

# Report on Bachelor / Master Thesis

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

<b>Student:</b>	<b>Vojtěch Korbélius</b>
<b>Advisor:</b>	<b>Mgr. Michal Paulus</b>
<b>Title of the thesis:</b>	<b>Life expectancy and its determining elements: A study for the Czech Republic at the beginning of the 21th century</b>

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

Thesis examines life expectancy in the Czech Republic at the beginning of the 21st century in seven chapters. The thesis begins with literature review, describes model and methodology for empirical analysis, which results are checked for robustness. Quite an interesting topic with a possible contribution is unfortunately presented with a lot of mistakes.

The thesis has a promising beginning with quite clear research question and logic behind it. Author promises great contribution and improvement of findings from previous studies, which is not delivered. Author also repeats the ideas behind his work again and again, which makes the reader a bit confused and interrupts the flow of the work. Moreover, author does not use commas and makes long and complex sentences which can cause misinterpretation.

When summarizing literature about determinants of life expectancy I would expect the author to provide also findings about determinants and connect studies to those determinants rather than in subchapter 2.5 state findings from a few studies separately. Literature should provide some intuition about possible explanatory variables and their effects. One would assume, that life expectancy is rather question of personal characteristics and medical history, but the thesis focuses only on factors from broader perspective and does not provide any evidence about results from micro level.

Mixing model and data together (Chapter Model and Methodology) makes the logic and any comments hard to follow. Logic behind usage of modified model for one females at the age of 65 is also missing. Next to it, I have found no reason for explanatory variable expressing share of people above 65 years of age to be separated by gender. Text mentions some data modifications, but precise steps undertaken are not described at all.

Interpretation of results in some cases is wrong and no table provides all results including explained significance levels. Robustness check follows after the interpretation of results which is strange.

What I see as a huge problem is author's clear misunderstanding of OLS method, which is used in the empirical part. Author obviously does not understand what linearity of OLS estimator means and what linearity of a model is. Assumptions of used model are tested but with lack of interpretation of the meaning.

### Key notes for the manuscript:

- THESIS TITLE: 21<sup>ST</sup>, not 21TH
- When using citation in parenthesis after a paragraph, there is no dot before, but after the bracket and no space: ...end of text (Author, 2014).
- In typography, there is a difference between minus sign, dash and hyphen!
- Page 1: "In that light 30 years..." – what does it mean?
- Page 1 and following: "points of views" – either "points of view" or "perspectives"
- Better use "thesis" than "paper" in the text to point to the actual text
- Spellcheck needed: "eleswhere", "focuse", "seoncd", "expectacny", "Facotrs" etc.
- Check for plural forms needed: "those number"
- Page 3: "We are will begin"
- Page 3: "we will able to have a" – missing a verb
- Page 3: Overview of the thesis should be brief, not almost a page and using whole name of chapters as well as chapter numbering
- Page 3: Reference should be "Chan & Kamala Devi (2012)"

# Report on Bachelor / Master Thesis

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

<b>Student:</b>	<b>Vojtěch Korbélius</b>
<b>Advisor:</b>	<b>Mgr. Michal Paulus</b>
<b>Title of the thesis:</b>	<b>Life expectancy and its determining elements: A study for the Czech Republic at the beginning of the 21th century</b>

- "So let us now proceed", "So lets continue", etc. – inappropriate language for an academic text
- Page 4: "As stated earlier it is one..." – it was not stated earlier in the text
- Page 4: "in later chapter (3.2)" – it should be a subchapter
- Page 4: "During the 20<sup>th</sup> century conditions in Europe and the..." – reference needed for such statement
- Page 5: "In total it is 40 years." – LE or growth of LE?
- Page 5: "... which have helped to create life expectancy gap..." – reference needed
- Page 5: "... agreed that democracy has...Both of those researchers..." – there are more
- Page 5: "... has been considered as the core foundation for increasing life expectancy" – reference needed
- Page 5: "... see its consequences on increase in life expectancy" – reference needed; there should be a modal verb before see; and "on" is a wrong preposition
- Page 6: "And if yes" – "and if so"
- Page 6: "...do not make any sense to research from an economic point of view" – reference or reasoning needed
- Page 6: "Life expectancy is calculated as follows" – this definition holds for the Czech Republic, is there any difference compared to the academia elsewhere?
- Page 6: Reference "LAU1" – incorrect
- Page 7: "Person... they knew" – he/she knew
- Page 7: "How we have pointed out" – "How have we..."
- Page 7: "But what does it mean? What bad can it cause?" – what is the purpose of those questions?
- Subchapter 2.4 is rather about causes of life expectancy and possible flaws in research
- Page 8: "... will be fired..." – inappropriate language
- Page 8: "... has recently decreased same as the number" – wrong grammar
- Page 8: "Competitions" – it has a different meaning than competition
- Page 9: "But it is necessary to understand that not everybody looks..." – so what are the perspectives?
- Page 9: "... life expectancy-environmental" – there should be a dash and spaces
- Page 10: "Most of those studies..." – why is a direct quotation used?
- Page 10: "... but there are still studies that are pointing..." – references needed
- Page 12: "even though his model is very complex we were able to go deeper" – adding some variable but not using others?
- Page 12: "... because our sample is more homogeneous." – but the author also argues, that using variables with aggregation on country level would be wrong
- Page 12: "... start their work..." – this should be in a literature review
- Page 12: "AGEDIST85" in the model but "AGEDIST1985" in text
- Page 12/13: what does as the dependent variable "LE97?" stands for?
- Page 14: "concrete factors" – concrete means solid
- Page 14: "When we look at...dummy variable" – there is no dummy variable in the presented model; following sentence is without any sense
- Page 14, model: not a proper way to write an econometric model; "Facotrs"
- Page 15: "The main highlight..." – is rather difference
- Page 15: "... a painful issue..." – we are not talking about medical procedures
- Page 15: "... as we know them now" – reader might not be familiar with the districts at all
- Page 15: "Other reason for complication" – firstly: "another", secondly: what intervals?
- Page 16: Table 3.1 – source [www.uzis.cz](http://www.uzis.cz) – too general and not in the bibliography
  - o Table also occurs out of the blue, no reference in the text nor detailed explanation, some of the variables are used only in the robustness check, which is not stated here
-

# Report on Bachelor / Master Thesis

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

<b>Student:</b>	<b>Vojtěch Korbélius</b>
<b>Advisor:</b>	<b>Mgr. Michal Paulus</b>
<b>Title of the thesis:</b>	<b>Life expectancy and its determining elements: A study for the Czech Republic at the beginning of the 21th century</b>

- Page 17, average wage: any differences between men and women taken into account if models are separate for both genders?
- Page 17: "For now we will present you" – change of narrator perspective
- Page 18: "lifestyle" and "life-style" – should be consistent within the thesis
- Page 18: "... we ... consider the life-style ... to be same on average" – a very strong assumption without any other reasoning and even no note in the conclusion.
- Page 18: "we can increase the economic growth through preserving the nature" – reference needed
- Page 18: "pollution factors to be so crucial" – reference needed
- Page 19: "... how has it has..." – too many verbs
- Page 19: "The biggest differences are exclusion" – only one difference mentioned
- Page 19: "Other attribute of the OLS is linearity. OLS is a linear regression which means that can model linear relationships but in case of non-linear relationship we will get very poor results. Very common way of dealing with this issue is transforming the collected data into linear form. The most common way is to use the logarithm." – misinterpretation of OLS estimator
- Page 20: Next paragraph starts with "The second benefit of using", yet the next one "Those two drawbacks of the OLS..."
- Page 20: "OLS in its basics is standing on few prerequisites that have to be fulfilled or controlled for before applying the OLS." – so called assumptions, some of them cannot be checked before (property of residuals)
- Page 20: "Let us look at our final model and its results, afterwards we will go through those so called assumption and see that all of them are fulfilled in our model, namely: linearity in parametrs, random sampling, no perfect collinearity, zero conditional mean, homoskedasticity, normality." – normality is not an assumption of OLS estimate
- Page 20: "To estimate an ideal model was not such an easy task as in the case..." – why?
- Page 20: Why is percentage of population above 65 separated by gender too?
- Page 20: Why do the model use logarithm of unemployment?
- Page 20/21: "Why model 3.4" is different?
- Tables 3.2 – 3.5: Signs of significance are not labeled, 4 different tables make it unreadable
- Descriptive stats are missing. What is natural level of poisonous oxides, what is too high, any references that would help us interpret dataset in that manner?
- Page 24: Table 3.6 – comments are needed;
- Page 24, Table 3.6: "Shapiro-Wilk W test for normal data" – "normal data" vs. "normality"
- Page 24: "It may be a little harder to test everything for four models so if needed to show something we are going to indicate each model by its dependent variable" – language
- Page 24: leverage points and outliers, what is the difference (stated high value of independent variable in both cases)
- Page 24: "Luckily Stata has..."
- Page 24: "...are not as odd as in other extreme cases..." – what are extreme cases, what was done with them?
- Page 25: normality makes it BUE, not BLUE
- Page 25: "multiple regressions" – multiple?
- Page 25: "The last assumption out of the basic five..." – only 3 described in the previous paragraphs (First, Second, something)
- Page 26: "Luckily both of them ended up in our favor in all of our cases" – language, but mainly null hypotheses for above mentioned test are not stated
- Page 26: Why do we need CLT a LLN?
- Page 26: "hypothesis" – hypotheses
- Page 26: last paragraph

# Report on Bachelor / Master Thesis

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

<b>Student:</b>	<b>Vojtěch Korbélius</b>
<b>Advisor:</b>	<b>Mgr. Michal Paulus</b>
<b>Title of the thesis:</b>	<b>Life expectancy and its determining elements: A study for the Czech Republic at the beginning of the 21th century</b>

- Page 27: where can we find the averages and other numbers?
- Page 28: Table 4.1 – first row has wrong columns
- Page 28: Interpretation is wrong – there is difference between percentage and percentage points! – thus increase of unemployment by only “0.1088 percentage points”.
- Page 28: “disastrous consequences” – rather strong words for a small effect
- Page 29: “let me talk about the males and I will put the results for females into brackets so”
- Page 29: “ten 10%”
- Page 29: “Unemployment rate as an economical variable does not seem to have any influence” – such variable is not even included in the model!
- Page 29: “Concretely” – “specifically”
- Page 31: “Country that went...deterministic factor” – factor for what?
- Page 31: “... pollution factors...” – only one factor in the model
- Page 31: “...as well as money invested ...” – such variable only in one model for females at age of 65
- Page 32: “Those factor”
- Page 32: “Statistically indefensible data” – what does it mean?
- Page 32: “... due to the high correlations...” – any specific numbers?
- Page 34: “it is our duty to give an explanation”
- Page 35: Hypothesis can be rejected or not rejected, but neither approved nor confirmed.
- Page 36: “... we were able to explain all the differences” – which differences?
- Page 36: “... holds in vacuum”
- Pages 37-38: No significance level are marked, thus any interpretation is impossible
  - o  $R^2$  for tables 6.3 and 6.4 in the first column are wrong
  - o Again model differences in table 6.4 are not explained
- Page 39: “IPoiOx reacts on change in llnvEnv and ISolid” – how?
- Page 39: “The last good news we can conclude from the pictures above is what...” – there are tables
- Page 40: “This widely used method are”
- Page 40: “There are clear outcome of longer or shorter life expectancy on the economy...”
- Page 40: “general luck”
- Page 41: “We have found two out of three hypotheses to be correct”
- Page 41: Just a quick glance at OECD data shows different methodology of calculation of consumption of pharmaceuticals compared to the Czech statistical office
- Page 41: (Tamakoshi 2010) – which statement the reference belongs to?
- Page 41: “This fact is represented by that in each of the four models we have found exactly one health care factor to be significant”
- Page 42: “One more propositions...”
- Page 42: “We were not able to cooperate” – any tries?

## References:

- Auster, Leveson – year is missing; what is the name of that work?
- Crimmins (2004) – publisher?
- CzechStatisticalOffice (2013) – Author should be in the text, not LAU1 (page 6)
- CzechStatisticalOffice (2014) – insufficient reference
- Institutueof...(2014) – insufficient reference
- OxfordUniversityPress – insufficient reference
- Tsang (1990) – publisher?
- Walker (2013) – publisher?
- Several tenses used with no reason – present simple, present continuous, past simple

# Report on Bachelor / Master Thesis

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

<b>Student:</b>	<b>Vojtěch Korbélius</b>
<b>Advisor:</b>	<b>Mgr. Michal Paulus</b>
<b>Title of the thesis:</b>	<b>Life expectancy and its determining elements: A study for the Czech Republic at the beginning of the 21th century</b>

Overall, the thesis needs several changes to achieve a proper level of academic work from the manuscript perspective, additionally the chapter about model should be corrected for mistakes.

Promising topic with a clear contribution and proper usage of methods for model testing is unfortunately turned down by the way the thesis is written and steps undertaken during the analysis are presented. The thesis overall is from content perspective very good, but author's writing style raise a question if the autor truly and fully understands what he has done or has just repeated some steps. That should be the key part of the defence in my opinion.

In the case of successful defense, I recommend “**dobře**” (**satisfactory, 3**).

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

<b>CATEGORY</b>	<b>POINTS</b>
<i>Literature</i> (max. 20 points)	14
<i>Methods</i> (max. 30 points)	15
<i>Contribution</i> (max. 30 points)	25
<i>Manuscript Form</i> (max. 20 points)	6
<b>TOTAL POINTS</b> (max. 100 points)	<b>60</b>
<b>GRADE</b> (1 – 2 – 3 – 4)	<b>3</b>

**NAME OF THE REFEREE: Petr Polák**

**DATE OF EVALUATION: 2. 6. 2014**



---

**Referee Signature**

**EXPLANATION OF CATEGORIES AND SCALE:**

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

Strong                  Average                  Weak  
20                          10                          0

**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.*

Strong                  Average                  Weak  
30                          15                          0

**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.*

Strong                  Average                  Weak  
30                          15                          0

**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.*

Strong                  Average                  Weak  
20                          10                          0

**Overall grading:**

TOTAL POINTS	GRADE		
81 – 100	<b>1</b>	= excellent	= výborně
61 – 80	<b>2</b>	= good	= velmi dobře
41 – 60	<b>3</b>	= satisfactory	= dobře
0 – 40	<b>4</b>	= fail	= nedoporučuji k obhajobě