

Report on Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Bc. Radka Doležalová
Advisor:	Mgr. Petr Polák MSc.
Title of the thesis:	What is my car worth? Hedonic price analysis of the German used car market

OVERALL ASSESSMENT (provided in English, Czech, or Slovak):

Please provide your assessment of each of the following four categories, summary and suggested questions for the discussion. The minimum length of the report is 300 words.

Contribution

The author claims that her work is original from the perspective of focusing on e-commerce sources of data, such as AutoScout24 for the German used car market which provided by far larger dataset (near 51 000 observations) than used in previous analyses on the similar topic. The author further claims that the use of Bayesian model averaging is also novel, but I have little means to verify that claim, while at the same time I consider the Bayesian model averaging a standard method. Further, the author claims that her work is novel for including some new characteristics of cars, which is correct to a certain degree, please read below on my comments on possible multicollinearity purposefully introduced in her model by including both engine capacity and engine power variables.

What I can verify though, is the impact and relevance of analysis of this type, since the car producing industry and the secondary market of used cars play a significant role in German economy, as well as in the Czech Republic and many other countries. A clear view on the role of certain characteristics of a car on its price on the market of used cars is of high importance for a vast number of people.

Methods

To the best of my knowledge, the author used methods which are covered by the standard curriculum on the master level at IES (in particular Bayesian model averaging).

I am disappointed with the minimum exposure to the hedonic price analysis theory which is advertised in the title of the thesis. Chapter 4 serves mostly as a literature review on this topic and the author gave up on the reader in this part and just points out to the literature. Despite having a potential to present a nice mathematical background with roots in the works by Rosen and Nash, this section mostly confused me by clear negligence of providing specifics, in particular, I am not sure if the author decided on a convention that vectors, unless specified otherwise, are rows (vector z , page 16). Similarly, I am not sure if the mapping p maps R^n to R and if the author assumes its continuous differentiability, or if it is enough to assume p being Lipschitz continuous, or just lower semicontinuous? These details are important for the properties of the optimization/equilibrium problems in the background of the hedonic price analysis and influences the choice of the algorithm to solve these optimization/equilibrium problems. Few notes on the assumptions listed on page 1: the one but last leads to Nash equilibrium concept and the last one is a nonsense. From the perspective of the optimization theory, set of optimal points is always a subset of a set of feasible points and thus, I suspect, that the actual assumption is that the feasible set must be nonempty, since any point of optimum is by definition feasible.

Similarly, I am disappointed by development in chapter 7 (containing a large concentration of typos, such as Baye's theorem instead of Bayes theorem – or perhaps the author is convinced that the original name is Baye?) and beyond. The author does not differentiate between K , the total number of alternative models, and k , an index running between 1 and K . On page 52, the author mentioned two Markov Chain Monte Carlo methods (Birth-death sampler and Reversible-jump sampler) but it is not

Report on Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Bc. Radka Doležalová
Advisor:	Mgr. Petr Polák MSc.
Title of the thesis:	What is my car worth? Hedonic price analysis of the German used car market

clear which one is described in the subsequent three step procedure (if any, in that case, these methods remain a black-box for the reader).

What bugs me the most is the total lack of discussion about the testing of assumptions of the models used. From the reported results it is clear that the author eventually used OLS, but there is no discussion on testing for homoscedasticity, no multicollinearity etc. In particular, I am strongly convinced that the lack of testing for no multicollinearity is a huge mistake and that the author included variables that do correlate despite strong advisements in the literature which the author even mentioned on page 44, lines 4 and 5 from above for the specific case of variables describing the engine capacity and engine power. Similarly, I would expect a strong correlation between number of doors and number of seats. The purposely omitted variables stated in Table 7.2 are just those which would cause perfect collinearity as an effect of treating qualitative variables with Boolean dummies, but that itself does not guarantee no multicollinearity.

For a master's thesis at IES, I expected a bit more focus on description of the advanced methods and loads of technical details about the analysis performed rather than lengthy exposure to literature review and justification of inclusion of some variables, most of which are absolutely logical and hard to object to.

Literature

The list of references follows near flawless uniform citation style and I also find the number of items on the list of references to be medium to large for the length and depth of this manuscript. Also, the author uses clear citation form in throughout the text.

On the other hand, despite section 5 called „Current literature review“ which contains a bit too many references on a little over 4 pages of the text with little to no viable details about the specific items, the author writes also chapters 2, 3, 4, and 7, and to the large extent also chapter 6 in the form of the literature review which is a bit too much for my taste and it became very early during my reading of the manuscript rather tiresome to read as many of the references are mentioned multiple times throughout these sections but each time focusing on a slightly different aspect. I would prefer not to produce a master's thesis where majority of the first 50 pages of the text out of 66 in total serves the role of the literature review on the expense of the text devoted to the details of the used methodology, presentation of the econometric model, discussion of the technical aspects of the authors numerical analysis such as verification of the assumptions of the model, and possible extensions of the first and only numerical results achieved. Many parts of the current manuscript could have been easily shortened by at least 10 pages which could have been used for a more relevant original material (or just to produce a shorter manuscript which would be more joyful to read).

Manuscript form

In my opinion, the table of contents is of too much detail and for the length of the manuscript could have been adjusted to a single A4 page. Moreover, in this form, it stresses out the uneven level of detail in each of the sections of the manuscript, where nearly one third of the whole manuscript is just the subsection 6.3.

Abstract written in Czech was obviously done as a quite sloppy translation of its English version.

The manuscript follows a logical structure, although it becomes lengthy at some parts, cf above the comments on too much of a literature-review style of the text.

Report on Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Bc. Radka Doležalová
Advisor:	Mgr. Petr Polák MSc.
Title of the thesis:	What is my car worth? Hedonic price analysis of the German used car market

The author has a decent command of English, though, occasionally writes systematically misspelled words (e.g. she frequently uses „to an extend“ instead of „to an extent“) or modifies incorrectly well known phrases (e.g. „go down to history“ instead of „go down in history“, page 15, line 5 from above; or „please refer to“ instead of „confront“ or „please see“ or „we refer the reader to“).

To cause some smiles on the faces of the members of the committee, I refer to a typo on page 40, line 9 from below.

Summary and suggested questions for the discussion during the defense

The results of the Urkund analysis **do not** indicate significant text similarity with other available sources. The Urkund analysis found just partial similarities to the earlier version of the author's manuscript.

In my view, the thesis fulfills the requirements for a master's thesis at IES, Faculty of Social Sciences, Charles University. Taking into account the above, I recommend the thesis for the defense and I suggest to award the thesis **grade D (67%)** depending on the performance during the thesis defense.

Suggested questions for the defense:

1/ On page 21, line 4 from below, you claim that one of the important factors is „buyer's reputation“. Shouldn't that be a seller's reputation?

2/ In subsection 6.2 you comment on restrictions for observations included in the dataset. Could you state, in per cents with respect to the whole population, how many observations did you disregarded and how many made it to the dataset? Further more, to confuse me more, you comment on additional filtering of the observations in sections 6.3.1 and beyond. How many data did you eventually filter out?

3/ Also in subsection 6.2, you state that you randomly selected 10% of the advertisements. Could you elaborate on the process of the random filter?

4/ One of the variables used is Cost on consumption in Eur/100 km. But this variable is highly unstable as it depends on the actual price of gas/fuel of a car, which, on top of that, is unstable geographically and I am pretty convinced that each Bundesland has different average prices of gas/fuel. Did you compensate in your thesis for that phenomenon? I believe this affects to a high degree the discussion on page 61.

5/ On page 33, line 2 from above, you claim that approximately one third of cars have driven **less** than 10 000 kilometers in general. With respect to Figure 6.4, should not that be that one third of cars have driven **more** than 10 000 kilometers?

6/ If I understand correctly the results reported in Table 7.3, both postcode regions 8 and 9 came out of the hedonic price analysis using the Bayesian model averaging as strong predictors. Nevertheless, you do not comment on significance of postcode region 9 (the northern part of Bavaria and related territory) in subsequent analysis, cf Table 7.4. Why?

7/ Does it really make sense to assume that e.g. age or mileage influence the market value of a car linearly? Did you try to include squares and/or higher powers of the considered variables in your models?

Report on Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Bc. Radka Doležalová
Advisor:	Mgr. Petr Polák MSc.
Title of the thesis:	What is my car worth? Hedonic price analysis of the German used car market

SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	21
<i>Methods</i> (max. 30 points)	13
<i>Literature</i> (max. 20 points)	16
<i>Manuscript Form</i> (max. 20 points)	17
TOTAL POINTS (max. 100 points)	67
GRADE (A – B – C – D – E – F)	D

NAME OF THE REFEREE: RNDr. Michal Červinka, Ph.D.

DATE OF EVALUATION: September 8, 2020

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F