

## “LEMON” PERCEPTIONS? A COMPARATIVE ANALYSIS OF SECOND-HAND CARS’ PRICES IN ROMANIA AND GERMANY

Vladimir GHITA<sup>1,2</sup>, Denis IORGA<sup>3</sup>, Andreea DUTULESCU<sup>1,4</sup>, Mihai DASCALU<sup>1,4</sup>,  
Gheorghe MILITARU<sup>2</sup>

<sup>1</sup> R&D Department, Global Resolution Experts S.A., Bucharest, Romania.  
E-mail: [vladimir.ghita@upb.ro](mailto:vladimir.ghita@upb.ro)

<sup>2</sup> Management Department, University Politehnica of Bucharest, Bucharest, Romania.  
ORCID: <https://orcid.org/0000-0003-4650-8427>  
E-mail: [gheorghe.militaru@upb.ro](mailto:gheorghe.militaru@upb.ro)

<sup>3</sup> Interdisciplinary School of Doctoral Studies, University of Bucharest, Bucharest, Romania.  
ORCID: <https://orcid.org/0000-0001-5595-6315>  
E-mail: [denis.iorga@drd.unibuc.ro](mailto:denis.iorga@drd.unibuc.ro)

<sup>4</sup> Computer Science Department, University Politehnica of Bucharest, Bucharest, Romania.  
ORCID: <https://orcid.org/0000-0002-4815-9227>  
E-mail: [mihai.dascalu@upb.ro](mailto:mihai.dascalu@upb.ro)  
E-mail: [andreea.dutulescu@stud.acs.upb.ro](mailto:andreea.dutulescu@stud.acs.upb.ro)

**Abstract:** “Lemon” perceptions refer to beliefs about the existence of low-quality second-hand cars in the market. Such beliefs, as part of a wider socio-economic phenomenon named adverse selection, manifest themselves through risk-aversion attitudes, for instance, a decreased willingness to pay high prices for any used car, regardless of its actual quality. In the current work, we argue that the prevalence of “lemon” perceptions can be explored by comparing the prices of similar second-hand cars in two different markets. To this extent, the current work advances a comparative analysis of the selling prices of similar second-hand cars in Romania and Germany by using data from online platforms dedicated to selling second-hand vehicles. Our results show, by examining the price differences of a diverse range of second-hand cars, that the price of the same car tends to be higher in the Romanian market than in the German market, irrespective of usage, but not of the brand. These results are both of practical value to stakeholders in the second-hand car market (i.e., buyers, sellers, policymakers) and of theoretical value as regards researching the adverse selection market mechanism.

**Keywords:** second-hand cars, adverse selection, prices, comparison, online

### INTRODUCTION

Akerlof [1] illustrates how buyers tend to be less willing to pay high prices due to their fear of buying “lemons” (i.e., low-quality cars) and lack of information regarding the quality of second-hand cars (i.e., uncertainty whether the car is “a lemon” or “a peach”) and, therefore, discourage the selling of “peaches” (i.e., high-quality cars) and eventually leading to market collapse. This socio-economic phenomenon defined as “adverse selection” has since then been extensively studied in various domains, such as the labor market [2], the healthcare domain [3], the management domain [4], the durable goods market [5], or the insurance market [6]. The current work expands on the adverse selection mechanism in the second-hand cars market by advancing a different angle for exploring the fear of buying “lemons.”

“Lemon” perceptions refer to beliefs about low-quality second-hand cars in the available market. Such perceptions lead to risk-aversion actions (i.e., the fear of buying “lemons”), for instance, a decreased willingness to pay high prices for any used car in the market. As this reduces the potential profit from selling high-quality cars, sellers of “peaches” (i.e., high-quality cars) are discouraged from selling their products. Therefore, the actual number of “lemons” (i.e., low-quality cars) in the market increases, reinforcing the buyers’ perception of a predominantly low-quality market. Consequently, the average price of used cars in the market tends to be lower, as buyers are hesitant to pay a premium due to the perceived risk of purchasing a lemon. This adverse selection dynamic contributes to a market equilibrium

with lower prices and a higher proportion of low-quality vehicles. The link between “lemon” perceptions, adverse selection, and lower prices represents the theoretical focus of the current work.

Based on the link between adverse selection and lower prices, it would be expected for second-hand car markets where adverse selection is more pronounced to display lower prices than second-hand car markets where adverse selection is less pronounced. The link between “lemon” perceptions and adverse selection further leads to the expectation that “lemon” perceptions are more prevalent among prospective buyers in second-hand car markets with lower prices than among those in second-hand car markets with higher prices. In other words, the prevalence of “lemon” perceptions in a second-hand car market is closely linked to the prices of goods circulating in the respective market. The current work explores this relationship by measuring price differences between two second-hand car markets.

Our work aims to understand whether similar second-hand cars have different prices in the Romanian and German second-hand car markets using data from online platforms dedicated to selling used vehicles. By comparing the prices of similar cars in Romania and Germany, we seek a proxy of the prevalence of “lemon” perceptions in the two markets. For instance, if the prices in Romania are significantly lower than those in Germany for the same second-hand cars, it might indicate a higher prevalence of “lemons” perceptions among Romanian buyers when compared with German buyers, given the link between prices, adverse selection, and “lemon” perceptions. However, the same results might indicate different strategies deployed in the two markets to compensate for the adverse selection mechanism. This topic covers a considerable proportion of the literature dealing with information asymmetries in the markets [7]. Nonetheless, both scenarios highlight the theoretical value of the current endeavor.

On the practical level, our work allows prospective buyers of second-hand cars to gain a deeper understanding of the price differentials and make informed decisions when purchasing used cars from either market. Similarly, sellers of second-hand cars can leverage the findings presented in the current work to optimize their pricing strategies and improve their competitiveness. Furthermore, policymakers can use the current work to identify potential market inefficiencies and implement measures to promote transparency and fair pricing practices. Finally, examining the price comparison can provide insights into the two markets' overall supply and demand dynamics. As such, the following section introduces the experimental procedure of the current work. Next, we introduce the results, followed by the discussion and conclusion sections.

## EXPERIMENTAL

The aim of this study is to compare whether prices of the same second-hand cars differ in Romanian and German markets. For this purpose, we used the dataset presented by Dutulescu et al. [8], which contains information from 30,264 second-hand car ads scraped from Autovit.ro and 1,308,575 second-hand car ads extracted from Mobile.de on March 2023. As data from the two online marketplaces for second-hand cars differed in terms of available information, we enforced a pre-processing procedure that sought to standardize information in both databases, which led to 22,046 second-hand car ads remaining in the Autovit.ro dataset and 1,108,045 second-hand car ads in the Mobile.de dataset. Both datasets contained information related to the car brand, model, year, usage/mileage/kilometers, cylindric capacity, horsepower, fuel type, type of transmission, and price.

The first step of the procedure involved identifying similar second-hand cars sold in the Romanian and German markets. We developed a script in Python to determine the most similar car ad in the German dataset for each car ad in the Romanian dataset. The most similar car ad in the German dataset for a given car ad in the Romanian dataset was defined as the car ad with the lowest  $d$  distance among car ads in the German dataset with the same model, maker, fuel, and transmission as the car presented in the Romanian ad. Distance  $d$  was defined as the Euclidian distance between the numerical features (i.e., kilometers, capacity, horsepower, and year) of two cars.

The above-presented procedure resulted in a dataset of 22,046 pairs of similar cars. From the resulting dataset of pairs, we further selected only pairs with a similar year and a maximum difference of 100 kilometers in terms of mileage. The 100 kilometers threshold was selected based on the distribution of the mileage difference between all car pairs. This filtering process yielded a dataset of 2,379 samples, where each sample represented a pair of similar car ads from the Romanian and German datasets, including the prices for both the Romanian and the German cars.

The second step of the experimental procedure aimed to determine whether the Romanian prices were higher than the German prices for each pair of similar second-hand cars. We initially applied the Shapiro-Wilk test [9] to assess the data distribution and determine the appropriate parametric or non-parametric statistical test. Based on the test results, we employed the non-parametric paired Wilcoxon test [10] to analyze the price differences. The choice of a paired test was made because our design assumed that the same car received a price in the Romanian and German markets. All the procedures were conducted using the *R* base package [11].

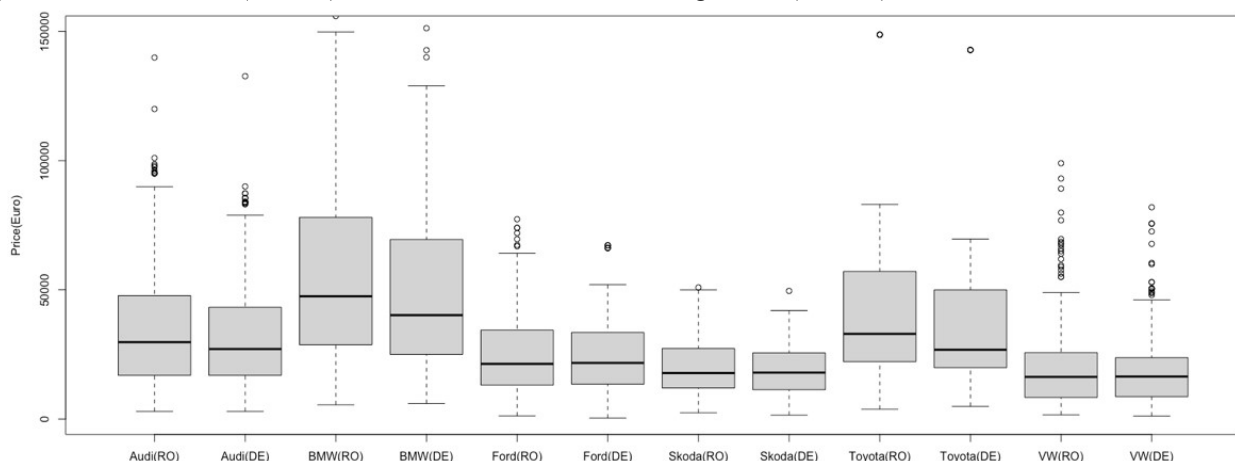
The third step of the analysis focused on conducting a similar price comparison while considering the usage and brand of the cars. We transformed the number of kilometers into distinct categories to categorize the mileage: almost new (0-10,000 km), good condition (10,001-50,000 km), decent condition (50,001-100,000 km), used (100,001-200,000 km), and very used (200,001-300,000 km). We excluded extremely used cars ( $n = 7$ ) with more than 300,001 km as we considered this to be a threshold that would leave no uncertainty regarding the quality of the car. Regarding the car brand, we selected only brands with more than 100 samples in the dataset, namely: Audi, BMW, Ford, Opel, Skoda, Toyota, and Volkswagen. Subsequently, we created separate datasets for each usage category and brand category and repeated the previously described procedure, including the Shapiro-Wilk test, paired Wilcoxon test, as well as boxplots. The decision to consider a difference in price as significant took into account the  $p$  values after the Bonferroni corrections applied on the .05 level. It should be noted that the price limit on the y-axis for the boxplot visual representations was set at 100,000 and 150,000 euro to ensure that the graphical representation was intelligible.

## RESULTS

This section introduces the results of the price comparisons of similar second-hand cars in the Romanian and German second-hand car markets. First, we analyzed the median differences and distribution differences between the Romanian and German prices for all similar second-hand cars. A statistically significant difference was identified between the prices for the same cars in the two markets ( $p < .001$ ), with a median difference of 986 euro. In other words, when comparing the prices of multiple similar cars sold in Romania and Germany, we would expect the prices in the Romanian market to be higher by approximately 1000 euro.

Next, we focused our analysis on price comparison by usage categories, as defined in the experimental section. Statistically significant differences between the median prices were found for each category ( $p < .001$ ). In all instances, Romanian median prices were higher than German median prices. The median difference for almost new cars ( $n = 896$ ) was 2,024 euro. The median difference for good cars ( $n = 327$ ) was 3,401 euro. Regarding second-hand cars defined as decent ( $n = 358$ ), the median difference was 2,127.5 euro. The median price difference for cars defined as used ( $n = 565$ ) was 600 euro. For cars defined as being very used ( $n = 226$ ), the median difference was 730 euro.

A similar analysis compared the median prices of similar cars by car brand (Figure 1). Statistically differences between the Romanian and German prices of similar cars were found for each considered brand ( $p < .005$ ), except Opel ( $n = 126$ ,  $p = .482$ ). Romanian prices were higher for Audi, BMW, and Toyota and lower for Ford, Skoda, and Volkswagen. The median difference for Audi second-hand cars ( $n = 223$ ) was 2,653 euro. The median difference for BMW second-hand cars ( $n = 123$ ) was 7,302.5 euro. Regarding Ford second-hand cars ( $n = 330$ ), the median price difference was 365.5 euro. The median price difference for Skoda second-hand cars ( $n = 201$ ) was 55 euro. The median difference for Toyota second-hand cars ( $n = 119$ ) was 6,152 euro, and for Volkswagen cars ( $n = 408$ ) was 148.5 euro.



**Figure 1:** Median Romanian (RO) and German (DE) prices for similar cars by brand type ( $n=1,407$ ).

## DISCUSSION

Our findings indicate that prices are higher in Romania than in Germany for similar second-hand cars. Based on the argument presented in the introductory section, this result suggests that the adverse selection phenomenon is more prevalent in the German second-hand car market compared to the Romanian second-hand car market. The lower prices for similar second-hand cars in the German market imply a higher level of perceived risk to buy a low-quality product associated with used cars in Germany. However, the higher prices for similar cars in Romania suggest that buyers in this market are relatively less concerned about the risk of purchasing a lemon. The results of the current work suggest that

buyers in the German market display more “lemon” perceptions than buyers in the Romanian market. Nonetheless, these results must be validated by directly measuring the prevalence of “lemon” perceptions in the two markets, as other confounding variables may explain the difference in prices (i.e., import costs).

On the one hand, the identified price difference for similar cars holds true irrespective of the car’s usage. Here, we observed that the price difference between Romanian and German markets is higher for almost new, good, and decent cars while being smaller for used and very used cars. The finding that the difference in prices is higher for cars with acceptable quality further supports the idea that “lemon” perceptions play a significant role, especially in the case of cars that can be sold as “peaches.” On the other hand, we found that prices for Audi, BMW, and Toyota vehicles were significantly higher in Romania compared to Germany by examining the price differences by brand. Conversely, prices for Skoda, Ford, and Volkswagen cars were lower in Romania, albeit with a smaller price difference. This observation suggests that the brand of the car has a certain influence on “lemon” perceptions. Buyers in the Romanian market may associate certain brands with higher quality and reliability, leading to an increased willingness to pay higher prices. In contrast, other brands may be perceived as having a higher risk of being lemons, resulting in lower prices.

## CONCLUSIONS

This study examined the adverse selection market mechanism by comparing prices of similar second-hand cars in the Romanian and German markets, using second-hand car ads data from Autovit.ro and Mobile.de. Our analysis conceptualized the differences in prices as indicators of the prevalence of “lemon” perceptions among the buyers in the markets under study. However, it is important to acknowledge the study’s limitations, including the lack of a validation measure for “lemon” perceptions and assumptions made about similarity based on available dataset features. In this sense, further work should consider variations in unrecorded factors such as add-ons and the color of cars and also focus on measuring the actual “lemon” perceptions in the two markets.

## ACKNOWLEDGMENT

This work was funded by the “Automated car damage detection and cost prediction – InsureAI” project, Contract Number 30/221\_ap3/22.07.2022, MySMIS code: 142909.

## REFERENCES

- [1]. Akerlof, G.A., *The market for “lemons”: Quality uncertainty and the market mechanism*. The quarterly journal of economics, 1970. **84**(3): p. 488-500.
- [2]. Greenwald, B.C., *Adverse selection in the labour market*. The Review of Economic Studies, 1986. **53**(3): p. 325-347.
- [3]. Frank, R.G., J. Glazer, and T.G. McGuire, *Measuring adverse selection in managed health care*. Journal of Health Economics, 2000. **19**(6): p. 829-854.
- [4]. Harrison, P.D. and A. Harrell, *Impact of “adverse selection” on managers' project evaluation decisions*. Academy of Management Journal, 1993. **36**(3): p. 635-643.
- [5]. Hendel, I. and A. Lizzeri, *Adverse selection in durable goods markets*. American Economic Review, 1999. **89**(5): p. 1097-1115.
- [6]. Cohen, A. and P. Siegelman, *Testing for adverse selection in insurance markets*. Journal of Risk and insurance, 2010. **77**(1): p. 39-84.
- [7]. Wilson, C., *Adverse selection*. Allocation, information and markets, 1989: p. 31-34.
- [8]. Dutulescu, A., et al., *Car Price Quotes Driven by Data-Comprehensive Predictions Grounded in Deep Learning Techniques*. 2023.
- [9]. Razali, N.M. and Y.B. Wah, *Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests*. Journal of statistical modeling and analytics, 2011. **2**(1): p. 21-33.
- [10]. Lam, F. and M. Longnecker, *A modified Wilcoxon rank sum test for paired data*. Biometrika, 1983. **70**(2): p. 510-513.
- [11]. R Core Team. *R: A Language and Environment for Statistical Computing*. 2023.

### Corresponding Author:

Name and surname, Title: Mr. Vladimir Ghita

Full address: 313 Splaiul Independentei, Bucharest, Romania

Email: [vladimir.ghita@upb.ro](mailto:vladimir.ghita@upb.ro)