results should be obtained with the PPI since its correlation with the new cars CPI is higher than in case of the real effective exchange rate, therefore it will be preferred as a supply shifter in the regression.

5.2.3 Trend and seasonality

Since in case of the dependent and several explanatory variables a time trend and a seasonality were observed, to provide the assumption of time series weak dependence is fulfilled, a linear time trend and set of dummy variables for particular months have to be added into the regression. To avoid a dummy variable trap, the dummy for January is omitted and its function is substituted by an intercept in the model.

5.3 Results

As derived in the theoretical part, the econometric model for new cars supply and demand consists of two simultaneous linear equations. Using the variables described in the previous section the model can be written as follows:

$$ln\ NewReg\_pc_t = \alpha_0 + \alpha_1\ New\_price_t + \alpha_2\ Used\_price_t + \alpha_3\ Fuel\_price_t + \alpha_4\ Real\_GDI_t + \alpha_5\ t + \sum_{i=6}^{17} \alpha_i Month_{ti} + u_{dt}$$

$$lNew\_price_t = \beta_0 + \beta_1\ lNewReg\_pc_t + \beta_2\ Producer\_price\_CZ_t + \beta_3\ t + \sum_{i=6}^{17} \alpha_i Month_{ti} + u_{st}$$

where the first equation represents the demand and the second one represents the supply. The variables are:

- $NewReg\_pc_t$ is the number of new cars registrations *per capita* in the Czech Republic in month t, aggregated for categories M1 and N1
- New_price_t is the Czech new cars basic CPI (2005=100) yoy % change in month t
- $Used\_price_t$ is the Czech used cars basic CPI (2005=100) yoy % change in month t
- $Fuel\_price_t$ is the Czech fuel basic CPI (2005=100) yoy % change in month t
- $Real\_GDI_t$ is the Czech real gross domestic income yoy % change in month t
- $Month_{ti}$ represent the monthly dummy variables $i = 6(February), \dots, 17(December)$

- $Producer\_price\_CZ_t$ is the PPI of Czech producers of motor vehicles yoy % change in month t
- $t = 1,2, \ldots 120$ (1/2004,…, 12/2013)
- u_{dt} and u_{st} are the random disturbances in month t

The SEM is identified since it meets the order- and rank conditions requirements. The demand equation, which estimates are in the point of interest, is just identified, hence, there is exactly one possible instrument for the endogenous variable in the demand equation, which is the new cars price.

5.3.1 OLS regression

As a baseline the demand equation was estimated by the Ordinary Least Squares method. Although the obtained estimates are subject to the SEM bias, they provide an initial information and they are comparable to another models estimated by OLS. The results are presented on the following page.

Table 5.5: Results of OLS regression (FGLS in the bracket)

ln New_reg_pc	Coefficient	Robust St. Error	*t* – statistics	*P* - value
constant	-8.2070 (-8.0586)	.4243 (.5080)	-19.34 (-15.86)	0.000 (0.000)
New_price	-.0223 (-.0197)	.0045 (.0053)	-4.94 (-3.70)	0.000 (0.000)
Used_price	.0110 (.0100)	.0024 (.0028)	4.62 (3.59)	0.000 (0.001)
Fuel_price	.0018 (.0016)	.0010 (.0012)	1.76 (1.37)	0.081 (0.175)
Real_GDI	.0130 (.0117)	.0130 (.0046)	3.40 (2.53)	0.001 (0.013)
February	-.0090 (-.0087)	.0515 (.0443)	-0.18 (-0.20)	0.861 (0.845)
March	.2662 (.2666)	.0491 (.0460)	5.42 (5.79)	0.000 (0.000)
April	.2977 (.2976)	.0483 (.0481)	6.16 (6.18)	0.000 (0.000)
May	.2915 (.2915)	.0423 (.0426)	6.90 (6.84)	0.000 (0.000)
June	.3409 (.3411)	.0430 (.0426)	7.93 (8.01)	0.000 (0.000)
July	.1223 (.1223)	.05644 (.0567)	2.17 (2.16)	0.033 (0.033)
August	.1010 (.1009)	.0416 (.0422)	2.43 (2.39)	0.017 (0.019)
September	.0874 (.0873)	.0447 (.0450)	1.96 (1.94)	0.053 (0.055)
October	.2231 (.2231)	.0511 (.0501)	4.37 (4.45)	0.000 (0.000)
November	.2388 (.2381)	.0387 (.0380)	6.17 (6.27)	0.000 (0.000)
December	.2279 (.2248)	.0419 (.0383)	5.44 (5.87)	0.000 (0.000)
Time	.0016 (.0015)	.0003 (.0004)	4.72 (3.80)	0.000 (0.000)
	n = 120 (120)	$R^2 = .6805\ (.7944)$		

Source: author's computations

To test a serial correlation of disturbances the Durbin-Watson statistics was computed. With a value of 1.685 it is lying between the lower and the upper bound and it did not allow to make a conclusion about the serial correlation. To find out whether a serial correlation is present the residuals were predicted and tested whether they follow an AR(1) process. According to results we do not have enough evidence to reject the null hypothesis of no autocorrelation at 10% significance level, but because of inconclusive results there were also computed Feasible Generalized Least Squares (FGLS) estimates using the Prais-Winsten method. The FGLS results must be interpreted very carefully, since the data are persistent (although not I(1)). Performing the Breusch-Pagan test for heteroskedasticity resulted in the F-statistics of 2.15 (2.00 in case of FGLS), which allowed us to reject the null hypothesis of no heteroskedasticity present at 5% significance level. Hence, the heteroskedasticity of error terms is concluded and therefore the heteroskedasticity robust standard errors were computed. The R-squared of the regression is 0.6805 in case of OLS and 0.7944 in case of FGLS, which tells us that the data explain the variation in the dependent variable quite well.

The F-test for the joint significance of the monthly dummy variables resulted in the F-statistics accounting for 18.07 (15.50 in case of FGLS). Therefore, the null hypothesis of no seasonality was rejected at 1% significance level and there is concluded that the seasonality is present. A similar result was obtained in case of the linear time trend. The t-statistics 4.72 (3.80 in case of FGLS) allowed to reject the null hypothesis of no time trend at the 1% significance level and assume an increasing linear time trend in the dependent variable.

The regression coefficients of all explanatory variables except the fuel price turned out to be statistically significant at 1% significance level. The fuel price seems to have only a very low impact according to the regression coefficient. Since both the OLS and the FGLS results might be suffering from SEM bias, the variable was not excluded from the model. Results are interpreted only for the OLS estimates, but the results obtained by FGLS are generally consistent with them.

The effect of new cars CPI can be interpreted as that a 1% yoy change in the price with respect to the previous year causes a change in the dependent variable by 2.23 % in the opposite direction. According to this result, new cars are price elastic ordinary goods. This is in variance with results of previous researches of similar structure, which presented cars to be price-inelastic. But it is necessary to interpret this result carefully since the indicated decrease of demand is specified only for new cars and it assumes that at least a part of the lost customers would buy a used car instead.

A 1% yoy change in the used cars CPI causes a change in the dependent variable by 1.10 % in the same direction. That confirms the expectations that used cars are gross substitutes to new cars.

A 1% yoy change in the fuel CPI causes a change in the dependent variable by 0.18 % in the same direction. This result is inconsistent with the theory, which says that the fuel price should be negatively correlated with the car demand since fuel and cars are complementary goods. Because the aggregated registrations do not take into account car size and consumption, the possible changeover of consumer preferences when the fuel price increases could not be captured. Similar results were obtained by Witt and Johnson (1986) or by Lee and Kang (2008).

If the real disposable income increases by 1 % in comparison to the same period of the previous year, it causes a change in the dependent variable by 1.3 % in the same direction. According to this result, in terms of the microeconomics theory new cars are normal goods. The estimated elasticity shows that cars are rather luxury goods and this outcome is consistent with results of other researches. Results would probably differ for various car classes and types.

5.3.2 2SLS regression

Recalling the SEM model the endogenous variable in the demand equation is the new cars price. Unfortunately, from the available supply shifters only the Czech PPI fulfilled the requirements to serve as a proper instrument for the new cars CPI. It is a rather poor instrument since its exogeneity is not assured, nevertheless, it should provide more precise results than the OLS estimation.

Table 5.6: Results of 2SLS regression

ln New_reg_pc	Coefficient	Robust St. Error	t – statistics	P - value
constant	-8.739763	.4840266	-18.06	0.000
New_price	-.0360217	.007979	-4.51	0.000
Used_price	.0159076	.0033635	4.73	0.000
Fuel_price	.0029051	.0011597	2.51	0.012
Real_GDI	.0177813	.0043512	4.09	0.000
February	-.0096909	.0485426	-0.20	0.842
March	.2658874	.0489149	5.44	0.000
April	.2995412	.0458833	6.53	0.000
May	.2933667	.0413027	7.10	0.000
June	.3416299	.0406371	8.41	0.000
July	.1237267	.054322	2.28	0.023
August	.1028095	.0409852	2.51	0.012
September	.0897715	.0436279	2.06	0.040
October	.2246628	.0488029	4.60	0.000
November	.2423317	.03773	6.42	0.000
December	.2306974	.0409112	5.64	0.000
Time	.0018692	.0003616	5.17	0.000
n = 120	$R^2 = .6542$	New_price instrumented		

Source: author's computations

To test a serial correlation of disturbances the residuals were predicted and tested whether they follow an AR (1) process. The F-statistics 1.60 and the corresponding p-value 0.21 did not provide enough evidence to reject the null hypothesis on no serial correlation. Therefore it is concluded that the serial correlation is not present. The Breusch-Pagan test for heteroskedasticity resulted in the F-statistics 2.44 and the p-value 0.0037. The null hypothesis of no heteroskedasticity could be rejected at 5% significance level and it is assumed that the heteroskedasticity is present. To avoid the resulting inaccuracies there were computed heteroskedasticity robust standard errors. The R-squared of the regression is 0.6542. In comparison with the OLS regression it slightly decreased, which is an expected consequence of using the instrumental estimation.

The test for a joint significance of the monthly dummies allowed to reject the null hypothesis of no seasonality at 1% significance level. The positive coefficient of the linear time trend showed to be significant as well, the null hypothesis of no linear time trend was rejected at 1% significance level. The p-value of the fuel price coefficient decreased to 0.012 and confirmed its importance in the model. All the remaining explanatory variables are significant at 1% level of significance.

Application of the 2SLS method confirmed the hypothesis that the OLS estimates are underestimated. The impact of all the explanatory variables increased and it can be interpreted as follows.

A 1% year on year increase of the new cars CPI causes a decrease of the new cars registrations by 3.6 % and a 1% increase of the used cars CPI causes an increase of the new cars registrations by 1.59 %. According to this result, new cars are price elastic ordinary goods and the used cars confirmed the expectations to be their gross substitutes. It is necessary to interpret this result as that the indicated decrease of demand is specified only for new cars and it assumes that at least a part of the lost customers would buy a used car instead. Moreover, since the prices of new and used cars showed to be correlated, it can be expected that a change in new cars prices would induce change in prices of used cars as well. The final impact of the price change is therefore only tentative.

A 1% year on year change in the fuel CPI causes a change in the dependent variable by 0.29 % in the same direction. This result is inconsistent with the theory, which says that the fuel price should be negatively correlated with the car demand since fuel and cars are complementary goods. Because the aggregated registrations do not take into account car size and consumption, the possible changeover of consumer preferences when the fuel price increases could not be captured. Also the causality of the fuel price change can be questioned since an increasing fuel price might be an impact of the rising disposable income and the price level.

If the real disposable income increases by 1 % in comparison to the same period of the previous year, it causes a change in the dependent variable by 1.78 % in the same direction. According to this result, in terms of the microeconomics theory new cars are normal goods. The estimated elasticity shows that cars are rather luxury goods. Results would probably differ for various car classes and types, but generally confirm the theory that when the income increases and the future expectations about the economy development are optimistic, consumers tend to buy more cars.

The largest impact on the quantity of new cars demanded has time of the year. In spring months and before the end of a year, the new cars registrations are significantly higher than in rest of the year. The lowest demand is regularly observed in January and February, when the monthly registrations represent approximately only 70 % of monthly registrations recorded in spring or before the end of a year. There may be several reasons for such a behavior. A combination of subjective preferences and a tax optimization is possible. Whereas private customers prefer to buy a new car in summer instead of winter, companies purchase their car fleets before the end of year to optimize their tax costs. A better knowledge could bring a future research of consumer preferences.

6 Conclusion

The main goal of this thesis is to estimate the new cars demand in the Czech Republic using a model constructed with respect to local market specifications. Although the model is inspired by previous researches on this topic, its main contribution is taking into account also the new cars supply as a force which also affects the traded quantity of new cars.

At first, factors affecting the new cars supply and demand are derived using the microeconomic and the macroeconomic theory. New cars price, used cars price, fuel price and consumer income are expected to be the most important demand shifters, whereas as supply affecting factors the new cars price, producer costs and the exchange rate are assumed. Since the supply and demand approach involves the need to estimate the simultaneous equations model (SEM), the SEM estimation theory is introduced. It is shown that the OLS estimates are inconsistent in case of the SEM model and that with use of the instrumental variable (IV) we are able to obtain the consistent results.

Data used to estimate the derived model represent the chosen explanatory factors in the space of the Czech market environment. The consistent monthly data are available for the years 2004 to 2013, hence this period is subject to the study. Important finding of the research is that the new passenger cars registrations in the Czech Republic suffer from bias caused by a change of the policy in April 2009. Until then, a large part of all passenger cars were registered as light utility vehicles because of the taxation policy, which did not allow to deduct the VAT in case of passenger cars. To ensure the unbiasedness of the measure of new cars registrations over the studied period, the aggregated registrations for passenger cars and light utility vehicles had to be taken into account.

Finally, the demand for cars is estimated by the OLS and 2SLS methods. Results confirm the hypothesis that the OLS estimates are underestimated in comparison with those obtained by the 2SLS. As expected, the number of new cars demanded by Czech consumers depends on the price of new cars, the price of used cars, consumer income, and the time of the year. It is shown that new cars are luxury goods, as the estimated income elasticity of demand is approximately 1.8. The demand for new cars also appears to be price elastic. In tight relationship with the new cars price is the price of used cars, which are considered as close substitutes to new cars. According to the results, the number of new cars demanded is highly sensitive to the new and used cars price ratio. More details about this relationship could be explained by future research. The estimated results are at variance with the expected impact of fuel price on the demand. New cars registrations do not seem to be negatively influenced by the price of fuel, but the research did not take into account particular car classes. A disaggregate approach, which would explain more about the fuel price impacts on consumer behavior, was not performed because of data inconsistency. The quantity demanded for new cars is probably influenced at most by the time of the year, since the spring and summer months reported up to 30% increase of registrations in comparison to the winter time. It is assumed that the main reason for such an increase is a subjective perception of consumers, but an additional research could discover more causes of this seasonality.

Bibliography

Automotive Industry Association (2014). *Statistics* [online]. Available from: http://www.autosap.cz/en/zakladni-prehledy-a-udaje/ [Accessed 2014-05-10]

Car Importers Association (2014). *Statistics* [online]. Available from: http://portal.sda-cia.cz/stat.php?m#str=nova [Accessed 2014-05-10]

Carlson, R. (1978). Seemingly Unrelated Regression and the Demand for Automobiles of Different Sizes, 1965-75. *The Journal of Business*, vol. 51, issue 2, pp. 243-62.

Carlson, R. & Umble, M. (1980). Statistical Demand Functions for Automobiles and Their Use for Forecasting in an Energy Crisis. *The Journal of Business*, vol. 53, issue 2, pp. 193-204.

Czech National Bank (2014). *ARAD – Time series data system* [online]. Available from: http://www.cnb.cz/docs/ARADY/HTML/index.htmEuropean Automobile Manufacturers [Accessed 2014-05-10]

Czech statistical office (2014). *Time series* [online]. Available from: http://www.czso.cz/eng/redakce.nsf/i/time_series [Accessed 2014-05-10]

CzechInvest (2014). *Automotive industry* [online]. Available from: http://www.czechinvest.org/en/1automotive-industry [Accessed 2014-05-10]

CzechInvest (2014). *Automotive Industry in the Czech Republic* [online]. Available from: http://www.czechinvest.org/data/files/auto-obalka-vnitrek-97-en.pdf [Accessed 2014-05-10]

Dougherty, C. (2011). *Introduction to econometrics*. 4th ed. New York: Oxford University Press.

European Automobile Manufacturers Association (2014). *The Automobile Industry Pocket Guide 2013* [online]. Available from: http://www.acea.be/uploads/publications/POCKET_GUIDE_13.pdf [Accessed 2014-05-10]

Eurostat (2014). *Short term business statistics* [online]. Available from: http://epp.eurostat.ec.europa.eu/portal/page/portal/short_term_business_statistics/data/database [Accessed 2014-05-10]

Gujarati, D. (2004). *Basic econometrics*. 4th ed., Tata McGraw-Hill ed. New Delhi: Tata McGraw-Hill.

Hamilton, J. (1994). *Time series analysis*. Princeton: Princeton University Press.

Johnston, J. (1997). *Econometric methods*. 4th ed. New York: McGraw-Hill.

Koubek, I. (2014), *Textbook to courses Microeconomics I. and Microeconomics II.* Institute of Economic Studies, FSV, Charles University in Prague [online]. Available from: http://samba.fsv.cuni.cz/~koubek/IES-FSV%20UK/UCEBNI%20TEXTY/ [Accessed 2014-05-10]

Lee, H., Kang, H. H. (2008). The Study of Brazilian Automobile Demand. *In Decision Sciences Institute Annual Meeting Procreedings 2008*. Atlanta, Georgia. Available from: http://www.decisionsciences.org/Proceedings/DSI2008/docs/143-8905.pdf

Mankiw, N. (2012). *Principles of Microeconomics*. 6th ed. Mason, OH: South-Western Cengage Learning.

McCarthy, P. (1996). Market Price and Income Elasticities of New Vehicle Demands. *The Review of Economics and Statistics*, Vol. 78, No. 3 ,pp. 543-547

Mikas, T. (2011). *Faktory ovlivňující poptávku po osobních automobilech v ČR*. Diploma thesis. Pardubice: University of Pardubice, Jan Perner Transport Faculty. Available from: http://dspace.upce.cz/bitstream/10195/39764/1/MikasT_FaktoryOvlivnujici_KP_2011.pdf [Accessed 2014-05-10]

Ministry of Transport (2013). *Transport Yearbook Czech Republic 2012* [online]. Available from: https://www.sydos.cz/cs/rocenka_pdf/Rocenka_dopravy_2012.pdf [Accessed 2014-05-10]

Portal of the Public Administration (2014). *Search in legislation* [online]. Available from: http://portal.gov.cz/app/zakony/?path=/portal/obcan/ [Accessed 2014-05-10]

PricewaterhouseCoopers (2014). *Vývoj českého automobilového trhu Q1 2014* [online]. Available from: http://portal.sda-cia.cz/clanky/download/2014_04_Analyza_PwC.pdf [Accessed 2014-05-10]

Rees, H. (2004). *Microeconomics*. 3rd ed. New York: Pearson Education.

Samuelson, P. (2005). *Economics*. 18th ed. Boston: McGraw-Hill/Irwin.

Suits, D. B. (1958). The Demand for New Automobiles in the United States 1929-1956, *The Review of Economics and Statistics*, vol. 40, No. 3, pp. 273-280.

Thomassen, O. (2007). *A Pure Characteristics Model of Demand for Cars*, University of Oxford. [online]. http://dev3.cepr.org/meets/wkcn/6/6652/papers/Thomassen.pdf/ [Accessed 2014-05-10]

Witt, S. & Johnson, R. (1986). An econometric model of new-car demand in the UK. *Managerial and Decision Economics*, vol. 7, issue 1, pp. 19-23.

Wooldridge, J. (2009). *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, OH: South Western, Cengage Learning.