CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Smoking ban: A data analysis of sales

Bachelor's thesis

Author: Tiep Luu Danh

Study program: Ekonomická teorie

Supervisor: Mgr. Petr Polák, MSc.

Year of defense: 2019

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 9, 2019

Tiep Luu Danh

Abstract

Smoking ban and VAT Fiscalization are the two most recent acts that were causing a nuisance for hospitality industry. In this thesis, the effect of the smoking ban on the restaurant sales is analyzed. The analysis research, whether sales were significantly affected based on monthly Czech Statistical Office time series and on daily VAT Fiscalization time series. That way, not only the smoking ban is inspected, but also the potential benefit of VAT Fiscalization as a data sampling tool for policy evaluation is taken into the question.

JEL Classification	C54, C81
Keywords	Smoking ban, Sales, VAT fiscalization, Czech Statistical office, Time Series Analysis
Title	Smoking ban: A data analysis of sales

Author's e-mail	tiep.luud@gmail.com
Supervisor's e-mail	polakpet@gmail.com

Abstrakt

Zákon o ochraně před návykovými látky, ve kterém je také obsažen zákaz kouření v restauracích a barech, a zákon o elektronické evidenci tržeb jsou dva nejnovější zákony, které způsobují dle slov podnikatelů nemalé komplikace v pohostinství. V této tezi se budeme zabývat, zda zákaz kouření ovlivňuje signifikantně tržby pohostinských zařízení na základě měsíční časové řady z Českého statistického úřadu a na denní časové řady shromážděné EET systémem. Tím nejen prozkoumáme zákaz kouření, ale také zjistíme, zda by se EET systémse mohl stát novým nástrojem na shromáždění dat pro hodnocení různých nařízení a zákonů.

Klasifikace JEL	C54, C81
Klíčová slova	Zákaz kouření, tržby, EET, Český statistický úřad, Analýza časové řady
Název práce	Zákaz kouření: Analýza tržeb

E-mail autora	tiep.luud@gmail.com
E-mail vedoucího práce	polakpet@gmail.com

Acknowledgments

The author is grateful to Mgr. Petr Polák, MSc. for his guidance throughout the thesis. The author would also like to express his gratitude towards Mr. Pham Minh Hiep, MSc. who helped him with difficulties with R programming during the modeling work. The author would also love to thank his family and his friends for the support throughout the years until this very moment.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Luu Danh, Tiep: *Smoking ban: A data analysis of sales.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2019, pages 96. Advisor: Mgr. Petr Polák, MSc.

Contents

List of Tables

List of Figures

Acronyms

VAT Value Added Tax

EET Elektronická evidence tržeb, in english Electronical Recording of Sales

CSO Czech Statistical Office

SI Sales Index

NACE Nomenclature statistique des Activités économiques dans la Communauté Européenne, Statistical Classification of Economic Activities in the European Community

ARMA Auto Regressive Moving Average

ARIMA Auto Regressive Integrated Moving Average

ARIMAX Auto Regressive Integrated Moving Average with Exogenous input

ITSA Interrupted Time Series Analysis

Bachelor's Thesis Proposal

Author	Tiep Luu Danh
Supervisor	Mgr. Petr Polák, MSc.
Proposed topic	Smoking ban: A data analysis of sales

Motivation In recent years, the Czech government passed particularly two acts, that are hindering hospitality industry, the VAT Fiscalization and the Smoking prohibition in restaurants and bars. Lots of heated debate not only among politicians but also among the society were raised. Smoking prohibition in restaurants and bars is a consequence of the European Union initiative in tobacco combat, specifically the smoke-free environment recommendation. Smoking has negative aftereffect not only for the smoker themselves but also for the people around him as a result of second-hand smoking. Results have shown that second-hand smoking is more toxic than ordinary smoking. Second-hand smoking can also influence the behavior of children and adolescents, as they can pick up smoking in the later phase of their lives due to being excessively exposed to it. Thus, reducing exposure to cigarette smoking seems like a rational way for increased health. However, there are also some drawbacks that need to be discussed. One of the most vocal ones are the claims about reduced sales in restaurants and bars. VAT Fiscalization is a tool implemented in several countries to elevate the efficiency of tax collection (VAT, income, corporate, etc.). It has many variations, mainly on the technical side. Some require special hardware specifications. Some, like in Croatia, requires software specifications. The main idea is that each sale process is registered and reported to the tax authorities to avoid any further tax evasion. The Czech model follows Croatia and developed a system of encryption and a key to distinguish each seller. It is called EET. The implementation was divided into three phases. Until now, two phases were implemented while the third one is still yet to be implemented due to law complications. Analyses on the value that EET is bringing through higher or more effective tax collection were made but there is no clear answer on this matter yet. However, no analyses were made with the data that were generated with the EET system. Considering the main idea and the concept of EET, it seems that they might serve as a perfect tool for the government to monitor the development of the economy or the impact of its policy. It is also not dependent on the sample period as the Czech Statistics Office or Tax Office thus it offers fast and almost immediate data extraction. One can see that maybe EET generated data about the smoking ban might provide an answer to the question, whether smoking ban in restaurant and bars reduced the sales in them. Or it might provide additional information from the dataset we would use if we did not have EET data such as data provided by the Czech Statistical Office. We will analyze this.

Hypotheses

Hypothesis #1: the Smoking ban is not significantly affecting sales of restaurants and bars according to data provided by the Czech Statistical Office.

Hypothesis #2: the Smoking ban is not significantly affecting sales of restaurants and bars according to data provided by the EET system.

Methodology This work will be using time series analysis with the data provided. Czech Statistical Office provided data are monthly time series while EET provided data are daily time series.

Expected Contribution This work shall evaluate the effect of the smoking ban on sales of restaurants and bars. Such evaluation is expected in order to stand behind the law in further discussion. This evaluation will provide an argument for either side from the economical point of view. Additionally, this work shall also evaluate the added value in terms of the data generating process of EET. Again, this evaluation shall provide an argument for either side from the point of data acquiring point of view. To my best knowledge, there is no such analysis done in the Czech Republic on either smoking ban evaluation in terms of economic point of view and on EET from the data generating point of view.

Outline

1. Introduction:
2. Literature review: a deeper review of past works done on the smoking ban and VAT Fiscalization
3. Model building: theory needed to understand and builds the time series models
4. Modelling work and results: how did we decide on the particular model and what does it imply

5. Discussion – Limitation, evaluation, further research
6. Conclusion

Core bibliography

1. Hajdušek, Tomáš a Milan Vodička. První zkušenosti s EET –elektronickou evidencí tržeb, Wolters Kluwer 2017. ISBN: 978-80-7552-589-5

2. GLANTZ, S A a L R SMITH. The effect of ordinances requiring smoke-free restaurants and bars on revenues: a follow-up.: a follow-up. American Journal of Public Health. American Public Health Association, 1997/10/01, 87(10), 1687-1693. DOI: 10.2105/AJPH.87.10.1687. ISSN 0090-0036.
Dostupné také z: https://doi.org/10.2105/AJPH.87.10.1687

3. GLANTZ, S A a L R SMITH. The effect of ordinances requiring smoke-free restaurants on restaurant sales. American Journal of Public Health. American Public Health Association, 1994/07/01, 84(7), 1081-1085. DOI: 10.2105/AJPH.84.7.1081. ISSN 0090-0036.
Dostupné také z: https://doi.org/10.2105/AJPH.84.7.1081

4. MCDOWALL, David, Richard MCCLEARY, Errol E. MEIDINGER a Richard A. HAY, JR. Interrupted time series analysis. United States of America: Sage Publications, 1980. ISBN 0-8039-1493-8.

5. BOX, George E. P., Gwilym M. JENKINS, Gregory C. REINSEL a Greta M. LJUNG. Time series analysis: forecasting and control. Fifth edition. Hoboken, New Jersey, [2016]. ISBN 978-1-118-67502-1.

Author Supervisor

Chapter 1

Introduction

Recent Czech government regulations, such as the introduction of VAT fiscalization system and stricter smoking bans were considered as an unnecessary burden for businesses in the Czech Republic The smoking ban is a controversial topic mainly because on the one hand, it is perceived as limiting smokers' freedom to smoke while on the other hand, it protects the non-smoker's health from secondhand smoking. In the Czech Republic, even though the law was met with fierce opposition and it is still discussed, there is no analysis of the effect that the smoking ban had on the sales of the affected businesses This thesis is inspecting the significance of the smoking ban's effect on hospitality sales, the industry considered to be the most affected. To conduct the analysis, the thesis will be using VAT fiscalization system provided daily time series and also monthly time series extracted from the Czech Statistical Office.

VAT fiscalization act was first implemented in restaurants and bars on the 1.12.2016 officially as: "Elektronická evidence tržeb (Electronic evidence of sales)", abbreviated as EET. VAT fiscalization was already established in some countries and there are works analyzing the added value of the VAT fiscalization in terms of VAT collected or VAT collecting efficiency. However, there is a lack of works that use the data extractable from the VAT fiscalization system for policy analysis. This thesis will try to use such extracted data to analyze the before-mentioned smoking ban. With such an analysis, the thesis intends to add another point of view to the discussion of the pros and cons of the EET system in the Czech Republic.

Using ARIMA time series forecasting and regression models, this work concludes that the smoking ban insignificantly affects the sales of restaurants and bars. Additionally, our study finds that the EET dataset is not at the moment more reliable source than CSO dataset, based on the data inspection and the results the dataset provided during the analysis.

To present the points above, the thesis is structured in this order: First, I will begin with Chapter 2 of literature review to highlight historical and prominent works, which have analyzed the smoking ban consequences in other parts of the world. This will give hindsight of what VAT fiscalization is, its benefits and drawbacks analysis and I will also describe what kind of data are we able to retrieve from the EET system in theory. Following the literature review, I will provide a description of the acquired dataset that will be used for this thesis in Chapter 3 Data Description. Chapter 4 Models will deal with model descriptions, the reasoning behind them and the hypotheses and the methodology to test them. Chapter 5 Results will consist of testing results presentation and interpretation. Finally, the thesis will conclude with a Chapter 6 Discussion and Chapter 7 Conclusion, summarizing and evaluating the findings, and suggesting further research ideas for the smoking ban and EET.

Chapter 2

Literature review

2.1 Smoking ban

Smoking ban is a public policy, which is enforced by using laws and regulations. Although Ireland comprehensive ban in 2004 was hailed as a historic move, the first smoking ban was in fact enacted in California in 1998. South Africa later joined the smoke-free environment movement with smoke-free public places with the exemption for bars and restaurants (Koh *et al.* 2007). According to the European Commission, tobacco consumption is the biggest preventable health risk, and the most significant root of premature death in the European Union and is responsible for nearly 700 000 deaths every year (Bogdanovica *et al.* 2011; European Council 2009). In order to address this situation, the European Union is trying to enforce several regulations and tool in order to decrease the impact of smoking and subsequently decrease smoking overall. These regulations and tools are:

- Regulating tobacco products e.g. packaging, labelling, ingredients etc.
- Restricting tobacco products advertisement
- **Promoting smoke-free environments**
- Tax measures

Based on the priorities and the European Council's recommendation on smoke-free environment enforcement, members of the European Union have slowly adopted smoke-free regulations over the years. Seventeen of twenty-eight member states already have a comprehensive smoking ban and among them, Ireland, the United Kingdom, Greece, Bulgaria, Malta, Spain, and Hungary have the strictest legislation (European Commission 2013).

In the Czech Republic, the smoke-free environment proposition was included in the Act 65/2017 On Health Protection Against Addictive Substances and has become valid since the 1st of June 2017. The most significant addition to the smoke-free environment policy, that Act 65/2017 brought, was the prohibition of smoking in hospitality services such as restaurants and bars. Prior to the act 65/2017, the ban was only imposed on the internal area of medical institutions, educational institutions, children related activities, events, whose majority attendee are below 18 years old, public transport and spaces (with exception of air travel, but even there is a restricted area for smoking) and in designated shops, that also offers products for below 18 years old customers (Parlament České republiky 2017).

The reasoning behind the smoke-free environment initiative was that all human beings deserve a high level of health protection and as smoke from smoking activities such as tobacco smoking is toxic, there should be regulations to defend the people who are not smoking. The largest concern of this topic is second-hand smoking. Especially children and adolescents are the most easily affected group as they do not only breath the dreadful smoke but also could eventually pick up smoking later due to the exposure. The European Commission published a document with recommendations, where it stated that member states are bound to develop and/or strengthen strategies and measures to reduce exposure to second-hand tobacco smoke of children and adolescents (European Council 2009). Following later on the publication of the recommendations, smoking was prohibited in public spaces as well as workplace (Parlament České republiky 2017). Several studies in various places in the world confirmed the harmfulness of second-hand smoking (Goláň 2007; Oberg *et al.* 2010). Note that this is not only the concern of the European Union but it is also a worldwide problem. World Health Organizations published several works and recommendations on this topic (World Health Organization and Tobacco Free Initiative 2007). Additionally, with the comprehensive ban for restaurants and bars, the health of employees is accounted into the equation as their well-being is often the most affected by the second-hand smoking in the hospitality industry (Bates *et al.* 2002). Furthermore, according to the study conducted in Marion County, Indiana, second-hand smoke causes many illnesses that need to be treated by the hospital and thus forms a significant part of the expense for health-care (Zollinger *et al.* 2004).

However, on the other hand, various research has reported negative consequences of the smoking ban.

Adams and Cotti mentioned in their work in 2008, that many heavy smokers were also drinkers. Smoking ban would cause them to travel to the neighbor town so that they would able to drink and smoke without worrying about the ban. The underlying theory was that this effect is similar to cross-border shopping. Due to cross-border shopping intuition, people are willing to go outside of their states so that they could have better deals in their point of view. All of that should cause more drunken drivers on the roads, which should lead to a higher rate of accidents. The work indeed found that banning smoking in the bars increased the fatal accident risk. Nonetheless, the work also warned about possible measurement errors and biases as many lucky drunken drivers will not cause any accident thus they would not be recorded (Adams & Cotti 2008). However, the ban in the Czech Republic is nationwide. Germany also has a nationwide smoking ban, Slovakia as well. Poland and Austria do have some exceptions. Looking on this possibility from the ordinary person, the barriers to go to the neighbour state, where the laws and language are different, in order to drink and smoke at the same time are higher than the barriers to go to the next town, which was the case in the United States, thus this argument might not be valid if we reconsider it on the scale of the Czech Republic.

The relationship between smoking and alcohol consumption was also researched by Koksal and Wohlgenant in 2016, where the authors use the theory of rational addiction and apply it as a base to analyze dairy data in the period of 2002-2008 from Customer expenditure survey. Using repeated cross-sections and pseudo-panel data analysis to avoid econometric difficulties such as measurement error, censoring, etc., the authors found out that smoking ban increases alcohol-consumption because non-smokers stay longer in the restaurant and drink more. In the meantime because smokers cannot smoke they are inclined to drink more. This offsets the benefits of a smoke-free environment for the health a bit as there is a trade-off between lungs and liver. The authors therefore also recommend that the smoking ban will come with some supplemental drinking regulation suggestions (Koksal & Wohlgenant 2016).

Another reason, which was often presented, is that the smoking ban would adversely affect sales of several business industries. Various association and communities, usually in hospitality industries, claimed that customer's behavior and choices would change and therefore would lead to a significant decrease in restaurant's and bar's revenue and in extreme cases force the closure of many of them.

This reason was later formulated into the hypothesis of the correlation between sales of the subjects and prohibition and it was widely researched and tested in the United States. Mr. Stanton Arnold Glantz, Ph. D. is a vocal voice in this field with many of his work served as a foundation for many researchers. The effect of Ordinances Requiring Smoke-Free Restaurants on Restaurant sales was the first work that investigated the relation of the ban and the sales. Using tax data for the first 15 US cities, that implemented smoke-free laws affecting restaurants and compare them to the tax data of 15 US cities, that had similar population, income, smoking prevalence and other factors as well. The author used two main metrics to measure and compare the change that occurred. The first metric was the fraction of restaurant sales over total retail sales. The hypothesis was, that the ratio of restaurant sales over total retail sales will drop after imposing the smoking. The second metrics was the fraction of restaurant sales in the city with the ban over comparison city sales without the ban. The hypothesis was, that the ratio between the restaurant sales in town with the ban and the restaurant sales in the city without the ban will decrease after the ban. The conclusion from this work was insignificant changes in the percentage of the ratio of sales and insignificant changes in the ratio of sales between cities with the ban and cities without the ban (Glantz & Smith 1994). A follow-up work, done by the same author, used a similar method but added a quadratic term in time to negate positive serial correlation in residuals. The conclusion this time was an increase in restaurant sales and one significant decrease in term of the comparison ratio (Glantz & Smith 1997). This author also conducted various research on the lobby of the tobacco industry (Dearlove *et al.* 2002).

Building upon previous Glantz works, many other studies were developed in the U.S. with a different approach. The work conducted by Stolzenberg and Allesio in 2007 in California used interrupted time-series auto-regressive integrative moving average (ARIMA) and observed 4% immediate drop in the beginning but then the numbers returned to normal. The work compared trends in revenues for non-alcohol serving restaurants and alcohol-serving restaurant before and after the ban and used the quarterly interval to eliminate yearly confounding historical effects. Using over 99-quarter period data with 3 nonequivalent dependent variables gave the result that there was initially 4% dip but then increased back to normal. However, the authors noted that 12.5% of the sample businesses closed during the 15 month period. Furthermore, the authors mentioned businesses in the research also tried to be creative with setting up out-door smoking areas, which was not contradicting the law (Stolzenberg & D'Alessio 2007).

Work done by Huang in El Paso, Texas used sales tax of restaurants and bars and mixed-beverage tax data during 12 years preceding and 1 year following the ban. Applying multiple linear regression, independent regressors that were considered were a dummy variable indicating, whether the ban was enforced, an ordinal variable to indicate the secular time and three dummy calendar variable to indicate in which quarter was the data collected. Both time series, sales tax of restaurants and bars and mixed beverage tax, did not show any statistically significant change (Huang 2004).

The paper done by Kim and Yoruk in 2015 analyzed the data from the confidential version of PSID (Panel Study of Income dynamics) and estimated the impact of the ban on dining out expenditures and discovered that there is 15.1% decrease in dining out expenditures in smoking households but 8.5% increase in non-smoking households. Pairing it with the fact that the majority of the United States does not smoke, according to the authors, the aggregate impact on dining out expenditure is positive but insignificant (Kim & Yörük 2015). Overall, studies in the United States of America are more inclined to the opinion that enacting the law is beneficial for the public health and not economically detrimental for the restaurants and bar revenues on the aggregate level.

To my best knowledge, there are very few works done regarding this topic from the European perspective. One of them is the research was done by Cornelsen in 2012 in the Republic of Ireland. The author uses ARIMAX (Autoregressive integrated moving average with external regressors) modeling by using data from the Central Statistics Office of 2 metrics, Volume Index of Retail Sales in bars and Aggregate Retail Sales Index. The paper found a small permanent reduction in sales volume that was significant (Cornelsen & Normand 2012). The same author then conducted a meta-analysis to assess the economic impact and came to the conclusion, that the ban did not bring any substantial economic loses and gains (Cornelsen *et al.* 2014). Another research was done in Austria to evaluate the smoking ban. The number of randomly selected restaurants and bars was 172, which were observed using standardized observation. Additionally, to evaluate the satisfaction of the customers, 372 randomly selected customers were interviewed using standardized questions. The work concluded that the partial ban, which is in Austria, was an inefficient solution to the smoke-free problematic (Reichmann & Sommersguter-Reichmann 2012).

In the Czech Republic, although the smoking ban has risen many debate and frustration, there is not much work concerning with it. One of the most vocal voices was the Czech Association of Hotels and Restaurants. They have been active already in 2014 when they released a statement on their websites, stating that the law at that time was sufficient to regulate the smoke-free environment. However, they did not provide any argument besides that even in Europe, the law is not unified (Stárek 2014). To my best knowledge, the Czech Republic there is a work that is analyzing the campaign side of the smoking ban such The analysis of communication of campaigns against smoking by Balatkova in 2013 (Balatková 2013). There is also a work concerning with hospitality employees, which is similar to work done by Bates in 2015 and also confirms its findings (Červenková 2016). But in the scope of the Czech Republic, there is yet a work concerning with the economic impact of the smoking ban.

2.2 VAT Fiscalization

VAT fiscalization is a method of tax control, that is designed to combat tax evasion for subjects, who charge cash for receipts such as small retailers, bars, restaurants, etc. VAT fiscalization law has been enforced in several countries such as Albania (2004), Austria (2016), Bosnia and Herzegovina (2008), Bulgaria (2006/2018), Croatia (2013), Czech Republic (2016), France (2018), Germany (2020), Hungary (2018), Italy (2018), Montenegro (2001), Poland (2022), Romania (1999), Serbia (2004), Slovak Republic (2019), Slovenia (2016). Each country has a slightly different way in application e.g. some countries use hardware solution to track the receipts, some use software solution with a certified key. (Fiscal Solutions d.o.o. 2019)

Certified cash registers were the first proposed method of VAT fiscalization in the Czech Republic. The act has been approved and it's validity had started from 1. July 2005 and first certified cash register should have been introduced in January 2007,(Parlament České republiky 2007) before the act was later abolished by then following Czech Minister of Finance, Miroslav Kalousek (Parlament České republiky 2005).

The current VAT fiscalization act and method was passed on 10. February 2016, proposed by then Minister of Finance, Andrej Babiš, with the name "Elektronická evidence tržeb" - abbreviated as **EET** or "e-tržby". The information campaign started on 1. June 2016, voluntary participation began in November 2016. The implementation of EET was divided into four phases.

The first phase was mandatory for catering, hospitality and accommodation industry and started on 1. December 2016. The second phase was mandatory for small and large retailers and began at 1. March 2017. The third and fourth phase, which were mandatory for other activities with some exceptions, was expected to launch in June 2018 but were delayed by Constitutional court in December 2017 (Ústavní soud České Republiky 2018).

Since the beginning, VAT fiscalization sparked many debates about its usefulness and overall impact. According to the official information sites about EET, it should promote a better entrepreneur environment as it will provide fairer conditions on the market and also better-targeted control resulting in less administration for the entrepreneurs and business entities. And mainly, it would help the government to become more effective in tax collection which would result in more transfer payments availability (Finanční správa 2016).

Analyses in foreign countries that implemented EET on VAT effectiveness collection are however inclined more to the opinion that it does not have any effect. In Tanzania, there were not found any significant coefficients on the relation between rolling out Electronic Fiscal Devices (EFDs), a VAT Fiscalization solution in Tanzania and VAT collection. Thus there is not enough evidence to link the effectiveness of VAT collection and EFDs (Chege *et al.* 2015). On the other hand, the analysis done by Andreja Katolik in Croatia concluded that by enforcing Fiscalization Law resulted in an increase of taxable deliveries. This analysis was done by analyzing only Taxable account on panel data of 3 years nonetheless it did not account for possible growth of the economy, inflation in the calculation thus we believe this analysis needs further review (Katolik *et al.* 2014). Another work done in Croatian scope by Šimovič, on the other hand, concluded that VAT efficiency did not increase by comparing the implicit tax rate and standard VAT efficiency indicators such as Efficiency Ratio (ER), C-Efficiency Ratio (CER) and VAT Revenue Ratio (VRR) (Šimović & Deskar-Škrbić 2015). In the Czech Republic, the reported contribution of EET on VAT collection in the Czech Republic reported by Alena Schillerová (Ministerstvo Financí České republicky 2019), current Minister of Finance, was objected by Vladimír Škop, Ph.D. and stated, that the analysis is misleading, as the methodology calculated was incorrect in the author's opinion (Škop 2019).

All the past work so far has been analyzing the main propagated added value, better, more efficient tax collection. But there is an externality arising from this system and that is data collecting process, which to my best knowledge was not considered in the works. Thanks to the technical implementation, it is relatively easy to extract the sales data. Each taxpayer is required to request access log-in name from the Tax authority. With this log-in name, the taxpayer can download a certificate from the EET site. with the certificate, the taxpayer is going to sign the packages of information, that is being sent to the tax authorities. This requires that the taxpayer has a program, that can process the certificate. The data are immediately sent with the amount of the sale, the identification code of the entity and the selling location, date and time (Finanční správa 2016). It is a software solution, in comparison to fiscal, registered, certified cash registers, which were initially considered in the Czech Republic in 1999 (Parlament České republiky 2005). Such hardware solution was in fact implemented in Montenegro, and Bosnia and Herzegovina (Fiscal Solutions d.o.o. 2019). The software solution offers a much more elegant and modern approach as it addresses several drawbacks of cash registers such as limited licenses to cash register sellers or purchasing costs. By providing the framework so that everyone is able to develop a software solution, this avoids the problem of giving out limited licenses to cash register sellers thus avoiding perfect competition for cash registers which would lead to low quality and overpriced products (Finanční správa 2016). Purchasing costs for a natural person are also low and thanks to the subsidiary for the first purchase of the hardware (Finanční správa 2016), it lifts the burden substantially.

The advantage of these EET data is that they are usually sent in real time and have a high level of categorization and granularity. Such data are flexible and can be used for deeper and region-focused analysis for various policies. o

Chapter 3

Data Description

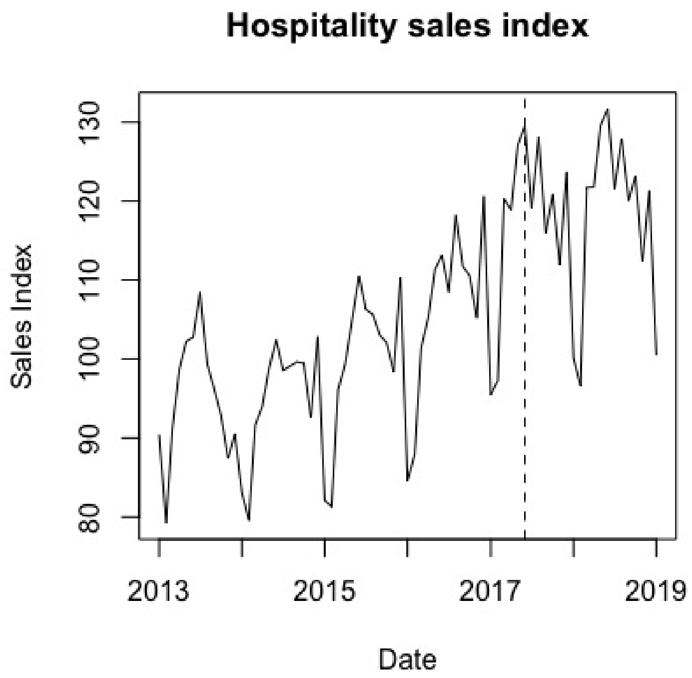
3.1 Dataset from the Czech Statistical Office

This dataset was retrieved via the Czech Statistical Office website. It is freely available for the public. The dataset shows sales index (SI) in percentage point throughout the months and years. The base 100 percent value is the average of SI in the year 2015. The granularity of our dataset is monthly time series from January 2013 until January 2019. Calculation of the monthly SI is sourced from "SP 1-12" survey with addendum specific for trade, accommodation industry, transportation and storage, information and communication and market services. Aggregate sales are observed monthly from the sale of products, services, and goods. The VAT is excluded (Czech Statistical Office 2019).

The year-on-year (y-o-y) sales indices are calculated for individual months. Then, year to year price deflators are used for each month of the current year. Absolute values of monthly sales at current prices are converted into the price level of the base year using the method of chain linked deflators. By summing these numbers we get y-o-y sales indices at a constant price of the base price. Thus in the base year, the sum of monthly sales in the current price in the base year should be equal to the sum of the constant price in the period. Additionally, the averaged value of the sum should be 100. This is known as constant prices and we will be using these numbers to evade inflation treatment. For our dataset, the constant price base is 2015 (Czech Statistical Office 2019).

We extracted **NACE 56 Food and beverage service activities** SI. These SI contains subcategories **ACE 56.1 Restaurants and mobile food service activities**, **NACE 56.2 Event catering and other food service activities**, **NACE 56.3 Beverage serving activities**. 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 as our proxy for the **NACE 56.3 Beverage serving activities**, which consists of restaurants and bars.

Figure 3.1: Yearly seasonality in the plot of hospitality SI of CSO dataset



Annual seasonality easily observable via visual inspection as the SI rises until in the middle of the year and then start to decrease periodically. Additionally, we can observe the trend as each year value tends to increase

Source: Author's own calculation based on CSO dataset

In the Figure 3.1 we can identify annual seasonality in the time series, as hospitality sales tend to increase until June, July and then starts to decrease until November. In December, there is a slight increase in SI, possibly caused by pre-Christmas events and New Years events. We can also identify a rising trend in SI. The reason might be that the people are dining out more in recent years. Table 3.1 shows some basic simple descriptive statistics of the time series.

We also extracted SI of **NACE 47 excl. 47.3 Retail sales excluding sales of automotive fuels** from the CSO for the same period. This time series

Table 3.1: Descriptive statistics of CSO data of hospitality sales

Hospitality	SI in the period	Pre-law period	Post-law period
Observations	73	53	20
Min	79.24	79.24	96.61
1st Q.	96.61	92.83	115.10
Mean	102.82	99.43	121.45
Median	105.43	100.35	118.89
3rd Q	118.23	106.30	124.77
Max	131.68	127.08	131.68
Var	182.77	120.09	101.15

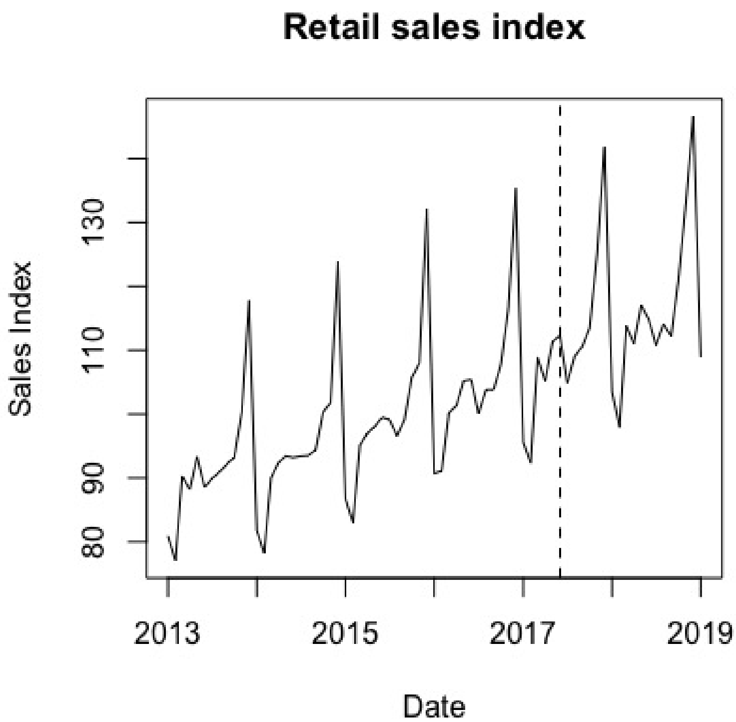
Source: Author's own calculation based on CSO dataset

Table 3.2: Descriptive statistics of CSO data of retail sales excl. sales of automotive fuels

Retail	SI in the period	Pre-law period	Post-law period
Observations	73	53	20
Min	77.10	77.10	97.9
1st Q.	93.20	91.10	110.2
Mean	100.30	96.60	112.9
Median	103.20	98.36	116.1
3rd Q	111.00	103.80	118.1
Max	146.60	135.40	146.6
Var	136.61	140.30	149.15

Source: Author's own calculation based on CSO dataset

Figure 3.2: Yearly seasonality in the plot of retail sales SI of CSO dataset



Annual seasonality and trend is easily observable via visual inspection as the SI rises then stagnate for a bit until right before the end of the year, when it starts to rise and peak for that year. The peak might be caused by Christmas shopping spree.

Source: Author's own calculation based on CSO dataset will serve as a regressor for one of our model. It is a proxy variable for the economic situation, similarly used in the work conducted in Ireland (Cornelsen & Normand 2012). In Figure 3.2 we can see annual seasonality as the retail sales often peaks in the period before the end of the year. It might be due to the shopping spree before Christmas. We can also easily see the rising trend in the time series as the values tend to increase each year. We do not observe visible significant change after the smoking ban, that might be caused by increased alcohol consumption outside of the restaurants and bars. In Table 3.2, we have simple descriptive statistics on the time series.

3.2 Dataset from the EET system

This dataset was retrieved by Mr. Marek Sušický by submitting a request and paying 5 000 CZK fee. The dataset was then distributed for free via his own created website *www.datazeet.cz* with some simple summaries of the dataset status and additional simple analysis using the dataset. The dataset period received is from 1.12.2016 to 24.11.2017. The data scheme is relatively simple as depicted in Table 3.3. According to Mr. Sušický, the Financial Administration is able to provide even smaller granularity, such as an hourly time series. This, however, requires approximately 150 000 CZK fee. The author also stated that the Financial Administration is unable to provide the EET data regularly because of the lack of manpower (Sušický 2018).

Table 3.3: EET dataset column name scheme and description

Name of the column	Explanation
DEN	Day
KOD_CINNOSTI_PROVOZOVNY	Code for econ. activity
KOD NACE	Statist. class. of econ. activities
NAZEV_CINNOSTI_PROVOZOVNY	Econ. activity name
KRAJ	Region
CELKOVA_TRZBA	Overall turnover
ZAKL_NEPODL_DPH	Not taxed base
DAN 21%	21% Tax
DAN 15%	15% Tax
DAN 10%	10% Tax

Source: Author's own calculations based on EET dataset

The dataset also has NACE subcategories thus we are able to filter out only **NACE 56.3 Beverage serving activities**. If the dataset provided was much longer, we could have made SI from this dataset to compare it to the SI in the CSO. The dataset, however, has many measurement errors. Many observations were negative millions (see Appendix A). For sales of one region and on one day, such negative amount is unthinkable. To some extent, negative sales can be possible due to refunds. However, with the amount of being in millions CZK, it is more reasonable to suspect that this might be more of a measurement error than millions of people demanding refunds in one region on one day. To fix this issue, those numbers were interpolated, although this procedure is not generally recommended Bisgaard & Kulahci (2011), but in our case, this is a severe measurement error which will be definitely affecting our analysis. Using interpolation, we shall get a more accurate measure of the sales on that day than the negative sales reported.

Table 3.4: Descriptive statistics of sales of hospitality from EET system before the treatment

Hospitality	Sales in the period	Pre-law period	Post-law period
Observations	359	182	177
Min	**-197.2**	**-197.2**	23.46
1st Q.	151.6	138.5	168.82
Mean	186.8	159.1	194.21
Median	181.0	167.1	195.36
3rd Q	213.5	199.4	224.31
Max	483.5	483.5	327.84
Var	2832.963	3699.853	1549.289

Notice the negative minimal value. For one single day in the whole country, negative sales of negative -197.2 millions CZK is unimaginable.

Source: Author's own calculations based on EET dataset

Before we treated the data, we see in Table 3.4 that there was a minimal value of sales in a day negative 192.2 million CZK, which is highly probably a substantial measurement error.

After the treatment, the values in the post-law period seem more balanced. There is a slight increase in means. The variance was decreased substantially. But from the basic intuition, these data have much more sense than

Table 3.5: Descriptive statistics of sales of hospitality from EET system after the treatment

Hospitality	Sales in the period	Pre-law period	Post-law period
Observations	359	182	177
Min	62.95	62.95	132.3
1st Q.	157.45	141.65	171.4
Mean	181.91	163.39	197.2
Median	187.32	175.29	199.7
3rd Q	216.74	203.12	225.3
Max	483.50	483.50	327.8
Var	2054.386	2525.278	1278.555

Notice the change in negative minimal value. Such statistics make much more sense in this regard.

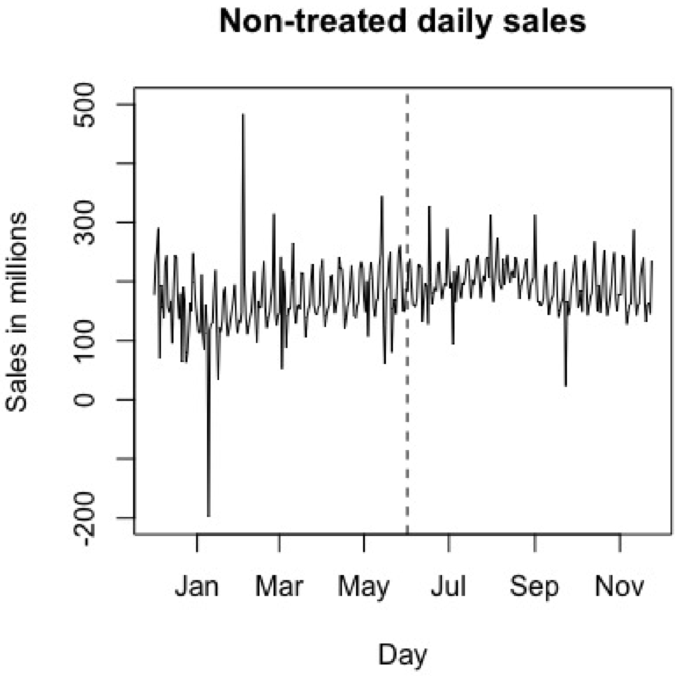
Source: Author's own calculations based on EET dataset

pre-treatment data, especially the millions of negative sales. Therefore, for further analysis, we will be working with treated data of EET dataset

The visual inspection of Figure 3.3 suggests, that there seems to be no drastic significant change after the smoking ban was enforced for both treated and non treated data.

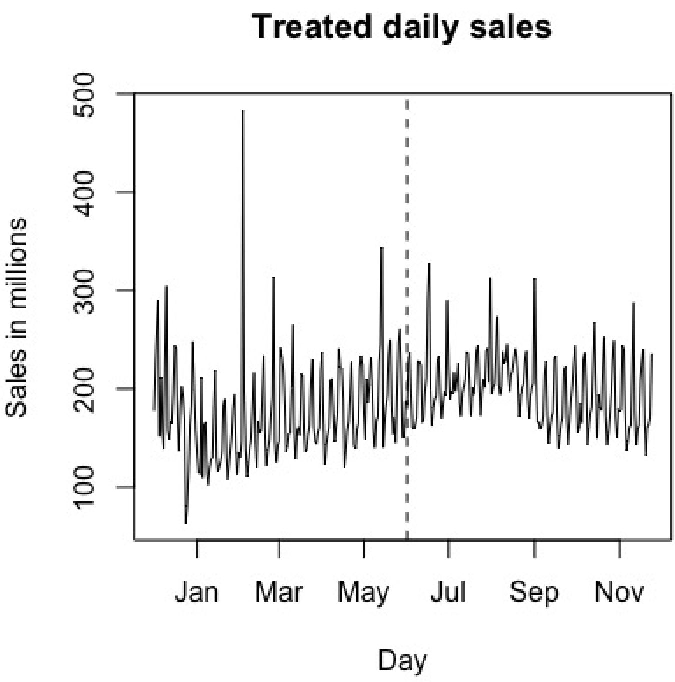
Similarly, we will be extracting the proxy for economic activity, the retail trade sales. In the Czech Republic, from **NACE 47 Retail trade, except of motor vehicles and motorcycles**, activities **NACE 47.2 - Retail sale of food, beverages and tobacco in specialized stores, NACE 47.4 - Retail sale of information and communication equipment in specialized stores ,NACE 47.5 - Retail sale of other household equipment in specialized stores, NACE 47.7 - Retail sale of other goods in specialized stores, NACE 47.8 - Retail sale via stalls and markets, NACE 47.9 - Retail trade not in stores, stalls or markets** were filtered. For the use of the model, that requires retail sales, we will shorten our dataset to begin from the mandatory period which is from 1.3.2017, as the number of volunteers is not so high thus creating a dimensional shift when the law was mandatory as seen in Figure 3.4. Because the data also have similar measurement errors, we will be using the same method for corrections as we did with hospitality sales.

Figure 3.3: EET hospitality series daily series



(a) *Non-treated hospitality sales displays no significant shift in the daily sales after smoking ban*

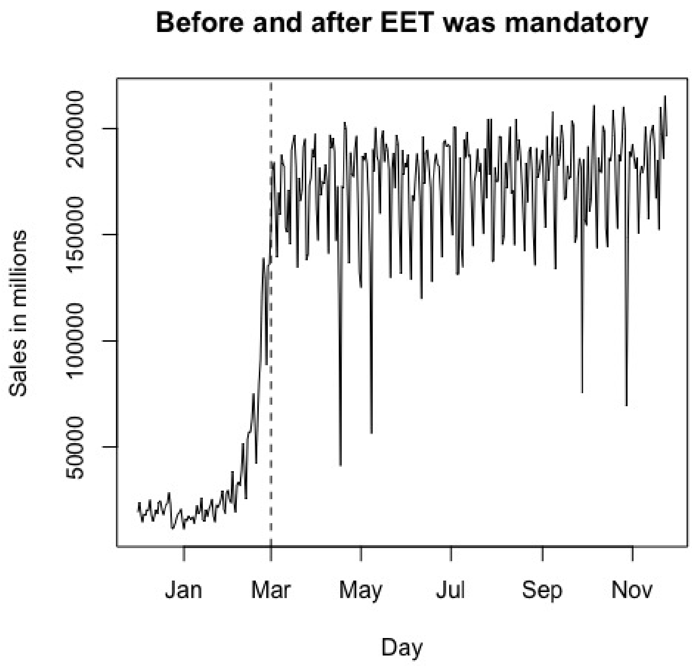
Source: Author's own calculation based on EET dataset



(b) *Treated hospitality sales displays no significant shift in the daily sales after smoking ban*

Source: Author's own calculation based on EET dataset

Figure 3.4: EET retail sales daily series



The shift between the sales before and after EET was mandatory for the retail sales is significant from visual inspection. The retail sales numbers almost doubled. We can observe some dips, which might be due to holiday closing policy.

Source: Author's own calculation based on EET dataset

Chapter 4

Model and hypothesis formulation

4.1 Time series problems

The definition of time series is specified as data collected from a system (process) over time. These data are usually not random, but have some sort of immobility and does not change quickly (Bisgaard & Kulahci 2011). Combined with the sampling frequency, ensuing observations tend to be correlated. This is called autocorrelation (Bisgaard & Kulahci 2011). This breaches the assumptions for a linear regression model, which would result in misleading or even useless estimations. Therefore there is a need to implement other methods and models to inspect time series. One of the most popular models is ARIMA using Box-Jenkins methodology (Box *et al.* 2016). An important requirement for time series is that they need to fulfill the minimum of stationarity so that the models can be developed.

4.1.1 Stationarity of a time series

The fundamental block of any time series analysis is a stationary time series. Stationarity is defined in (Bisgaard & Kulahci 2011) as: joint probability distribution of any n observations of the time series that is $\{y_{t+1}, y_{t+2}, ..., y_{t+n}$ has the same joint probability as another set of n observations of the same time series shifted by k time units that is $\{y_{t+1+k}, y_{t+2+k}, ..., y_{t+n+k}\}$. This is called strict stationarity. Weak stationarity is defined as a finite variance process, whose mean and variance remain the same through the time and the correlation among the observations depends on the distance in time units between them (also called lag dependent). Mathematically, we define time series y_t as weakly stationary, if $E(y_t|t) = \mu$ and $Cov(y_{t_1}, y_{t_2}) = Cov(y_{t_1+h}, y_{t_2+h})$ for all

h. This attribute allows us to develop models and forecast as it serves us as an anchor that remains the same. However, in areas, which are in the field of interest for time series analysis, such as business, economics, industrial applications, most time series are non-stationary. A solution to this problem is not to look at the raw data but at the change between the successive observations. These called differenced time series are often stationary. Thanks to that, we are able to develop models and forecasts on the changes between the observation and then apply it to the raw data (Bisgaard & Kulahci 2011).

4.1.2 ARMA process

ARMA (p,q) or AutoRegressive Moving Average process is the base for ARIMA. It is a combination of AR(p) and MA(q) processes.

AR(p) process is defined as (McDowall *et al.* 1980):

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + u_t, \qquad u_t \sim N(0, \sigma^2)$$

and displays relation of the current observation y_t to the past observations y_{t-1}, $...y_{t-p}$.

MA(q) process is defined as (McDowall *et al.* 1980):

$$y_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - ... - \theta_q u_{t-q}$$

and illustrates how the "averages" of past and present noise terms move.

Combining them together, we obtain ARMA(p,q) which is written as (McDowall *et al.* 1980):

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - ... - \theta_q u_{t-q}$$

u_t component is called white noise and has a normal distribution with zero mean and variance σ^2 or $u_t \sim N(0, \sigma^2$. This is the foundation for all ARMA models and subsequently devised models such as ARIMA, ARIMAX.

4.1.3 ARIMA process

Rarely there are processes that are stationary in the raw state as mentioned in Subsection 4.1.1. Often to obtain the stationary process, one needs to transform (integrate) non-stationary process into stationary by e.g. differencing. To simplify the explanation of the ARIMA process, one can picture the ARIMA

Figure 4.1: ARIMA model process

White noise term u_t enters the ARIMA "filters" to become ARIMA. The order of the "filters" determine the parameters (p,d,q) of the ARIMA process.

Source: Interrupted time series analysis (McDowall *et al.* 1980)

Figure 4.2: Reversed model process

To analyze time series using ARIMA models, we need to identify the parameters (p,d,q) such that after entering the identified order of the "filters", the beginning residuals will become white noise.

Source: Author's own calculation based on Interrupted time series analysis (McDowall *et al.* 1980)

process as white noise entering and passing through a system of "filters" such as Integration, Autoregressive and Moving Average to become ARIMA process as depicted in Figure 4.1 (McDowall *et al.* 1980).

Each of the "filters" will modify the white noise. The task is to identify the parameters p, d, q, so that if our time series enters the "reversed" process, the result will be the white noise. as depicted in Figure 4.2.

4.1.4 ARIMAX process

Building upon ARIMA is ARIMAX(p,d,q) which Auto Regressive Integrated Moving Average with Exogenous input. It is written as:

$$y_t = \beta_1 x_1 t + \beta_2 x_2 t + ... + \beta_k x_k t + \eta_t, \quad \eta_t \sim ARIMA(p, d, q)$$

with following requirements:

- y_t is a stationary process
- x_t is a stationary process

The model is also used as a Regression model with ARIMA errors. As mentioned before Subsection 4.1.1, many processes are non-stationary and becomes stationary by differencing. It is also desired to maintain the relationship between regressand and regressor. Thus if any variable needs differencing, it is advised to difference the others too. This will become the "model in differences" in comparison of "model in level" that is often encountered in classical linear regression (Hyndman & Athanasopoulos 2019).

4.1.5 Seasonal models

Much common time series have some kind of seasonality e.g. Sales on current Saturday depends more on how much the sales were on the last Saturday than on sales on the previous day, Friday. The seasonal model has syntax ARIMA(p,d,q)(P,D,Q)[n], where n is the periodicity e.g. monthly data will have $n = 12$, quarterly $n = 4$ and daily $n = 7$.

Parameter D indicates seasonal nonstationarity, where time series might trend or drift in yearly steps such. With parameters $D = 1$ and $n = 12$, the process is defined as (McDowall *et al.* 1980):

$$y_t - y_{t-12} = \theta_0, \quad \theta_0 \text{ is stationary}$$

Parameter P indicates seasonal autoregression, which means that the current value may depend on the corresponding previous period observation. With parameters P = 1 and n = 12 the process is defined as (McDowall *et al.* 1980):

$$y_t = \beta_{12} y_{t-12} + u_t, \quad u_t \sim N(0, \sigma^2)$$

Parameter Q indicates seasonal moving average, which means that the current value may be affected by random shocks from the previous period. With parameters Q = 1 and n = 12 the process is defined as (McDowall *et al.* 1980):

$$y_t = u_t + \theta_{12} u_{t-12}, \quad u_t \sim N(0, \sigma^2)$$

4.1.6 Criteria for model selection

Lots of models with different parameters will seem to fit our datasets. From them, we will try to select the one, that fit the most to our data by using model fitting criterion.

The most popular criteria are Akaike's information criterion and Bayesian's information criterion. The former one is defined for n observations as:

$$AIC = -2\ln(\text{maximized likelihood} + 2r \approx n\ln(\hat{\sigma}_u^2) + 2r \qquad (4.1)$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of the residual variance σ_u^2, r is the number of parameters estimated in the model including a possible constant term. The latter one is defined as:

$$BIC = -2\ln(\text{maximized likelihood}) + r\ln(n) \approx n\ln(\sigma_u^2) + r\ln(n) \qquad (4.2)$$

To select the best models, we choose to minimize either AIC or BIC or if the results allow, both. Adding additional parameter will reduce the residual variance $\hat{\sigma}_u^2$ thus decrease in term $n\ln(\hat{\sigma}_u^2)$ but will also increase r which in result act as a penalty term. Adding extra parameter to a current model is a benefit at the moment when the decrease in the variance of the residuals is not offsetted by the added parameter. However, it turns out that AIC tends to overestimate autoregressive models AR(p) hence overestimating p. On the other hand, as seen from the definition of the BIC, the penalty term $r\ln(n)$ is much more severe in comparison with AIC. This leads to the fact that BIC often selects simpler and parsimonious models than AIC but falls to capture the model enough if the number of parameters required is high (Bisgaard & Kulahci 2011). We will prefer AIC as our primary if the sum of models parameters is higher than 3 and BIC if the sum of the models parameters is lower or equal to 3.

4.2 Hypotheses for CSO dataset

The ban in the Czech Republic was introduced throughout the whole country which is similar to the ban that was analyzed in the work by Cornelsen in the Ireland (Cornelsen & Normand 2012) in contrast with the United States, where the bans were gradually introduced just in some areas (Glantz & Charlesworth 1999). The former work also utilized data from the Central Statistics Office in Ireland, an institution equivalent of Czech Statistical Office. Thus we will be implementing similar intuition to form our logic.

The logic behind the proposed models are:

1. If the sales of restaurants were severely affected by the smoking ban, there should be a fundamental shift in the sales index from June 2017 onward.

This implies that if we forecasted sales indexes from the data until June 2017 and compare them to the actual data from June 2017, we should see a significant difference.

2. If the sales of restaurants were heavily affected by the smoking ban the relation between the ban and sales indexes should be significant and negative.

3. Based on the work done in the Ireland (Cornelsen & Normand 2012), the sales of restaurants will be more affected by the economic situation in the country than on the introduction of the smoking prohibition in restaurants and bars.

From that we form following hypotheses:

1. Given common sense, that the smoking ban did not affect the sales of restaurants, forecasted and recorded values should not be significantly different therefore the null hypothesis to test is

   H_0: Significant number of observations of the recorded values does not break out of the forecasted confidence interval values.

2. Based on the initial thought, the model is formulated as:

$$Hospitalitysales_t = \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(p,d,q)$$

   with $Hospitalitysales_t$ is sales index of hospitality, law_t is a dummy variable for law implementation, η_t are the errors that follows ARIMA process. Then the null hypothesis to test is:

$$H_0 : \delta = 0$$

3. Based on the initial thought, the model is formulated as:

$$Hospitalitysales_t = \beta retailsales_t + \delta law_t + \eta_t \quad \eta_t \sim ARIMA(p,d,q)$$

   with $Hospitalitysales_t$ is sales index of hospitality, $retailsales_t$ is sales index of retail sales, law_t is a dummy variable for law implementation, η_t are the errors that follows ARIMA process. Then the null hypothesis to test is:

$$H_0 : \delta = 0$$

4.3 Hypotheses for EET dataset

Past published works have rarely worked with data with daily granularity, probably due to their unavailability, as it was technologically and practically inconvenient or even impossible to sample such data. With the EET system, this is less of a problem. Lower granularity sales data as daily or hourly granularity time series might be immensely powerful tools for investigating public policies such as their impact and consequences. Although our data has daily granularity, which is fairly flexible, the disadvantage of this data source is its shortness, which might cause troubles when determining seasonality of the time series, and somewhat not cleaned data, where we encountered several possible measurement errors. To build our models we can use the same intuition as in Section 4.2. Thus following hypotheses are formed:

1. Given common sense, that the smoking ban did not affect the sales of restaurants, forecasted and recorded values should not be significantly different therefore the null hypothesis to test is

   H_0: Significant number of observations of the recorded values does not break out of the forecasted confidence interval values.

2. The relation can be written in a model as:

   $$Hospitalitysales_t = \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(p, d, q)$$

   with $Hospitalitysales_t$ is sales index of hospitality, law_t is a dummy variable for law implementation, η_t are the errors that follows ARIMA process. Then the null hypothesis to test is:

   $$H_0 : \delta = 0$$

3. Based on the initial thought, the model is formulated as:

   $$Hospitalitysales_t = \beta retailsales_t + \delta law_t + \eta_t \quad \eta_t \sim ARIMA(p, d, q)$$

   with $Hospitalitysales_t$ is sales index of hospitality, $retailsales_t$ is sales index of retail sales, law_t is a dummy variable for law implementation, η_t are the errors that follows ARIMA process. Then the null hypothesis to test is:

   $$H_0 : \delta = 0$$

4.4 Methodology for testing hypotheses

4.4.1 Forecast and comparison with recorded value

With the below described steps, we will be testing the null hypothesis

H_0: Significant part of the recorded values does not break out of the forecasted confidence interval values.

1. We identify the parameters of the most fitted ARIMA model on the pre-law period
2. We conduct residuals diagnostics to ensure that they follow white noise
3. We forecast with the identified parameters the post-law period
4. We visually inspect the plot of the fitted model, forecast and recorded value and compare

4.4.2 Using ARIMA with law dummy variable

With the below-described steps, we will be testing the null hypothesis

$$Hospitalitysales_t = \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(p, d, q)$$
$$H_0 : \delta = 0$$

i. Interrupted time series analysis approach (=ITSA) (McDowall *et al.* 1980)

   1. We identify the parameters of the most fitted ARIMA model on the pre-law period
   2. We regress the time series on the law dummy variable on the whole time period with parameters estimated in the previous step
   3. We inspect the coefficients for significance and value

ii. Auto Regressive Integrated Moving Average with Exogeneous Input regression (=ARIMAX) (Hyndman & Athanasopoulos 2019)

   1. We identify the parameters of the most fitted ARIMAX model on the whole time period
   2. We conduct residuals diagnostics to ensure that they follow white noise
   3. We inspect the coefficients for significance and value

4.4.3 ARIMAX with regressors dummy variable law and retail sales variable

With the below-described steps, we will be testing the null hypothesis

$$Hospitalitysales_t = \beta retailsales_t + \delta law_t + \eta_t \quad \eta_t \sim ARIMA(p,d,q)$$
$$H_0 : \delta = 0$$

1. We identify the parameters of the most fitted ARIMAX model on the whole time period
2. We conduct residuals diagnostics to ensure that they follow white noise
3. We inspect the coefficients for significance and value

Chapter 5

Results

In this chapter, we will be presenting the results of each hypothesis testing with different methods (Forecasting, ITSA approach with dummy variable law and ARIMAX with regressor dummy variable law, and ARIMAX with regressors dummy variable law and retail sales variable) and with both datasets (CSO, EET). From the results of the testing, we shall be able to draw a conclusion on the significance of the effect of the smoking ban on hospitality retail sales.

5.1 Forecast

Using forecasting, we will be testing the null hypothesis

H_0: Significant part of the recorded values does not break out of the forecasted confidence interval values.

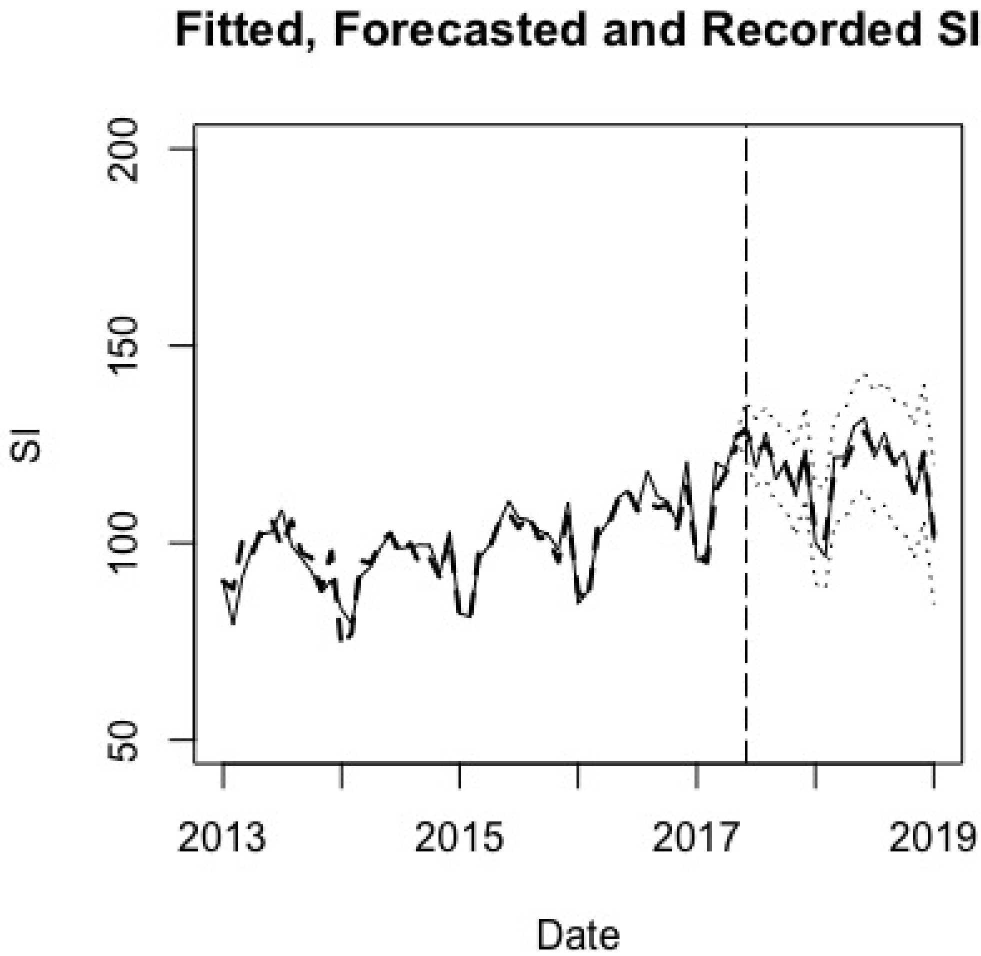
Forecasting is the prediction of the post-law period based on the pre-law period. If the forecast is an inaccurate or a significant number of recorded values break out of the confidence interval values of the forecast, a potential reason might be that the law has a significant effect on the sales.

5.1.1 CSO dataset

The most fitted model for forecasting CSO dataset is ARIMA(0,1,1) (see Subsection B.1.1 for step by step).

As seen in Figure 5.1, the forecast is relatively accurate and in the Figure 5.2 one can see that all recorded values are contained in the confidence interval values of the forecast. Thus there is not a significant number of observations

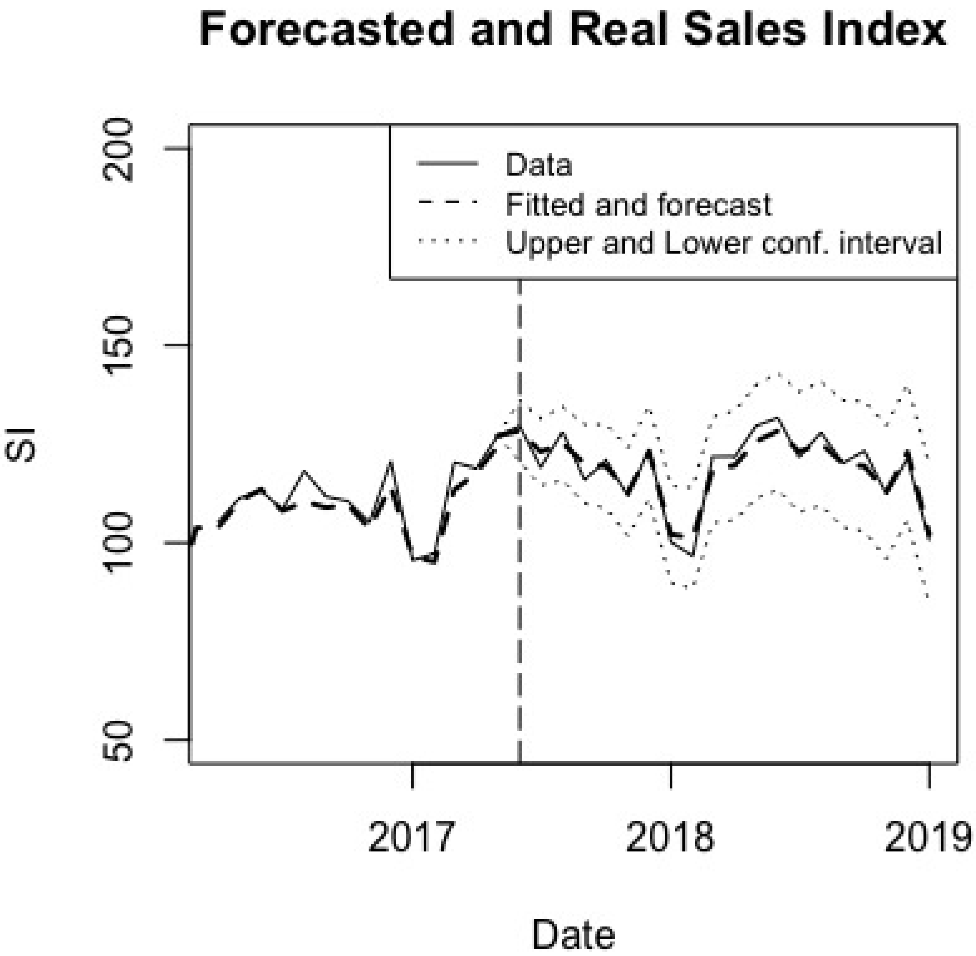
Figure 5.1: Fitted and forecast in comparison with CSO data



The model fits fairly well as we see small differences between the fitted (dashed line) and the recorded (solid) line.

Source: Author's own calculation

Figure 5.2: Fitted and forecast in comparison with CSO data in the post-law period



In the post-law period, the recorded values (solid line) are contained in the confidence interval values (dotted line) of the forecast.

Source: Author's own calculation

that breaks out of the confidence interval values. Therefore, we are unable to reject the null hypothesis

H_0: Significant part of the recorded values does not break out of the forecasted confidence interval values.

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using forecast on CSO dataset.

5.1.2 EET dataset

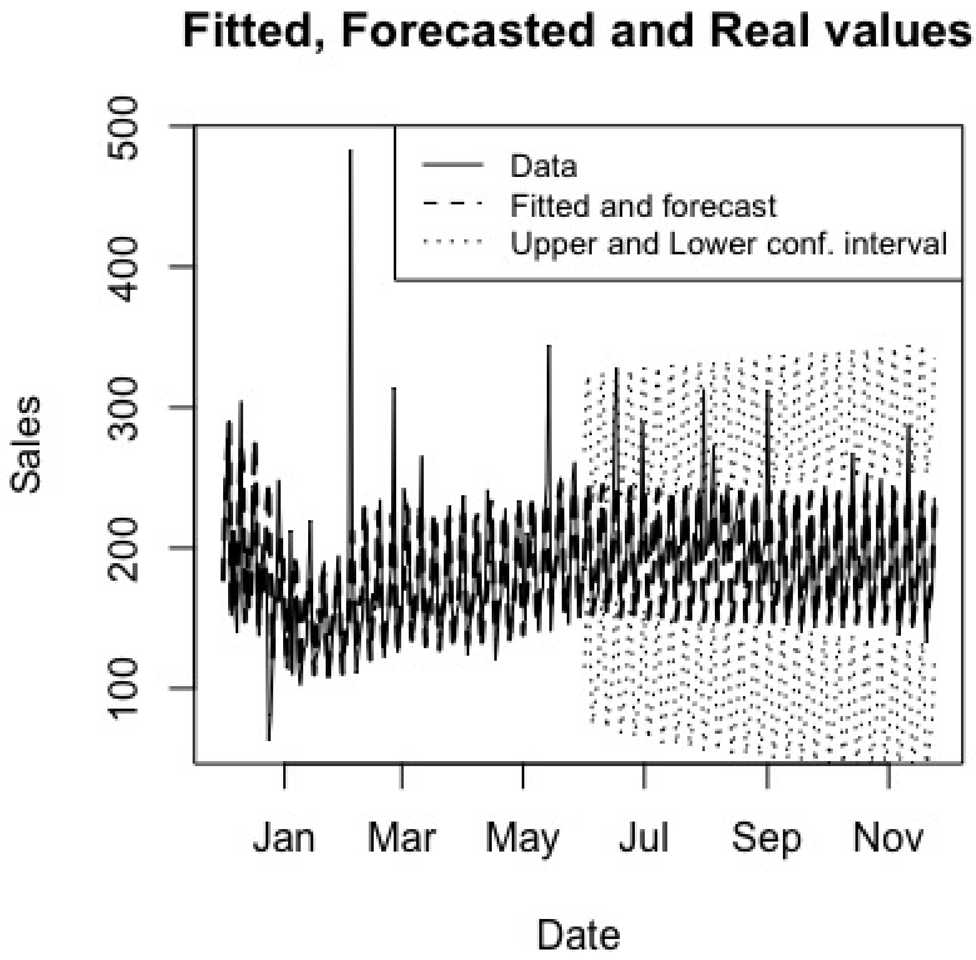
The most fitted model for forecasting EET dataset is ARIMA(1,0,1)(0,1,1)[7] (see Subsection B.2.1 for step by step).

Visual inspection of the plot of the forecast and fitted model depicted in Figure 5.3 shows that the model used is relatively fit and accurate. Inspecting the values in the post-law period in the Figure 5.4, majority of the recorded values lies inside of the confidence interval, apart from 2 observations, which lies outside of the confidence interval and 2 observations, which lies on the confidence interval value (see Figure B.11. The model forecasted 177 observation, from which 4 were out of the confidence interval values, which is approximately 2% of observations out of the confidence interval, which is insignificant. Thus we can not reject the null hypothesis

H_0: Significant part of the recorded values should not break out of the forecasted confidence interval values.

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using forecast on EET dataset.

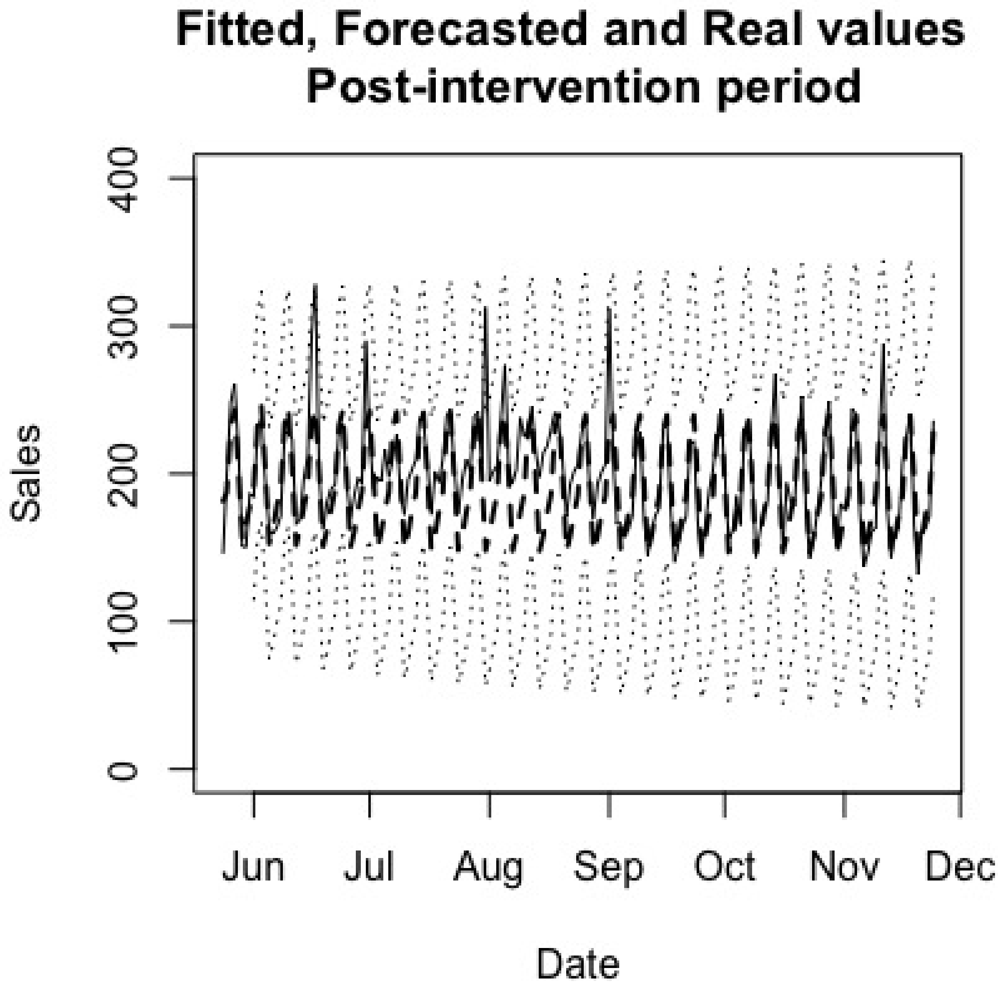
Figure 5.3: Fitted and forecasted values in comparison with EET data



The model fits fairly well as we can observe a small difference between fitted (dashed line) and recorded (solid line) values, apart from the outliers spiking moments. This might be due to high discount sales and people bought more thus there is another underlying parameter that we might not have taken into account. But apart from that, the fitted model represents the graph relatively well enough.

Source: Author's own calculation

Figure 5.4: Fitted and forecasted values in comparison with EET data in the post-law period



In the post-law period, the recorded values (solid line) are fairly in the bound of confidence interval values (dotted line) apart from some observations, that are spiking out of the upper confidence interval

Source: Author's own calculation

5.2 ITSA approach with dummy variable law

Using ITSA approach, we will be testing the null hypothesis of the model

$$Hospitalitysales_t = \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(p,d,q)$$
$$H_0 : \delta = 0$$

Identifying the most fitted model on the pre-law period and then run regression on the law dummy variable for the whole period and check whether the coefficient for the law dummy variable is significant or not. If it is not then the law does not have a significant effect on the sales of restaurants and bars.

5.2.1 CSO dataset

The most fitted model on the pre-law period is found to be ARIMA(0,1,1) as we have discovered in Subsection 5.1.1. Our model is then defined as:

$$\begin{aligned} sales_t &= \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(0,1,1) \\ sales'_t &= \delta law'_t + u_t + \alpha u_{t-1} \\ u_t &\sim N(0, \sigma^2), \quad {}' \text{ is first differencing} \end{aligned}$$

With this model identification, we proceed with the regression on the whole time series with the law as dummy variable regressor. Such regression with given parameters generate coefficients estimates reported in Table 5.1.

The coefficient estimate of *law* dummy variable variable in *law* in Table 5.1 is insignificant thus we are unable to reject the second null hypothesis

$$H_0 : \ \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using ITSA approach on CSO dataset.

Table 5.1: Coefficients estimates of ARIMAX(0,1,1) on law ITSA approach based on CSO dataset

	Coefficients estimate:
α	-0.568^{***}
	(0.086)
δ	$\mathbf{-0.671}$
	(2.558)
Note:	*p<0.1; **p<0.05; ***p<0.01

α is the coefficient of the moving average term, and δ is the coefficient of law variable term and also our main focus. We can see that α is significant (indicated by the stars next to the estimate) and is different from zero but δ is not significant even at 10% level of significance.

Source: Author's own calculation

5.2.2 EET dataset

The most fitted model on pre-law period is ARIMA(1,0,1)(0,1,1)[7], as we have identified in Subsection 5.1.2. The model is then written as:

$$\begin{aligned} sales_t &= \delta law_t + \eta_t, \eta \sim ARIMA(1,0,1)(0,1,1)[7] \\ sales'_t &= \delta law'_t + \beta\eta'_{t-1} + u_t + \alpha u_{t-1} + \Theta(u_{t-7} + \alpha u_{t-8}) \\ u_t &\sim N(0, \sigma^2), \quad ' \text{ is seasonal differencing} \end{aligned}$$

The estimates of the coefficients of such model is reported in Table 5.2.

The Table 5.2 reports, that the coefficient estimate of the *law* dummy variable is not significant therefore we can not reject the null hypothesis

$$H_0: \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using ITSA approach on EET dataset.

Table 5.2: Coefficient estimates of ARIMAX(1,0,1)(0,1,1)[7] on law ITSA approach based on EET dataset

	Coefficients estimates:
β	0.983^{***}
	(0.018)
α	-0.881^{***}
	(0.031)
Θ	-1.000^{***}
	(0.031)
δ	**5.124**
	(14.622)
Note:	*p<0.1; **p<0.05; ***p<0.01

α is the coefficient of the moving average term, ϕ is the coefficient of the autoregressive term, Θ is the coefficient of the seasonal moving average term, δ is the coefficient of law variable term and also our main focus. We can see that all α, ϕ, Θ are significant (indicated by the stars next to the estimate) and are different from zero but δ is the only one that is not significant even at 10% level of significance.

Source: Author's own calculation

5.3 ARIMAX with regressor dummy variable law

Using Regression with ARIMA errors approach, we will be testing the null hypothesis of the model

$$Hospitalitysales_t = \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(p,d,q)$$
$$H_0 : \delta = 0$$

Regression with ARIMA errors approach means that we will be regressing the time series with suspected models. We will then proceed with residuals diagnostics to check whether the residuals follow the white noise. If the residuals follow the white noise, our model is legitimate and we can then inspect the coefficients estimates.

5.3.1 CSO dataset

The most fitted and legitimate model is ARIMAX(0,1,1) (see Subsection B.1.2 for step by step). The model is then written as:

$$sales_t = \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(0,1,1)$$
$$sales'_t = \delta law'_t + u_t + \alpha u_{t-1}$$
$$u_t \sim N(0, \sigma^2), \quad ' \text{is first differencing}$$

The coefficients estimates are reported in Table 5.3.

The coefficient estimate of law dummy variable in Table 5.3 is not significant thus we are unable to reject the null hypotheses

$$H_0 : \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using ARIMAX with regressor dummy variable law on CSO dataset.

Table 5.3: Coefficient estimates of ARIMAX(0,1,1)on law, Regression with ARIMA errors approach based on CSO dataset

	Dependent variable:
α	-0.568^{***}
	(0.086)
δ	$\mathbf{-0.053}$
	(3.064)
Note:	*p<0.1; **p<0.05; ***p<0.01

α is the coefficient of the moving average term, δ is the coefficient of law variable term and also our main focus. We can see that both α is significant (indicated by the stars next to the estimate) and is different from zero but δ is not significant even at 10% level of significance.

Source: Author's own calculation

5.3.2 EET dataset

The most fitted model is ARIMAX(1,0,2)(2,1,2)[7] (see Subsection B.2.3 for step by step). The model is then written as:

$$
\begin{aligned}
sales_t &= \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(1,0,2)(2,1,2)[7] \\
sales'_t &= \delta law'_t + \phi\eta'_{t-1} \\
&\quad + \Phi_1(\eta'_{t-7} - \phi\eta'_{t-8}) + \Phi_2(\eta'_{t-14} - \phi\eta'_{t-15}) \\
&\quad + u_t + \alpha_1 u_{t-1} + \alpha u_{t-2} \\
&\quad + \Theta_1(u_{t-7} + \alpha_1 u_{t-8} + \alpha u_{t-9}) \\
&\quad + \Theta_2(u_{t-14} + \alpha_1 u_{t-15} + \alpha u_{t-16}) \\
u_t &\sim N(0,\sigma^2), \quad {}' \text{ is weekly differencing}
\end{aligned}
$$

The regression with such model parameters return the coefficients estimates reported in Table 5.4.

Table 5.4: Coefficients estimates of ARIMAX(1,0,2)(2,1,2)[7] on law, Regression with ARIMA errors approach based on EET dataset

		Coefficient estimate:			
ϕ	0.984^{***}	$alpha_1$	-0.780^{***}	α_2	-0.127^{**}
	(0.017)		(0.057)		(0.054)
Φ_1	-0.433^{**}	Φ_2	0.167^{***}	Θ_1	-0.489^{**}
	(0.217)		(0.057)		(0.214)
Θ_2	-0.511^{**}	δ	**4.977**		
	(0.213)		**(14.660)**		
Note:			*p<0.1; **p<0.05; ***p<0.01		

α_n are the coefficients of the moving average terms, ϕ is the coefficients of the autoregressive terms, Φ_n are the seasonal autoregressive term, Θ_n are the seasonal moving average, δ is the coefficient of law variable term and also our main focus. We can see that all but Θ_2, δ are significant (indicated by the stars next to the estimate).

Source: Author's own calculation

Estimate of δ, the coefficient for the dummy variable law reported Table 5.4

is not significant therefore we are unable to reject the null hypothesis:

$$H_0 : \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using ARIMAX with regressor dummy variable law on EET dataset.

5.4 ARIMAX with regressors dummy variable law and retail sales variable

Using ARIMAX, we will be testing the null hypothesis of the model

$$Hospitalitysales_t = \beta retailsales_t + \delta law_t + \eta_t, \quad \eta_t \sim ARIMA(p, d, q)$$
$$H_0 : \delta = 0$$

ARIMAX means that we will be regressing the time series with suspected models. We will then proceed with residuals diagnostics to check whether the residuals follow the white noise. If the residuals follow the white noise, our model is legitimate and we can then inspect the coefficients estimates.

5.4.1 CSO dataset

The most fitted model is ARIMAX(0,1,1) based on the AIC and residuals check (see Subsection B.1.4 for step by step). The model is then written as:

$$\begin{aligned} sales_t &= \delta law_t + \beta retail_t + \eta_t, \quad \eta_t \sim ARIMA(0, 1, 1) \\ sales'_t &= \delta law'_t + \beta retail'_t + u_t + \alpha u_{t-1} \\ u_t &\sim N(0, \sigma^2), \quad {}' \text{is first differencing} \end{aligned}$$

The coefficient estimates are reported in the Table 5.5

Coefficient estimate in Table 5.5 of *law* dummy variable is not significant, thus we are unable to reject the null hypothesis

$$H_0 : \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and

Table 5.5: Coefficients estimates of ARIMAX(0,1,1) on law and retail based on CSO dataset

	Dependent variable:
α	-0.591^{***}
	(0.097)
β	0.616^{***}
	(0.180)
δ	$\mathbf{-1.152}$
	(2.786)

Note: *p<0.1; **p<0.05; ***p<0.01

α is the coefficient of the moving average term, β is the coefficient of retail sales, δ is the coefficient of law variable term and also our main focus. We can see that both α, β are significant (indicated by the stars next to the estimate) and are different from zero but δ is not significant even at 10% level of significance.

Source: Author's own calculation

bars' sales significantly according to the testing using ARIMAX with regressor law dummy variable and retail sales variable on CSO dataset.

5.4.2 EET dataset

The most fitted model is ARIMAX(1,0,1)(2,1,1)[7] based on the AIC and residuals diagnostics (see Subsection B.2.3 for step by step). The model is then written as:

$$
\begin{aligned}
sales_t &= \delta law_t + \beta retail_t + \eta_t, \quad \eta_t \sim ARIMA(1,0,1)(2,1,1)[7] \\
sales'_t &= \delta law'_t + \beta retail'_t + \phi\eta'_{t-1} \\
&\quad + \Phi_1(\eta'_{t-7} - \phi\eta'_{t-8}) + \Phi_2(\eta'_{t-14} - \phi\eta'_{t-15}) \\
&\quad + u_t + \alpha u_{t-1} + \Theta(u_{t-7} + \alpha u_{t-8}) \\
u_t &\sim N(0,\sigma^2), \quad {}' \text{ is seasonal differencing}
\end{aligned}
$$

The coefficients estimates are reported in the Table 5.6.

Table 5.6: Coefficients estimates of ARIMAX(1,0,1)(2,1,1)[7] on law and retail based on EET dataset

	Coefficient estimate:				
ϕ	0.976***	α	−0.876***	Φ_1	0.070
	(0.028)		(0.048)		(0.064)
Φ_2	0.254***	Θ	−1.000***	β	0.004
	(0.065)		(0.052)		(0.003)
δ	−1.169				
	(11.272)				

Note: *p<0.1; **p<0.05; ***p<0.01

α is the coefficients of the moving average term, ϕ is the coefficient of the autoregressive term, Θ is the coefficient of the seasonal moving average term, Φ_n ate the coefficient of the seasonal term, β is the coefficient of retail sales, δ is the coefficient of law variable term and also our main focus. We can see that ϕ, α, Φ_2, Θ are significant at the level 5% (indicated by the stars next to the estimate), δ, β, Φ_1 are not significant even at 10% level of significance.

Source: Author's own calculation

Coefficient estimate in Table 5.6 of *law* dummy variable is not significant, thus we are unable to reject the null hypothesis

$$H_0 : \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the testing using ARIMAX with regressor law dummy variable and retail sales variable on CSO dataset.

5.5 Summary of the results

The Table 5.7 summarizes all the testing and results we have conducted. As we can see, all results state, that we are not able to reject the null hypotheses. The null hypotheses, in general, meant, that the smoking ban effect is not significant on the sales of the restaurants and bars. Because the null hypothesis was not rejected in all cases, therefore we can say that the smoking ban did not affect the sales of restaurants and bars significantly.

Table 5.7: Summary of the hypotheses tests

	CSO	EET
Forecasting	H_0 not rejected	H_0 not rejected
ITSA approach ~law	H_0 not rejected	H_0 not rejected
ARIMAX ~law	H_0 not rejected	H_0 not rejected
ARIMAX ~law, retail	H_0 not rejected	H_0 not rejected

Source: Author's own calculation

Chapter 6

Discussion and further research

6.1 Models and data source evaluation

For this thesis, I chose the ARIMA forecasting method and regression with ARIMA errors for the analysis and using methodology from various sources (Box *et al.* 2016; Bisgaard & Kulahci 2011; McDowall *et al.* 1980; Hyndman & Athanasopoulos 2019), which were initially out of the scope of this thesis. There are also other models, that might be used for the time series data analysis such as models with transfer function or with Fourier terms (Box *et al.* 2016; Hyndman & Athanasopoulos 2019; Bisgaard & Kulahci 2011). Such models are however far beyond the scope of this bachelor's thesis. Additionally, after observing the initial models of ARIMA forecasting and regression with ARIMA errors, we see that they are fairly accurate and for the thesis of this scope sufficient enough.

6.1.1 CSO models and data

The models developed with the provided data by CSO were reliable and legitimate as the residuals from the models follow white noise. With the models developed, I concluded that the smoking ban does not have a significant effect on hospitality sales at 5% significance level.

However, there are some drawbacks of this data source. Because I could not filter out only the NACE, that was affected thus we needed to use a proxy time series. Additionally CSO dataset only allows us to make a policy analysis on the aggregate level this the author can not make analyze the effect of regional policies. The values are often presented as an index and not in absolute values.

Another drawback of these data is that they need time to be sampled and thus data extraction is not immediate.

6.1.2 EET models and data

The models that we have developed from the treated data provided by the EET system were reliable and legitimate as the residuals follow the white noise. Based on the conclusion of the models, the author concluded that the smoking ban does not have a significant effect on hospitality sales at 5% significance level.

However, as mentioned before, the raw data provided from the system are not cleaned and has many errors e.g. negative sales in millions for a region in one single day is definitely a measurement error or systematic one, that might have occurred during the sampling of the data. the author treated these data as missing and interpolated the values based on the region. That way the author minimizes the effect of missing data and the author do not affect the aggregate level too much. But the fact that such dataset is flawed with so many measurement errors is suspicious and might require a more thorough investigation. Was it really a measurement error or a sampling error? Are these errors further investigated by the Financial Administration etc?

6.2 Smoking ban

Based on hypothesis testing utilizing both datasets, the author came to the conclusion that the smoking ban did not have a significant effect on sales numbers in the hospitality industry in the Czech Republic on the aggregate level. Thus economically, the smoking ban is not detrimental for the Czech Republic.

For further research on the topic of consequences of the smoking ban in the Czech Republic, research on whether there is increased alcohol consumption and increase of drunken driver on the roads might be needed in order to have the evaluation as comprehensive as possible.

6.3 VAT Fiscalization

The EET system in the current state is not bringing much value as data collecting tool. The data provided are with measurement errors. The Financial

Administration is not able to distribute the data frequently and at the present, it is charging a substantial fee for providing even the uncleaned data.

On the other hand, if sampled correctly, cleaned and without measurement errors, the EET system offers very high granularity data such as daily or even hourly. This provides much more flexibility as these data can be aggregated to lower granularity such as weekly or monthly, the same granularity as CSO data, thus providing an almost equivalent to them. The values are in absolute terms thus can help in quantifying the research. The data extraction, in theory, can be immediate as these data are sampled continuously. The additional feature that these data can give is the region classification. That allows for region-specific analysis, which is not achievable with CSO data.

However, with such features, the skills required to be able to conduct proper data analysis are very high. The data analyst needs to have in-depth knowledge about time series and panel data analysis to not misinterpret the data. The question is, whether such skilled individuals can be found on the municipality, regional level so that the advantage of the EET system can be fully utilized.

Therefore, although the EET system might be a powerful tool to collect the data, there are many questions, that need to be addressed, such as dirtiness of the data, the real utility of the data for analysis, etc. Before that, it is hard to state that the EET system provides helpful and unique data for analysis.

For further research for the topic EET system as a data collecting tool there might be some policies that might be worth to look into such as Hazard prohibition in Chomutov, Holiday closing policy for retailers with more than 200 square meters.

Chapter 7

Conclusion

Based on data analysis and modeling, the key finding of the thesis is that the smoking ban does not have a significant effect on the sales of restaurants and bars. Our objective was to inspect the effect of the smoking ban on the sales of the hospitality industry. To examine it as comprehensively as possible, we used daily time series from EET dataset and monthly time series from CSO dataset for the analysis. and we used ARIMA forecasting and regression with ARIMA errors (ARIMAX) models. The contribution of this thesis in the literature is our analysis of the smoking ban's effect from an economic point of view. This adds onto the ongoing discussion whether or not to cancel the smoking ban from an economic point of view. The ban does not have to be canceled as it has an insignificant effect on businesses' revenues.

The thesis also finds that the EET dataset is not a more reliable data source than CSO and did not provide additional valuable information in the sales analysis. The conclusion drawn from the EET dataset is identical to CSO dataset, thus we can say that both datasets cross-confirmed each other's findings. However, the daily time series is flawed with many measurement errors in observations such as millions of negative sales in one day in one region. Therefore in this analysis, the monthly time series from CSO dataset seems to be more reliable than daily time series from the EET. To my best knowledge, this is the first work to utilize the VAT fiscalization data for a sales analysis like this. The contribution of this thesis is that it opens up another possible discussion about the broader utility of the EET system in the Czech Republic.

For further research on the topic of the smoking ban, additional sales analysis using more advanced methods such as transfer function or Fourier terms might be considered.

For further research on the topic of the EET datasets utility, assuming that the measurement errors will be fixed, the EET system might become a powerful tool for data extraction for policy analysis as it boasts superior granularity. However, the availability of these datasets is very uncertain, as the Financial Administration does not have the capacity to distribute such datasets regularly. In comparison to CSO, whose datasets are being shared systematically, standardized and for no fee at all online. Analysis of the progress in the availability and cleanness in the future might reveal more about EET utility.

Bibliography

Adams, S. & C. Cotti (2008): "Drunk driving after the passage of smoking bans in bars." *Journal of Public Economics* **92(5)**: pp. 1288–1305.

Balatková, T. (2013): *Analýza komunikace kampaní proti kouření.* Diplomová, Univerzita Karlova, Praha.

Bates, M. N., J. Fawcett, S. Dickson, R. Berezowski, & N. Garrett (2002): "Exposure of hospitality workers to environmental tobacco smoke." *Tobacco Control* **11(2)**: pp. 125–129.

Bisgaard, S. & M. Kulahci (2011): *Time series analysis and forecasting by example.* Hoboken, N.J.: Wiley, 1st ed. edition.

Bogdanovica, I., F. Godfrey, A. McNeill, & J. Britton (2011): "Smoking prevalence in the european union: a comparison of national and transnational prevalence survey methods and results." *Tobacco Control* **20(1)**: pp. e4–e4.

Box, G. E. P., G. M. Jenkins, G. C. Reinsel, & G. M. Ljung (2016): *Time series analysis.* Hoboken, New Jersey: Wiley, fifth edition edition.

Červenková, L. (2016): *Servírky a číšníci jako oběti pasivního kouření.* Bakalářská, Západočeská Univerzita v Plzni, Plzeň.

Chege, A., N. Kiragu, C. Lagat, & G. Muthoni (2015): "Effect of electronic fiscal devices on vat collection in tanzania: A case of tanzania revenue authority." *European Journal of Business and Management* **7(33)**: pp. 125–133.

Cornelsen, L., Y. McGowan, L. M. Currie-Murphy, & C. Normand (2014): "Systematic review and meta-analysis of the economic impact of smoking bans in restaurants and bars." *Addiction* **109(5)**: pp. 720–727.

Cornelsen, L. & C. Normand (2012): "Impact of the smoking ban on the volume of bar sales in ireland – evidence from time series analysis." *Health Economics* **21(5)**: pp. 551–561.

Czech Statistical Office (2019): "Czech statistical office." https://www.czso.cz.

Dearlove, J. V., S. A. Bialous, & S. A. Glantz (2002): "Tobacco industry manipulation of the hospitality industry to maintain smoking in public places." *Tobacco control* **11(2)**: pp. 94–104.

European Commission (2013): "Report on the implementation of the council recommendation of 30 november 2009 on smoke-free environments (2009/c 296/02)." .

European Council (2009): "Council recommendation of 30 november 2009 on smoke-free environments." pp. 4–15.

Finanční správa (2016): "Kdo, co a odkdy?" https://www.etrzby.cz/cs/kdo-co-a-kdy.

Fiscal Solutions d.o.o. (2019): "Fiscal solutions." http://www.fiscal-solutions.net.

Glantz, S. A. & A. Charlesworth (1999): "Tourism and hotel revenues before and after passage of smoke-free restaurant ordinances." *Jama* **281(20)**: pp. 1911–1918.

Glantz, S. A. & L. R. Smith (1994): "The effect of ordinances requiring smoke-free restaurants on restaurant sales." *American Journal of Public Health* **84(7)**: pp. 1081–1085.

Glantz, S. A. & L. R. Smith (1997): "The effect of ordinances requiring smoke-free restaurants and bars on revenues: a follow-up." *American Journal of Public Health* **87(10)**: pp. 1687–1693.

Goláň, L. (2007): "Vliv kouření na morfologii a funkci kardiovaskulárního aparátu." *Interní medicína pro praxi* **9(9)**: pp. 386–388.

Huang, P. (2004): "Impact of a smoking ban on restaurant and bar revenues — el paso, texas, 2002." *Morbidity and Mortality Weekly Report* **53(7)**: p. 150.

Hyndman, R. J. & G. Athanasopoulos (2019): *Forecasting principle and practice*. United States of America: Otexts, 2nd edition.

Katolik, A. *et al.* (2014): "Impact of fiscalization at the height of calculated value added tax." *Interdisciplinary Management Research* **10**: pp. 545–556.

Kim, D. & B. K. Yörük (2015): "The impact of restaurant smoking bans on dining out expenditures: Evidence from panel data." *Journal of Urban Economics* **88**: pp. 38–49.

Koh, H. K., L. X. Joossens, & G. N. Connolly (2007): "Making smoking history worldwide." *New England Journal of Medicine* **356(15)**: pp. 1496–1498.

Koksal, A. & M. K. Wohlgenant (2016): "How do smoking bans in restaurants affect restaurant and at-home alcohol consumption?" *Empirical Economics* **50(4)**: pp. 1193–1213.

McDowall, D., R. McCleary, E. E. Meidinger, & J. Richard A. Hay (1980): *Interrupted time series analysis*. United States of America: Sage Publications, Inc., 1st ed. edition.

Ministerstvo Financí České republicky (2019): "Fiskální přínosy eet po dvou letech fungování." https://www.mfcr.cz/cs/aktualne/aktuality/2019/fiskalni-prinosy-eet-po-dvou-letech-fung-34419.

Oberg, M., M. Jaakkola, A. Pruss-Usun, A. Peruga, A. Woodward *et al.* (2010): "Global estimate of the burden of disease from second-hand smoke." *World Health Organization* .

Parlament České republiky (2005): "Zákon č. 215/2005 sb."

Parlament České republiky (2007): "Vládní návrh na vydání zákona o stabilizaci veřejných rozpočtů."

Parlament České republiky (2017): "Zákon č. 65/2017 sb."

Reichmann, G. & M. Sommersguter-Reichmann (2012): "The austrian tobacco act in practice–analysing the effectiveness of partial smoking bans in austrian restaurants and bars." *Health Policy* **104(3)**: pp. 304–311.

Šimović, H. & M. Deskar-Škrbić (2015): "Efficiency of vat in croatia." In "European Entrepreneurship Forum-Year IX:" Efficiency in the Private and the Public Sector"," .

Stárek, V. (2014): "Stanovisko ahr čr ke kouření - zákonu č. 379/2005 sb." https://www.ahrcr.cz/novinky/stanovisko-ahr-cr-ke-koureni-zakonu-c-3792005-sb/.

Stolzenberg, L. & S. J. D'Alessio (2007): "Is nonsmoking dangerous to the health of restaurants? the effect of california's indoor smoking ban on restaurant revenues." *Sage Journals* pp. 75–92.

Sušický, M. (2018): "Data z eet." https://www.datazeet.cz/.

World Health Organization and Tobacco Free Initiative (2007): *Protection from exposure to second-hand tobacco smoke: policy recommendations*. World Health Organization.

Zollinger, T. W., R. M. S. Jr, A. D. Overgaard, S. J. Jay, A. M. Holloway, & S. F. Cummings (2004): "Estimating the economic impact of secondhand smoke on the health of a community." *American Journal of Health Promotion* **18(3)**: pp. 232–238.

Ústavní soud České Republiky (2018): "Nález č.8/2018sb." https://www.zakonyprolidi.cz/cs/2005-215.

Škop, M. (2019): "Efekt eet a kontrolních hlášení na výběr daní za roky 2016 a 2017." https://i.iinfo.cz/files/podnikatel/54/analyza-metodiky-prinosu-eet-a-kontrolnich-hlaseni-1.pdf.

Appendix A

Tables and graphs

Table A.2: EET observation with negative retail sales

	DEN	CELKOVA_TRZBA	KRAJ
1	2016-12-01	-589	liberecký
2	2016-12-08	$-341,138$	Praha
3	2016-12-09	$-3,020$	Vysočina
4	2016-12-14	-25	jihomoravský
5	2016-12-15	-240	jihomoravský
6	2016-12-22	$-3,092$	královéhradecký
7	2016-12-24	-678	Praha
8	2016-12-27	$-3,730$	Vysočina
9	2017-01-05	$-397,144$	karlovarský
10	2017-01-06	-3	jihomoravský
11	2017-01-09	$-5,437$	plzeňský
12	2017-01-10	-330	středočeský
13	2017-01-10	$-19,831$	jihomoravský
14	2017-01-13	-965	zlínský
15	2017-01-17	$-836,836$	jihomoravský
16	2017-01-17	$-133,544,355$	olomoucký
17	2017-01-17	-1	Praha
18	2017-01-18	$-6,429$	plzeňský
19	2017-01-19	-469	moravskoslezský
20	2017-01-20	-1	Vysočina
21	2017-01-21	$-1,509$	Praha
22	2017-01-23	-398	Praha

23	2017-01-24	-40	plzeňský
24	2017-01-25	$-462,738$	Praha
25	2017-01-26	-1	Praha
26	2017-01-30	$-3,912,706,117$	zlínský
27	2017-01-30	$-663,355,388$	moravskoslezský
28	2017-01-31	$-3,868$	jihomoravský
29	2017-02-02	$-3,709,824,353$	zlínský
30	2017-02-02	$-3,410,362$	královéhradecký
31	2017-02-03	-426	karlovarský
32	2017-02-04	-2	pardubický
33	2017-02-05	$-5,714$	karlovarský
34	2017-02-06	-500	karlovarský
35	2017-02-06	-70	jihomoravský
36	2017-02-06	-675	zlínský
37	2017-02-09	-534	pardubický
38	2017-02-11	-192	jihomoravský
39	2017-02-12	-33	jihomoravský
40	2017-02-13	$-2,672,351,555$	královéhradecký
41	2017-02-13	$-97,149,665$	královéhradecký
42	2017-02-13	$-1,240,157,157$	moravskoslezský
43	2017-02-14	$-72,316,847$	jihomoravský
44	2017-02-15	-162	jihomoravský
45	2017-02-16	-819	Vysočina
46	2017-02-18	$-149,938,382$	Praha
47	2017-02-19	-23	jihočeský
48	2017-02-19	-38	zlínský
49	2017-02-21	$-15,776$	olomoucký
50	2017-02-23	$-46,992$	zlínský
51	2017-02-26	$-1,729$	plzeňský
52	2017-02-26	$-4,231,482$	moravskoslezský
53	2017-02-27	$-210,074,469$	jihočeský
54	2017-02-27	$-213,746,281$	moravskoslezský
55	2017-02-28	$-1,717,778,782$	královéhradecký
56	2017-02-28	$-12,164$	zlínský
57	2017-03-06	$-118,857,994$	Praha
58	2017-03-14	$-8,660,689,493$	liberecký
59	2017-03-17	$-31,596,999$	moravskoslezský

60	2017-03-23	$-7,557$	nelze určit
61	2017-03-24	$-2,881,393$	jihomoravský
62	2017-03-25	$-100,000$	nelze určit
63	2017-03-31	$-17,787,136$	Vysočina
64	2017-04-01	$-125,454,864$	karlovarský
65	2017-04-02	$-26,514$	jihočeský
66	2017-04-03	$-24,675,928$	karlovarský
67	2017-04-03	$-1,262,085,752$	ústecký
68	2017-04-06	$-532,802$	Vysočina
69	2017-04-12	$-308,864,079$	středočeský
70	2017-04-15	$-9,255,604,681$	moravskoslezský
71	2017-04-16	$-1,360,294$	plzeňský
72	2017-04-17	$-8,140$	olomoucký
73	2017-04-18	$-6,771,275$	moravskoslezský
74	2017-04-19	$-75,957,951$	moravskoslezský
75	2017-04-20	$-211,706,399$	Praha
76	2017-04-22	$-437,255,861$	moravskoslezský
77	2017-04-28	$-35,882,561$	jihočeský
78	2017-04-28	$-3,937,669,427$	olomoucký
79	2017-05-01	$-1,667,681$	liberecký
80	2017-05-08	$-788,399,829$	středočeský
81	2017-05-10	$-3,717,647,812$	plzeňský
82	2017-05-11	$-705,092,659$	Praha
83	2017-05-14	$-22,091,306$	středočeský
84	2017-05-16	$-432,800,534$	středočeský
85	2017-05-19	$-18,939,362$	královéhradecký
86	2017-05-21	$-340,450$	plzeňský
87	2017-05-31	$-356,399$	jihočeský
88	2017-06-03	$-2,611,098$	pardubický
89	2017-06-06	$-233,680$	plzeňský
90	2017-06-07	$-3,065,192,386$	Vysočina
91	2017-06-07	$-4,913,586,382$	jihomoravský
92	2017-06-08	$-1,943,884,801$	Vysočina
93	2017-06-15	$-36,514,515$	Praha
94	2017-06-15	$-130,688,127$	karlovarský
95	2017-06-15	$-4,465,176,361$	plzeňský
96	2017-06-15	$-4,425,561,203$	ústecký

97	2017-06-15	$-117,695,004$	moravskoslezský
98	2017-06-16	$-545,376,165$	královéhradecký
99	2017-06-29	$-54,280$	ústecký
100	2017-06-29	$-821,164$	jihomoravský
101	2017-07-03	$-31,720,163$	plzeňský
102	2017-07-07	$-195,926,465$	jihočeský
103	2017-07-07	$-538,180$	ústecký
104	2017-07-08	$-191,791,315$	jihočeský
105	2017-07-08	$-1,493,973$	Vysočina
106	2017-07-08	$-3,022,445,347$	plzeňský
107	2017-07-11	$-829,596$	plzeňský
108	2017-07-15	$-11,817,219$	královéhradecký
109	2017-07-18	$-726,644$	Vysočina
110	2017-07-20	$-1,748,463$	pardubický
111	2017-07-26	$-11,736,560$	jihočeský
112	2017-07-31	$-369,813,645$	Praha
113	2017-08-02	$-1,176,326,641$	zlínský
114	2017-08-14	$-71,928,127$	středočeský
115	2017-08-14	$-3,522,273,576$	ústecký
116	2017-08-20	$-18,492$	Vysočina
117	2017-08-26	$-524,567$	pardubický
118	2017-09-05	$-2,562,155$	liberecký
119	2017-09-07	$-1,703,055$	pardubický
120	2017-09-28	$-7,479,977$	zlínský
121	2017-10-02	$-122,041,649$	středočeský
122	2017-10-02	$-268,150,851$	jihomoravský
123	2017-10-07	$-191,353,285$	jihočeský
124	2017-10-09	$-54,651$	jihočeský
125	2017-10-12	$-14,794,739$	pardubický
126	2017-10-19	$-278,762,261$	liberecký
127	2017-10-29	$-57,206,106$	jihomoravský
128	2017-10-30	$-163,629,469$	středočeský
129	2017-10-31	$-251,662,083$	středočeský
130	2017-11-03	$-85,686,867$	jihomoravský
131	2017-11-05	$-354,952$	jihomoravský
132	2017-11-08	$-19,859,246$	Vysočina
133	2017-11-13	$-31,974,001$	Praha

134	2017-11-21	$-101,340,827$	středočeský
135	2017-11-22	$-2,090,440,477$	Vysočina

Table A.1: EET observations with negative hospitality sales

	DEN	CELKOVA_TRZBA	KRAJ	Law
1	2016-12-03	−1, 520, 049.000	Vysočina	0
2	2016-12-05	−76, 426, 529.000	královéhradecký	0
3	2016-12-06	−5, 792, 897.000	plzeňský	0
4	2016-12-10	−52, 598, 835.000	královéhradecký	0
5	2016-12-14	−64, 976, 400.000	liberecký	0
6	2016-12-15	−2, 818, 677.000	zlínský	0
7	2016-12-21	−75, 493, 040.000	Praha	0
8	2016-12-28	−7, 528, 988.000	moravskoslezský	0
9	2017-01-06	−71, 231, 812.000	ústecký	0
10	2017-01-07	−422, 587.200	zlínský	0
11	2017-01-09	−289, 111, 089.000	moravskoslezský	0
12	2017-01-16	−66, 106, 955.000	jihomoravský	0
13	2017-01-18	−4, 446, 659.000	ústecký	0
14	2017-01-27	−2, 168, 160.000	zlínský	0
15	2017-02-13	−18, 387, 546.000	Vysočina	0
16	2017-03-03	−82, 082, 476.000	karlovarský	0
17	2017-03-03	−27, 832, 767.000	jihomoravský	0
18	2017-03-06	−24, 844, 053.000	ústecký	0
19	2017-03-20	−24, 437, 233.000	pardubický	0
20	2017-04-02	−2, 308, 230	karlovarský	0
21	2017-05-03	−2, 415, 246.000	ústecký	0
22	2017-05-04	−31, 886, 230.000	jihomoravský	0
23	2017-05-15	−2, 232, 805.000	pardubický	0
24	2017-05-16	−51, 839, 596.000	Praha	0
25	2017-05-21	−91, 815, 930.000	ústecký	0
26	2017-06-07	−247, 892.800	ústecký	1
27	2017-06-12	−7, 692, 054.000	středočeský	1
28	2017-06-15	−2, 446, 814.000	olomoucký	1
29	2017-06-16	−93, 127, 368.000	Vysočina	1
30	2017-06-16	−350, 588.200	olomoucký	1
31	2017-06-22	−1, 066, 609.000	pardubický	1
32	2017-07-04	−84, 744, 035.000	středočeský	1
33	2017-07-06	−7, 039, 249.000	jihomoravský	1
34	2017-07-12	−440, 217.200	liberecký	1
35	2017-08-02	−28, 468, 523.000	pardubický	1
36	2017-08-08	−5, 219, 133.000	Vysočina	1
37	2017-08-10	−10, 366, 540.000	královéhradecký	1
38	2017-08-17	−5, 510, 184.000	zlínský	1
39	2017-09-23	−188, 829, 617.000	plzeňský	1
40	2017-09-26	−3, 492, 076.000	zlínský	1
41	2017-10-05	−1, 246, 731.000	plzeňský	1
42	2017-10-18	−14, 187, 283.000	ústecký	1
43	2017-11-06	−1, 495, 955.000	pardubický	1
44	2017-11-23	−19, 103, 496.000	liberecký	1

Appendix B

Modelling process and selection

B.1 CSO dataset modeling

B.1.1 Forecasting SI

We decomposed the hospitality industry sales data from January 2013 to January 2019 to inspect the seasonality and trend in the time series.

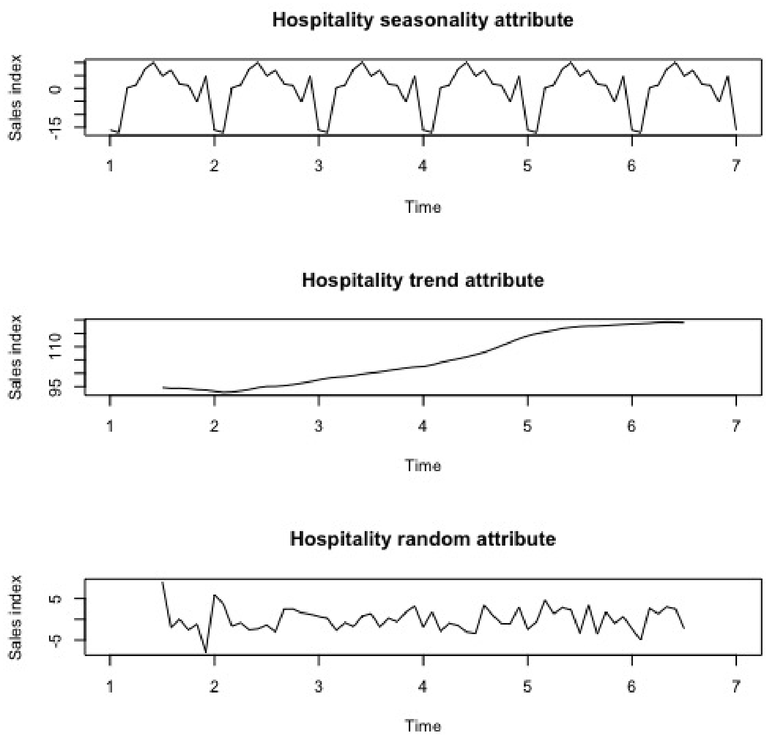
In Figure B.1 the seasonality is depicted clearly as a year to year seasonality, trend attribute shows an increasing tendency. For further work, deseasonalizing the time series by subtracting the seasonal component and detrending by taking the appropriate differences is necessary to obtain the stationary time series, which is a base requirement for any time series analysis (see Subsection 4.1.1).

Deseasonalized (seasonally adjusted) and differenced time series seems to be stationary by visual inspection of the plot (see Figure B.2 as with visual inspection the time series seems to have zero mean and constant variance through the time after one differencing already and second differencing does not change the distribution by much. Augmented Dickey-Fuller test statistics p-value is lower than 0.05 and KPSS test statistics p-value is higher than 0.05 for differenced seasonally adjusted hospitality sales index time series thus assuring its stationary. We can continue to the identification of the parameters. Because we have deseasonalized the time series, parameters for seasonality models are by default 0.

We used the first tool, auto.arima() function in the R to identify the parameters. We received a suspected model with parameters ARIMA (2,1,0).

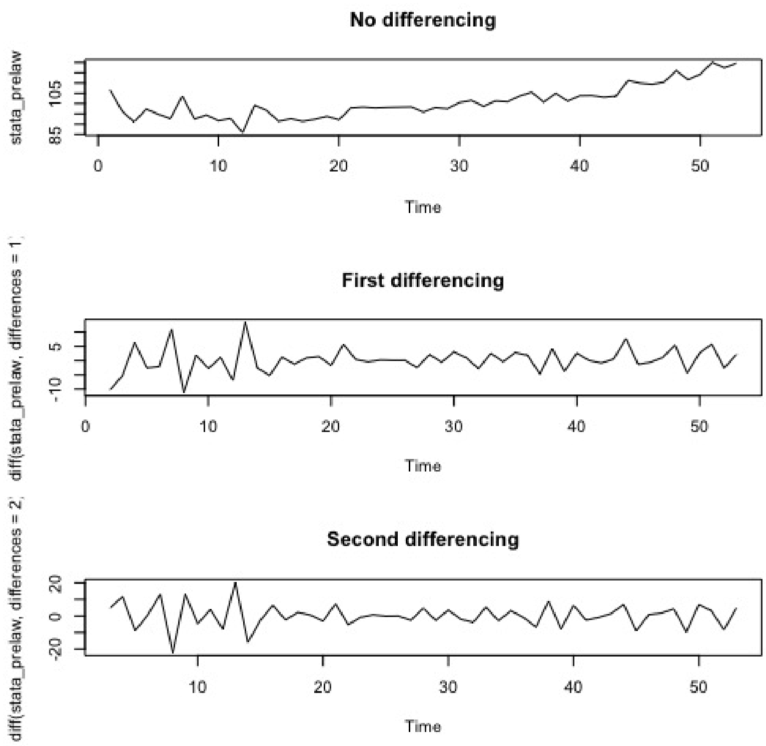
Then we run our custom search function, constructed based on recommendation (Hyndman & Athanasopoulos 2019) with $\{max.p = 2, max.q = 2d = 1, max.P = 0, max.Q = 0, D = 0, \}$ and we discarded models with Ljung-Box

Figure B.1: Hospitality decomposition



Source: Author's own calculation

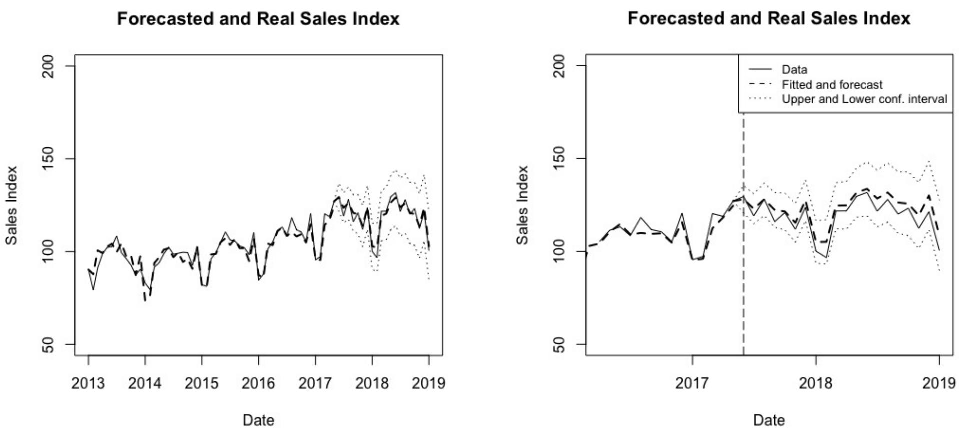
Figure B.2: Hospitality sales times series



Source: Author's own calculation

test statistics for residuals below 0.05, which would means that their residuals does not follow white noises. Because our filtered models have a low number of parameters, we use BIC as the main criterion to not overestimate the model. The model with the lowest BIC is ARIMA(0,1,1).

Figure B.3: Forecast and real value index based on CSO data



(a) Forecasted and Real Sales Indexes (b) Post-law period

Source: Author's own calculation

As seen in Figure B.3a, the forecast is relatively accurate and in the Figure B.3b one can see that the real values are inside the confidence interval values. Thus we did not observe a significant shift in the hospitality sales index after the introduction of the smoking ban in restaurants and bars. Therefore, we are unable to reject null hypotheses

H_0: Significant part of the recorded values does not break out of the forecasted confidence interval values. Therefore, we can say that the smoking ban did not affect hospitality sales significantly.

B.1.2 ITSA approach with dummy variable law

By using ITSA approach, we utilize the parameters that were found in Subsection B.1.1 to be the most fitted model on the pre-intervention period. With this model identification, we proceed with the regression on the whole time series with the law as dummy variable regressor. Such regression with given parameters generate coefficients estimates reported in Table B.2.

Table B.1: Coefficients estimates of ITSA approach

	Dependent variable:
ma1	-0.568^{***}
	(0.086)
law	-0.053
	(3.064)
Note:	*p<0.1; **p<0.05; ***p<0.01

Source: Author's own calculation

The coefficient estimate of *law* dummy variable variable in *law* in Table B.1 is insignificant thus we are unable to reject the second null hypothesis $H_0 : \delta_t = 0$. Therefore, we can say that the smoking ban did not affect the hospitality sales significantly.

B.1.3 ARIMAX with dummy variable law

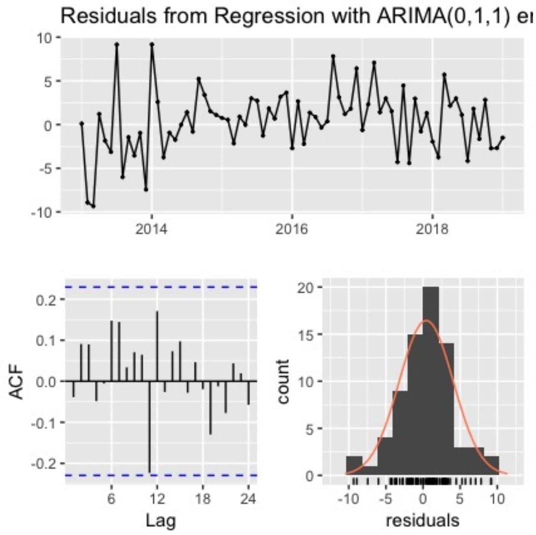
We need to inspect the stationarity of the dummy variable *law*. Augmented Dickey-Fuller test statistics p-value below 0.05 and KPSS test statistics above 0.05 is reported for differenced time series of dummy variable *law* thus it is stationary. For the raw data, the Augmented Dickey-Fuller test statistics p-value is above 0.05 and KPSS test statistics p-value is below 0.05 thus it is not stationary so we are not in danger of over differencing.

Auto.arima() returned suspected model ARIMAX(0,1,1)(1,0,0)[12].

Then we run our custom search function with $\{max.p = 2, max.q = 2d = 1, max.P = 0, max.Q = 0, D = 0, \}$ and we discarded models with Ljung-Box test statistics for residuals below 0.05, which would means that their residuals does not follow white noises. Because our filtered models have a low number of parameters, we use BIC as the main criterion to not overestimate the model. The model with the lowest BIC is ARIMA(0,1,1).

Residuals diagnosis reveals that the ACF plot of the residuals of the model is contained in the confidence interval. This means that the model does not have unnecessary larger confidence interval (Hyndman & Athanasopoulos 2019).

Figure B.4: Residuals of ARIMAX(0,1,1) with dummy variable law



Source: Author's own calculation

The regression with the model returned coefficients estimates reported in Table B.2.

Table B.2: Coefficient estimates of ARIMAX(0,1,1) with law dummy variable

	Dependent variable:
ma1	-0.568^{***}
	(0.086)
law	-0.053
	(3.064)
Note:	*p<0.1; **p<0.05; ***p<0.01

Source: Author's own calculation

The coefficient estimate of *law* dummy variable in Table B.2 is not significant thus we are unable to reject the null hypotheses

$$H_0 : \delta_t = 0$$

Therefore, we can say that the smoking ban did not affect the hospitality sales significantly.

B.1.4 ARIMAX with regressors dummy variable law and retail sales variable

Stationary check for variable retail sales is essential for further work.

Augmented Dickey-Fuller test statistics is below 0.05 and KPSS test statistic is above 0.05 for differenced time series. We have seasonally adjusted the retail time series as we have done with hospitality time series thus we will not be using seasonal terms of the ARIMA models.

Auto.arima() function returned a suspected model ARIMAX(1,0,1)(0,0,1)[12]. This is a fairly strange result because we expected at least one order of differencing as our base time series are not stationary. However, there is one exception and that is when non-stationary variables are co-integrated (Hyndman & Athanasopoulos 2019). However, because we have already deseasonalized the seasonal term occurring indicates possible over-identification.

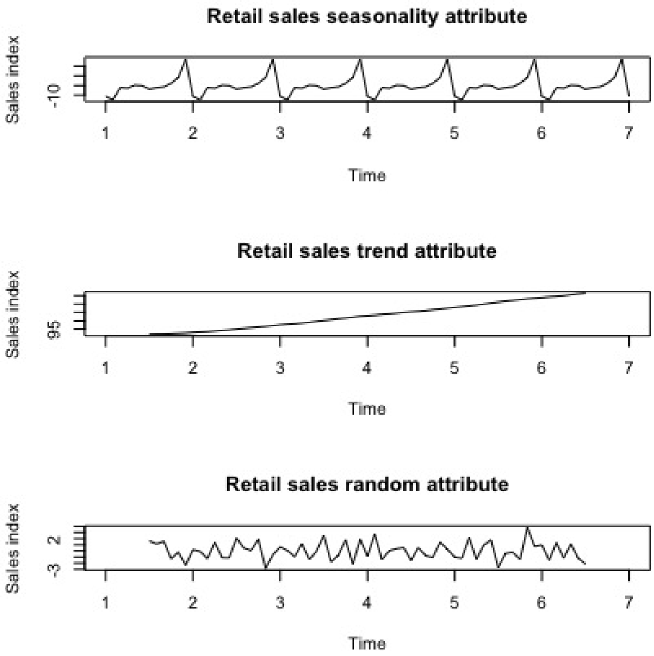
Then we run our custom search function with $\{max.p = 2, max.q = 2, d = 1, max.P = 0, max.Q = 0, D = 0, \}$ and we discarded models with Ljung-Box test statistics for residuals below 0.05, which would means that their residuals does not follow white noises. Because our filtered models have a low number of parameters, we use BIC as the main criterion to not overestimate the model. The model with the lowest BIC is ARIMA(0,1,1).

Residuals diagnosis reveals that the ACF plot of the residuals of the model is contained in the confidence interval. This means that the model does not have unnecessary larger confidence interval (Hyndman & Athanasopoulos 2019). The regression with the model returned coefficients estimates reported in Table B.3.

Model ARIMAX(0,1,1) has the coefficients reported in Table B.3

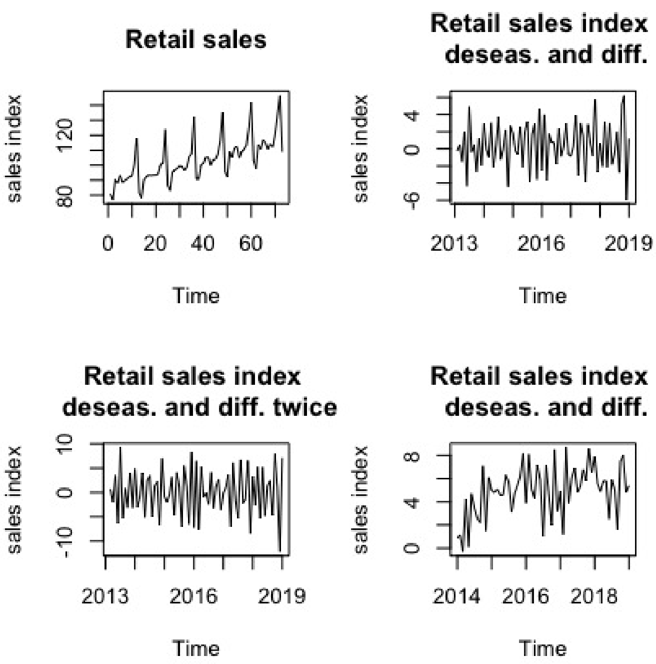
Coefficient estimate Table B.3 of *law* dummy variable is not significant, thus we are unable to reject the null hypothesis H_0: $\delta_t = 0$. The coefficient estimate of retail sales, on the other hand, is significant and has a relatively

Figure B.5: Visual inspection of retail SI time series of CSO



(a) Retail sales decomposition

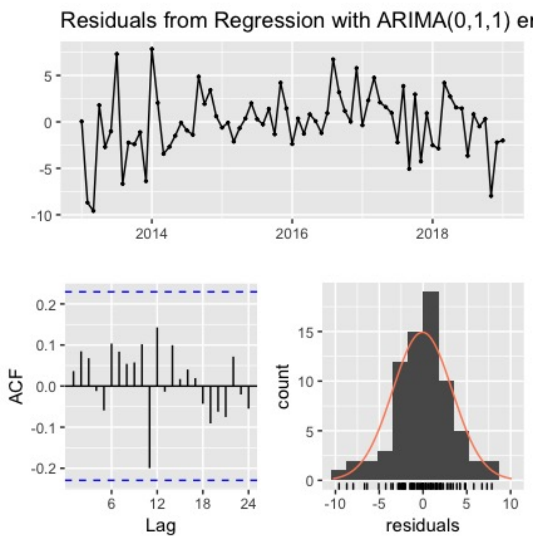
]]



(b) Retail sales variously differenced and lagged time series

Source: Author's own calculation

Figure B.6: Residuals of ARIMAX(0,1,1) with retail SI and dummy variable law



Source: Author's own calculation

Table B.3: Coefficients estimates of ARIMAX(0,1,1) with retail SI and law dummy variable

	Dependent variable:
ma1	-0.591^{***}
	(0.097)
seasadjretail	0.616^{***}
	(0.180)
law	-1.152
	(2.786)
Note:	*p<0.1; **p<0.05; ***p<0.01

Source: Author's own calculation

high influence. Therefore, we can say that the smoking ban did not affect hospitality sales significantly.

B.1.5 Conclusion from CSO dataset

We conducted hypothesis testing:

1. Given common sense, that the smoking ban did not affect the sales of restaurants, forecasted and recorded values should not be significantly different therefore H_0: Significant part of the recorded values should not break out of the forecasted confidence interval values.

2. Based on the initial thought, the model is formulated as:

$$Hospitalitysales_t = \delta_t law_t + \eta_t. \quad \eta_t \sim ARIMA(p, d, q)$$

   with $Hospitalitysales_t$ is sales index of hospitality, law_t is a dummy variable for law implementation, eta_t is the errors that follows ARIMA process. Then the null hypothesis to test is: H_0: $\delta_t = 0$

3. Based on the initial thought, the model is formulated as:

$$Hospitalitysales_t = \beta_t retailsales_t + \delta_t law_t + \eta_t \quad \eta_t \sim ARIMA(p, d, q)$$

   with $Hospitalitysales_t$ is sales index of hospitality, $retailsales_t$ is sales index of retail sales, law_t is a dummy variable for law implementation, eta_t is the errors that follows ARIMA process. Then the null hypothesis to test is: H_0: $\delta_t = 0$

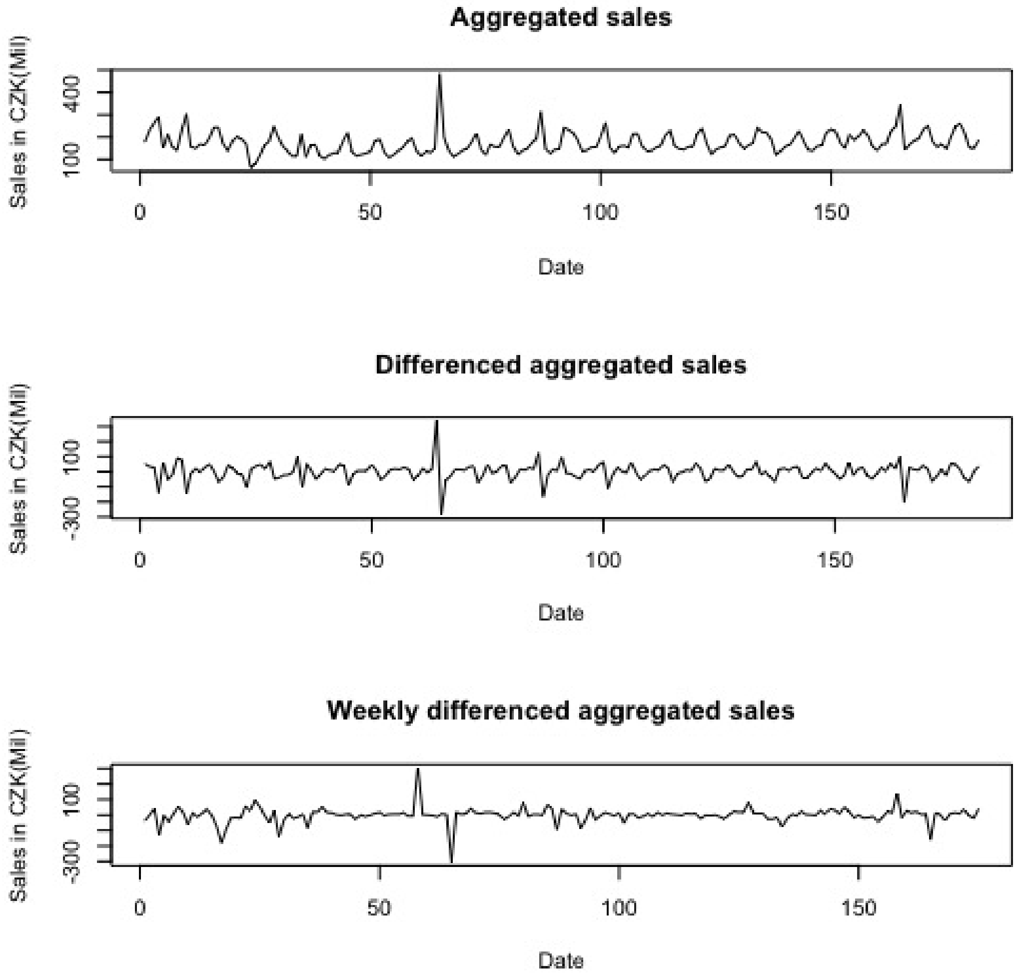
and we could not reject any one of those null hypothesis. We can therefore say that the smoking ban did not have any significant effect on sales in hospitality based on the data provided by CSO.

B.2 EET dataset modeling

B.2.1 Forecasting sales

Inspecting the data in the pre-law period, based on the visual inspection of Figure B.7 it seems that the EET time series are stationary after first difference and after weekly difference already. Augmented Dickey-Fuller test statistics for

Figure B.7: EET variously differenced time series



Differenced and weekly differenced time series seems to be stationary as their mean seems to be 0 and the variance seems to be stable all time series has p-value below 0.05 and KPSS test has test statistics p-value for all time series above 0.05 thus the time series are confirmed to be stationary. We can proceed to determine the most fitted model for forecasting.

Auto.arima() returned suspected model ARIMA(4,1,3). Then we run our custom search function, constructed based on recommendation (Hyndman & Athanasopoulos 2019) with $\{\{max.p = 2, max.q = 2, d = 0, max.P = 1, max.Q = 1, D = 1\}, \{max.p = 2, max.q = 2, d = 1, max.P = 1, max.Q = 1, D = 0\}$ and we discarded models with Ljung-Box test statistics for residuals below 0.05, which would means that their residuals does not follow white noises. Because our filtered have also seasonal parameters, we use AIC as we do not want to under identify the model. The most fitted model identified is (1,0,1)(0,1,1)[7].

Visual inspection of the plot of the forecast and fitted model in Figure B.8 and the comparison with the recorded value seems like a relatively well-fitted model and forecast. Inspecting the post-law period, we see that most of the real values lie inside of the confidence interval, apart from 2 values, which lies outside of the confidence interval and 2 values, which lies directly on the confidence interval limits. Given that the model forecasted 177 observations, from which 2 were out of the confidence interval, this is relatively good for the forecast. Because the number of forecast lying outside of the confidence interval is 4 which is approximately 2% of the forecasted values, we can not reject the null hypothesis

H_0: Significant part of the recorded values does not break out of the forecasted confidence interval values.

B.2.2 ITSA approach with dummy variable law

Because in Subsection B.2.1 we identified the most fitted model on the pre-law period, we can use those parameters to make a Regression model with ARIMA errors. The estimates of the coefficients of such model are:

The Table B.4 report, that the coefficient estimate of the *law* dummy variable is not significant therefore we can not reject the null hypothesis

$$H_0 : \delta = 0$$

B.2.3 ARIMAX with dummy variable law

First we need to examine stationarity of our regressor, the dummy variable **law**. We know that our regressand time series is stationary when differenced

Figure B.8: Forecast based on EET data visualized

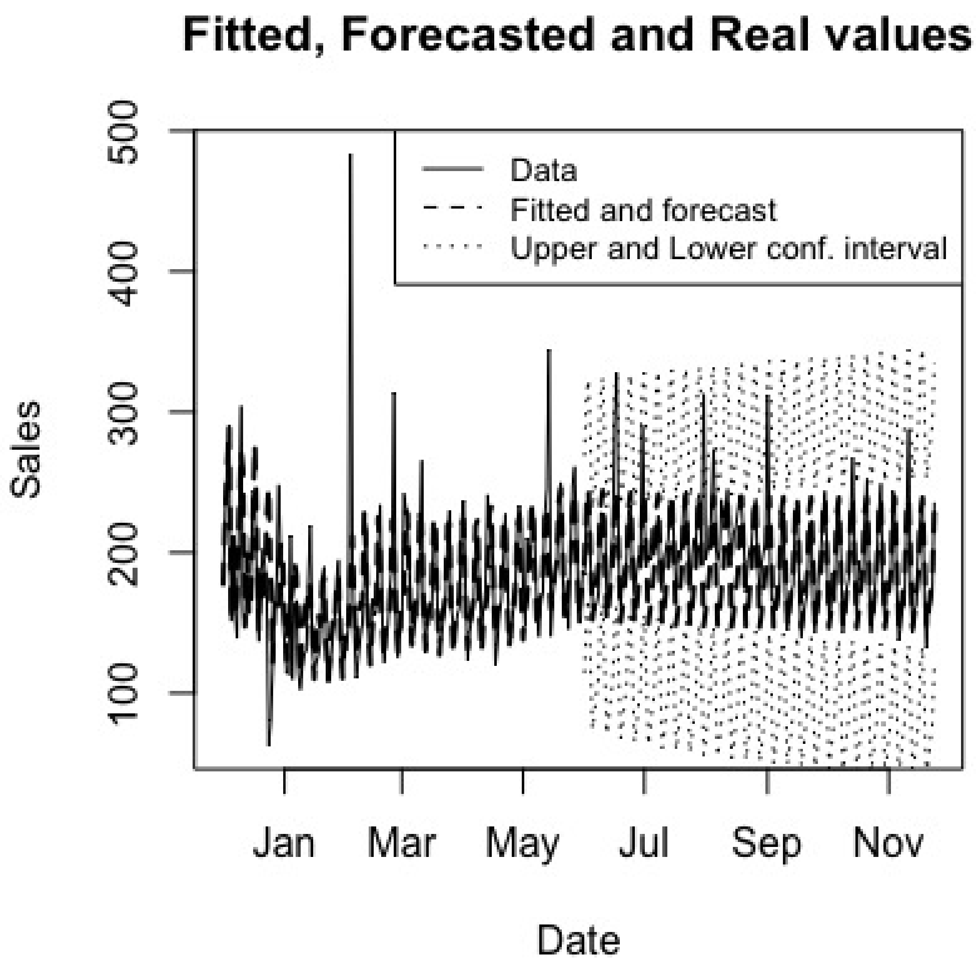


Figure B.9: Forecast and fitted model

Source: Author's own calculation

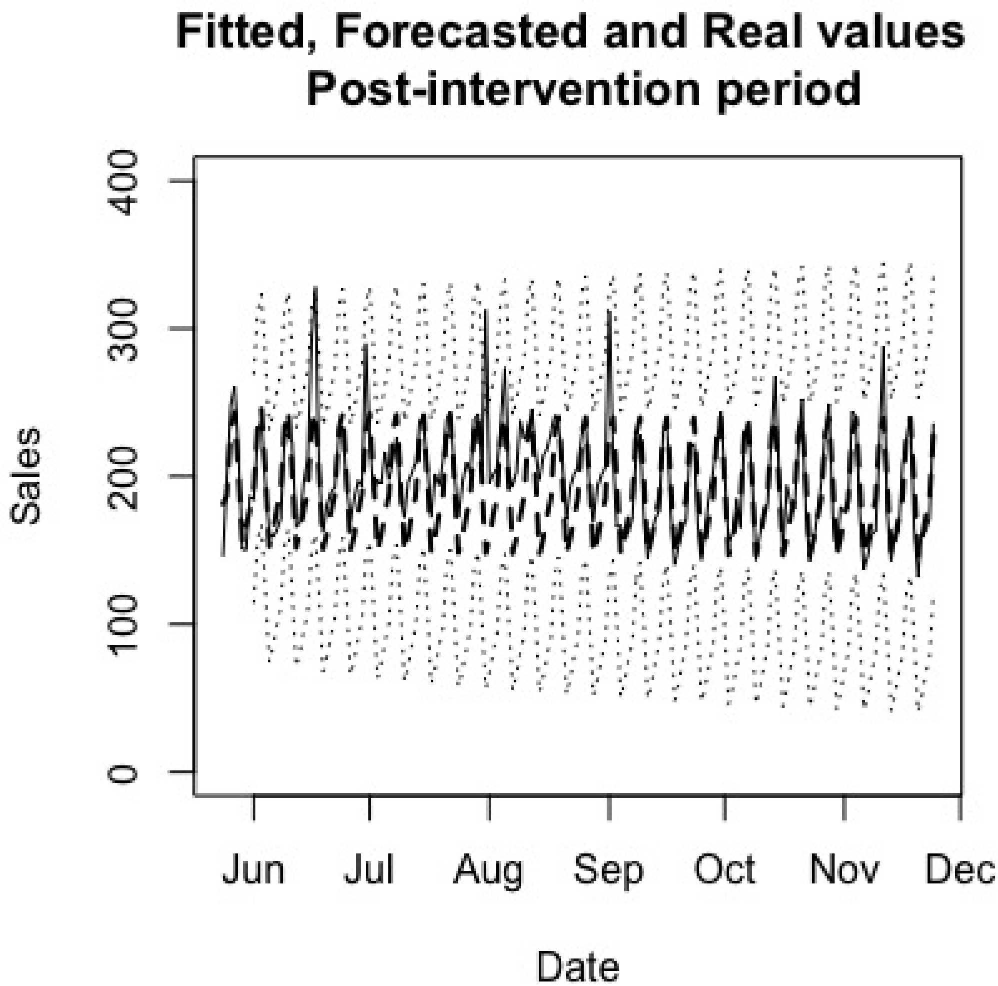


Figure B.10: Post-law period

Source: Author's own calculation

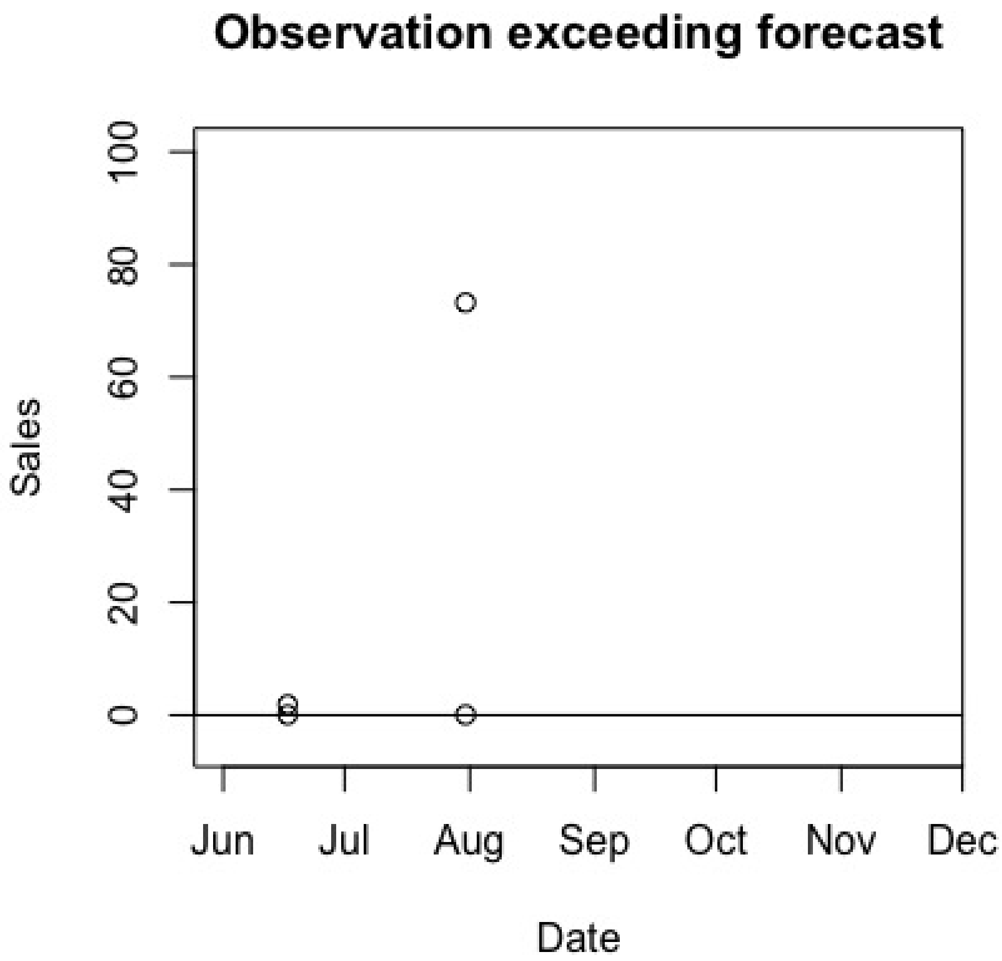


Figure B.11: Observations that exceeded confidence interval

Source: Author's own calculation

Table B.4: Coefficients estimates of ITSA approach

	Dependent variable:
ar1	0.983^{***}
	(0.018)
ma1	-0.881^{***}
	(0.031)
sma1	-1.000^{***}
	(0.031)
law	5.124
	(14.622)
Note:	*p<0.1; **p<0.05; ***p<0.01

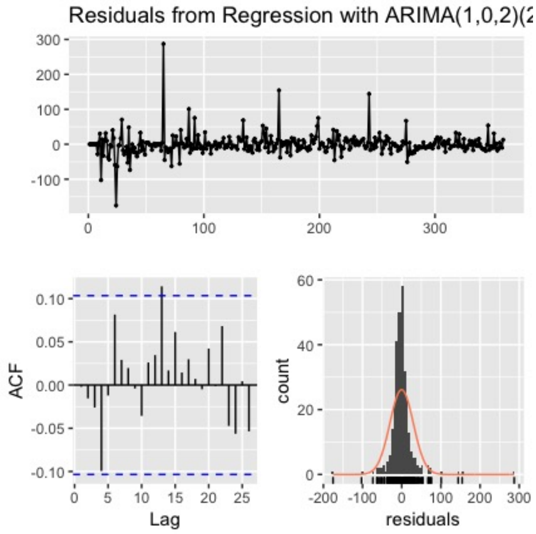
Source: Author's own calculation

and weekly differenced. It is desired to maintain the relationship between regressand and regressors (Hyndman & Athanasopoulos 2019). Therefore, we inspect the stationarity of differenced and weekly differenced regressors time series. Augmented Dickey-Fuller test statistics for all time series p-value is lower than 0.05 and KPSS test statistics for all time series p-value is higher than 0.05 thus assuring us that the time series is stationary after the mentioned transformations

Running auto.arima() returned a suspected model: ARIMA(4,0,0), We see that the auto.arima() function returned us an anchor 4. Our custom search model will run with parameters $\{max.p = 4, max.q = 4, d = 1, max.P = 2, max.Q = 2, D = 0, \}$, $\{max.p = 4, max.q = 4, d = 0, max.P = 2, max.Q = 2, D = 1, \}$ and we discarded models with Ljung-Box test statistics for residuals below 0.05, which would means that their residuals does not follow white noises. Because our filtered models have also seasonal parameter, we use AIC as main criterion to not under identify the model. The most fitted model is ARIMAX(1,0,2)(2,1,2)[7].

Residuals diagnosis reveals that the ACF plot of the residuals of the model is not contained in the confidence interval. This means that the model have

Figure B.12: Residuals of ARIMAX(1,0,2)(2,1,2) with dummy variable law



Source: Author's own calculation

unnecessary larger confidence interval (Hyndman & Athanasopoulos 2019) however none of the models inspected does not have the spike. Thus we conclude that this model is the most fitted. The regression with the model returned coefficients estimates reported in Table B.2.

Because the residuals diagnostics confirmed that residuals follow white noise, our model is legitimate. Coefficient estimates for dummy variable *law* in Table B.5 is not significant therefore we are unable to reject null hypothesis:

$$H_0 : \delta_t = 0$$

Hence, we can say that the smoking ban did not affect hospitality sales significantly.

B.2.4 ARIMAX with regressors dummy variable law and retail sales variable

We shortened the datasets for regression to start from 01.03.2017, when the EET was mandatory for the retail segment because, in the period before, the participation was only volunteer for retail segment thus the data are skewed as it can be seen in the plot.

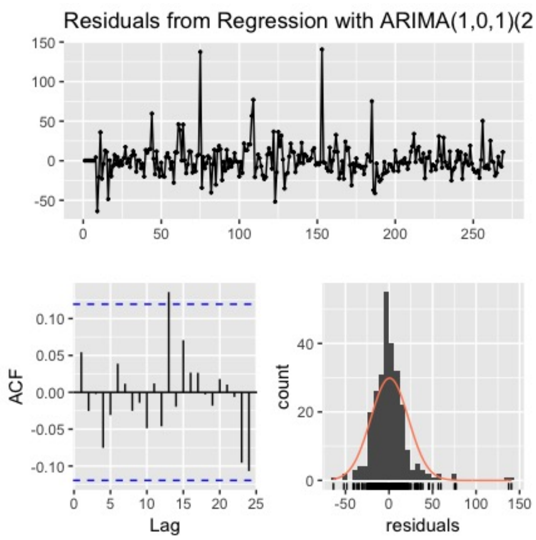
Table B.5: Coefficients estimates of ARIMAX(1,0,2)(2,1,2)[7] with law dummy variable

	Dependent variable:
ar1	0.984***
	(0.017)
ma1	-0.780^{***}
	(0.057)
ma2	-0.127^{**}
	(0.054)
sar1	-0.433^{**}
	(0.217)
sar2	0.167***
	(0.057)
sma1	-0.489^{**}
	(0.214)
sma2	-0.511^{**}
	(0.213)
law	4.977
	(14.660)
Note:	*p<0.1; **p<0.05; ***p<0.01

Source: Author's own calculation

Auto.arima() returned suspected model: ARIMA(4,0,0). Then we run our custom search function, constructed based on recommendation (Hyndman & Athanasopoulos 2019) with $\{max.p = 5, max.q = 5, d = 1, max.P = 2, max.Q = 2, D = 0,\}$, $\{max.p = 4, max.q = 4, d = 0, max.P = 2, max.Q = 2, D = 1,\}$ and we discarded models with Ljung-Box test statistics for residuals below 0.05, which would means that their residuals does not follow white noises. Because our filtered have also seasonal parameters, we use AIC as we do not want to under identify the model. The most fitted model identified is (1,0,1)(2,1,1)[7].

Figure B.13: Residuals of ARIMAX(1,0,1)(2,1,1) with retail sales variable and dummy variable law



Source: Author's own calculation

Residuals diagnosis reveals that the ACF plot of the residuals of the model is not contained in the confidence interval. This means that the model have unnecessary larger confidence interval (Hyndman & Athanasopoulos 2019) however none of the models inspected does not have the spike. Thus we conclude that this model is the most fitted.

Coefficient estimate in Table B.6 of *law* dummy variable is not significant, thus we are unable to reject the null hypothesis

$$H_0 : \delta = 0$$

Therefore, we can say that the smoking ban did not affect restaurants' and bars' sales significantly according to the results from the ARIMAX Regression with ARIMA errors approach regressing on law and retail sales testing based on EET dataset.

Table B.6: Coefficients estimate of ARIMAX(1,0,1)(2,1,1)[7] with retail sales and law dummy variable

	Dependent variable:
ar1	0.976***
	(0.028)
ma1	−0.876***
	(0.048)
sar1	0.070
	(0.064)
sar2	0.254***
	(0.065)
sma1	−1.000***
	(0.052)
retail	0.004
	(0.003)
law	−1.169
	(11.272)
Note:	*p<0.1; **p<0.05; ***p<0.01

Source: Author's own calculation

B.2.5 Conclusion from EET data

We conducted two hypothesis testing:

1. Given common sense, that the smoking ban did not affect the sales of restaurants, forecasted and recorded values should not be significantly

different therefore H_0: Significant part of the recorded values should not break out of the forecasted confidence interval values.

2. The relation can be written in a model as:

$$Hospitalitysales_t = \delta_t law_t + \eta_t. \quad \eta_t \sim ARIMA(p, d, q)$$

with $Hospitalitysales_t$ is sales index of hospitality, law_t is a dummy variable for law implementation, eta_t is the errors that follows ARIMA process. Then the null hypothesis to test is: H_0: $\delta_t = 0$

3. Based on the initial thought, the model is formulated as:

$$Hospitalitysales_t = \beta_t retailsales_t + \delta_t law_t + \eta_t \quad \eta_t \sim ARIMA(p, d, q)$$

with $Hospitalitysales_t$ is sales in hospitality, $retailsales_t$ is sales index in retail sales, law_t is a dummy variable for law implementation, eta_t is the errors that follows ARIMA process. Then the null hypothesis to test is: H_0: $\delta_t = 0$

and we could not reject any of the two null hypothesis. Therefore, we can conclude that based on data provided by EET system, smoking ban does not have significant effect on sales in restaurant and bars.

Appendix C

Content of Enclosed Website

The empirical data and source code for R are available on the author GitHubs git `https://github.com/tomasluu/smokingban-in-Czech`.

- OBU01B_B_M.xlsx base time series extracted from the CSO site for the retail industry
- OBU02B_B_M.xlsx base time series extracted from the CSO site for the hospitality industry
- datazeet.zip is a folder containing the dataset, Mr. Sušický analysis and NACE names
- Statak.csv is time series created from the time series extracted from the CSO site for modeling use
- EET.csv is time series created from the dataset from Mr. Sušický dataset
- Model EET.R is an R script of the command used to analyze the time series of EET.csv
- Model Stata.R is an R script of the command used to analyze the time series of Statak.csv