

Report on Bachelor / Master Thesis

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

Student:	Tiep Luu Danh
Advisor:	Mgr. Petr Polák, MSc.
Title of the thesis:	Smoking ban: A data analysis of sales

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

Contribution

This thesis brings new insights into the discussion about the appropriateness of a smoking ban that was introduced in the Czech Republic in May 2017. Before and after an introduction of the smoking ban, there was a broad public discussion about the necessity of the law and its consequences. The impact on public health and hospitality sales were two main points. The argument against the smoking ban that the hospitality sales will decrease, and moreover a lot of small locals will have to close, was discussed longtime after the introduction of the law. This work evaluates the impact of the smoking ban on hospitality sales. As the author states, there are no similar studies in the current literature, which is surprising taking into account the previous broad discussion of the topic. Moreover, the analysis employs two datasets from two different sources. Hence, apart from the answer to the main question, we can make conclusions about the reliability of the data where one dataset is a standardized metric of sales and the second dataset contains the recordings from electronic evidence of sales (EET). The thesis employs the data that are recently available as the EET was introduced in 2016 and is a new and basically undiscovered source of data. Briefly, the contribution of this work is twofold. It statistically answers the question about the impact of the smoking ban on hospitality sales and it inspects a recently available data source - records from EET.

Methods

The author very clearly introduces the question the thesis should answer and formulates the hypotheses to be tested together with the precisely described procedure of testing (methodology). As I mentioned above, one of the nice features of the thesis is that it uses two sources of data, and the author can confirm or not the results comparing two datasets and also investigate the usage and usefulness of relatively new type of data (EET). The time series methods to analyse the impact of the smoking ban on hospitality sales are appropriate and well described (ARMA, ARIMA, ARIMAX processes). However, I have a few remarks, perhaps questions, regarding the methodology that should be answered in the thesis.

Table 3.1 presents the descriptive statistics of hospitality sales in pre-law and post-law period. The author does not comment on the findings at all, although these might be interesting. In the post-law period, all the statistics are higher than in the pre-law period (min, median, mean, max), even though the data is defined relatively with price basis in 2005 so the inflation should not influence the findings. In perspective that the thesis concludes that the smoking ban does not have any significant impact on hospitality sales, it would be interesting to discuss the findings of descriptive analysis.

The work with EET data was more complicated. The nonsense values seem to me to be a noticeable problem, and I would try to question the source of the data (datazeet.cz) as it should be familiar with the problem and maybe solved it in their own data processing. When listing all the problematic values in the Appendix, it would be helpful to add the new values resulting from the interpolation so the reader can see how the data problem is solved and magnitude of changes made. More importantly, if the data suffers from measurement errors for negative values, is not it possible that also extremely positive values are measurement errors (Figure 3.3)?

Concerning the results of the third hypothesis testing, it is interesting that here, although the conclusions are the same for CSO and EET data, the significance of the beta coefficient is different (p.41 and 43). This coefficient should show the relationship between hospitality sales and retail sales. This is one of the main differences between the results coming from two datasets.

Finally, it is unclear why the author uses the ARIMA process in section 4.4.2 (i) on the pre-law period only (instead of the whole period) if we do not deal with forecasting.

Report on Bachelor / Master Thesis

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

Student:	Tiep Luu Danh
Advisor:	Mgr. Petr Polák, MSc.
Title of the thesis:	Smoking ban: A data analysis of sales

Literature

The literature review describes both the smoking ban and related literature together with EET and related literature. It introduces well the topic and is sufficient. I would recommend discussing a bit some of the information the author is reporting. Is the Tanzania relevant source of information for our country regarding EET? On page 5, the author is describing that the smoking ban might increase alcohol consumption, and in the following paragraph, the author immediately describes that hospitality sales can decrease after the smoking ban. This contradiction could be more developed.

Manuscript form

The overall manuscript form is unfortunately rather average. I would recommend reconsidering the size of Tables and mainly Figures in the author's next work. They are extremely big in some cases, and it is common to use text wrapping above and below figures. Some parts of the text are unnecessarily repeated in the whole work. The hypotheses of EET data are just the same as for the CSO dataset. Thus it is not necessary to rewrite the entire page once again. The same holds for Appendix B. I appreciate the extra information that is given there, but parts of the text are just a repetition of what is described above in Chapter 5. The better description of NACE in general (meaning, definition, structure) would give better insight to CSO data. The NACE 47 is more explained on page 17, although we work with this notion since page 12. Similar imprecisions are present in the text, and I give few examples below. However, I have to mention that the text is coherent and the shortcomings do not worsen the comprehension of the work.

To cite a few smaller problems:

- typos such as "dairy data" on page 5
- form of citations: some citations are in parenthesis when it is not supposed to be and vice versa, there is a missing citation of the paper of Bates (2015) that is mentioned on page 8 and is missing in the bibliography; I would recommend citing the paper the author is explaining in the whole paragraph at the beginning of this paragraph instead of at the end or in the middle (even though you talk about the authors, it is common to add the citation immediately)
- equation (4.1) is not correctly defined and described
- different sizes of Figures that are displaying similar features (e.g., Figure B.5 and B.7)
- Appendix starts with Table A.2
- it is better to explicitly state that the parentheses contain the standard errors in Tables with coefficients results (e.g., Table 5.5).

Summary and suggested questions for the discussion during the defense

This thesis brings a very discussed topic into the new perspective, employs recently available data and thus represents a considerable contribution to current literature. The method of the analysis is appropriate and advanced for a bachelor student. Some of the steps needed more comments. However, the main drawback of the thesis is manuscript form as the text contains few imprecisions, even though these are small. Therefore, I recommend this thesis for a defense and suggest grade B.

Questions that might be discussed during the defense:

- 1) The introduction of the smoking ban in 2017 was indeed new for restaurants (before this law, it was only compulsory to indicate whether smoking is allowed in the restaurant or not). The thesis states that: "*The ban was enforced only on the NACE 56.3 Beverage serving activities. Because the CSO does not offer lower granularity, we use these data [NACE 56 Food and beverage service activities] as our proxy for the NACE 56.3 Beverage serving activities, which consists of restaurants and bars.*" In the case of EET data, the author extracts the sales only for NACE 56.3. to be included in the analysis as the hospitality sales. Why did you choose such an approach instead of analysing the food and beverage sector as a whole?
- 2) If I understand correctly, the EET data is region-specific (Table A.1). How do you address this feature in your analysis?

Report on Bachelor / Master Thesis

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

Student:	Tiep Luu Danh
Advisor:	Mgr. Petr Polák, MSc.
Title of the thesis:	Smoking ban: A data analysis of sales

- 3) In the literature review (page 6), the author mentions that in California 12.5 % of the hospitality business from the sample closed during the first 15 months after the smoking ban. Have you tried to find such information for the Czech Republic? The closure of restaurants is one of the arguments against an introduction of the smoking ban and might complement the author's findings.

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

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	30
<i>Methods</i> (max. 30 points)	24
<i>Literature</i> (max. 20 points)	19
<i>Manuscript Form</i> (max. 20 points)	10
TOTAL POINTS (max. 100 points)	83
GRADE (A – B – C – D – E – F)	B

NAME OF THE REFEREE: Šarlota Smutná

DATE OF EVALUATION: 2.6.2019



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.*

<i>Strong</i>	<i>Average</i>	<i>Weak</i>
30	15	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.*

<i>Strong</i>	<i>Average</i>	<i>Weak</i>
30	15	0

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

<i>Strong</i>	<i>Average</i>	<i>Weak</i>
20	10	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.*

<i>Strong</i>	<i>Average</i>	<i>Weak</i>
20	10	0

Overall grading:

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