Data Mining Application for Used-Car Market Analysis

Valerijs SKRIBANS
Riga Technical University,
Kalnciema street 6, office 216, Riga, LV1048, Latvia

Scott W. HEGERTY
Northeastern Illinois University,
5500 N. Saint Louis Ave. Chicago, IL 60625, USA

ABSTRACT

The monitoring of used-car markets began on January 3, 2018, when the Federal Administrative Court in Leipzig allowed cities and communities in Germany to impose bans on diesel cars in order to reduce the level of nitric oxide (NOx) in the air. This decision can radically change the used-car market in the EU, because Germany exports large quantities of used cars to Eastern Europe. This article analyses the influence of the German Court decision on the used-car market in Latvia, as part of the EU market. For this analysis, data for offers on 258,600 used cars in the Latvian market and nearly 800,000 cars in the German market were collected. Based on these data, an econometric model was developed to describe the relation between the price of these used cars and their main characteristics, which include 51 important elements such as issue year, mileage and so on. The model describes 89% of the variance in market prices (adjusted $R^2 = 0.89$). This article concludes that in the short-run, the export of used cars is profitable for both Germany and Eastern Europe. However, Eastern European countries do not pay enough attention to this "recycling" of cars In the long run, which may lead to a request for financial assistance from the EU to solve environmental problems.

Keywords: Data Mining, Used-Car Market, Econometric Model, Transport, Export, Environmental Problems, Prices.

1. INTRODUCTION. DATA MINING AND DESCRIPTION

The monitoring of used-car markets was started when Federal Administrative Court in Leipzig allowed cities and communities in Germany to impose bans on diesel cars in order to reduce the content of nitric oxide in the air. This decision can radically change used car market in EU. The article estimates the influence of this decision on the used-car market in EU.

The article's hypothesis is that with this ban on diesel cars, their use has become expensive and ineffective. Accordingly, the price of cars with diesel motors should decrease. Prices for these cars will reach the point where any associated penalties from using them will not compensate for any price differentials versus gasoline cars. The next influencing factor is the prices of used cars in other countries. The possibility of exporting used cars to other countries can smooth out any negative effects on the domestic market. Among the potential countries for exports are countries in Central and Eastern Europe, the Baltic, the Balkan region, and the former Soviet Union.

To conduct this research, monitoring of the used car market was done in Germany and Latvia from January until October, 2018. Latvia was chosen as an example of potential countries for export.

In this used-car-market monitoring, the largest used- car sales web-services in Germany were analyzed (mobile.de [1] and autoscout24.de [2]) and the transport sales section from biggest

advertising site in Latvia (ss.lv [3]). In one period for Germany, nearly 800 thousand passenger cars offers were analyzed, which is nearly 2% of the 45 million German used cars, according to European Automobile Manufacturers' Association data [4]. For Latvia, in this period we analyzed nearly 17,000 passenger-cars offers, which is nearly 2.6% of the 680,000 Latvian used cars (according to European Automobile Manufacturers' Association data [4]). The collected database is published in a data repository and in an associated data paper [5].

The first discussion question is: how representative are these data? There is a difference between a car in use and a car in the market; for example, if the average life of car in Germany is 9 years, it will be offered for a market deal at least one time in this 9-year period (before utilization). Accordingly, the number of cars in the market can be nine times less than the number of cars in use, or nearly 5 million. In this case, our research analyzes 16% of the market and can considered to be representative. But in reality, we must take in to account the fact that during a car's lifetime it is s sold more than one time, in which case our research data represent not less than 25-30% of market. Why is this question typical? The authors underline that mobile.de or autoscout24.de data are not market data, but rather data only for these sites, but after that data checking, selection and aggregation, conclusions about market representation must be done. It is important to note that the aforementioned sites have more offers than are analyzed in this research. For example, mobile.de has 1.3 million offers in Germany (for cars issued after 1996). Data from these sites were cleaned, with "fake" offers (for example new car for 1 Euro), some duplicate offers, and offers with missing data removed.

In the case of Latvia, which is analyzed here, the data are more representative, and it is possible to tell that our data represent the full market. It is related to the specifics of local market. In Latvia, there is no other alternative for car offers, selling and advertising. According to the Latvian Vehicle Trade Register Information System [6], which was established on the basis of the Latvian Cabinet of Ministers Regulations No.876 from 18 December 2007 "Regulations for Trading of Vehicles and Their Numbered Units" [7], in the Latvian market at the same time were a little bit less than 6,000 car offers (including new cars). The official trading registry doesn't represent the market situation, since research analyzes three times more offers.

In the car market, we collected information about the following characteristics: car price, producer, model name, auto first registration date, motor cubic capacity, fuel type, gearbox type, mileage, color, vehicle type, the date of the next technical inspection, vehicle condition, the number of seats, door count, the presence of an emissions sticker, the number of vehicle owners, airbags, interior design, and features.

Additional limitations to this research were as follows: This research does not consider super cars, or cars with price higher than 99,900 Euro, does not consider retro and vintage cars, and limits the first production year to 1996. This limitation was related to the statistic, for example, the in Latvian market, the proportion of cars produced before 1996 is not more than 1% of the total.

2. CAR MARKET ANALYSIS

Car markets in Germany and most countries in the post-Soviet Eastern Bloc countries are very different. Figure 1 shows offers in the German and Latvian car markets by car issue year.

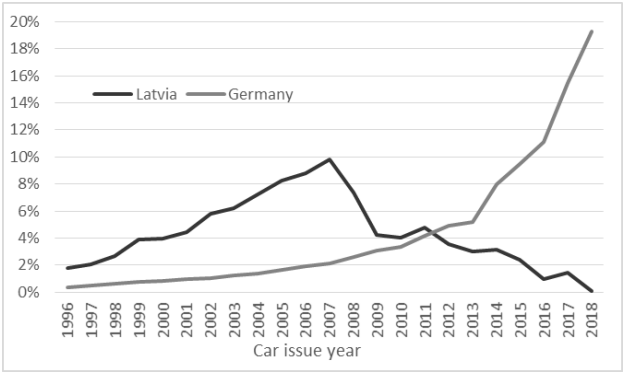

Fig. 1. Proportion of cars by issue year in Latvia and Germany in October 2018.

In the first figure, it is visible that relatively new cars (under 4 years old) make up the biggest proportion of cars in Germany, and the corresponding proportion exponentially decreases for each year. It is possible to tell that this situation is absolutely normal; nearly 20% of cars are sold on the primary market (new cars), and after cars' amortization they are sold on the secondary market and leave the market.

The situation in the Latvian market is completely different. The amount of new cars in the market is very small. The proportion of cars in the last issue year constant increases in a linear fashion over 12 years. This means that the market has an external source that adds used cars to market. According to LR CSB data, Germany supplies 30% of car imports into Latvia; the next two biggest importers are Estonia (18%) and Poland (16%) [8]. It is highly likely that Germany not only exports cars to Latvia directly, but also uses re-export possibilities (for example, from Poland and/ or Lithuania). In Latvia, 12 years is a breakpoint for car imports and utilization. After this, the proportion of old cars decreases each year.

There are different reasons for the decrease of used-cars imports to Latvia: First of all, it is an absolutely normal situation, when there is any limit, after which people do not buy used cars, and a second import source also is limited. However, research data show that there is an economic reason, which is shown Figure 2

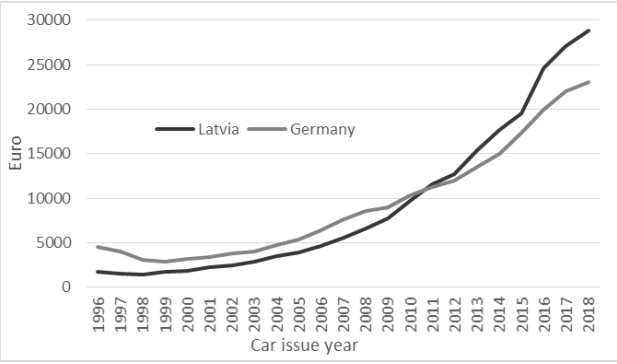

Fig. 2. Average car price comparison in Latvia and in Germany by cars issue year in October 2018.

If we compare used car prices in Latvia and in Germany by issue year, we see that for small and medium used cars (up to 11 years old), prices are higher in Latvia. This is the main reason why Germany exports used cars to Latvia. A similar situation is for other former Eastern Bloc countries. We also see that for older cars, prices in Germany are higher. It is therefore absolutely logical that a car should be able to be returned (re-exported) to Germany. However, there are no reverse flows. It follows that the key factors behind this include not only the price, but also the quality of cars, which is compared in Figure 3.

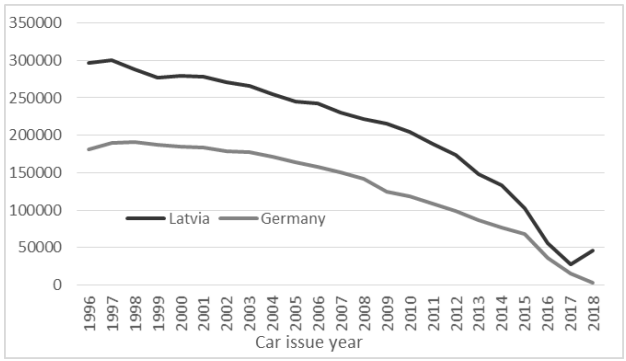

Fig. 3. Average mileage comparison in Latvia and in Germany by cars issue year in October 2018.

How can we understand these differences in quality? If cars have the same producer, model, color, issue year, and condition (incidents), to find cars with close mileage is hard. But mileage is important indicator, which characterizes the quality of cars. Research data show (seen in Fig. 3), that in Germany, cars with mileage over 200,000 km are considered to have near zero value. In Latvia, a car with similar mileage can serves additional 15 years. It is provided that based on quality, car re-exports from Latvia to Germany are not possible.

3. MODEL DEVELOPMENT. RESEARCH

Differences in automobile quality in Germany and Latvia underline the question: How much will Latvians be willing to pay in Latvia to accept cars from Germany? To answer this question, the following econometric model was developed, as shown in Equation (1):

$$P_i = a_1 I_i + a_2 M_i + a_3 MTD_i + a_4 MTB_i + a_5 MTG_i + \\ + a_6 MTH_i + a_7 G_i + a_8 MI_i + \sum_{k=1}^{13} a_n C_{ki} + a_{22} ME_i + \\ + \sum_{k=1}^{9} a_n T_{ki} + \sum_{k=1}^{21} a_n P_{ki} + a_{53} TI_i + a_{54} SET_i \quad (1)$$

Where, for all cars *i*:
P_i – price;
I_i – Issue year;
M_i – Motor cubic capacity;
MTD_i – Dummy variable for cars with a diesel motor type;
MTB_i – Dummy variable for cars with a benzoyl motor type;
MTG_i – Dummy variable for cars with a gas system;
MTH_i – Dummy variable for cars with a hybrid system;
G_i – Dummy variable for cars with an automatic gearbox;
MI_i – Mileage;
C_{ki} – Dummy variable for cars with *k* color (13 colors in total);
ME_i – Dummy variable for cars with "metallic" option in color;
T_{ki} – Dummy variable for cars with *k* type (9 types: convertible / roadsters; mini; minivan; off road / pickup; others; sedan; sport / coupe; universal; unknown);
P_{ki} – Dummy variable for cars from *k* producer (21 producers in total, for example, Audi, BMW and so on);
TI_i – Time (in months) to the next technical inspection for the car *i*;
SET_i – Set (features) (car complication – no complication 0 class; only one, for example AirBug – 1 class, and so on, till 98).

Developed model quality estimation data are following:

```
Residual standard error: 2643 on
21796 degrees of freedom
Multiple R-squared:  0.8916,
Adjusted R-squared:  0.8913
F-statistic:  3320 on 54 and 217
96 DF,  p-value: < 2.2e-16
```

Our model concludes that for every year on the road, a car loses 522 Euros of value; that the price of each cubic centimetre of motor capacity is 1172 Euros; that the difference in price for cars with different motors and fuel types, or colors, is minimal; that the price for cars with automatic gearboxes are 475 Euros higher for manual; that each kilometre of increased mileage decreases the price of a car by 0.0123 Euros; that the type and producer of car mostly are important for the consumer, but that characteristics have different influence; that the time before the next technical inspection costs 10 Euros per month (for the actual situation in Latvia, it is similar to technical inspection cost calculated to one month); and that in some conditions set (features) of the car in total can increase price up to 6500 euros (that is real for full set or luxury cars).
From the model that we estimate R^2=0.89, indicating good fit. We also tested several alternative models, including only with main characteristics, such as issue year and mileage. Simpler models do not insure higher quality. This shows that the market in question depends not only on quantitative details, but also on consumer preferences.

4. RESULTS AND CONCLUSIONS

First, we apply our model for the German case, as is seen in Figure 4.

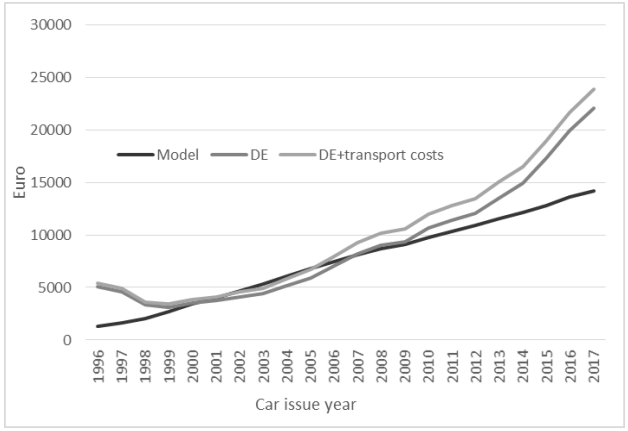

Fig. 4. Application of the Latvian used-car price model for the German market

In our results, we see that car prices in Germany are similar to the Latvian model for 10-20 year old cars or for cars in price range

from 2500 Euros to 9000 Euros. In Latvia, vintage autos issued before 1999 but still in good technical condition are very rare; cars in this group are not in demand, and in Germany their prices are significantly higher. For small used cars, or for cars with prices over 10,000 Euros, our model does not work because this is the border where used cars start to compete with new cars in Latvia. Cars priced at a little more than 11,000 Euros have different producers in Latvia, such as Hyundai, Škoda and so on.

Based on our research data, it can be concluded that the prohibition of old diesel cars (with high CO_2 emissions) in Germany do not change the used car market in EU. Because there is not a similar limitation in other EU countries, it is possible to export an unwanted car from Germany to another EU country without losing any value of the cars, for example in former Eastern Bloc countries.

However, this potential or possibility does not ensure that such an export will take place. To clarify the question, we must see whether the diesel car problem applies to used-car owners in Germany, and what changes was in Latvian market.

In Figure 5, we compare used diesel car markets in Latvia and Germany.

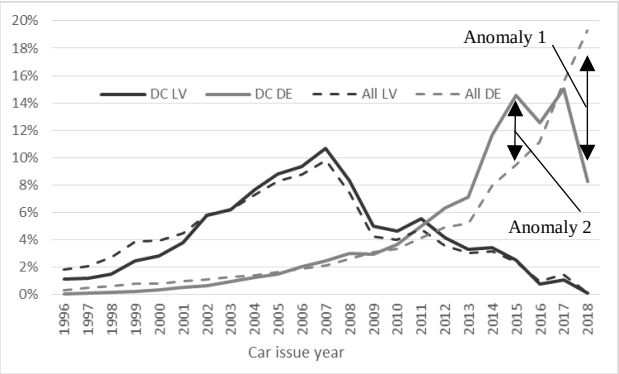

Fig. 5. Proportion of diesel cars by issue year in Latvia and Germany in October 2018.

In this figure, we see that in Latvia the used diesel car market is very similar to the entire used car market; the difference is insignificant. However, in Germany there are anomalies. In this market, the proportion of diesel cars produced in the last year (2018) was roughly 11% less than the proportion of all cars (anomaly 1). This means that with a decrease in diesel car production, this car's proportion of the market decrease too. An additional increase in this market's proportion of diesel cars produced 2-6 years ago (anomaly 2) confirms, that diesel cars have low demand in Germany. Consumers try to get rid of them as soon as possible. In this situation in the market, there are only two possibilities: export these cars or sell them on the domestic market after a price cut. It is most likely that with a decrease in prices in the domestic market, the export potential increases and the cars will be exported.

The decrease in diesel cars it is an established and well-known trend in the EU; statistics about it are shown in Table 1.

Table 1.
Diesel cars' proportions in several EU countries

	2015	2016	2017
EU15	52.1	49.9	44.8
DE	47.7	45.8	38.7
LV	49.8	51.8	58.0

Note: EU15, DE and 2015 year for LV are ACEA data [9]. 2016 and 2017 year for LV estimated by authors. DE 2017 year full data matches with author estimation (38.743%) in 2018 (in October 2018).

From Table 1 it is visible, that the proportion of diesel cars on the market decrease for Germany and EU15, but increase in Latvia. Multidirectional dynamics confirm the assumption that cars are not utilized, but are exported, partly to Latvia. Germany and the developed countries of Western Europe at make decisions on environmental protection the legislative level. For the countries of Eastern Europe, environmental protection issues are secondary issues. The biggest problems of Eastern Europe are associated with economic growth and national security (reducing the influence of Russia in the region). In these conditions, the principle of solidarity in the EU does not work. Due to economic efficiency, Eastern European countries are turning into a landfill. Figure 6 depicts the newest data about

the proportion of diesel cars in the Latvian market.

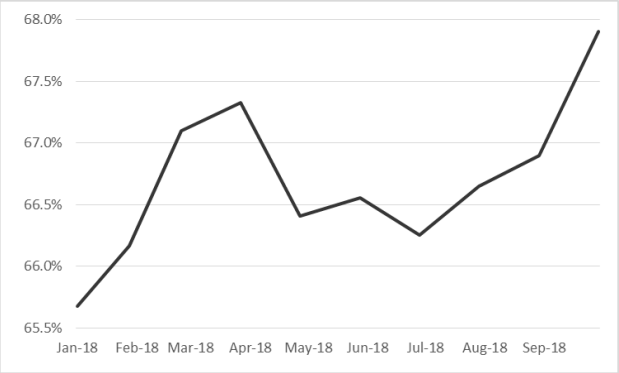

Fig. 6. Dynamics of the proportion of diesel cars in the Latvian used car market.

We see in Figure 6 that in 2018, the proportion of diesel cars in the Latvian used-car market reached its historical maximum. This is explained by both changes in the German used-car market and the desire of Latvian consumers to use cars that are less expensive to operate.
Despite the benefits on both sides, the situation is very dramatic. Eastern European countries do not pay enough attention to the recycling of cars. In the long run, this may lead to a request for financial assistance from the EU to solve environmental problems. Accordingly, despite the immediate benefits for both sides, the situation is not profitable in the long run. To eliminate this situation, uniform standards of quality and environmental protection are needed for all EU members.

5. ACKNOWLEDGMENTS

This study was funded by European Regional Development Fund (ERDF), Measure 1.1.1.5 “Support to international cooperation projects in research and innovation of RTU”. Project No. 1.1.1.5/18/I/008.

6. REFERENCES

[1] Mobile.de, Germany's Biggest Vehicle Marketplace Online. mobile.de, 2018 (accessed 10 November 2018).

[2] AutoScout24, a European automotive classifieds leader. autoscout24.de, 2018 (accessed 10 November 2018).

[3] Sludinājumi, Advertising website. https://www.ss.com/lv/transport/cars/, 2018 (accessed 10 November 2018) (in Latvian).

[4] ACEA, ACEA Report: Vehicles in use Europe 2017. https://www.acea.be/uploads/statistic_documents/ACEA_Report_Vehicles_in_use-Europe_2017.pdf, 2018 (accessed 10 November 2018).

[5] V. Skribans, Used-car market dataset for Latvia 2018, Data in Brief (2019), https://doi.org/10.1016/j.dib.2018.12.075i

[6] LVTRIS, Latvian Vehicle Trade Register Information System. https://e.csdd.lv/tltr/, 2018 (accessed 10 November 2018).

[7] LR MK, Ministru kabineta noteikumi Nr.876 Transportlīdzekļu un to numurēto agregātu tirdzniecības noteikumi (Cabinet Ministry Regulation No.876 Regulations for the sale of vehicles and their numbered assemblies). https://likumi.lv/doc.php?id=168576, 2007 (accessed 10 November 2018) (in Latvian).

[8] LR CSB, ATD200. Eksports un imports pa valstīm, valstu grupām un teritorijām (euro, KN 2 zīmēs) (Exports and imports by country, group of countries and territories) http://data1.csb.gov.lv/pxweb/lv/atirdz/atirdz__detalizeta__2zim/ATD200.px/table/tableViewLayout1/?rxid=a39c3f49-e95e-43e7-b4f0-dce111b48ba1, 2018 (accessed 10 November 2018) (in Latvian).

[9] ACEA, Trends in fuel type of new cars between 2016 and 2017, by country https://www.acea.be/statistics/article/trends-in-fuel-type-of-new-cars-between-2016-and-2017-by-country, 2018 (accessed 10 November 2018).