Report on Bachelor / Master Thesis

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

Student:	Iveta Tučková
Advisor:	Milan Ščasný
Title of the thesis:	Analysing Factors of Physical Activity Across Different European Countries

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

Short summary

In her thesis Iveta Tučková studies the factors associated with individual physical activity. The author relies on individual level survey data collected in five European countries, i.e. Portugal, Spain, Czech Republic, Lithuania, and the United Kingdom. The author deploys the Heckman model and the seemingly unrelated regression framework and finds that higher education and age have a positive effect on the probability that an individual will participate in physical activities, while smoking, living in an urban area, and being female decrease the probability.

Overall, it is difficult to evaluate the submitted thesis as the manuscript seems to be a work-in-progress draft rather than a completed thesis. The manuscript lacks a complete interpretation of the results. It is therefore difficult to assess the author's ability to conduct and correctly interpret scientific research.

Contribution

The main contribution of the thesis seems to be the analysis of a previously unused dataset collected by Zvěřinová et al. (pg. 12). While a rigorous analysis of a novel dataset would constitute a sufficient contribution for a bachelor thesis, it is difficult to judge the contribution in this case as the thesis seems to be unfinished. Only a portion of the results are discussed and interpreted while *„The rest of the results can be found in the appendix"* (pg. 25). The author discusses only the results from the participation equation of the Heckman model and leaves results from further analysis (included in Table A.2, Table A.3, and Table A.4) unexplored. This significantly limits the contribution.

Methods

The author correctly identifies the issues of non-response, and censoring associated with analysis of survey data. She deploys the Heckman model to deal with the censoring of the outcome variable which I believe to be the correct approach. She further relies on the seemingly unrelated regression framework. Though the selected framework seems correct there remain multiple aspects of the analysis which are not addressed in the thesis.

The analysis is based on survey data collected in five European countries. The author presents results from a single model based on pooled data. This raises the issue of whether the results should be interpreted as representative for the specific countries, whole Europe, or specific subsets of the population. This issue is especially pressing for the analysis based on seemingly unrelated regressions which seems to be based on 3,630 observations relative to the full sample size of 10,364 observations which implies significant truncation. Furthermore, the author does not provide any insights for how the sample sizes change for the respective countries. As the discussion of results is mostly lacking, these issues receive little to no attention.

Related to the previous issues, the author accommodates for pooling of data from different countries by including dummy variables for country of origin for each of the considered countries. This seems insufficient as the effect of control variables likely varies between countries. Consequently, a model specification including interaction terms or separate models for each of the countries might be more suitable.

Report on Bachelor / Master Thesis

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

Student:	Iveta Tučková
Advisor:	Milan Ščasný
Title of the thesis:	Analysing Factors of Physical Activity Across Different European Countries

Some choices regarding model specification seems to be unclear. For example, when modelling vigorous physical activity, the author controls for age, while when modelling moderate physical activity, she controls for both age and age squared, and when modelling walking, she controls only for age squared while omitting the linear term (Table A.1).

Treatment of missing observations in control variables is unclear. The description of the treatment of missing observations in the income variable is ambiguous (pg. 15). It is unclear, whether the author imputed 0 or the average wage or the minimum wage for households not reporting income. Furthermore, the author states that *„respondents could choose from 12 intervals, in which they would include their net monthly income"* (pg. 15). However, in the reported regression results she includes only one variable for income.

Literature

The author provides an in depth overview of the literature on the determinants of physical activity. She further presents the definition of physical activity, which is extremely useful for readers not familiar with the topic. The citations do not seem to follow any specific citation pattern. These thus range from author (date) (e.g. Guthold et al. (2018) on *pg. 8*), author (Nikitara et al. on *pg. 11*), to miscellaneous reference forms (e.g. Nikitara et al. from 2021 on *pg. 7*, Gerovasili et al. from 2015 on *pg. 7*, Mayo et al. of 2019 on *pg. 8*). Furthermore, some of the references in the text are not included in the list of references (e.g. *Zvěřinová et al.* on pg. 12).

Manuscript form

The manuscript form constitutes the principal weakness of the thesis and significantly hinders reader's understanding of the presented research and its contribution.

First, the thesis seems to be incomplete. As noted in the contribution section of this report a significant portion of the results is not discussed in the thesis and is only contained in tables in the appendix.

Second, the thesis would benefit from a thorough proof-reading. The quality of the text is significantly below of what one might expect in a finished bachelor thesis. At this stage the thesis contains phrases which the author likely intended to complete but did not, e.g. *„From original 10364 observations, only XX were used in the main analysis, due to extensive data cleaning"* on *pg. 12*; *„… similar significant differences also for men (s10; Marques 2016)"* on *pg. 9.*; *„Multiple studies examined by Abu-Omar (..)"* on *pg. 8*; using both PA and PIA as shortcuts for physical activity as shown in the section defining acronyms on *pg. 3.*.

Third at some points the text is difficult to follow as the author used acronyms and terminology which have not yet been defined, e.g. *„with aim to reduce a PIA by 15%"* on *pg. 4* or *„physical activity and correlation between all three types of activities"* on pg. 5.

Overall evaluation and suggested questions for the discussion during the defense

Report on Bachelor / Master Thesis

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

Student:	**Iveta Tučková**
Advisor:	**Milan Ščasný**
Title of the thesis:	**Analysing Factors of Physical Activity Across Different European Countries**

In my view, the thesis does not fulfil the requirements for a bachelor thesis at IES, Faculty of Social Sciences, Charles University, I do not recommend it for the defense and suggest a **grade F**.

Questions for defense:
- Significant portion of your results are not discussed in your thesis and are only included in the appendix. Could you provide a discussion of these results and explain their contribution to the available literature?
- Can you comment on how you reached the final sample and what is its representativeness? How do the final sample sizes vary between countries? Can this affect your results?
- You rely on a single model based on pooled data for all five countries. Could you provide justification for this empirical specification?
- Could you discuss in more detail how the possible mechanisms can vary between countries and how this can affect your results?

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

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

CATEGORY		POINTS
Contribution	*(max. 30 points)*	15
Methods	*(max. 30 points)*	15
Literature	*(max. 20 points)*	10
Manuscript Form	*(max. 20 points)*	2
TOTAL POINTS	*(max. 100 points)*	**42**
GRADE	**(A – B – C – D – E – F)**	**F**

NAME OF THE REFEREE: Marek Šedivý

DATE OF EVALUATION: 27.8.2022

Digitálně podepsáno (27.8.2022)
Marek Šedivý

_____

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