

# Report on Bachelor Thesis

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

<b>Student:</b>	<b>Veronika Mertová</b>
<b>Advisor:</b>	<b>Mgr. Petr Polák, MSc. Ph.D.</b>
<b>Title of the thesis:</b>	<b>The influence of watching videogame streams on purchase decisions of gamers and their willingness to pay, evidence from the Czech Republic</b>

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

### **Short summary**

The thesis aims at studying the relations between watching the game streams and the actual purchase decisions of players, and their willingness to pay for the video games. What it actually does is that it studies the impact of a large number of socio-demographic characteristics and behavioral variables related to watching streams, playing games, and purchasing. The author designs a simple online survey which she makes public via social media among the Czech gaming community. A huge number of hypotheses which is actually hard to digest are studied and interpreted. Logistic regression is used to study purchasing decisions and OLS regression to study the willingness to pay. Such selection of econometric models is correct for the studied problems and appropriate at the bachelor level. Still, their econometric elaboration is poor, with potential issues and flaws decreasing the credibility of presented results. The two most serious issues to my mind about which I will be more specific below, are that the author does not cite or cites incorrectly many sources and that the author neither deals with nor even mentions OLS assumptions. At the same time, the OLS regression takes up about half of the econometric analysis.

### **Contribution**

While using the own online survey might be criticized from many different angles by anybody with some background in statistics, I find this part of work potentially the most contributive. The author collects 1020 responses and seems well aware of the potential biases from which the self-selection bias is perhaps the most serious one. Also, while the author collects the data about the willingness to pay by the most trivial and the less reliable method (simply the stated preference in the form of an amount in CZK), she at least discusses other possible more complicated alternatives in 2.4.2.

What I find, however, alarming is the „acknowledgment“ that *„the questionnaire was developed based on intuition and limited previous research, which might have led to the omission of certain relevant factors“* (pg. 51). I believe a statement like this should never appear in any bachelor thesis! Also, the underrepresentation of women is serious: it reaches circa 10% in the sample compared to statistical data cited by the author herself, reporting circa 40% representation of women in the gaming community. Both issues seriously decrease the credibility of presented results.

The topic is generally economically interesting, especially after the covid period when the digital segment gained unprecedented importance. The results might bring some value-added, especially in the Czech context. The main takeaways are that watching streams is generally positively associated with purchasing decisions. A profit-maximizing price is about CZK 500 for independently developed titles and up to CZK 1500 for professionally developed famous games. That makes intuitive economic sense and potentially can help the developers to optimize their marketing and boost profits.

Especially Fig. 5.3 (pg. 37) raises doubts about the quality of the estimated model, which is poorly trained for predicting zeros and does it, in fact, randomly for independent titles (18:17).

### **Methods**

As suggested above, the methodology is generally correctly selected, but the resulting analysis is poorly and insufficiently, in some cases incorrectly elaborated. Starting with the dataset, the author states that she undertook a „cleaning process“ (about 3% of data deleted), but no details are given. Even after this „cleaning“, some respondents report more than 15 hours of playing a day (Graph 3.2) which I suppose is impossible in long term. The main issue I can observe is the coding of some variables: e.g., *Education* is coded as 1 for primary, 2 for high school,... 4 for university. This implicitly

# Report on Bachelor Thesis

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

<b>Student:</b>	<b>Veronika Mertová</b>
<b>Advisor:</b>	<b>Mgr. Petr Polák, MSc. Ph.D.</b>
<b>Title of the thesis:</b>	<b>The influence of watching videogame streams on purchase decisions of gamers and their willingness to pay, evidence from the Czech Republic</b>

dictates a strong assumption that the impact of university education is exactly 4x larger than primary education in the linear model for the willingness to pay. Three dummy variables should have been used instead similarly as it is done, e.g., for *Occupation*. The same imperfection appears for more variables and, most importantly, for *Income* coded 1-6 based on completely arbitrary thresholds. I believe the *Income* variable (in a sort of demand/pricing equation) can keep its collected values in CZK, providing intuitive interpretation.

50 (!) working hypotheses (pg. 29-31) are almost impossible to digest, similar to their interpretation (pg 39-41 and 45-47). While often intuitive, no explanation regarding the hypotheses is provided, and many of them are stated so that they cannot be econometrically evaluated by rejecting  $H_0$ .

Logistic regression is correctly explained and described, but the OLS regression is not explained at all! Also, one of the severe drawbacks at the bachelor level already highlighted above is that the OLS assumptions are not tested; not even their existence is mentioned in the text. I also suspect the results based on so many variables in a regression might be affected by the omission of the multicollinearity assessment.

Finally, the author states that the selection of the specific evaluated models is done based on the statistical significances of the explanatory variables starting with the „full model“ and using the training set and tested against the test set. This is perhaps the most interesting part of the analysis. Still, almost no details are reported, and the reported 796 out of 990 observations do not even match the ratio 80:20 reported in the Conclusion. The author definitely should take advantage of the defense and explain the steps taken in detail to the committee. Last but not least, SEs are not reported in the tables with results.

## Literature

The literature section is large and must have taken a significant effort to elaborate. I especially appreciate the wide range of topics discussed and the overall general overview of the field the author demonstrates. In fact, this is one of the additional important contributions of the thesis. I also understand that many sources are electronic compared to the more mainstream economic topics where academic articles prevail. On the other hand, some large sections are only based on none (Sec. 2.2.1) or one single citation (Sec. 2.2.5). Also, Sec. 2.2.4 about the production process is unnecessarily detailed, perhaps because a lot of information was simply accessible. At the same time, it does not seem to have any important connection to the methodological part of the work based on consumer data.

Nonetheless, there is the second unpardonable flaw because learning how to work with and cite academic literature is one of the main aims of elaborating the bachelor thesis. Put simply: many sources are not cited at all (Czech Game Developers Association; Pavel Šebor; Twitch 2021; Eurobarometer 2013) or cited incorrectly in the Bibliography (e.g., GDA, 2020; Forbes 2021; VironIT 2018, The Escapist 2009) if I only focus on pg. 6-8. Even a direct 5-line long citation taken from the web (pg. 9) is not cited in the Bibliography.

## Manuscript form

The thesis is written in decent English, standardly structured, and typeset in Word for which one can hardly expect miracles wrt the overall design. The flaws in citing the sources are described above. Referencing tables and figures is done standardly in the text, the tables are reasonably labeled and described, and the figures are self-contained. The graphics are standard, easy to look through and understand. I'm not too fond of empty pages such as 21 or 31. The lists of hypotheses and their interpretations are inappropriate. I also believe that statements like „*This may be caused by the*“

# Report on Bachelor Thesis

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

<b>Student:</b>	<b>Veronika Mertová</b>
<b>Advisor:</b>	<b>Mgr. Petr Polák, MSc. Ph.D.</b>
<b>Title of the thesis:</b>	<b>The influence of watching videogame streams on purchase decisions of gamers and their willingness to pay, evidence from the Czech Republic</b>

women's natural affinity for shopping and buying new things" (pg. 50), although definitely funny, are not appropriate for academic texts.

## Overall evaluation and suggested questions for the discussion during the defense

My opinion is that apparently incorrect citing of resources and not even discussing the OLS methodology on which half of the analysis is based are such serious flaws at the bachelor level that push the thesis to the edge of defensibility at the IES. Notwithstanding that the work is generally interesting, contributing and its elaboration must have taken considerable effort; these are two core skills a bachelor student should master during the studies and demonstrate in the final thesis. If the student manages to persuade the committee that she understands the OLS methodology well, that the model selection and elimination of variables were made correctly, and to explain the Bibliography issue (which, I am sure, was not done in bad faith), I would suggest the grade weak D at best.

The results of the Urkund analysis do not indicate significant text similarity with other available sources.

Additional topics for the discussion:

- Is there any econometric method to reduce the self-selection bias? Can you think of how such an approach can be implemented to your online survey (ignoring the practical issues...)?
- Especially the model for the willingness to pay is rather trivial: all variables appear in the linear functional form (Eq. 4.8). How can we implement some potentially nonlinear relationship, e.g. wrt *Age*, *Hours*, *DaysWatching*, or some other where relevant?

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

<b>CATEGORY</b>	<b>POINTS</b>
<i>Contribution</i> (max. 30 points)	20
<i>Methods</i> (max. 30 points)	15
<i>Literature</i> (max. 20 points)	12
<i>Manuscript Form</i> (max. 20 points)	14
<b>TOTAL POINTS</b> (max. 100 points)	<b>61</b>
<b>GRADE</b> (A – B – C – D – E – F)	<b>D</b>

**NAME OF THE REFEREE:** Jiří Kukačka

**DATE OF EVALUATION:** 27. 8. 2021

---

**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