VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

Fakulta financí a účetnictví

Katedra veřejných financí

Studijní obor: Zdanění a daňová politika

Je daňová politika u cigaret v kontextu zemí EU účinná?

Autor diplomové práce: Bc. Markéta Lichterová
Vedoucí bakalářské práce: Ing. Jana Tepperová, Ph.D.
Rok obhajoby 2016

Čestné prohlášení:

Prohlašuji, že jsem diplomovou práci na téma „Je daňová politika u cigaret v kontextu zemí EU účinná?“ vypracovala samostatně a veškerou použitou literaturu a další prameny jsem řádně označila a uvedla v přiloženém seznamu.

V Praze dne 26. 08. 2016

podpis

Poděkování

Touto cestou bych ráda vyjádřila své poděkování Ing. Janě Tepperové, Ph.D. za čas, věnovaný konzultacím nad touto závěrečnou prací, které mi pomohly při zpracování tématu mé diplomové práce.

ABSTRAKT

Tato práce se zabývá otázkou, zda výše zdanění ovlivňuje spotřebu cigaret. Pro účely analýzy byly nejprve prozkoumány faktory ovlivňující spotřebu cigaret, jako je socioekonomický status obyvatelstva nebo používané politiky státu omezující spotřebu cigaret. Analýza samotná je založena na sestavení ekonometrického modelu regresní analýzy na panelových datech. Panelová data v sobě obsahují informace za státy Evropské unie za dobu 2003 až 2015. Cílem diplomové práce je zjistit, zda daňová politika států v oblasti zdanění cigaret je v kontextu zemí EU účinná, tj. zda zdanění ovlivňuje spotřebu cigaret. Z provedené analýzy vyplývá, že proměnná zdanění cigaret je statisticky nevýznamná. Z toho vyplývá, že výše zdanění neovlivňuje spotřebu cigaret.

KLÍČOVÁ SLOVA:

zdanění cigaret, spotřeba cigaret, ekonometrická analýza, regresní analýza, panelová data, Evropská unie, spotřební daň, daň z přidané hodnoty

ABSTRACT

This thesis is based on the question, whether a taxation have an effect on consumption of cigarettes. At the first there was needed to find factors which affect the cigarettes consumption. These factors are socioeconomic status of the population or using regulatory politics for tobacco products. Analysis is based on estimation econometrics model of regression analysis on panel data set. The panel data set consist from information about member states of European union for years 2003 to 2015. The goal of the thesis is to find an answer whether the tax policy of member states in according to taxation of cigarettes is effective. It means whether taxation of cigarette have an impact on cigarettes consumption. The analysis show that the explanatory variable for taxation is statistically inefficient. That say that the level of taxation does not affect the cigarettes consumption.

KEY WORDS:

taxation of cigarettes, cigarettes consumption, econometrics analysis, regression, panel data, European union, excise duties, VAT

Obsah

Úvod

„Přestat kouřit je ta nejsnazší věc na světě – já to musím vědět, přestával jsem už asi tisíckrát." – Mark Twain (EUROPEAN COMMISSION, 2016a).

Spotřeba cigaret se liší stát od státu. Vlády jednotlivých států se snaží snižovat spotřebu cigaret vhodnými stimuly, kdy přijímají různá opatření, nicméně někdy s neurčitým výsledkem. Vlivy jednotlivých politik nejsou účinné bezprostředně po zavedení konkrétních opatření, účinnost použitých opatření ukáže jen čas a post analýzy.

Zhruba 5,8 trilionů cigaret se na světě spotřebovalo během roku 2014, spotřeba cigaret nadále roste. Zatímco ve vyspělých státech světa spotřeba cigaret klesá, například v Číně vzrostla. Čínský trh zkonzumuje více cigaret než jiné nízko až středně příjmové země dohromady. Roste také světová spotřeba jiných tabákových produktů. Od roku 2000 se celosvětová spotřeba doutníčků, které jsou menší než doutníky, ale větší než cigarety, více než zdvojnásobila, spotřeba baleného tabáku a tabáku do dýmek se až ztrojnásobila. Tento růst může být způsoben tím, že tyto druhy tabákových výrobků mají ve většině zemí nižší zdanění než cigarety, proto zákazníci ve snaze ušetřit přecházejí na levnější druh. (ERIKSEN et al. 2015)

Dle údajů dostupných v The Tabbacco Atlas (ERIKSEN et al. 2015), je vyšší spotřeba cigaret spojena s nižším socioekonomickým statusem, dokonce i s nižší či střední příjmovou třídou. Tyto poznatky by bylo vhodné využít v boji proti kouření. Na druhé straně je třeba si uvědomit, že proti zavedení skutečně účinných opatření stojí nejrůznější tabákové lobby a v neposlední řadě je třeba vzít v úvahu i významnou fiskální funkci daní z tabákových výrobků, kdy se jedná o naplnění státního rozpočtu ze zboží se stálou spotřebou.

Jedna z hlavních linií boje proti kouření je cílena na mladé lidi, kteří ještě nezačali kouřit. Existuje velmi mnoho důvodů, proč mladí lidé začínají kouřit. Chtějí si připadat vyspělí, cool, socializovat se, nebo se jen domnívají, že cigarety pomáhají čelit stresu a pomáhají vyřešit problémy s nadváhou (ERIKSEN et al. 2015). Je prokázáno, že mladí lidé jsou více citliví na cenu tabákových výrobků (MLČOCH a DOLEŽAL, 2014). Vyšší ceny je proto mohou odradit od pravidelné konzumace tabákových výrobků. Jedním z užitečných

nástrojů k prevenci je stanovení minimálního věku, od kterého si mohou mladí lidé koupit cigarety v obchodech, a jeho přísné dodržování. Například v České republice v roce 2011 pravidelně kouřilo cigarety 31 % dětí ve věku 13 – 15 let a 17 % dětí, kteří kouřili jiné tabákové výrobky (ERIKSEN et al. 2015). Téměř polovina těchto regulérních kuřáků si obvykle kupovala cigarety v obchodech.

Cigarety jsou jednou z mála legálních drog na světovém trhu. Negativně ovlivňují zdraví i život lidí, a to nejen kuřáků, ale i osob v jejich okolí. V posledních letech se brojí proti tomuto nešvaru a mnoho světových vlád přijímá různá opatření k omezení kouření. Ať už tato opatření mají chránit osoby od pasivního kouření (například kouření v restauracích, veřejných budovách apod.) nebo odradit kuřáky od kouření.

Tento celospolečenský problém mě přivedl na myšlenku zpracovat diplomovou práci na téma, zda výše zdanění ovlivňuje spotřebu cigaret. Za posledních dvanáct let se zvýšila minimální spotřební daň u cigaret na více než trojnásobek (CELNÍ SPRÁVA, 2016).

Cílem mé diplomové práce je zodpovědět otázku, zda je daňová politika u cigaret v kontextu zemí Evropské unie účinná. Budu se zabývat řešením problému, zda zvyšování zdanění cigaret příznivě ovlivňuje spotřebu cigaret nebo zda se jedná jen o skrytou fiskální funkci této daně. Z ekonomické teorie vím, že spotřeba cigaret je spíše cenově neelastická (MLČOCH a DOLEŽAL, 2014). Vzhledem k tomu by zvýšení ceny cigaret nemělo ovlivnit jejich spotřebu. Tento poznatek hovoří spíše proti ochranitelské teorii daňové politiky u cigaret. Řešení budu hledat prostřednictvím analýz. Jako dílčí úkol jsem si stanovila zjistit, jaké faktory a jakým způsobem ovlivňují spotřebu cigaret.

Zpočátku jsem uvažovala o provedení analýzy na časových řadách, ale pro potřeby korektní analýzy nebyly časové řady dostatečně dlouhé. Na základě těchto úvah jsem se rozhodla sestavit ekonometrický model regresní analýzy na panelových datech zemí Evropské unie. Panelová data mají tři rozměry a to rozměr času, vysvětlujících proměnných a počet pozorování. Jde vlastně o spojení časových řad s cross sectional daty.

Jako dílčí úkol jsem si stanovila zjistit, jaké faktory a jakým způsobem ovlivňují spotřebu cigaret. Mojí prvotní myšlenkou bylo zabývat se tímto problémem pouze na datech za Českou republiku, ale z důvodu mála pozorování, jež by způsobovalo neprůkaznost

statistik, jsem se rozhodla odstranit tento nedostatek použitím dat za Evropu. Přesněji použít data za Evropskou unii, protože země Evropské unie mají harmonizované daňové zákony. Tento stav zajišťují směrnice zavazující zdanit tabákové výrobky určitým způsobem a určují minimální sazby spotřební daně. Omezení analýzy na Evropskou unii neznamená použitelnost výsledků jen v tomto prostředí. Dle výsledků této analýzy lze postupovat i v jiných částech světa, a to z toho důvodu, že zhruba 21 % celosvětové spotřeby cigaret (data za rok 2013) připadá na Evropu (STATISTICA, c2015). Jde o region s druhou nejvyšší spotřebou cigaret na světě. Zhruba polovina celosvětové spotřeby cigaret je v regionu Západního Pacifiku a o třetí místo se s 11 % dělí Amerika s Jihovýchodní Asií (STATISTICA, c2015).

Pro potřeby diplomové práce jsem vybrala předpoklad, že spotřebu cigaret ovlivňuje především cena cigaret, vzdělání obyvatelstva, životní styl, míra blahobytu, věk kuřáků, zda se jedná o postkomunistickou zemi a zda určitá země přistoupila k omezení kouření v restauracích, barech a na veřejných místech. Může se jednat o úplný zákaz kouření, jinou formu regulace nebo o oddělené prostory pro kuřáky. Jmenované faktory budu dále specifikovat v následujících kapitolách mé diplomové práce.

Zajímavé bude porovnání regresních parametrů na konci analýzy. Pokud mají jednotlivé proměnné různé jednotky, ve kterých jsou měřené, regresní parametry nejsou srovnatelné. Srovnatelné odhady regresních parametrů lze vypočítat na základě beta koeficientů nebo normovaných veličin vysvětlujících proměnných. Nejedná se o matematickou úpravu, která by změnila výsledky analýzy, jde jen o transformaci dat tak, aby byly údaje srovnatelné.

Při vyhodnocování výsledků regresní analýzy je třeba mít na paměti, že prokázání statistické závislosti neznamená automaticky kauzalitu. Existence lineární závislosti mezi dvěma proměnnými může být způsobena korelací s proměnnou třetí. Pak se může stát, že dané proměnné jsou na sobě velmi závislé, přitom je však jejich změna způsobena změnou třetí, nepozorované, proměnné.

V první kapitole se zaměřím na vývoj spotřeby cigaret ve světě. Na základě vědeckých článků, zpráv Evropské unie a Světové zdravotnické organizace budu sledovat faktory ovlivňující spotřebu cigaret.

V druhé kapitole budu prezentovat ekonometrickou teorii týkající se regresní analýzy na panelových datech. Uvedu hlavní hypotézu, zda zdanění cigaret ovlivňuje jejich spotřebu. Tuto hypotézu budu testovat t testem o statistické významnosti regresních parametrů. Pokud bude regresní parametr týkající se zdanění různý od nuly, prokáže se hypotéza, že zdanění ovlivňuje spotřebu cigaret. Regresní analýza na panelových datech je poněkud odlišná od regresní analýzy na cross sectional datech. Je to způsobeno hlavně povahou panelových dat. Modely, které budu popisovat v druhé kapitole, využívají výhod panelových dat nebo odstraňují jejich nedostatky.

Třetí kapitola bude plně věnována samotné analýze. Analýza bude provedena v programu R, který má open-source licenci. Tento program je volně dostupný na stránkách www.r-project.org a umožňuje uživatelům přizpůsobit uživatelské prostředí svým potřebám. Většina packetů je napsána přímo uživateli. Pro účely práce s panelovými daty byl vytvořen packet plm, který podporuje veškeré modely používané pro panelová data a popsané v mém hlavním knižním zdroji Introductory Econometrics (Wooldridge, 2013).

Nedílnou součástí třetí kapitoly bude příprava dat tak, aby byla použitelná v programu R. To zahrnuje i transformaci kategoriálních proměnných pro použití jako dummy vysvětlující proměnné. Na připravených datech budu odhadovat jednotlivé modely a testovat splnění jednotlivých předpokladů modelů. Pokud zjistím nesplnění určitého předpokladu, musí dojít ke korekci modelu. Nejčastěji dochází k heteroskedasticitě náhodné složky, která se dá odstranit robustními odhady korelační matice, která se vypořádá s heteroskedasticitou. Po těchto nezbytných úkonech mohu přejít k porovnání jednotlivých modelů a vyhodnocení nejvhodnějšího modelu.

V závěru diplomové práce odpovím na otázku, zda jsem provedenou analýzou zjistila statisticky významný vliv zdanění na spotřebu cigaret v zemích Evropské unie. Provedená analýza dokáže odpovědět i na další otázku, a to, zda existuje vhodnější proměnná, kterou lze ovlivňovat politikou státu, a tím omezit škodlivé kouření. Posoudím vlivy jednotlivých vysvětlujících proměnných a porovnám je. Na základě porovnání odhadů regresních parametrů posoudím, jaká proměnná má největší vliv na spotřebu cigaret a jaká proměnná by byla nejvíce vhodná pro účinnou politiku omezující kouření.

1 Faktory ovlivňující spotřebu cigaret

Ve vyspělých státech sice spotřeba cigaret pomalu klesá, zatímco v jiných částech světa se jejich spotřeba za posledních několik let i znásobila (ERIKSEN et al. 2015). Zvýšení daní z cigaret může způsobit přesun spotřeby na tabákové výrobky s nižším zdaněním. Bylo dokázáno, že někteří kuřáci při zvýšení ceny cigaret přecházejí k cigaretám s vyšším obsahem nikotinu a dehtu (CZUBEK a JOHAL 2010). Tím vykouří méně cigaret, ovšem účinek na zdraví je mnohonásobně horší.

Spotřeba cigaret se liší ve všech částech světa. Ovlivňují ji socioekonomické aspekty i politiky států. Vlivy jednotlivých opatření státu mají různé výsledky, většinou se však projevují až v dlouhodobém horizontu.

V různých studiích (ERIKSEN et al. 2015, MLČOCH a DOLEŽAL, 2014 a další) je spojována vyšší spotřeba cigaret s nižším socioekonomickým statusem, nižším vzděláním a dokonce i nižším příjmem. Nejohroženější skupinou jsou děti a dospívající. Na ně působí velmi negativně reklama a jsou velice náchylní na vysokou cenu balíčku cigaret (MLČOCH a DOLEŽAL, 2014). Z tohoto důvodu se na tuto skupinu zaměřují doporučení Světové zdravotnické organizace, jejíž doporučení přijala Evropská komise k zajištění nekuřáckého prostředí.

V této kapitole na základě rozboru vědeckých článků a odborných publikací zjistím, jaké další proměnné ovlivňují spotřebu cigaret. Budu se zabývat i problematikou cenové elasticity poptávky po cigaretách a shrnu používané politiky omezující kouření v rámci Evropské unie.

1.1 Cenová elasticita

Cenová elasticita poptávky po cigaretách uvádí, jak se změní poptávané množství cigaret při změně jejich ceny a nezměněných ostatních faktorech. Problémem cenové elasticity poptávky po cigaretách se zabývá mnoho vědeckých článků (MLČOCH a DOLEŽAL, 2014), dle nich se pohybuje krátkodobá cenová elasticita okolo – 0,40 a dlouhodobá elasticita okolo – 0,44. Pokud tedy vzroste cena cigaret o 1 %, poptávané množství klesne o 0,4 % v krátkém období a o 0,44 % v dlouhém období (MLČOCH a DOLEŽAL, 2014). Cenová

elasticita se u různých skupin kuřáků liší. Například dospělí lidé mají o 16 % méně elastickou poptávku. Mladí dospělí mají oproti tomu poptávku elastičtější, a to o 13 %, adolescenti dokonce o 89 %, jejich cenová elasticita činí – 1,46 (MLČOCH a DOLEŽAL, 2014).

Zajímavá je i křížová elasticita, pokud spotřeba jednoho statku ovlivňuje spotřebu druhého statku. Takové statky mohou být komplementy či substituty. Doplňujícím statkem pro cigarety by mohl být alkohol. Dle studie S. Decker A. Swartz (2000) provedené na datech za USA je alkohol zároveň substitutem i komplementem cigaret. Při zvýšení ceny alkoholu se sníží jeho spotřeba a zároveň se sníží i spotřeba cigaret, zatímco vyšší ceny cigaret mají tendenci snížit spotřebu cigaret a zároveň zvýšit spotřebu alkoholických nápojů. Díky těmto poznatkům by politika států, i ta daňová, mohla být efektivnější. Pokud by se podařilo pomocí státní politiky snížit spotřebu alkoholu, mohlo by to vést i ke snížení spotřeby cigaret.

Zavedení vyššího zdanění cigaret může způsobit přesun spotřeby k levnějším druhům cigaret nebo přechod k vlastnímu balení cigaret z tabáku ke kouření, který mívá nižší zdanění. Z tohoto důvodu se může opatření týkající se cigaret zdát účinným, avšak ve skutečnosti tomu tak není. Některé studie dokonce zjistili (CZUBEK a JOHAL 2010), že kuřáci mohou reagovat na změnu zdanění cigaret přechodem na levnější druhy cigaret, které mají větší obsah dehtu a nikotinu, a tím je jejich vliv na zdraví jedince mnohem horší.

1.2 Strategie států Evropské unie v boji proti kouření

Evropská unie se zavázala přijmout zákony ohledně zajištění nekuřáckého prostředí na základě doporučení Světové zdravotnické organizace. V současné době přijalo 17 zemí z celkových 28 států komplexní soubor zákonů zajišťující nekuřácké prostředí (EUROPEAN COMMISSION, 2016b).

Jedním druhem restriktivní politiky je zvyšování spotřební daně z cigaret a jiných tabákových výrobků. Zvýšením spotřební daně se zvýší i daňový výběr z těchto výrobků pomocí DPH, která je vypočítaná z ceny výrobku včetně spotřební daně. Tím se zvyšuje cena, která má odradit spotřebitele od kouření. Průměrné daňové zatížení v Evropské unii

činí 77 % z ceny cigaret (vlastní výpočet na základě údajů z Excise duties tables za jednotlivé roky).

Cigarety jsou jedním z nejvíce uniformovaných produktů, které jsou lehce vyrobitelné s nízkými náklady. Cenu cigaret určuje tabákový trh. Bylo zjištěno, že 10% nárůst ceny cigaret způsobí propad spotřeby mezi 2 až 8 % (ERIKSEN et al. 2015). Zhruba polovina poklesu je způsobena omezením spotřeby u stávajících kuřáků a polovina poklesem počtu mladých začínajících kuřáků.

1.2.1 Používané politiky omezující kouření

V listopadu roku 2009 přijala Evropská unie doporučení ohledně zajištění nekuřáckého prostředí v členských státech (EUROPEAN COMMISSION, 2016a). Tato doporučení byla inspirována Rámcovou úmluvou WHO (World Health Organization) o kontrole tabáku. Doporučení vyzývá členské státy přijmout zákony k zajištění nekuřáckého prostředí nejdéle do roku 2012.

Dle zprávy Evropské Komise (2013) všechny členské státy přijaly protikuřácké zákony, které mají za úkol ochránit obyvatele od tabákového kouře ve vnitřních prostorech na pracovištích, ve veřejných dopravních prostředcích a veřejných prostorech jako jsou úřady, školy apod. Nejpřísnější opatření k zajištění nekuřáckého prostředí přijalo Irsko, Velká Británie, Bulharsko, Španělsko, Malta a Řecko (EUROPEAN COMMISSION, 2016b). Tyto zákony zcela zakazují kouření v uzavřených veřejných prostorech, na pracovištích a ve veřejných dopravních prostředcích.

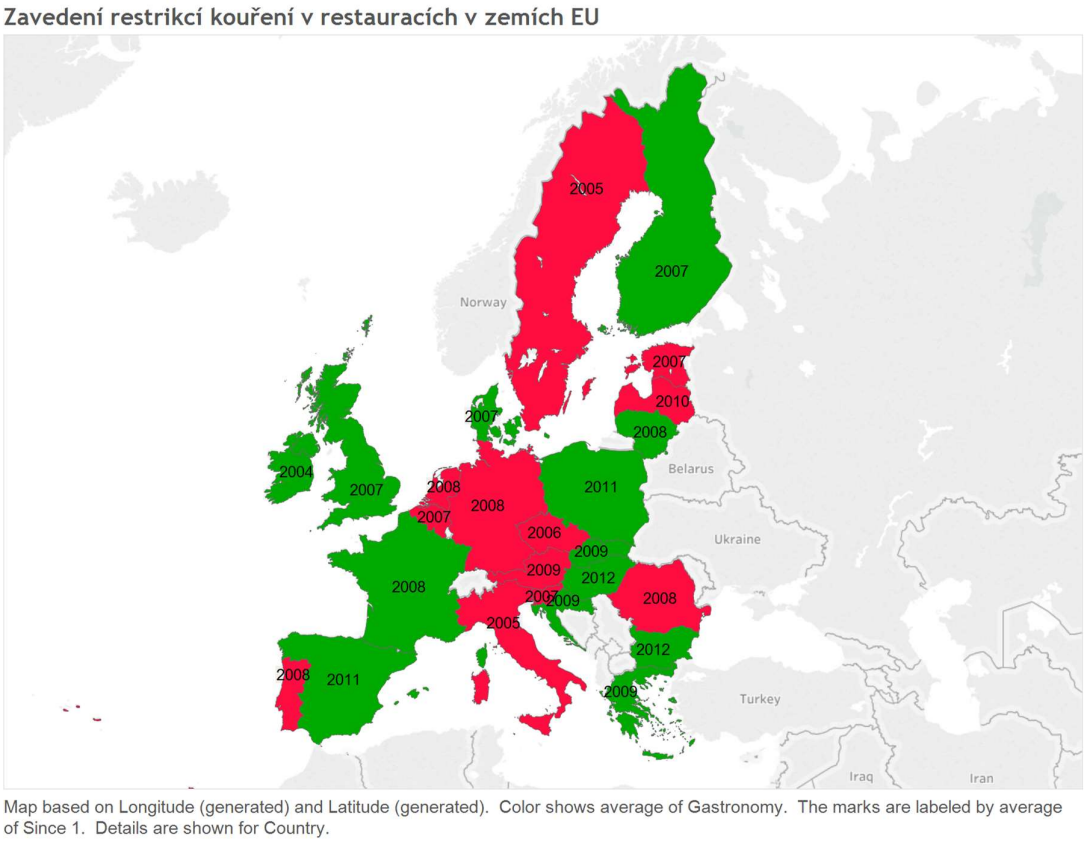
Druhy protikuřáckých opatření v Evropské unii:

- úplný zákaz kouření v restauracích, oddělené prostory pro kuřáky a nekuřáky nebo označení celé restaurace, baru za kuřácké,
- zákaz kouření ve školách – úplný zákaz, úplný zákaz ve vzdělávacích prostorech nižšího stupně a oddělené prostory pro kuřáky pro vyšší stupně vzdělávacích institucí,
- úplný zákaz kouření v dopravních prostředcích veřejné dopravy – některé státy jako Finsko, Litva a Dánsko povolují kouření ve speciálních prostorech pro kuřáky na dálkových osobních lodích,

- zákaz kouření v nemocnicích – úplný nebo v oddělených prostorech pro kuřáky,
- zákaz kouření v hotelových pokojích – úplný zákaz nebo označené kuřácké pokoje, některé státy zavedly úplný zákaz kouření v ložnicích (Kypr, Rakousko a Bulharsko),
- všechny členské státy mají nějaký druh zákazu kouření na pracovištích a veřejných prostorech,
- všechny členské státy aplikují určitý zákaz reklamy tabákových výrobků a varování na krabičkách cigaret (EUROPEAN COMMISSION, 2013).

Často diskutovaným tématem je zákaz kouření v restauracích a barech. Dle zprávy Evropské komise (2013) všechny členské státy zavedly určitou formu restrikcí kouření v restauracích a barech. Následující obrázek ukazuje roky[1], kdy jednotlivé členské země zavedly určitou formu restrikce kouření v restauracích a jiných zařízeních. Země jsou rozděleny do dvou skupin dle stupně zákazu kouření v restauracích a barech. Červeně jsou označeny země, kde je povoleno kouření v oddělených prostorech apod., zeleně jsou značeny země, kde je zaveden velmi striktní zákaz kouření v restauracích a barech.

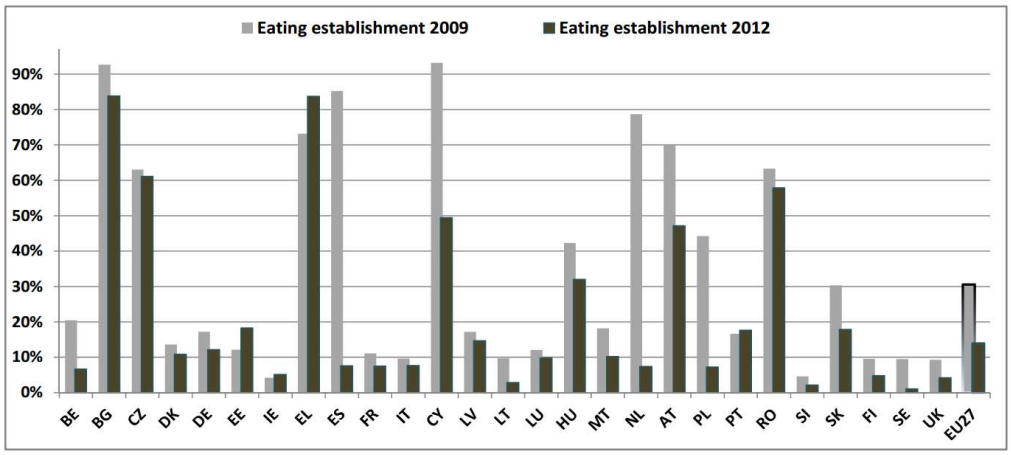
[1] Španělsko zavedlo zákaz kouření na všech pracovištích, barech a restauracích již 1. ledna 2006. Jelikož se tento zákon nedodržoval, zavedlo Španělsko nový zákon zakazující kouření uvnitř veřejně přístupných míst včetně restaurací, barů a kaváren od začátku roku 2011 (Zákaz kouření podle zemí, 2001-b). Proto jsem se rozhodla použít pro účely analýzy účinné zavedení zákona v roce 2011.



Obrázek 1: Zavedení restrikcí kouření v restauracích v zemích EU, vlastní zpracování, výstup z programu Tableau. 1 – plný zákaz kouření, 2 – kouřní povoleno v oddělených částech.

Jak je vidět na grafu (obrázek číslo 2), ve většině zemí Evropské unie se v rozmezí let 2009 až 2012 snížilo vystavení osob pasivnímu kouření v restauracích a barech, a to díky zavedení zákonů plně zakazujících nebo omezujících kouření v těchto zařízeních.

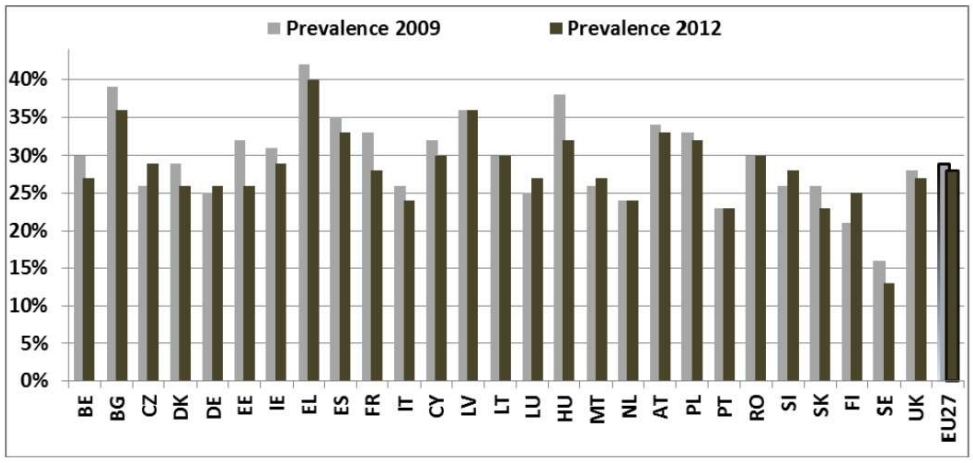
Figure 3b - Exposure to second hand tobacco smoke in EU-27 in 2009 and 2012



Obrázek 2: Vystavení pasivnímu kouření v restauračních zařízeních, zdroj: EUROPEAN COMMISSION, 2013. Report on the implementation of the Council Recommendation of 30 November 2009 on Smoke-free Environments

Během let 2009 až 2012, tedy po zavedení většiny zákazů kouření v restauracích, klesl počet kuřáků téměř ve všech státech Evropské unie (viz. Obrázek číslo 3). Tento pokles samozřejmě nemusí být způsoben jen zavedením zákazu kouření v restauracích, ale i jinými faktory. Přesto je tato souvislost natolik zajímavá, že jsem se rozhodla pomocí statistické analýzy testovat vliv zavedení zákazu kouření v restauracích na spotřebu cigaret.

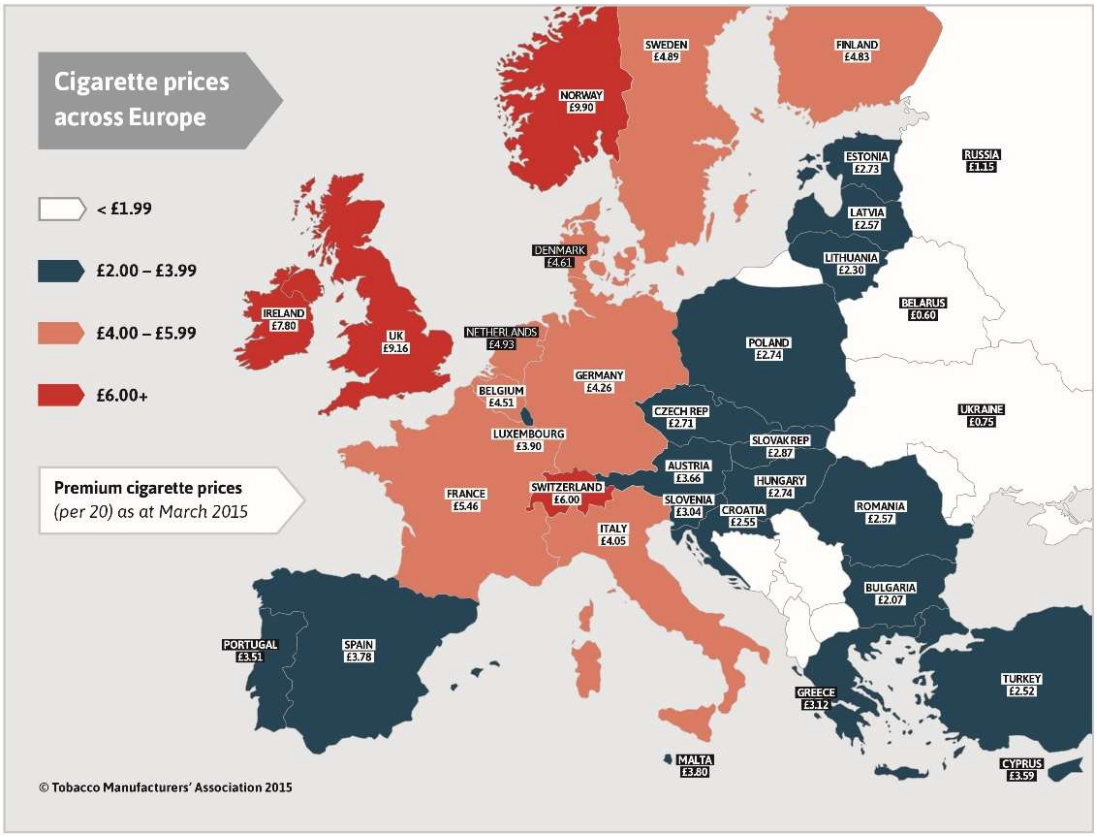
Figure 4 – Prevalence in EU 27 2009 -2012



Obrázek 3: Počet kuřáků v EU, zdroj: EUROPEAN COMMISSION, 2013. Report on the implementation of the Council Recommendation of 30 November 2009 on Smoke-free Environments

1.2.2 Daňové výnosy z cigaret a náklady kouření

V celé Evropské unii je díky harmonizaci daňových zákonů systém zdanění tabákových výrobků stejný. Na cigarety je uvalena spotřební daň a DPH, které se vypočítává z ceny cigaret včetně spotřební daně. Vzhledem k tomu, že Evropská unie ukládá jen minimální sazby obou daní, může se daňové zatížení i ceny cigaret lišit stát od státu.



Obrázek 4: Ceny cigaret v Evropě za rok 2015 v librách, zdroj: EU cigarette prices: Cigarette prices across Europe, 2015

V České republice byla v březnu roku 2015 cena prémiové značky cigaret 2,71 liber, což je dle přepočtu 101,90 Kč (KURZYCZ, 2015) za krabičku. V té tobě bylo průměrné celkové zatížení cigaret 78,1 % z maloobchodní ceny (EUROPEAN COMMISSION, 2015). Průměrné daňové zatížení v Evropské unii u cigaret je 77 % z ceny cigaret. Přínos do státního rozpočtu z výše uvedené krabičky cigaret se pohyboval okolo 80 Kč. Daňový výnos ze spotřebních daní bývá velice stabilní. Pokud má navíc produkt, na nějž je daň uvalena, stálou poptávku, je tento produkt naprosto ideální pro stabilizování státního rozpočtu.

Daňové výnosy z cigaret a tabákových výnosů bývají značné, ovšem náklady kouření jsou mnohonásobné. Například v roce 1999 se na daních z tabákových výrobků vybralo v České republice 20,2 mld. Kč. Přitom v tom samém roce byly odhadnuty náklady na zdravotní péči poskytnutou v lůžkových zařízeních na nemoci způsobené kouřením na 23 miliard Kč. Ty představují jen 35 - 50 % celkových nákladů na zdravotní péči. Pokud do celkových nákladů kouření započítáme i náklady způsobené vinou předčasných úmrtí a náklady spojené s vlivem cigaretového kouře na nekuřáky, dostáváme se k číslu 75,3 mld. Kč. Celkové náklady kouření převyšují výnosy z daní téměř čtyřikrát. (KRÁLÍKOVÁ, Eva, 2011)

2 Regrese na panelových datech a popis modelu

Regresní analýza se zabývá získáváním informací o vztahu mezi vysvětlovanou proměnnou a vysvětlujícími proměnnými. Pokud je mezi těmito proměnnými regresní vztah, můžeme ho použít například pro určení vlivu jedné vysvětlující proměnné na vysvětlovanou proměnnou, využít regresní model pro predikce apod. Regresní analýza stojí na následující premise: studujeme, jak se mění *y* při změnách *x ceteris paribus* (při nezměněných ostatních parametrech), kde proměnná *y* (vysvětlovaná) a *x* (vysvětlující) reprezentují určitou populaci (WOOLDRIDGE, 2013). Důležitým krokem je správná interpretace výsledků, protože nalezení regresního vzatu mezi proměnnými neznamená nutně kauzalitu. Pokud tedy najdeme určitý lineární vztah mezi proměnnými, nelze z toho usuzovat, že jedna proměnná ovlivňuje druhou. Může to být způsobeno i korelací s proměnnou třetí.

Regresní analýzu lze provádět na různých typech dat. Základním datovým souborem jsou cross sectional data, která jsou sbírána od náhodně vybraných subjektů (jednotlivců, firem, zemí apod.) ve stejný časový okamžik. Pooled cross sectional data jsou data o náhodně vybraných subjektech v různých časových okamžicích (nejčastěji různých letech), přičemž subjekty jsou v každý okamžik vybírány náhodně, liší se. Panelová data jsou data, která v sobě kombinují cross sectional data a časové řady. Tento datový soubor v sobě obsahuje informace o stejných subjektech (jednotlivci, města, země) napříč časem (WOOLDRIDGE, 2013). Hlavní rozdíl je tedy v tom, že sledujeme stále stejný náhodně vybraný vzorek z populace. Pro panelová data existují speciální metody regresní analýzy, které využívají jejich výhod nebo naopak řeší jejich nedostatky.

2.1 Předpoklady modelu a jeho proměnné

Model použitelný pro danou analýzu by měl být schopen efektivně pracovat s panelovými daty sesbíranými pro jednotlivé státy Evropské unie za určité časové období. Tento model by se měl vyrovnat s relativně malým počtem pozorování a krátkým časovým úsekem, protože potřebná data nejsou v delším časovém horizontu k dispozici. Cílovým, sledovaným vztahem je vztah mezi vysvětlující proměnnou, tj. spotřeba cigaret za jednotlivé země, a vysvětlující proměnnou reprezentující zdanění v jednotlivých

státech. V modelu nadále budou obsaženy i vysvětlující proměnné, u kterých očekávám zajímavý vliv na vysvětlující proměnnou a také takové, které by mohly ovlivňovat vysvětlující proměnnou. Kdyby nebyly tyto vlivy zachyceny v modelu, byly by součástí chybové složky, a to by mohlo vést k vychýlenosti regresních parametrů.

2.1.1 Hypotéza

Vytvořím model, kde jako závislá proměnná bude vystupovat počet cigaret vykouřených v jednotlivých státech za rok přepočtených na počet obyvatel. Klíčovou vysvětlující proměnnou bude zdanění. Ostatní vysvětlující proměnné by měli očistit model od dalších vlivů, takže estimátor reprezentující zdanění by měl být nestranný a konzistentní.

Hypotéza je následující. Pokud výše zdanění ovlivňuje spotřebu cigaret, jeho koeficient by měl být různý od nuly.

$$\boldsymbol{H_0: \beta_{tax} = 0}$$

$$\boldsymbol{H_1: \beta_{tax} \neq 0}$$

K testování této hypotézy slouží t test, který testuje statistickou významnost jednotlivých parametrů a udává správné hodnoty, pokud má náhodná složka modelu normální rozdělení nebo je alespoň asymptoticky normální.

2.1.2 Proměnné použité v modelu

Pro analýzu vztahu mezi spotřebou cigaret a jejich zdaněním by bylo vhodné použít následující proměnné. U každé proměnné je uveden zdroj, typ proměnné a její tvar.

Vysvětlovaná proměnná reprezentuje spotřebu cigaret v jednotlivých státech Evropské unie. Bohužel v databázi chybí údaje za Kypr, Maltu a Lucemburk. Pro lepší srovnatelnost, protože jednotlivé státy mají jiný počet obyvatelstva a kuřáků, jsem se rozhodla používat počet spotřebovaných cigaret za rok na osobu. Data jsem získala z databáze Passport (EUROMONITOR INTERNATIONAL, c2016a).

Vysvětlující proměnné jsou takové proměnné, u kterých lze předpokládat, že nějakým způsobem ovlivňují spotřebu cigaret. Výběr vysvětlujících proměnných je velice důležitý, protože může zásadně ovlivnit výsledky analýzy. Pokud do modelu zahrneme irelevantní proměnnou, tedy proměnnou, která nemá vliv na vysvětlující proměnnou, nemá to žádný

vliv na nestrannost odhadů regresních parametrů. Pokud však do modelu nezařadíme relevantní proměnnou, potom mohou být odhady regresních parametrů vychýlené (vlastní poznámky k předmětu Úvod do ekonometrie LS 2014/2015).

Cílovou vysvětlující proměnnou, jejíž vliv na spotřebu cigaret sleduji, je ***zdanění***. Zdanění cigaret je v zemích Evropské unie dvojí. Cigarety jsou zdaněny spotřební daní, která má dvě složky, specifickou a valorickou. Další daní uvalenou na cigarety je daň z přidané hodnoty (DPH), která se vypočítává z ceny cigaret včetně spotřební daně.

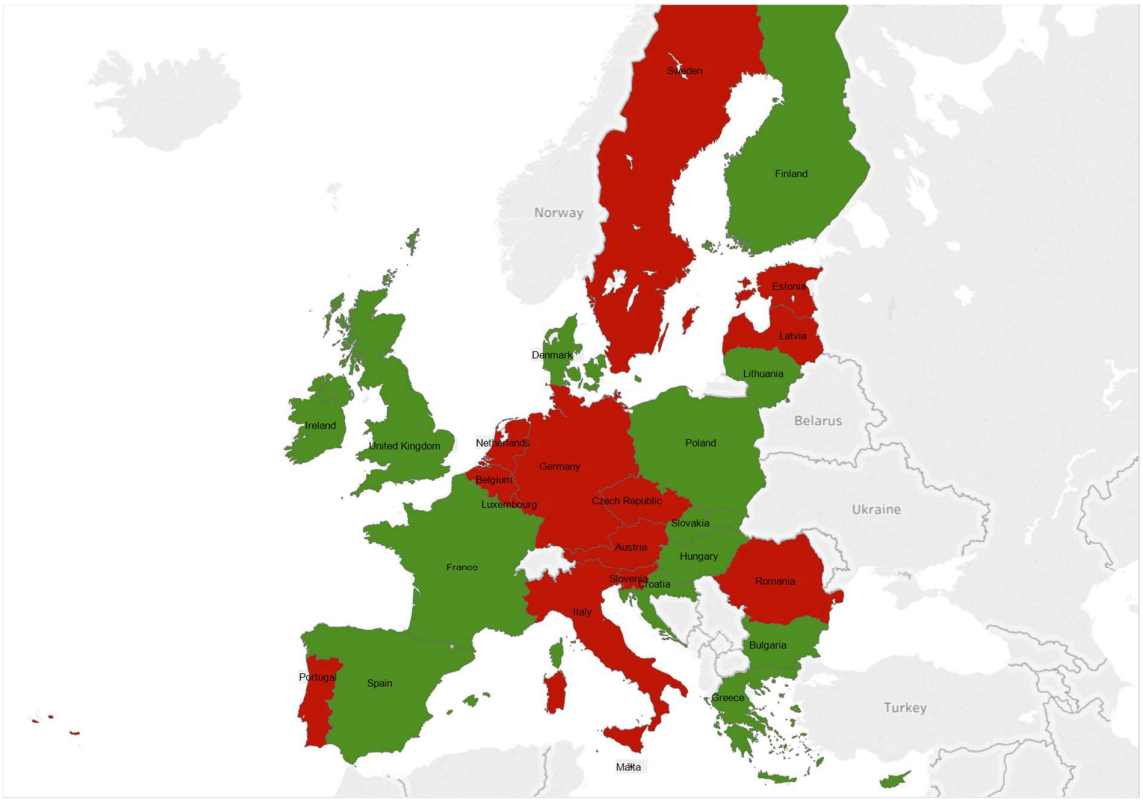
Pokud bychom použili jednotlivé sazby daně v jejich původním vyjádření, v pevné částce u specifické části spotřební daně a relativní částce u valorické části spotřební daně a DPH, nebyly by vlivy jednotlivých složek zdanění porovnatelné. A to právě kvůli jiným použitým jednotkám. Existuje více informačních zdrojů, kde jsou informace o zdanění různých komodit k dispozici, např. databáze Passport. Bohužel, jsou data v této databázi roztříštěna do několika kategorií a nejsou srovnatelná. Jejich filtrace by byla příliš náročná a výsledek by nemusel být správný. Pro zjednodušení jsem se rozhodla použít relativní vyjádření všech daní uvalených na cigarety vůči ceně. Tyto informace jsou dostupné z Excise duties tables pro jednotlivé roky publikované Evropskou komisí. Na stránkách Evropské komise jsou k dispozici jen tabulky s aktuálními informacemi a jejich databáze se mi nepodařila nalézt. Z různých zdrojů se mi podařilo najít tyto tabulky v různé podobě za roky 2003 až 2016 s chybějícími daty za rok 2004 a 2008. V tom neshledávám pro analýzu velký problém, vzhledem k tomu, že panelová data nemusí nutně obsahovat po sobě jdoucí roky.

Během let Evropská komise změnila metodiku výpočtu daně na cenu. Do roku 2010 aplikovala relativní vyjádření celkového zdanění na Maloobchodní prodejní cenu obsahující všechny daně (v originále Tax included retail selling price, dále jen TIRSP). Od roku 2011 je relativní zdanění vztaženo k Vážené průměrné ceně (v originále Weighted average price, dále jen WAP). Přesto, že se změnila metodika, nepředpokládám, že se změnila natolik, aby nějak výrazně ovlivnila výsledky analýzy. Proto nebudu data nijak upravovat.

Cenu nemohu použít jako vysvětlující proměnnou kvůli problému multikolinearity. Multikolinearita je stav, kdy mezi regresory (vysvětlujícími proměnnými) existuje

perfektní kolinearita (vlastní poznámky k předmětu Úvod do ekonometrie LS 2014/2015). Dané proměnné jsou vysoce korelované, je tudíž možné jednu vysvětlující proměnnou nahradit druhou bez velké ztráty informace (variability) v ní obsažené. Stupeň korelace, kdy dané proměnné splňují tuto podmínku, není pevně stanoven, nejčastěji se uvádí 0.9 a více (WOOLDRIDGE, 2013).

Velmi zajímavým a hodně diskutovaným nástrojem ke snížení negativního vlivu kouření je ***zákaz kouření v restauracích.*** V každé zemi Evropské unie je zavedena nějaká forma zákazu kouření v restauracích. Někde je zaveden úplný zákaz kouření (kuřáci mohou kouřit venku před barem/restaurací a na terasách), např. Velká Británie, Irsko a Kypr. V některých státech je povoleno kouření v oddělených místnostech/prostorech pro kuřáky, ve speciálně upravených kuřáckých místnostech apod., např. ve Švédsku, Slovinsku nebo Lucembursku. Pro účely analýzy jsem jednotlivé země rozdělila do dvou kategorií. V první kategorii existuje úplný zákaz kouření a v druhé kategorii je částečný zákaz, kdy lze kouřit v oddělených prostorách. V následujícím obrázku jsou zeleně zobrazeny země, kde je zaveden úplný zákaz kouření v restauracích a červeně, kde je zaveden částečný zákaz.

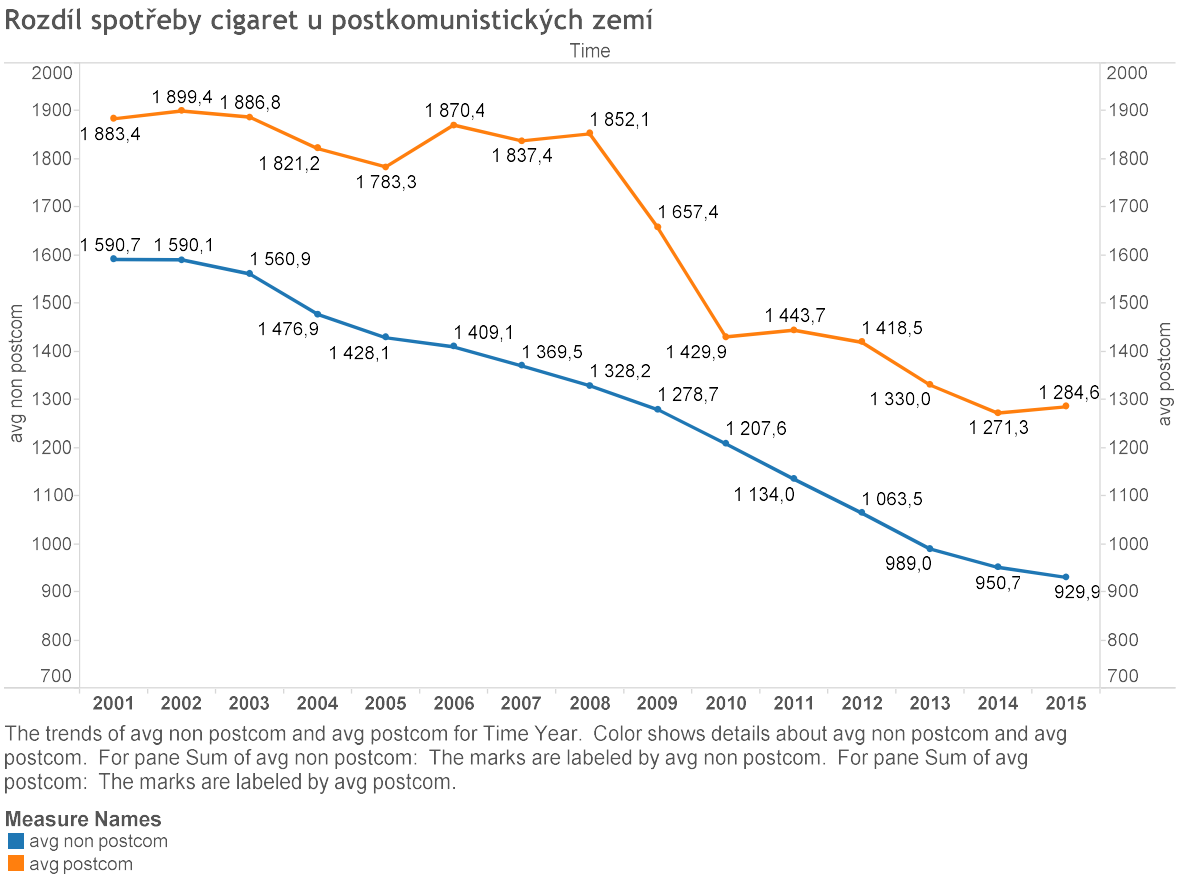


Map based on Longitude (generated) and Latitude (generated). Color shows average of Gastronomy. The marks are labeled by Country. Details are shown for Country.

Obrázek 5: Typ zákazu kouření v restauracích, úplný zákaz (zeleně), částečný zákaz (červeně), vlastní zpracování, výstup z programu Tableau

Pro účely analýzy, bude vytvořena dummy proměnná, která bude reprezentovat účinnost určité formy zákazu kouření v restauracích v jednotlivých letech. Data jsou sesbírána z různých zdrojů.

Dle studie z Harvardské university (Shleifer & Treisman, 2014, str. 9, vlastní překlad) *„se v **postkomunistických zemích** kouří a pije více, než je typické pro země se stejným HDP na obyvatele.“* Tato myšlenka mě přivedla ke grafické analýze vývoje spotřeby cigaret postkomunistických států Evropské unie. Na obrázku číslo 6 je vidět, že rozdíl mezi průměrnou spotřebou cigaret na osobu za rok tu opravdu je.

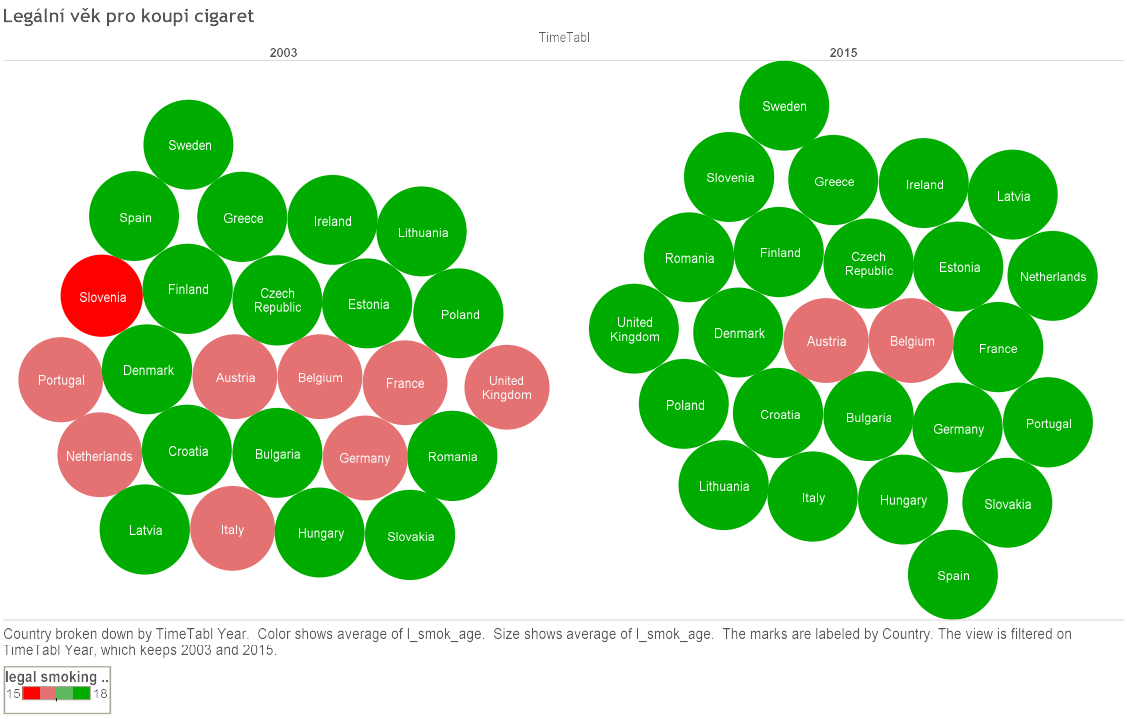


Obrázek 6: Rozdíl spotřeby cigaret u postkomunistických zemí, vlastní zpracování, výstup z programu Tableau

Na základě grafické analýzy jsem se rozhodla zařadit proměnnou reprezentující postkomunistické státy Evropské unie do analýzy ve tvaru dummy proměnné, která se rovná jedné pro postkomunistické země. Vzhledem k tomu, že výsledky analýzy poskytují standartní chyby odhadu, lze jednoduše otestovat výše zmíněnou hypotézu o statistické významnosti rozdílu ve spotřebě cigaret mezi postkomunistickými státy Evropské unie a ostatními členskými státy. Německo jsem zařadila do zemí, které nebyly komunistické, jelikož východní Německo bylo integrováno do západního Německa a jejich politika a ekonomika vykazuje jiné tendence než u většiny ostatních postkomunistických zemí.

Různé zdroje (ERIKSEN et al. 2015 nebo MLČOCH a DOLEŽAL, 2014) hovoří o velkém problému kouření mladistvých. Proto jsem se rozhodla do analýzy použít i vysvětlující proměnnou pro ***legální věk pro koupi cigaret***. Tento věk se v průběhu let 2001 až 2015

zvyšoval v několika státech Evropské unie, většinou z 16 let na 18 (EUROMONITOR INTERNATIONAL, c2016b). Tento jev je vidět na obrázku číslo 7. V roce 2015 již většina států Evropské unie používá legální věk pro koupi cigaret 18 let, jen Rakousko a Belgie mají sníženou hranici na 16 let. Zajímavé je, že v roce 2013 Slovinsko používalo nejnižší hranici legálního věku pro koupi cigaret a to 15 let.



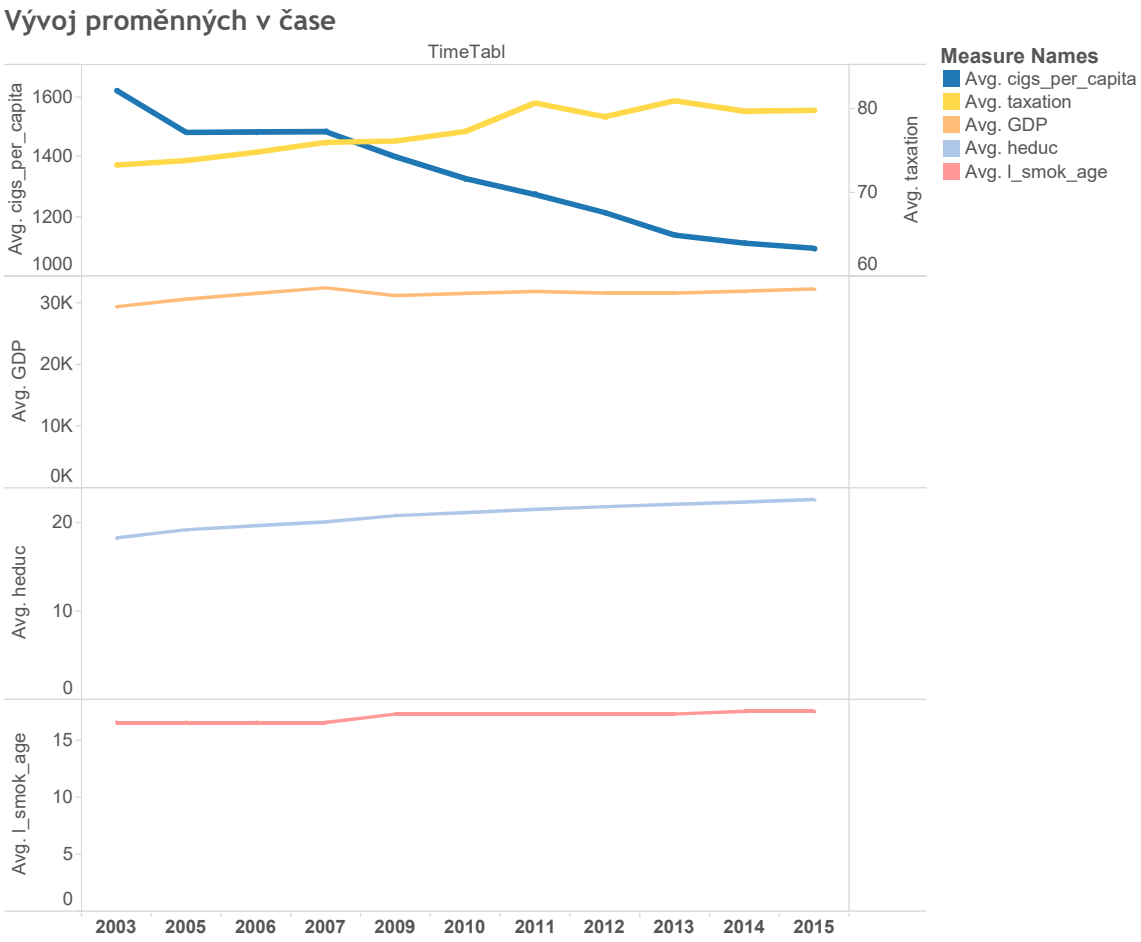
Obrázek 7: Změna legálního věku koupi cigaret, vlastní zpracování, výstup z programu Tableau

Dle Tobacco Atlas (ERIKSEN et al. 2015) se ve vyspělých státech snižuje počet kuřáků. Tovede k domněnce, zda výše ***GDP*** jako proxy proměnná pro blahobyt, může ovlivňovat spotřebu cigaret. Aby byla data srovnatelná, použiji HDP na obyvatele v reálných cenách (pevných cenách roku 2015) s pevným kurzem (roku 2015) z databáze Passport (EUROMONITOR INTERNATIONAL, c2016d).

Studie ukazují, že počet kuřáků klesá s vyšším vzděláním (ERIKSEN et al. 2015). A protože se úroveň ***vzdělání*** v populaci mění v průběhu let, měla by být zahrnuta do modelu. Nepřítomnost proměnné *educ* reprezentující vzdělání by mohla vést k vychýlenosti regresních parametrů. Pro účely analýzy jsem vybrala procento obyvatel starších 15-ti let s vyšším vzděláním. Vyšší vzdělání v sobě obsahuje vzdělání na univerzitách a vyšších

odborných školách a obdobných školách, které vyžadují jako minimální podmínku přijetí úspěšné dokončení středoškolského vzdělání nebo doklad o dosažení rovnocenné úrovně znalostí. Informace byly získány z databáze Passport (EUROMONITOR INTERNATIONAL, c2016e).

Neposledně i ***nezaměstnanost*** může ovlivňovat spotřebu cigaret. Dle The Tobbaco Atlas (ERIKSEN et al. 2015) je u nezaměstnaných větší pravděpodobnost, že začnou kouřit, znovu se vrátí ke kouření nebo zvýší počet vykouřených cigaret za den. Do analýzy bude zařazena proměnná *unem* jako procento nezaměstnanosti vztažená k ekonomicky aktivnímu obyvatelstvu získaná z databáze Passport (EUROMONITOR INTERNATIONAL, c2016f).



Obrázek 8: Vývoj proměnných v čase, průměrné hodnoty za všechny země, vlastní zpracování, výstup z programu Tableau

Obrázek číslo 8 prezentuje vývoj všech proměnných, kromě dummy proměnné pro postkomunistické země a zákazu kouření v restauracích. Jak je vidět z grafu, proměnné se vyvíjejí v čase více méně lineárně, proto jsem se rozhodla pro analýzu použít jen lineární vztah mezi vysvětlujícími proměnnými a vysvětlovanou proměnnou.

2.2 Metody analýzy na panelových datech

Regresní analýza aplikovaná na panelových datech se liší od klasické regresní analýzy na průřezových datech. Hlavním důvodem je přidání dalšího rozměru, času. V panelových datech máme dohromady tři rozměry, rozměr vysvětlujících proměnných, rozměr jednotlivých pozorování a čas. Na panelová data můžeme aplikovat i metody používané

u pooled cross sectional dat, ale výsledky pak mohou být zkreslené a parametry vychýlené. Klasický lineární regresní model s jednou vysvětlující proměnnou vypadá takto

$$y_i = \beta_0 + \beta_1 x_i + u_i, \qquad [2.1]$$

kde y_i je vysvětlovaná proměnná, x_i je vysvětlující proměnná, koeficient β_0 je intercept, koeficient β_1 je dílčí regresní parametr a proměnná u_i je náhodná chybová složka, která je různá napříč jednotlivými pozorováními. Vzhledem k tomu, že v panelových datech máme přidaný další rozměr, musíme náš model pozměnit na

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}. \qquad [2.2]$$

Model je obdobný, jen index *t* značí přidaný rozměr, čas, a proměnná a_i reprezentuje nepozorovaný efekt konstantní v čase. Máme tedy dva indexy, index *i = 1, …, n*, který označuje jednotlivce napříč pozorováním, a index *t = 1, …, T,* který označuje čas. Pro ekonometrickou analýzu panelových dat nemůžeme předpokládat, že pozorování jsou náhodně (nezávisle) rozloženy napříč časem (WOOLDRDGE, 2013, cap. 13). To znamená, že jedna vlastnost pozorovaného jedince (v čase neměnná) může ovlivňovat hodnoty jiné proměnné v různých časech. Z tohoto důvodu byly vytvořeny určité metody k analýze panelových dat.

V následujících subkapitolách popíši jednotlivé metody používané na panelových datech. Každá metoda je trochu jiná, ale všechny metody mají stejnou vlastnost, snaží se odstranit z modelu nepozorovatelný efekt, nebo také fixní efekt a_i, který je konstantní napříč časem a ovlivňuje y_{it} (WOOLDRDGE, 2013, cap. 13.1).

2.2.1 First differencing model

Tento model je skvělým odrazovým můstkem k pochopení analýzy panelových dat, jejich úskalí a jejich výhod. Nyní objasním základní vlastnosti panelových dat a práci s nimi.

Metoda první diference není nijak složitá a výrazně se neliší od ostatních metod. Díky své jednoduchosti a podobnosti s analýzou na cross sectional data je vhodná pro práci i s ne příliš sofistikovaným statistickým softwarem. Pokud uživatelé nevlastní software, který ovládá práci s modely s panelovými daty, může po jednoduché úpravě dat použít

klasický OLS[2] model. Stačí jen udělat první diferenci napříč pozorováními a tento diferencovaný model poté odhadnout pomocí OLS. Pokud bychom použili model OLS na panelová data, kde nepozorovaný efekt je korelovaný s vysvětlujícími proměnnými, byl by OLS odhad regresních parametrů nekonzistentní (CROISSANT a MILLO, 2008). Konzistentní odhad parametru je takový, který se s rostoucím N limitně blíží ke skutečné hodnotě odhadovaného parametru (WOOLDRIDGE, 2013).

Nejprve popíši unobserved effect modelu s jednou vysvětlující proměnnou, pro panelová data s T = 2. Máme model (WOOLDRIDGE, 2013, cap. 13.3)

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it}. \qquad [2.3]$$

Jak již dříve bylo řečeno, index *i* představuje jednotlivé pozorování napříč daty (pro jednotlivce) a index *t* označuje čas, v našem případě *t = 1, 2*. Proměnná $\delta_0 d2_t$ označuje dummy proměnnou, která je rovna nule, pokud *t = 1,* a jedné, pokud *t = 2*. Proměnná β_0 je intercept. Zahrnutí interceptu do modelu nám zajistí nulovou střední hodnotu náhodné složky (vlastní poznámky k předmětu Regrese 2014/15). Proměnná u_{it} je náhodná chybová složka, která je různá pro každé jednotlivé pozorování a různý časový okamžik. Náhodná chybová složka u klasického lineárního modelu pro cross sectional data (tedy pro model 2.1) by měla splňovat slabou sadu předpokladů, ty jsou následující:

- chybová složka má nulovou střední hodnotu, $E(u_i) = 0$, pro $i = 1, \ldots, n$,
- chybová složka má konečný konstantní rozptyl (homoskedasticita), $D(u_i) = \delta^2 < \infty$, pro $i = 1, \ldots, n$,
- chybové složky jednotlivých pozorování jsou nekorelované, jsou náhodné, $Cov(u_i, u_j) = 0$, pro $i = 1, \ldots, n, j = 1, \ldots, n$ a $i \neq j$ (vlastní poznámky k předmětu Regrese 2014/15).

Slabá sada předpokladů nestačí pro panelová data, a proto si musíme stanovit ještě další předpoklady proto, aby odhady regresních parametrů byly konzistentní a nevychýlené, tyto předpoklady budou následovat dále v textu.

[2] OLS neboli Ordinary Least Squares (Metoda nejmenších čtverců) je metoda, kterou se odhadují parametry regresního modelu. Tato metoda minimalizuje sumu čtverců reziduí (WOOLDRIDGE, 2013).

Proměnná a_i v sobě zachycuje nepozorovaný efekt, nebo také fixní efekt, který ovlivňuje vysvětlující proměnnou y_i a který se nemění v čase. Wooldridge (2013, cap. 13.3) často zdůrazňuje problém vychýlených regresních parametrů z jednoduché regresní analýzy způsobené nezahrnutím vysvětlujících proměnných. Tento problém je možné vyřešit zahrnutím více vysvětlujících do modelu, což nemusí být vždy jednoduché, protože některé faktory nelze kontrolovat, natož je sledovat. Další alternativou je zahrnutí zpožděné vysvětlované proměnné do modelu, která by mohla kontrolovat nepozorované efekty. Můžeme však také využít výhod panelových dat a sledovat faktory, které ovlivňují vysvětlovanou proměnnou. Tyto faktory jsou dva, jeden zmíněný již dříve, chybová složka a nepozorovaný efekt. První se mění v čase a druhý zůstává konstantní. Pokud bychom měli model, kde vysvětlovaná proměnná by byla počet vykouřených cigaret za rok na osobu a vysvětlující proměnou byla celková daň v relativním vyjádřená na cigaretu, pak by proměnná a_i reprezentovala nepozorovaný efekt, který ovlivňuje počet vykouřených cigaret, a který se nemění v čase. Například faktory, kterými jsou predispozice obyvatelstva ke kouření, vnímání kuřáků, historický kontext, společenskou přípustnost kouření a další faktory.

Pokud bychom model 2.3 odhadovali pomocí OLS, získali bychom vychýlené odhady parametrů a standardní chyby odhadů by byly nesprávné kvůli sériové korelaci. Jelikož předpokládáme, že nepozorovaný efekt, a_i, je korelovaný s vysvětlujícími proměnnými, musíme se této proměnné zbavit, jinak získáme vychýlené odhady parametrů. To provedeme tak, že diferencujeme data napříč časem.

Rovnici $y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2}, pro\ t = 2$

odečteme od rovnice $y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1}, pro\ t = 1,$

dostaneme tak rovnici $$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i, \qquad [2.4]$$

kde symbol Δ značí diferenci. Tím, že je fixní efekt v čase neměnný, z rovnice vypadne. Rovnice 2.4 se nazývá first differencing model, kde nám zůstal i intercept. Intercept v tomto případě reprezentuje rozdíl mezi vysvětlovanou proměnnou v čase *t = 1* a v čase *t = 2*. (WOOLDRIDGE, 2013, cap. 13.3)

Model 2.4 již můžeme modelovat pomocí OLS a za splnění následujících předpokladů získáme konzistentní a nevychýlené odhady regresních parametrů. Dalšími předpoklady pro model první diference jsou:

- diferencovaná chybová složka Δu_i splňuje striktní exogenitu, je tedy nekorelovaná s Δx_i v jakémkoliv čase, tj. $\mathrm{Cov}(\Delta u_i, \Delta x_i) = 0,$ [2.5]
- Δx_i musí mít variabilitu napříč *i,*

pokud se Δx_i nemění v čase nebo se mění vždy o stejný díl, nemáme dostatek variability pro odhadnutí regresního parametru. Takovou vysvětlující proměnnou může být například dummy proměnná pro pohlaví apod. Tyto dva předpoklady platí s menšími úpravami i pro následující modely. (WOOLDRIDGE, 2013, cap. 13.3)

Extrahování fixního efektu z rovnice je velká výhoda first differencing modelu, nemusíme tak vytvářet velmi komplexní model s mnoha vysvětlujícími proměnnými jen proto, abychom zamezili ponechání některých nepozorovaných vlivů, které se v čase nemění, v chybové složce a způsobili tak vychýlení odhadů regresních parametrů. Má to však i své nevýhody. Jednou z nich je, že provedením diference můžeme ztratit část variability vysvětlovaných proměnných (WOOLDRIDGE, 2013, cap. 13.3). Tak například x_i má velkou variabilitu, ale Δx_i ji už mít nemusí. Malá variabilita může vést k velké standardní chybě a velkému rozptylu intervalového odhadu regresních parametrů.

2.2.2 Fixed effect model

Fixed effect model je velmi obdobný first differencing modelu. Mohl by se považovat i za více využívaný model, jelikož za určitých předpokladů má lepší výsledky než first differencing.

Máme model s jednou vysvětlující proměnnou na panelových datech,

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \text{ pro } t = 1, 2, \ldots, T. \quad [2.6]$$

Nyní uděláme průměr pro všechny hodnoty *i* přes všechna *t* a dostaneme tuto rovnici

$$\bar{y_i} = \beta_1 \bar{x_i} + a_i + \bar{u}_i, \quad [2.7]$$

kterou odečteme od rovnice 2.5. Tím dostaneme

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it}. \quad [2.8]$$

Tento proces se nazývá fixed effect transformace. Rovnici 2.8 můžeme odhadnout pomocí OLS za splnění stejných předpokladů, které platí pro first differencing model. Tento model také nedokáže odhadnout regresní parametr u vysvětlující proměnné, která nemá nebo má konstantní variabilitu napříč časem. Avšak tyto proměnné můžeme spojit interakcí s proměnou, která se mění v čase, nejčastěji s dummy proměnnou pro různé roky. Tím zajistíme určitou variabilitu proměnné a můžeme ji použít v modelu. (WOOLDRIDGE, 2013, cap. 14.1)

First differencing a fixed effect model jsou velmi podobné modely, je však dokázáno, že za určitých předpokladů může mít jeden model lepší výsledky než druhý model. Tyto případy se pokusím přehledně vysvětlit v následujícím textu.

- Pokud máme T = 2, fixed effect a first differencing model dávají stejné výsledky, musíme však mít stejné proměnné. Pokud tedy v modelu first differencing použijeme intercept, poté bychom měli do fixed effect zakomponovat dummy proměnnou pro t = 2.
- Fixed effect model je efektivnější pro panelová data, kde je velké N a malé T a kde chybová složka u_{it} je sériově nekorelovaná.
- Pokud je chybová složka, u_{it}, silně pozitivně sériově korelovaná, je typu random walks, potom první diference náhodné složky, Δu_{it}, bývá sériově nekorelovaná.
- Pokud máme velké T a N není moc velké, pak musíme být opatrní, když používáme fixed effect model. Tento model je velmi náchylný na porušení jakéhokoliv předpokladu při velkém T ale malém N. (WOOLDRIDGE, 2013, cap. 14.1)

2.2.3 Random effect model and Correlated Random effect model

Fixed effect i first differencing model jsou zkonstruovány tak, abychom se zbavili nepozorovaného efektu a_i, který je korelovaný s jednou nebo více vysvětlujícími proměnnými. Pokud však předpokládáme, že nepozorovaný efekt není korelovaný s žádnou vysvětlující proměnnou, pak je použití těchto modelů nesmyslné. (WOOLDRIDGE, 2013, cap. 14.2)

Máme-li stejný model jako dříve s nepozorovaným efektem a_i,

$$y_{it} = \beta_0 + \beta_1 x_{it} \ + \cdots + \beta_k x_{itk} + a_i + u_{it} \qquad [2.9]$$

s více vysvětlujícími proměnnými a s interceptem. Přičemž předpokládáme, že nepozorovaný efekt je nekorelovaný s jakoukoliv vysvětlující proměnou v jakémkoliv čase,

$$Cov(x_{itj}, a_i) = 0 \text{ pro } t = 1, \dots, T, i = 1, \dots, N \text{ a } j = 1, \dots, k, \qquad [2.10]$$

pak můžeme použít pro odhad regresních parametrů **random effect model**. (WOOLDRIDGE, 2013, cap. 14.2)

Tento model na rozdíl od cross sectional modelu nebo pooled OLS modelu dává nestranné a konzistentní odhady regresních parametrů, ale také počítá správně standardní chyby odhadů. Ty jsou nesmírně důležité pro korektní testovací statistiky a vedou k řádným závěrům. Standardní chyby odhadu jsou nesprávné u cross sectional modelu a pooled OLS modelu z toho důvodu, že ignorují sériovou korelaci obsaženou ve složené chybové složce, která je definována takto:

$$v_{it} = a_i + u_{it.} \qquad [2.11]$$

Protože je nepozorovaný efekt a_i obsažen ve složené chybové složce ve všech časech, způsobuje sériovou korelaci v_{it}. Tato pozitivní korelace může být podstatná, a proto jsou OLS standardní chyby nekorektní. (WOOLDRIDGE, 2013, cap. 14.2)

Další výhodou random effect modelu je ta, že můžeme do modelu zahrnout i vysvětlující proměnné, které jsou konstantní v čase, na rozdíl od fixed effect a first differencing modelu. To je možné díky tomu, že předpokládáme, že fixní efekt, a_i, je nekorelovaný s jakoukoliv vysvětlující proměnnou v jakémkoliv čase bez ohledu na to, jestli je tato vysvětlující proměnná v čase konstantní nebo ne. (WOOLDRIDGE, 2013, cap. 14.2)

Problematika random effect modelu je mnohem komplexnější, ale jelikož nepředpokládám, že v mé analýze by mohl být prakticky použit, proto se jím nadále nebudu do hloubky věnovat. Ale poznatky z pochopení tohoto modelu jsou zásadní k pochopení navazujícího modelu, Correlated random modelu, který by mohl být použit pro mou analýzu. Předpokládám, že nepozorovaný efekt, v mém případě reprezentující neviditelný a v čase konstantní efekt predispozic obyvatelstva jednotlivých zemí ke kouření, je korelován s jednotlivými vysvětlujícími proměnnými, jako jsou legální věk koupi cigaret, GDP a další.

Correlated random effect model v sobě kombinuje výhody fixed effect a first differencing modelu s random effect modelem. V obou modelech s random effectem je fixní, nepozorovatelný, individuální efekt a_i, považován za náhodnou veličinu ve smyslu nezávislosti na vysvětlujících proměnných. Na rozdíl od random effect modelu umožňuje correlated random effect model, aby mezi a_i a střední hodnotou vysvětlujících proměnných mohla být korelace. Střední hodnotou vysvětlující proměnné se rozumí aritmetický průměr proměnné napříč časem pro jednotlivé i. Tedy $\bar{x}_i = \frac{1}{T}\sum_{i=1}^{T} x_{it}$. Můžeme tedy napsat, pokud předpokládáme model s jednou vysvětlující proměnnou [2.5], že mezi a_i a $\bar{x}_i$ existuje regresní vztah ve tvaru

$$a_i = \alpha + \gamma \bar{x}_i + r_i. \qquad [2.12]$$

Pokud je $\gamma \neq 0$, pak jsou a_i a $\bar{x}_i$ korelované. Pokud dosadíme rovnici [2.12] do rovnice [2.6], pak dostaneme rovnici

$$y_{it} = \alpha + \beta_1 x_{it} + \gamma \bar{x}_i + r_i + u_{it}. \qquad [2.13]$$

Tento model se nazývá correlated random effect model. I když zaměníme nepozorovaný efekt za regresní vztah vůči střední hodnotě vysvětlujících proměnných, stále splňujeme dva hlavní předpoklady modelu. A to přesně předpoklad neexistujícího lineárního vztahu mezi nepozorovaným efektem a vysvětlujícími proměnnými [2.10] a mezi náhodnou složkou, u_i, a vysvětlujícími proměnnými [2.5]. Pokud splníme všechny předpoklady, tento model nám dává konzistentní a nevychýlené odhady regresních parametrů. (WOOLDRIDGE, 2013, cap. 14.3)

Jak již bylo řečeno, correlated random effect model je velmi podobný fixed effect modelu. Prakticky se od sebe liší jen složkou $\gamma \bar{x}_i$, která vysvětluje část variability vysvětlované proměnné, kterou nese fixní, nepozorovatelný efekt a jenž je korelován se střední hodnotou vysvětlované proměnné. Díky tomu lze velice jednoduše zhodnotit, jaký model je pro daná data lepší. A to tak, že otestujeme t statistikou, zda regresní odhad parametru $\hat{\gamma}$ je statisticky významný či nikoliv. Pokud je statisticky významný, je vhodnější použít fixed effect model (v podobě CRE modelu), a naopak, pokud je statisticky nevýznamný, je vhodnější correlated random effect model. (WOOLDRIDGE, 2013, cap. 14.3)

Další výhodou correlated random effect modelu je, že do něho můžeme zahrnout i vysvětlující proměnné, které jsou konstantní v čase anebo mají jen velmi malou variabilitu okolo průměru. Je dokázáno, že correlated random effect poskytuje stejné odhady regresních parametrů jako fixed effect model, pokud je odhad parametru $\hat{\gamma}$ je statisticky nevýznamný. Takže i když bychom měli při analýze použít správně fixed effect model, můžeme použít cíleně correlated random effect, abychom mohli do modelu zahrnout i v čase neměnné vysvětlující proměnné. V tomto případě musíme být opatrní při interpretaci odhadnutého regresního parametru, nemusí tu nutně existovat kauzální vztah.

Všechny tyto modely jsou podporovaný jedním package (pml) ve statistickém softwaru R. Díky tomu není složité vyzkoušet přímo na datech, jaký z modelů je pro analýzu nejvhodnější.

2.2.4 Předpoklady pro Fixed effect a Random effect model

1. předpoklad FE.1

Pro všechna *i*, máme model $y_{it} = \beta_1 x_{it} \ + \cdots + \beta_k x_{itk} + a_i + u_{it}, pro\ t = 1, \dots, T$, kde β_j jsou lineární kombinací odhadovaných regresních parametrů a a_i je fixní, nepozorovaný efekt.

2. předpoklad FE.2

Výběrový soubor je získán z populace náhodným výběrem.

3. předpoklad FE.3

Všechny vysvětlující proměnné se mění v čase (alespoň pro nějaká *i*) a neexistuje perfektní lineární kombinace mezi dvěma a více vysvětlujícími proměnnými.

4. předpoklad FE.4

Náhodná složka, u_{it}, splňuje podmínku striktní exogenity. To znamená, že není korelovaná s žádnou vysvětlující proměnnou ani fixním efektem přes všechna *i* a *t*, $E(u_{it}|\boldsymbol{X}_i, a_i) = 0$.

První čtyři předpoklady jsou stejné i pro first differencing model. Pokud jsou splněny, pak jsou odhady regresních parametrů **nestranné**. Pokud jsou splněny první čtyři předpoklady, pak jsou FE odhady **konzistentní** s pevným *T* a *N* →∞.

5. předpoklad FE.5

Rozptyl náhodné složky je konstantní a konečný, tzv. homoskedasticita. $Var(u_{it}|\boldsymbol{X}_i, a_i) = Var(u_{it}) = \delta_u^2$, pro všechna t.

6. předpoklad FE.6

Náhodná složka není autokorelována, $Cov(u_{it}, u_{is}|\boldsymbol{X}_i, a_i) = 0$ pro všechna $t \neq s$.

Pokud platí předpoklady FE.1 až FE.6, pak je FE odhad regresních parametrů ten nejlepší ze všech lineárních nevychýlených odhadů (BLUE – best linear unbiased estimator).

7. předpoklad FE.7

Náhodná složka je normálně rozdělená se střední hodnotou 0 a rozptylem δ_u^2 a je nezávislá na vysvětlujících proměnných.

Pokud je splněn předpoklad FE.7, pak má FE odhad **normální rozdělení** a t a F statistiky mají přesně t a F rozdělení. Díky tomu můžeme používat t a F statistiky k testování hypotéz. Pokud tento předpoklad nesplníme, můžeme spoléhat na asymptotické vlastnosti odhadu, kdy se rozdělení limitně blíží t a F rozdělení, ovšem jen za předpokladu, že máme malé *T* a velké *N*.

Pro **Random effect model** platí obdobné předpoklady, jako pro FE model s následujícími obměnami.

RE.1

Za podmínku FE.3 nahradíme: Neexistuje perfektní lineární kombinace mezi dvěma a více vysvětlujícími proměnnými.

RE.2

K podmínce FE.4 připojíme následující: očekávaná hodnota a_i na všech vysvětlujících proměnných je konstantní, tj. $E(a_i|\boldsymbol{X}_i) = \beta_0$.

RE.3

K podmínce FE.5 přidáme: rozptyl a_i na všech vysvětlujících proměnných je konstantní, tj. $Var(a_i|\boldsymbol{X}_i) = \delta_a^2$. (WOOLDRIDGE, 2013, appendix 14A)

3 Praktická část – Analýza a výsledky

Všechny modely popsané v kapitole 2 jsou podporované jedním *package (pml)* ve statistickém softwaru R. Díky tomu není složité vyzkoušet přímo na datech, jaký z modelů je pro analýzu nejvhodnější. Nejprve posoudím všechny modely a popíši jejich parametry a poté na základě různých testovacích statistik, rozhodnu, jaký model je podle mého názoru pro daná data nejvhodnější.

3.1 Příprava dat

Pro práci s *plm package* musí být data připravena v určité podobě. Datový soubor by měl v každém řádku obsahovat informace o specifickém jedinci pro daný časový okamžik (CROISSANT a MILLO, 2008). Datový soubor má tedy na každém řádku informace o všech proměnných pro jednotlivce v jednom čase. Může vypadat takto:

id	*time*	*y*	*x*
1	1	37	8
1	2	23	3
2	1	14	12
2	2	35	7
⋮	⋮	⋮	⋮

Tabulka 1: Datový soubor panelových dat, vlastní zpracování

Takto strukturovaná data je pak velice jednoduché převést tak, aby je software R četl jako panelová data. Existuje funkce *pdata.frame*, která převede klasický *data.frame* na panelová data. Argumenty funkce jsou následující: *pdata.frame (name_data.frame, indexes = c("name_id", "name_time")*, kde pro index použiji vektor hodnot obsahující jméno proměnné použité pro id (označení jednotlivce) a jméno proměnné označující čas (HEISS, c2016, cap. 13.3).

V použitém datasetu je 25 jedinců (n, počet použitých států EU) s informacemi za 13 let (bohužel, proměnná taxation nemá známé hodnoty pro roky 2004 a 2008, ostatní proměnné jsou k dispozici). Na následujícím obrázku je ukázka, jak vypadají data v programu R.

```
> head(pdata[, c(-2, -3, -6, -7, -8, -9)]) # použité proměnné
        Code Time cigs_per_capita l_smok_age inflation taxation postcomunistic     GDP unem heduc rest
AT-2003   AT 2003          1864.7         16       1.4     74.6              0 35398.9  4.8  11.5    0
AT-2004   AT 2004          1745.9         16       2.1       NA              0 36167.8  5.5  11.6    0
AT-2005   AT 2005          1688.1         16       2.3     75.4              0 36677.3  5.6  11.6    0
AT-2006   AT 2006          1655.6         16       1.4     75.4              0 37663.1  5.2  11.6    0
AT-2007   AT 2007          1652.9         16       2.2     75.4              0 38892.0  4.9  11.6    0
AT-2008   AT 2008          1695.3         16       3.2       NA              0 39324.7  4.1  11.5    0
>
```

Obrázek 9: Ukázka pdata.frame, vlastní zpracování, výstup z programu R

Velmi užitečnou funkcí je funkce *pvar(name_of_pdata.frame)*, která poskytuje informaci o proměnných, které se nemění v čase nebo pro jednotlivce (HEISS, c2016, cap. 14.2). Použitím této funkce na zkoumaná data lze zjistit, že proměnná *postcomunistic* je v čase neměnná. Tato informace je velice důležitá pro použití FD a FE modelu, protože takovou proměnnou musím z analýzy odstranit nebo použít interakci s v čase proměnlivou proměnnou.

```
> #jaké proměnné se mění a jaké ne
> pvar(pdata)
no time variation   :  Code Country postcomunistic
no individual variation :  TimeTabl Time NA
>
```

Obrázek 10: Jaké proměnné nejsou proměnlivé v čase, vlastní zpracování, výstup z programu R:

Obrázek číslo 10 udává, jaké proměnné se v čase neliší. Proměnná *postcomunistic* je dummy proměnnou označující země, které jsou postkomunistické. To je důvod, proč je tato proměnná v čase neměnná. I když se proměnné legálního věku pro koupi cigaret (*l_smok_age*), a zákazu kouření v restauraci (*rest*) vyvíjí v čase, nemají velkou variabilitu, proto by se s nimi mělo zacházet velmi opatrně a nezařazovat je do modelů, kde je potřeba dostatečná variabilita proměnných, jako jsou modely fixed effect a first differencing effect.

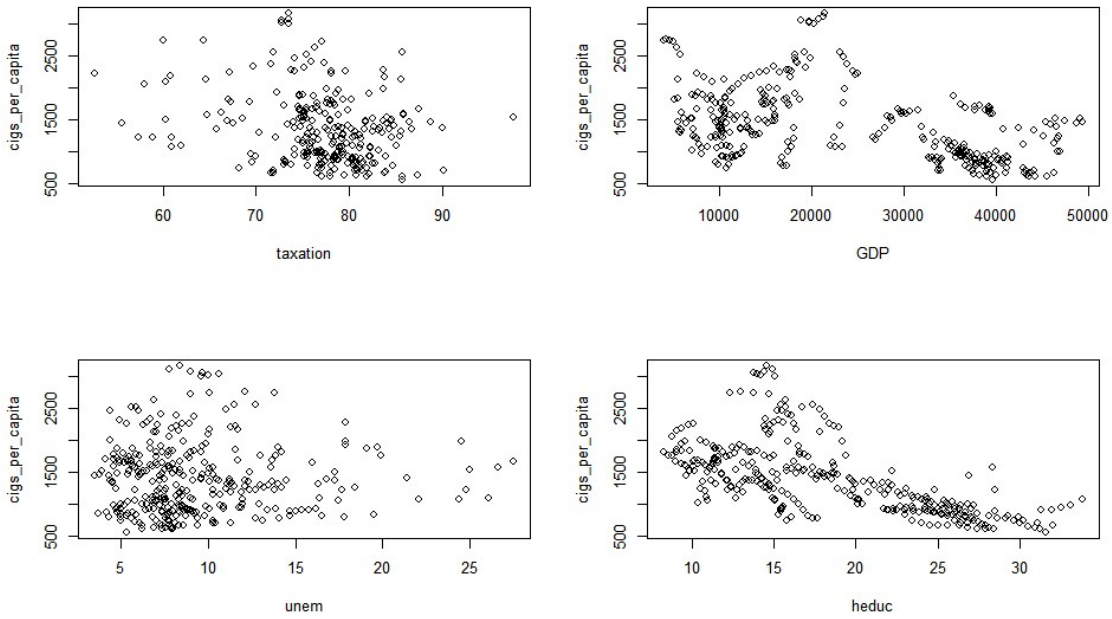
Proměnné *postcomunistic* a *rest* jsou dummy proměnné. Aby s nimi uměl program R pracovat jako s dummy proměnnou, musím změnit jejich charakter na factor příkazem *pdata$postcomunistic <- factor(pdata$postcomunistic, levels = c(0,1), labels = c("0", "posc"))* (obdobně i pro rest).

V následující tabulce jsou popisné statistiky proměnných.

```
> ## Popisné statistiky proměnných
> summary(Y)     > summary(X)
cigs_per_capita     taxation       l_smok_age    postcomunistic      rest
 Min.   : 560.8   Min.   :52.70   Min.   :15.00   Min.   :0.00   Min.   :0.0000
 1st Qu.: 933.9   1st Qu.:75.00   1st Qu.:18.00   1st Qu.:0.00   1st Qu.:0.0000
 Median :1342.2   Median :77.70   Median :18.00   Median :0.00   Median :1.0000
 Mean   :1401.6   Mean   :77.11   Mean   :17.55   Mean   :0.44   Mean   :0.9292
 3rd Qu.:1709.7   3rd Qu.:80.60   3rd Qu.:18.00   3rd Qu.:1.00   3rd Qu.:2.0000
 Max.   :3161.2   Max.   :97.60   Max.   :18.00   Max.   :1.00   Max.   :2.0000
                  NA's   :59

      GDP             unem            heduc
 Min.   : 3946   Min.   : 3.500   Min.   : 8.30
 1st Qu.:11116   1st Qu.: 6.700   1st Qu.:13.30
 Median :19681   Median : 8.300   Median :16.60
 Mean   :23963   Mean   : 9.363   Mean   :18.39
 3rd Qu.:37046   3rd Qu.:11.000   3rd Qu.:24.20
 Max.   :49399   Max.   :27.500   Max.   :33.80
```

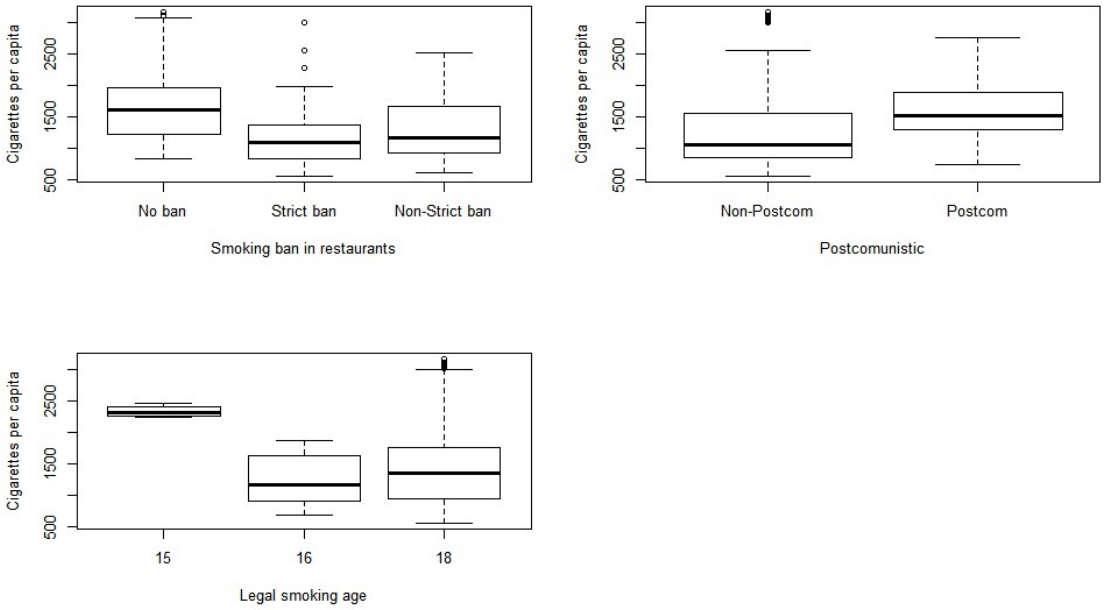
Tabulka 2: Popisné statistiky proměnných, vlastní zpracování, výstup z programu R

Zásadním krokem je určení tvaru závislosti vysvětlované proměnné na vysvětlujících proměnných. Takový tvar může být lineární, kvadratický nebo polynomický. Obecně se doporučuje nepoužívat polynomy vyššího stupně, ne výše než třetího stupně. Odhadnutí tvaru závislosti lze provést grafickou analýzou. Na obrázku číslo 11 jsou zobrazeny vysvětlované proměnné na vysvětlující proměnné (*cigs_per_capita*). U proměnných *GDP*, *unem* a *heduc* je zřejmé, že nevhodnější je proložit data přímkou. U proměnné *taxation* to není tak zřejmé. Přesto se domnívám, že nejvhodnější je také přímka. Složitější tvar závislosti by mohl být nakonec více nepřesný než aproximace lineárního vztahu.



Obrázek 11: Grafická analýza závislosti Y na X, vlastní zpracování, výstup z programu R

Na obrázku číslo 12 jsou zobrazeny boxploty vysvětlované proměnné v závislosti na méně variabilních vysvětlujících proměnných. I když je proměnná legal smoking age (*l_smok_age*) diskrétní kvantitativní proměnnou, chová se poněkud jako kategoriální proměnná, jelikož nemá velkou škálu hodnot. Jak je vidět, určité rozdíly mezi skupinami vysvětlujících proměnných tu jsou. Například se zdá, že zavedení přísného zákazu kouření v restauracích má větší vliv na kouření než méně přísný zákaz. Ovšem na tuto interpretaci si musím dávat pozor, protože množství vykouřených cigaret může záviset na jiných okolnostech (jiných vysvětlujících proměnných), než jsou zobrazené proměnné v grafech. Uvedené rozdíly nemusí mít nic společného s vlivem zobrazené vysvětlující proměnné. Na otázku, zda kategoriální vysvětlující proměnné (i kvantitativní proměnná legal smoking age) ovlivňují spotřebu cigaret, může odpovědět regresní analýza.



Obrázek 12: Boxplots Y na X (méně variabilní proměnné), vlastní zpracování, výstup z programu R

3.2 Modely v R

V následujících podkapitolách jsou zobrazeny a okomentovány jednotlivé výsledky modelů.

3.2.1 First differencing model

Jak již bylo řečeno výše, FD model je možné odhadnout i pomocí OLS odhadu na diferencovaných hodnotách pro každé x_i. Package *plm* poskytuje možnost odhadnout FD model bez nutnosti transformace dat, a to pomocí funkce *plm(…, model = „fd“)* (HEISS, c2016, cap. 13.5). FD model neumí pracovat s proměnnými, které se neliší v čase, nebo s proměnnými s malou variabilitou okolo průměru. Proto musím tyto proměnné z modelu odstranit. Vysvětlující proměnné pro FD a zároveň i FE model jsou *taxation, GPD, unem, heduc.*

FD model má jednu nevýhodu, provedením diference ztrácíme pozorování pro *t = 1*, to může být problém, pokud máme malé *T (Woodridge, 2013, cap. 13.3).*

Následující tabulka ukazuje výsledky first differencing modelu. Jak je vidět z výsledků analýzy, mám Unbalanced data, ale s těmi si program umí poradit. Model jako celek je statisticky významný, to znamená, že alespoň jeden odhad regresních parametrů je různý od nuly. Proměnné *unem* a *heduc* jsou statisticky významné na 5% hladině významnosti, ostatní proměnné jsou statisticky nevýznamné. Tyto testy jsou založeny na hodnotách standardních chyb, ty jsou správné, pokud jsou splněny všechny předpoklady uvedené v kapitole 2.2.4. Pokud nejsou splněny předpoklady modelu, mohou být standardní chyby odhadů chybné. To ovlivňuje většinu testů, které nemusejí vést ke správným výsledkům. Proto před interpretací výsledků musím provést diagnostiku modelu.

```
> # výsledky First differencing modelu
> summary(firstdiff)
Oneway (individual) effect First-Difference Model
Call:
plm(formula = Y ~ X2, data = pdata, model = "fd")
Unbalanced Panel: n=25, T=3-11, N=266
Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -624.0   -34.9    16.1    55.1   307.0

Coefficients :
               Estimate  Std. Error t-value  Pr(>|t|)
(intercept) -32.5878534  11.3814229 -2.8632   0.00457 **
X2taxation   -3.2335919   1.8285089 -1.7684   0.07828 .
X2GDP        -0.0104663   0.0093254 -1.1223   0.26286
X2unem      -29.9616497   4.6231596 -6.4808 5.272e-10 ***
X2heduc     -44.5359511  17.2695664 -2.5789   0.01052 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares:    4059000
Residual Sum of Squares: 3044900
R-Squared:      0.24985
Adj. R-Squared: 0.24466
F-statistic: 19.6505 on 4 and 236 DF, p-value: 5.647e-14
```

Tabulka 3: Tabulka výsledků FD modelu, vlastní zpracování, výstup z programu R

3.2.2 Fixed effect (Within effect) model

Fixed effect model odstraňuje nepozorovaný, fixní efect a_i odečtením průměrné hodnoty napříč časem pro jednotlivce. Tím nedochází k snížení počtu pozorovaní. Pokud má však určitá proměnná malou variabilitu napříč časem, může odečtením individuálního průměru o svou variabilitu přijít. Proto je nutné, aby měli vysvětlující proměnné dostatečnou variabilitu.

FE model se odhaduje v R pomocí funkce *plm (…, model = „within")*, který na rozdíl od OLS odhadu na transformovaných datech poskytuje správné stupně volnosti. Díky tomu jsou správné odhady směrodatných chyb a statistiky. (HEISS, c2016, cap. 14.1)

Pokud chci použít v čase neměnné proměnné, tak musím do analýzy zařadit jejich interakci s proměnnou, která je v čase proměnlivá (Wooldridge, 2013), jinak je musím z analýzy vypustit. Takovou proměnnou bývá nejčastěji dummy proměnná reprezentující čas. Pak by jako vysvětlující proměnná vystupovala interakce *year * poscomunistic* (obdobně i pro proměnnou *rest*). Rozhodla jsem se pro vypuštění proměnných s žádnou nebo malou variabilitou okolo individuálního průměru v čase. Pokud jsou tyto proměnné natolik důležité pro analýzu, mohu použít Correlated random effect model, který poskytuje stejné odhady regresních parametrů jako fixed model, a do kterého lze zařadit v čase neměnné proměnné.

V tabulce 4 jsou zobrazené výsledky fixed effect modelu. Proměnné *taxation, unem* a *heduc* jsou statisticky významné na 5% hladině významnosti, stejně jako celý model. Výsledky testů jsou správné, pokud model splňuje předpoklady. Pokud by byl tento model správný, zvýšení procenta zdanění o jeden procentní bod by vedlo ke snížení počtu spotřebovaných cigaret za rok na osobu o přibližně 13 cigaret.

```
> # výsledky Fixed effect modelu
> summary(fixed)
Oneway (individual) effect Within Model
Call:
plm(formula = Y ~ X2, data = pdata, model = "within")
Unbalanced Panel: n=25, T=3-11, N=266
Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-535.00  -89.90    2.01   99.50  529.00
```

```
Coefficients :
             Estimate Std. Error t-value  Pr(>|t|)
X2taxation -12.837665   2.735744 -4.6926 4.563e-06 ***
X2GDP       -0.014781   0.014559 -1.0153     0.311
X2unem     -45.367733   5.893046 -7.6985 3.713e-13 ***
X2heduc    -56.739441   9.092035 -6.2406 1.991e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares:    23706000
Residual Sum of Squares: 9353200
R-Squared:      0.60545
Adj. R-Squared: 0.53944
F-statistic: 90.9209 on 4 and 237 DF, p-value: < 2.22e-16
```

Tabulka 4:Tabulka výsledků FE modelu, vlastní zpracování, výstup z programu R

3.2.3 Random effect model

Random effect model předpokládá, že nepozorovaný efekt a_i je nekorelovaný s jakoukoliv vysvětlující proměnnou v jakémkoliv čase. Tím odpadá důvod proč používat FD nebo FE model. Ačkoliv by se mohlo zdát, že RE model je zbytečný a taková data by se dala odhadnout pomocí OLS, RE odhad bývá více efektivní, tj. mívá menší rozptyl. (HEISS, c2016, cap. 14.2)

I tento model je podporován *package plm*. Mohu jej zadat funkcí *plm (…, model = „random“)* (HEISS, c2016, cap. 14.2).

Do random effect modelu mohu zahrnout i v čase neměnné proměnné, proto teď použiji všechny vysvětlující proměnné a budu sledovat jejich vliv na vysvětlovanou proměnnou.

```
> # výsledky Random effect modelu

> summary(random)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)
Call:
plm(formula = Y ~ X2 + factor(postcomunistic) + factor(rest),
    data = pdata, model = "random")
Unbalanced Panel: n=25, T=3-11, N=266
Effects:
                   var  std.dev share
idiosyncratic  38011.9    195.0 0.249
individual    114540.4    338.4 0.751
```

```
theta  :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6844  0.8289  0.8289  0.8269  0.8289  0.8289

Residuals :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-505.00 -116.00   -8.29    0.17  105.00  585.00

Coefficients :
                              Estimate  Std. Error t-value  Pr(>|t|)
(Intercept)                 3.8855e+03  3.5264e+02 11.0183 < 2.2e-16 ***
X2taxation                 -9.9166e+00  2.7012e+00 -3.6712 0.0002934 ***
X2GDP                      -1.5026e-02  9.3701e-03 -1.6036 0.1100361
X2unem                     -4.1845e+01  5.2003e+00 -8.0467 3.114e-14 ***
X2heduc                    -4.3354e+01  7.9872e+00 -5.4279 1.314e-07 ***
factor(postcomunistic)posc -2.3907e+02  2.4306e+02 -0.9836 0.3262429
factor(rest)strict         -1.2640e+02  4.6261e+01 -2.7324 0.0067211 **
factor(rest)not-strict     -1.2363e+02  4.4110e+01 -2.8027 0.0054518 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares:    25453000
Residual Sum of Squares: 10038000
R-Squared:      0.60563
Adj. R-Squared: 0.58742
F-statistic: 56.601 on 7 and 258 DF, p-value: < 2.22e-16
```

Tabulka 5: Tabulka výsledků RE modelu, vlastní zpracování, výstup z programu R

Model jako celek je statisticky významný, také proměnné *taxation*, *rest*, *unem*, *heduc* a *rest* (jako dummy proměnná) jsou statisticky významné na 5% hladině významnosti. Testové statistiky podávají správné výsledky, pokud jsou splněny předpoklady modelu. Pokud je tento model správný, pak zvýšení daňového zatížení cigaret o jeden procentní bod vede v průměru ke snížení spotřeby cigaret na osobu za rok o zhruba 10 cigaret.

3.2.4 Corelated random effect

Pro provedení CRE modelu v R je nutné nejdříve vypočítat individuální průměr napříč *t = 2003, …, 2015* pro každé *i*. K tomu slouží funkce between, která vypočítá $\overline{x_i} = \frac{1}{T_i}\sum_{t=1}^{T_i} x_{it}$. Funkce Between pak tuto hodnotu vezme a zreplikuje ji T_i krát tak, aby vznikl řetězec dat o velikosti *N = Tn* použitelný při analýze. Modelu CRE dosáhnu

tím, že do modelu pro Random effect model přidám proměnné Between pro v čase proměnlivé proměnné (HEISS, c2016, cap. 13.4).

Také do CRE modelu lze zařadit v čase se neměnící proměnné jako je dummy proměnná pro postkomunistické zemně, zákaz kouření v restauracích a legální věk pro koupi cigaret, které nemají žádnou nebo velice malou variabilitu, nemohu vypočítat Between efekt, jejich průměrem by byla proměnná sama. Proto do CRE modelu nezařadím jejich Between hodnotu. V tomto případě však v čase neměnné proměnné musí splňovat podmínku nekorelovanosti s nepozorovaným efektem a_i.

```
> # Výsledky Correlated random effect modelu
> summary(cre)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)
Call:
plm(formula = Y ~ X2 + factor(postcomunistic) + factor(rest) +
    +Between(taxation, na.rm = TRUE) + Between(GDP) + Between(unem) +
    Between(heduc), data = pdata, model = "random")
Unbalanced Panel: n=25, T=3-11, N=266
Effects:
                  var std.dev share
idiosyncratic 38673.0   196.7 0.321
individual    81944.7   286.3 0.679
theta  :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6313  0.7972  0.7972  0.7949  0.7972  0.7972

Residuals :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-521.00 -111.00   -7.37    0.95  107.00  592.00

Coefficients :
                                  Estimate  Std. Error t-value  Pr(>|t|)
(Intercept)                     1.4122e+03  1.9178e+03  0.7364 0.4621949
X2taxation                     -1.0668e+01  2.8200e+00 -3.7831 0.0001932 ***
X2GDP                          -7.4128e-04  1.5478e-02 -0.0479 0.9618402
X2unem                         -3.9472e+01  6.1686e+00 -6.3989 7.490e-10 ***
X2heduc                        -4.5896e+01  9.8858e+00 -4.6427 5.519e-06 ***
factor(postcomunistic)posc     -1.9095e+01  2.8623e+02 -0.0667 0.9468638
factor(rest)strict             -1.3951e+02  4.8386e+01 -2.8833 0.0042719 **
factor(rest)not-strict         -1.2389e+02  4.7283e+01 -2.6202 0.0093183 **
Between(taxation, na.rm = TRUE)  1.9239e+01  2.3373e+01  0.8231 0.4112154
```

```
Between(GDP)                    -2.6999e-03  2.0071e-02 -0.1345 0.8930991
Between(unem)                    6.7990e+01  2.8701e+01  2.3689 0.0185887 *
Between(heduc)                   3.5879e+00  1.5620e+01  0.2297 0.8185071
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares:    26150000
Residual Sum of Squares: 10084000
R-Squared:      0.61441
Adj. R-Squared: 0.58669
F-statistic: 36.7906 on 11 and 254 DF, p-value: < 2.22e-16
```

Tabulka 6: Tabulka výsledků CRE modelu, vlastní zpracování, výstup z programu R

Model je statisticky významný na 5% hladině významnosti, stejně jako proměnné *taxation*, *unem*, *heduc* a *rest.* Uvedené statistické testy jsou platné jen za splnění předpokladů modelu. Pokud je tento model správný, pak by zvýšení daňového zatížení cigaret v úhrnu o jeden procentní bod snížilo spotřebu cigaret o necelé 2 cigarety na osobu za rok. Zdá se to sice málo, ale v úhrnu to může tato změna odradit několik desítek kuřáků.

3.3 Splnění předpokladů a testování hypotéz

První dva předpoklady - **FE.1** a **FE.2** - jsou splněné. Všechny použité modely jsou lineární v parametrech.

Třetím důležitým předpokladem - **FE.3** - je, aby mezi proměnnými nevznikla kolinearita, tedy aby mezi proměnnými nebyla perfektní lineární kombinace. Z obrázku 13 je vidět korelační matice vysvětlujících proměnných. Nejvyšší korelace je mezi proměnnými *GDP* a *postcomunistic*. Tyto proměnné mají korelaci téměř – 0,85. To je sice vysoký stupeň korelace, ale nemusí způsobovat velké výkyvy v odhadech. Přesto je vhodná při hodnocení výsledků modelů jistá opatrnost. Tento jev je pochopitelný, postkomunistické země většinou nemívají vysoké HDP. Ostatní proměnné nevykazují vysoký stupeň korelace.

```
> cor(X3, method = "pearson", use = "na.or.complete")
                taxation   l_smok_age postcomunistic         GDP        unem        heduc
taxation       1.00000000  0.086613181     -0.1224609  0.09223783  0.19085519  0.178687217
l_smok_age     0.08661318  1.000000000      0.3612869 -0.33828662  0.30788968 -0.006879794
postcomunistic -0.12246091  0.361286943      1.0000000 -0.84590782  0.10026325 -0.427643440
GDP            0.09223783 -0.338286623     -0.8459078  1.00000000 -0.37555509  0.533144643
unem           0.19085519  0.307889681      0.1002633 -0.37555509  1.00000000 -0.070733316
heduc          0.17868722 -0.006879794     -0.4276434  0.53314464 -0.07073332  1.000000000
```

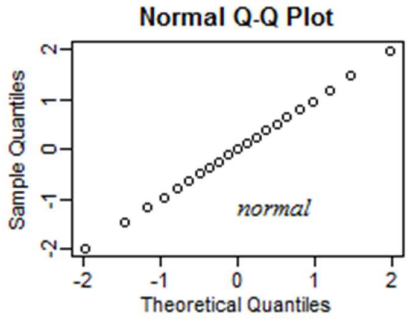
Obrázek 13: Kolinearita mezi proměnnými, vlastní zpracování, výstup z programu R

Testování normálního rozdělení náhodné složky - **FE.7** - lze provést několika způsoby. Jedna možnost je otestovat jej Shapiro-Wilk Normality Testem, který má jako nulovou hypotézu normální rozdělení náhodné složky. Tento test je jednoduchý na vyhodnocení, protože poskytuje p-hodnoty. Z tabulky číslo 7 vyplývá, že residua ani jednoho z testovaných modelů nemají normální rozdělení. To není nijak neobvyklé.

```
> #7###
Normality test of Residuals

> ## Shapiro test Normality
> shapiro.test(resid(firstdiff))

        Shapiro-Wilk normality test

data:  resid(firstdiff)
W = 0.88293, p-value = 1.099e-12

> shapiro.test(resid(fixed))

        Shapiro-Wilk normality test

data:  resid(fixed)
W = 0.98411, p-value = 0.004692

> shapiro.test(resid(random))

        Shapiro-Wilk normality test

data:  resid(random)
W = 0.98414, p-value = 0.004753

> shapiro.test(resid(cre))

        Shapiro-Wilk normality test

data:  resid(cre)
W = 0.98502, p-value = 0.006946

> # reálná data mívají málokdy přesně N-rozdělení
```

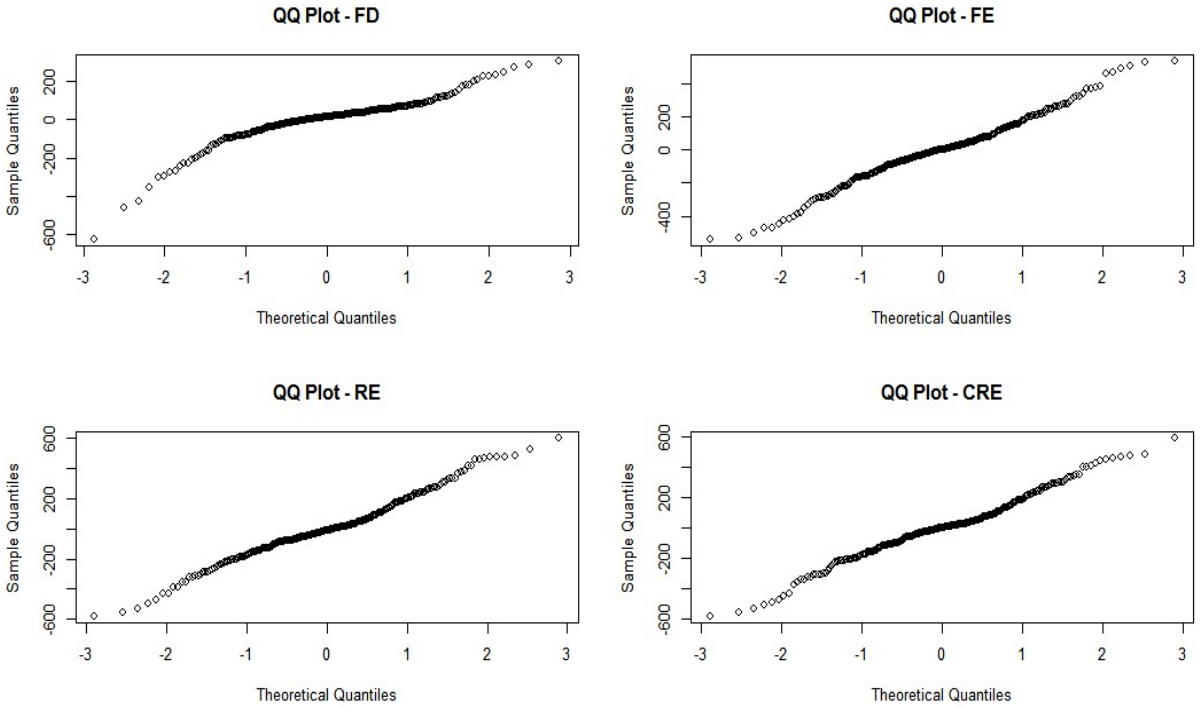
Tabulka 7: Shapiro-Wilks test normality residuí, vlastní zpracování, výstup z programu R

Reálná data nemají většinou přesně normální rozdělení. Proto se častěji používá QQ Plot, který zobrazuje hodnoty reziduí v závislosti na kvantily normálního rozdělení. Pokud má rozdělení residuí charakter normálního rozdělení, potom by měl vypadat takto.



Obrázek 14: Normální rozdělení residuí, zdroj: How to interpret a QQ plot, 2014

Na následujícím obrázku jsou vidět QQ Ploty zobrazující residua pro všechny použité modely.



Obrázek 15: QQ Plot normálního rozdělení residuí pro všechny modely, vlastní zpracování, výstup z programu R

Z grafů je vidět, že všechna residua jsou velmi blízko normálnímu rozdělení. Všechny modely mají přibližně normální rozdělení s lehkými chvosty, přičemž nejvýrazněji je tento jev vidět na first differenced effect modelu (QQ Plot – FD).

Testování homoskedasticity - FE.5 - provedu Breusch-Pagan Testem, kde nulová hypotéza zní, že residua splňují podmínku konstantního rozptylu náhodné složky. Tento test je citlivý na splnění podmínky normálního rozdělení náhodné složky, proto se používá jeho Koenkerova varianta se studentovým rozdělením (Woodridge, 2013, nápověda programu R k funkci bptest).

```
> #5###         Homoskedasticity test       > bptest(random)

> bptest(pooling)                                   studentized Breusch-Pagan test
        studentized Breusch-Pagan test
data:  pooling                              data:  random
BP = 56.404, df = 7, p-value = 7.853e-10    BP = 47.472, df = 7, p-value = 4.515e-08

> bptest(firstdiff)
        studentized Breusch-Pagan test      > bptest(cre)
data:  firstdiff                                  studentized Breusch-Pagan test
BP = 26.222, df = 4, p-value = 2.855e-05    data:  cre
> bptest(fixed)                             BP = 69.147, df = 11, p-value = 1.774e-10
        studentized Breusch-Pagan test      > # heteroskedasticita > robustní odhady na heterosked
data:  fixed
BP = 26.222, df = 4, p-value = 2.855e-05
```

Tabulka 8: Breuch-Pangan test Homoskedasticity, vlastní zpracování, výstup z programu R

Všechny p-hodnoty jsou moc malé, proto zamítám nulovou hypotézu o homoskedasticitě rozptylu náhodné složky. Důsledkem heteroskedasticity jsou špatné standardní chyby odhadů regresních parametrů. Tím je neplatná většina statistik, hlavně t a F testy. Kvůli tomu použiji robustní odhady regresních parametrů, které se dokážou vypořádat v heteroskedasticitou.

Breusch-Godfrey test autokorelace náhodné složky je všeobecný test, který lze použít pro všechny modely podporované *package plm.* Nulová hypotéza říká, že náhodná složka není sériově korelovaná, zatím co alternativní hypotéza připouští autokorelaci náhodné složky. (CROISSANT a MILLO, 2008)

```
> ## Serial correlation test for others models
> pbgtest(firstdiff, order = 2)

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  Y ~ X2
chisq = 6.8888, df = 2, p-value = 0.03192

alternative hypothesis: serial correlation in idiosyncratic errors

> pbgtest(fixed, order= 2)

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  Y ~ X2
chisq = 114.32, df = 2, p-value < 2.2e-16

alternative hypothesis: serial correlation in idiosyncratic errors

> pbgtest(random, order = 2)

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  Y ~ X2 + factor(postcomunistic) + factor(rest)
chisq = 129.56, df = 2, p-value < 2.2e-16

alternative hypothesis: serial correlation in idiosyncratic errors

> pbgtest(cre, order = 2)

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  Y ~ X2 + factor(postcomunistic) + factor(rest) + +Between(taxation,
               na.rm = TRUE) + Between(GDP) + Between(unem) + Between(heduc)
chisq = 129.36, df = 2, p-value < 2.2e-16

alternative hypothesis: serial correlation in idiosyncratic errors

> # všechny modely mají sériovou korelaci náhodné složky
                              > transformace dat a robustní odhady
```

Tabulka 9: Breusch-Godfrey test autokorelace náhodné , vlastní zpracování, výstup z programu R

Z uvedené tabulky vyplývá, že náhodné složky modelů obsahují sériovou korelaci. Proto použiji robustní odhady kovarianční matice odhadů, které se vypořádají s autokorelací náhodné složky.

3.4 Robustní odhady standardních chyb odhadu

Heteroskedasticita nebo přítomnost sériové korelace náhodné složky způsobuje neplatnost odhadů standardních chyb odhadu, potažmo i veškerých testovacích statistik založených na těchto parametrech, jako jsou t a F testy. S těmito nedostatky modelu

se lze vypořádat robustním odhadem kovarianční matice, tím získám správné odhady standardních chyb odhadu. V *package plm* existuje velice jednoduchá forma výpočtu robustních odhadů pomocí funkce *vcovHC*, která dokáže pracovat se všemi modely podporovanými *plm package*. Tato funkce může být použitá jako vstup pro regresní tabulky i různé testy hypotéz (HEISS, c2016).

V následující tabulce je porovnání odhadů regresních parametrů first differencing modelu a fixed modelu s jejich robustními variantami. Jak je vidět, bodové odhady regresních parametrů se nezměnily, jen jejich standardní chyby a tudíž i t statistiky hodnocení statistické významnosti se změnily. Z výsledků lze usoudit, že standardní chyby odhadu byly v původních modelech podhodnoceny. Takže některé proměnné, které byly na hranici statistické významnosti, teď již nejsou statisticky významné na 5% hladině významnosti.

```
> stargazer (firstdiff, fixed,
+           coeftest(firstdiff, vcovHC_firstdiff),
+           coeftest(fixed, vcovHC_fixed),
+            type = "text",
+            model.names = FALSE,
+            keep = c("tax", "l_smok_age", "postcomunistic", "rest", "GDP", "unem", "heduc"),
+            column.labels = c("firstdiff","fixed", "Robust firstdiff","Robust fixed"))


================================================================================================
                                          Dependent variable:
             ---------------------------------------------------------------------------
                                          Y
                       firstdiff                fixed          Robust firstdiff Robust fixed
                          (1)                    (2)                (3)            (4)
------------------------------------------------------------------------------------------------
X2taxation              -3.234*              -12.838***            -3.234       -12.838**
                        (1.829)                (2.736)            (2.331)        (5.799)

X2GDP                   -0.010                 -0.015             -0.010         -0.015
                        (0.009)                (0.015)            (0.007)        (0.024)

X2unem                -29.962***             -45.368***         -29.962***     -45.368***
                        (4.623)                (5.893)            (3.937)        (13.256)

X2heduc                 -44.536**              -56.739***         -44.536*       -56.739**
                        (17.270)               (9.092)            (25.734)       (23.141)


------------------------------------------------------------------------------------------------

Observations              241                    266
```

```
R2                    0.250                    0.605
Adjusted R2           0.245                    0.539

F Statistic  19.650*** (df = 4; 236) 90.921*** (df = 4; 237)
==========================================================================================
Note:                                                           *p<0.1; **p<0.05; ***p<0.01
```

Tabulka 10: Porovnání výsledků FD a FE modelu s robustními odhady, vlastní zpracování, výstup z programu R

Taktéž v random a correlated random modelu byly standardní chyby odhadu podhodnoceny, rozdíl výsledků je markantní.

```
> stargazer (random, cre,
+           coeftest(random, vcovHC_random),
+           coeftest(cre, vcovHC_cre),
+           type = "text",
+           model.names = FALSE,
+           keep = c("tax", "l_smok_age", "postcomunistic", "rest", "GDP", "unem", "heduc"),
+           column.labels = c("random","CRE", "Robust random","Robust CRE"))

==========================================================================================================
                                                    Dependent variable:
                              ----------------------------------------------------------------------------
                                                            Y
                                random                  CRE                 Robust random Robust CRE
                                  (1)                   (2)                      (3)           (4)
----------------------------------------------------------------------------------------------------------
X2taxation                      -9.917***             -10.668***               -9.917       -10.668*
                                (2.701)                (2.820)                 (6.333)       (6.324)

X2GDP                            -0.015                -0.001                  -0.015        -0.001
                                (0.009)                (0.015)                 (0.014)       (0.023)

X2unem                          -41.845***            -39.472***              -41.845***    -39.472***
                                (5.200)                (6.169)                (13.549)      (13.551)

X2heduc                         -43.354***            -45.896***              -43.354***    -45.896*
                                (7.987)                (9.886)                (15.282)      (23.396)


factor(postcomunistic)posc      -239.074              -19.095                 -239.074      -19.095
                               (243.065)             (286.225)               (398.212)     (398.799)


factor(rest)strict             -126.405***           -139.514***              -126.405      -139.514
                                (46.261)              (48.386)                (95.645)      (99.252)

factor(rest)not-strict         -123.627***           -123.889***              -123.627      -123.889
                                (44.110)              (47.283)                (77.692)      (82.518)


Between(taxation, na.rm = TRUE)                       19.239                                19.239
                                                     (23.373)                              (23.551)

Between(GDP)                                          -0.003                                -0.003
                                                      (0.020)                               (0.031)

Between(unem)                                         67.990**                              67.990*
                                                     (28.701)                              (39.595)

Between(heduc)                                         3.588                                 3.588
                                                     (15.620)                              (26.729)
```

```
-----------------------------------------------------------------------------------------------------
Observations                                266                    266
R2                                          0.606                  0.614
Adjusted R2                                 0.587                  0.587
F Statistic                       56.601*** (df = 7; 258) 36.791*** (df = 11; 254)
=====================================================================================================
Note:                                                                 *p<0.1; **p<0.05; ***p<0.01
```

Tabulka 11: Porovnání výsledků RE a CRE modelu s robustními odhady, vlastní zpracování, výstup z programu R

3.5 Porovnání modelů a výběr nejvhodnějšího modelu pro daná data

K porovnání vhodnosti modelů existují určité testy porovnávající odhady regresních parametrů a vyhodnocují jako vhodnější model ten model, který má přesnější odhady, ve smyslu menšího rozptylu, vydatnosti odhadu a podobně. Neexistuje jediný test, který by dokázal jednoznačně říct jaký model je nejvhodnější, ale existují určité testy porovnávající většinou dva modely. Postupně tedy lze vyhodnotit, jaký model je pro daná data nejvhodnější.

Nejprve bych vyhodnotila, zda máme s daty pracovat jako s panelovými daty nebo zda je možné použít pooling OLS odhad a nerespektovat tak strukturu panelových dat.

Pooling vs. Fixed

pFtest dostupný z package plm porovnává výsledky dvou plm objektů, a to fixed effect modelu a pooling OLS modelu, zda existuje v datech nepozorovaný efekt. Nulová hypotéza říká, že je vhodnější použití pooling OLS modelu, alternativní zase říká, že v datech je signifikantní nepozorovaný efekt.

```
> ###########  LM test for FE vs OLS
> pFtest(fixed, pooling)

        F test for individual effects

data:  Y ~ X2
F = 38.604, df1 = 21, df2 = 237, p-value < 2.2e-16

alternative hypothesis: significant effects
> # pro fixed effect model

> # robustní test
> pFtest(fixed, pooling, vcov = c(vcovHC_fix, vcovHC_pooling))
```

```
        F test for individual effects

data:  Y ~ X2
F = 38.604, df1 = 21, df2 = 237, p-value < 2.2e-16

alternative hypothesis: significant effects
> # FE je lepší
```

Tabulka 12: FE vs. pooling OLS, vlastní zpracování, výstup z programu R

V tabulce číslo 12 jsou zobrazeny výsledky testů jak s původními modely, tak i robustními odhady kovarianční matice. Z obou testů vyplývá, že je vhodnější použití fixed effect modelu.

Pooling vs. Random

Lagrange Multiplier test individuálních a/nebo časových efektů, dokáže otestovat, zda jsou v datech obsažené časové efekty, individuální efekty nebo oba efekty. Já použiji testování přítomnosti individuálního efektu a vyhodnotím, zda je vhodnější pooling OLS model či random effect model.

```
> ########### LM test for RE vs OLS
> plmtest(pooling)

        Lagrange Multiplier Test - (Honda)

data:  Y ~ X
normal = 27.164, p-value < 2.2e-16

alternative hypothesis: significant effects
> # pro random effect

>
> ## robustní test
> plmtest(pooling, vcov = vcovHC_pooling)

        Lagrange Multiplier Test - (Honda)

data:  Y ~ X
normal = 27.164, p-value < 2.2e-16

alternative hypothesis: significant effects
> # pro random effect
```

Tabulka 13: RE vs. pooling OLS, vlastní zpracování, výstup z programu R

Z obou testů pro robustní odhady kovarianční matice i pro původní modely jsou testy jednoznačné, označují přítomnost individuálního efektu. Proto je vhodnější použít random effect model více než pooled OLS model.

Fixed vs. Random

Hausman test testuje, jaký z RE a FE modelu je efektivnější, tj. který model poskytuje odhady s menším rozptylem, za předpokladu, že oba dávají konzistentní odhady. Tento test porovnává odhady parametrů a zjišťuje, jestli se liší. Nulová hypotéza říká, že RE a FE odhady parametrů jsou si natolik podobné, že je prakticky jedno, jaký model použiji. Naproti tomu alternativní hypotéza říká, že hlavní RE předpoklady neplatí, je vhodnější použít FE test (Wooldridge, 2013, cap. 14.2). Hausman test je dostupný v *package plm* pod funkcí *phtest(fixedeffect, randomeffect)* (HEISS, c2016, cap. 14.2).

```
> ########### Hausman test for FE vs RE model
> phtest(random, fixed) # p>alfa... odhady obou modelů jsou si natolik podobné, že je jedno
 jaký použijeme, za přepokladu konzistentních odhadů

        Hausman Test

data:  Y ~ X2 + factor(postcomunistic) + factor(rest)
chisq = 6.3114, df = 4, p-value = 0.1771

alternative hypothesis: one model is inconsistent
>
> ## robustní test
> phtest(random, fixed, vcov = c(vcovHC_fix, vcovHC_rand))

        Hausman Test

data:  Y ~ X2 + factor(postcomunistic) + factor(rest)
chisq = 6.3114, df = 4, p-value = 0.1771

alternative hypothesis: one model is inconsistent
> # odhady modelů jsou si natolik podobné, že je jedno jaký použijeme
```

Tabulka 14: RE vs. FE model, vlastní zpracování, výstup z programu R

Na 5% hladině významnosti se nepodařilo zamítnout nulovou hypotézu. Odhady obou modelů jsou si natolik podobné, že je prakticky jedno jaký z nich použiji.

Random vs. CRE

Jak již bylo řečeno v kapitole 2.2 correlated random effect model podává stejné výsledky jako FE model, ale má jednu velkou výhodu. Mohu do něho zařadit vysvětlující proměnné, které se nemění v čase nebo mají velmi malou variabilitu. Proto je vhodnější použití CRE modelu.

Rozdíl mezi correlated random effect a random effect modelem je jen ten, že v CRE modelu jsou obsaženy navíc proměnné střední hodnoty časově proměnlivých vysvětlujících proměnných pro všechna i. Random effect model předpokládá, že tato složka je rovna nule, zatímco CRE předpokládá, že je různá od nuly. A protože CRE model podává standardní chyby odhadu u těchto středních hodnot, mohu tuto hypotézu velice snadno otestovat. Pokud by byla přítomna jen jedna v čase proměnlivá proměnná, stačil by jednoduchý t test. Jelikož je více těchto proměnných, musím použít F test, který bude testovat, zda jsou střední hodnoty použité v analýze rovny nule či nikoliv.

```
> ########### Test for CRE vs RE
> #  H0: random je lepší
>
> linearHypothesis(cre, matchCoefs(cre, "Between"))
Linear hypothesis test

Hypothesis:
Between(taxation, na.rm = TRUE) = 0
Between(GDP) = 0
Between(unem) = 0
Between(heduc) = 0

Model 1: restricted model
Model 2: Y ~ X2 + factor(postcomunistic) + factor(rest) + +Between(taxation,
    na.rm = TRUE) + Between(GDP) + Between(unem) + Between(heduc)

  Res.Df Df  Chisq Pr(>Chisq)
1    258
2    254  4 9.9071    0.04202 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # 0.04<alfa...H1... CRE je lepší
>
```

```
> ##robustní test
> linearHypothesis(cre, matchCoefs(cre, "Between"), vcov. = vcovHC_cre)
Linear hypothesis test

Hypothesis:
Between(taxation, na.rm = TRUE) = 0
Between(GDP) = 0
Between(unem) = 0
Between(heduc) = 0

Model 1: restricted model
Model 2: Y ~ X2 + factor(postcomunistic) + factor(rest) + +Between(taxation,
    na.rm = TRUE) + Between(GDP) + Between(unem) + Between(heduc)

Note: Coefficient covariance matrix supplied.


  Res.Df Df  Chisq Pr(>Chisq)
1    258
2    254  4 4.3133     0.3653
> # 0,36 > alfa … random je lepší
```

Tabulka 15: CRE vs. RE model, vlastní zpracování, výstup z programu R

Zatímco bez použití robustních odhadů kovarianční matice lze předpokládat, že je vhodnější použít CRE model. Robustní test, který používá správné hodnoty standardních chyb odhadů, se přiklání k nulové hypotéze. Na 5% hladině významnosti se nepodařilo zamítnout nulovou hypotézu o statistické nevýznamnosti regresních parametrů reprezentujících střední hodnotu v čase proměnlivých proměnných pro všechna *i*. Na základě vyhodnocení tohoto testu je vhodnějším modelem random effect model.

3.6 Výsledný model

Z předešlé analýzy vyplývá, že konečným modelem k interpretaci je random effect model s robustními odhady parametrů. Tento model předpokládá, že nepozorovaný efekt a_i, konstantní v čase a různý pro jednotlivé státy, je nekorelovaný s vysvětlujícími proměnnými. Tím odpadá problém jeho specifikace a odstranění jeho vlivu z modelu. Nepozorovaný efekt tak vstupuje do složené náhodné složky a způsobuje tak její sériovou korelaci.

V následující tabulce jsou zobrazeny odhady regresních parametrů tohoto modelu. Pro ukázku jsou v tabulce zahrnuty i odhady původního modelu bez robustních odhadů korelační matice.

```
> ###################################################################
> #             Výsledný model
> ###################################################################
> stargazer (random, coeftest(random, vcovHC_random),
+            type = "text",
+            model.names = FALSE,
+            keep = c("tax", "l_smok_age", "postcomunistic",
+                     "rest", "GDP", "unem", "heduc"),
+            column.labels = c("random","Robust random"))

===================================================================
                                    Dependent variable:
                               -----------------------------------
                                             Y
                                    random          Robust random
                                      (1)                (2)
-------------------------------------------------------------------
X2taxation                          -9.917***          -9.917
                                     (2.701)            (6.333)

X2GDP                                -0.015             -0.015
                                     (0.009)            (0.014)

X2unem                              -41.845***         -41.845***
                                     (5.200)            (13.549)

X2heduc                             -43.354***         -43.354***
                                     (7.987)            (15.282)

factor(postcomunistic)posc           -239.074           -239.074
                                     (243.065)          (398.212)

factor(rest)strict                  -126.405***         -126.405
                                     (46.261)           (95.645)

factor(rest)not-strict              -123.627***         -123.627
                                     (44.110)           (77.692)

-------------------------------------------------------------------
```

Observations	266
R2	0.606
Adjusted R2	0.587
F Statistic	56.601*** (df = 7; 258)
Note:	*p<0.1; **p<0.05; ***p<0.01

Tabulka 16: Výsledný RE model, vlastní zpracování, výstup z programu R

Odhady regresních parametrů se nezměnily, ale jejich směrodatné chyby, potažmo statistiky ano. Z analýzy vyplývá, že jen dvě vysvětlující proměnné jsou statisticky významné, a to procento nezaměstnanosti a procento vysokoškolsky vzdělaných osob. Nezaměstnanost a vzdělání má dle výsledků analýzy negativní vliv na spotřebu cigaret. Tyto dvě proměnné mají stejné jednotky, proto jsou porovnatelné bez nutnosti přepočtu. Pokud stoupne procento nezaměstnanosti o jeden procentní bod, pak se sníží spotřeba cigaret o přibližně 42 cigaret na osobu za rok ceteris paribus (ostatní faktory jsou neměnné). Obdobně je to i se vzděláním, pokud se zvýší procento vysokoškolsky vzdělaného obyvatelstva o jeden procentní bod, sníží se spotřeba cigaret o přibližně 43 cigaret na osobu za rok, ceteris paribus.

Hlavní sledovaná proměnná, zdanění, vychází jako statisticky nevýznamná. To znamená, že na základě provedené analýzy se mi nepovedlo prokázat lineární závislost střední hodnoty spotřeby cigaret na zdanění. To není tak nečekaný výsledek. Cenová elasticita poptávky po cigaretách bývá spíše neelastická. Zdanění v průměru představuje 77% část ceny cigaret. Taktéž se neprokázal vliv zavedení restriktivních opatření na kouření v restauracích.

Z analýzy vyplývá, že postkomunistické státy nevykazují statisticky významný rozdíl ve spotřebě cigaret ve srovnání s ostatními evropskými státy. Určitý rozdíl zde může být, ovšem není tak významný, aby překračoval hodnoty normálních odchylek.

Závěr

Cílem diplomové práce bylo zodpovědět otázku, zda je daňová politika u cigaret v kontextu zemí EU účinná. Zabývala jsem se myšlenkou, zda vyšší zdanění, a s tím spojená vyšší cena cigaret, dokáže odradit kuřáky od kouření.

Problematice spotřeby cigaret jsem se věnovala v první kapitole. Z různých zdrojů jsem čerpala informace o faktorech, které ovlivňují spotřebu cigaret. Dle The Tobacco Atlas (ERIKSEN et al. 2015) je vyšší spotřeba tabákových výrobků spojena s nižším socioekonomickým statusem, nižším vzděláním i nižší příjmovou třídou. Zabývala jsem se cenovou elasticitou poptávky po cigaretách. Dlouhodobá elasticita se v průměru pohybuje okolo - 0,44 (MLČOCH a DOLEŽAL, 2014). Je zajímavé, že cenová elasticita je různá pro různé věkové kategorie, přičemž nejvíce elastická je pro skupinu adolescentů. Evropská unie se zavázala přijmout doporučení Světové zdravotnické organizace ohledně zajištění nekuřáckého prostředí ve členských státech. Na základě těchto doporučení existuje ve všech členských státech určitá forma zákazu reklamy na tabákové výrobky a zákaz kouření ve veřejných prostorech a veřejných dopravních prostředcích. Všechny členské státy také přijmuly zákony buď úplně zakazující, nebo omezující kouření v restauracích a barech.

V teoretické části diplomové práce, v druhé kapitole, jsem shrnula potřebné teoretické znalosti pro provedení analýzy. Nejprve jsem musela určit, jaké proměnné v analýze použiji. Jako vysvětlovanou proměnnou jsem vybrala spotřebu cigaret na obyvatele za rok. Vysvětlujícími proměnnými se pak staly takové proměnné, u kterých jsem předpokládala, že ovlivňují spotřebu cigaret. Těmito proměnnými bylo zdanění, HDP jako proxy proměnná pro blahobyt, nezaměstnanost, vzdělání, legální věk pro koupi cigaret za jednotlivé státy v jednotlivých letech a dummy proměnné pro postkomunistické země a zákaz kouření v restauracích. Při výběru vysvětlujících proměnných jsem vycházela z různých vědeckých článků, statistik a výsledků analýz. Zajímalo mě porovnání spotřeby cigaret v postkomunistických zemích v rámci Evropské unie, proto jsem v analýze použila i tuto proměnnou.

Dále jsem v druhé kapitole shrnula metody analýzy s panelovými daty. Pro daná data se nabízelo hned několik modelů. Předem nebylo jasné, jaký model by byl nejvhodnější.

Předpokládala jsem, že výsledným modelem bude fixed effect model, který předpokládá korelaci mezi vysvětlovanými proměnnými a nepozorovaným efektem a_i. Nepozorovaný efekt, jak už název napovídá, je efekt, který ovlivňuje spotřebu, ale není známý. Může to být souhrn několika efektů dohromady, který je různý pro jednotlivé státy, ale konstantní v čase. Nepozorovaným efektem na mnou zkoumaných datech může být například predispozice obyvatelstva ke kouření, vnímání kuřáků, historický kontext, společenská přípustnost kouření a další faktory. Fixed effect model má ovšem jednu nevýhodu, nedokáže pracovat s proměnnými, které jsou v čase neměnné nebo mají jen velmi malou variabilitu okolo průměru. Jelikož jsem vybrala i proměnné, které nejsou dostatečně variabilní, a to dummy proměnné pro postkomunistické země a zákaz kouření v restauracích nebo málo variabilní proměnnou legálního věku pro koupi cigaret, potřebovala jsem použít jiný model, který umožňuje zahrnutí těchto proměnných. Tím se stal model Correlated random effect, který podává stejné výsledky jako fixed effect model a mohou se do něj zahrnout i nedostatečně variabilní proměnné.

V praktické části diplomové práce, v třetí kapitole, jsem sestavila ekonometrické modely regresní analýzy na panelových datech, které jsem popsala v druhé kapitole. Jelikož nebylo možné předem rozhodnout jaký model je nejvhodnější, musela jsem vyzkoušet všechny modely a provést jejich diagnostiku. Diagnostikou náhodné složky jsem zjistila, že data vykazují sériovou korelaci a heteroskedasticitu. Proto bylo nutné provést robustní odhady korelační matice a přepočítat standardní chyby odhadů regresních parametrů.

Po získání správných standardních chyb odhadů jsem mohla přikročit k poslednímu kroku analýzy, a to zjistit jaký model je nejvhodnější. Na základě různých testů hypotéz jsem mohla postupně vyřadit modely, které nebyly vhodné. Neočekávala jsem, že výsledným modelem bude random effect model. Random effect model předpokládá, že náhodná složka a_i je nekorelovaná s vysvětlovanými proměnnými. Tento model však umožňuje použití málo variabilních proměnných v analýze. Takže všechny proměnné mohly být použity.

Jen dvě vysvětlující proměnné vyšly jako statisticky významné. Těmito proměnnými je vzdělání a nezaměstnanost. Obě proměnné mají pozitivní efekt na snižování spotřeby cigaret. Z analýzy vyplývá, že zvýšení procenta nezaměstnaných o jeden procentní bod sníží spotřebu cigaret o přibližně 42 cigaret na osobu za rok, ceteris paribus (při fixních

ostatních faktorech). Zvýšení počtu vysokoškolsky vzdělaného obyvatelstva snižuje spotřebu cigaret o přibližně 43 cigaret na osobu za rok, ceteris paribus. Tato čísla se nezdají být vysoká, ovšem nesmíme zapomenout, že analýza je založená na vlivu proměnných na vysvětlovanou proměnnou, kterou je počet spotřebovaných cigaret na obyvatelstvo za rok a ne všichni občané země kouří.

Původně jsem měla v úmyslu porovnat v závěru analýzy regresní koeficienty odhadů parametrů pomocí beta koeficientů nebo normovaných veličin regresorů. Jelikož jediné dva statisticky významné regresní koeficienty mají stejné jednotky, není tato transformace pro porovnatelnost vlivů nutná.

Ostatní proměnné vycházejí jako statisticky nevýznamné. Statisticky nevýznamnými proměnnými jsou HDP jako proxy proměnná pro blahobyt, legální věk pro koupi cigaret a dummy proměnné pro zákaz kouření v restauracích a postkomunistické země. To znamená, že se mi nepodařilo prokázat lineární závislost střední hodnoty spotřeby cigaret na těchto proměnných.

Statistická nevýznamnost zdanění není tak překvapivá, protože zdanění tvoří v průměru 77 % konečné ceny cigaret. Cenová elasticita poptávky po cigaretách je spíše neelastická. Z toho vyplývá, že vliv zdanění na spotřebu cigaret je velmi malý až mizivý.

Také vliv proměnné reprezentující zákaz kouření v restauracích vychází statisticky nevýznamný. Výsledky vedou k závěru, že tento zákaz neodrazuje kuřáky od kouření. Hlavním cílem zavedení zákazu kouření v restauracích, barech a veřejných místech nebo ve veřejné dopravě je uchránit obyvatelstvo od pasivního kouření.

Dummy proměnnou pro postkomunistické země jsem se rozhodla zařadit do analýzy, protože mě zajímalo, jestli se prokáže rozdíl ve spotřebě cigaret mezi postkomunistickými zeměmi Evropské unie a ostatními státy. Tento rozdíl se dle výsledků analýzy nepodařilo prokázat.

Dle The Tabbaco Atlas (ERIKSEN et al. 2015) se ve vyspělých státech snižuje počet kuřáků. Proto jsem se rozhodla použít proměnnou HDP jako proxy proměnnou pro blahobyt. Tato proměnná však nevychází jako statisticky významná. Z analýzy tak vyplývá, že blahobyt neovlivňuje spotřebu cigaret. Je však možné, že je to způsobeno tím, že jsem

nevybrala vhodnou proxy proměnnou pro blahobyt. Tento problém by bylo zajímavé rozvést v další analýze.

Výsledky analýzy se přiklánějí k názoru, že zvyšování spotřebních daní na tabákové výrobky neodradí kuřáky od kouření. Daňový výnos ze spotřebních daní bývá velice stabilní. Pokud má navíc produkt, na nějž je daň uvalena, stálou poptávku, je tento produkt naprosto ideální pro stabilizování státního rozpočtu. Nabízí se úvaha, že možným hlavním cílem může být fiskální náplň této daně. Jinak řečeno spotřební daň na tabákové výrobky má jako hlavní úkol naplnit státní rozpočet. Ale to je sporné, protože celkové náklady kouření (údaje za rok 2009 za ČR) jsou skoro čtyřikrát větší než je výběr daní z cigaret (KRÁLÍKOVÁ, 2011).

Spotřeba cigaret a samotné kouření je způsobeno mnoha faktory a většinou jde o faktory ovlivňující chování spotřebitelů. Proto by mohla být efektivnější analýza na panelových datech za spotřebitele. U každého jedince (obyvatele) by byly sledovány určité znaky jako vysvětlující proměnné a ty by se poté vyhodnocovaly. Taková data však není snadné shromáždit, protože vyžadují rozsáhlý dotazníkový výzkum v průběhu několika let, přičemž subjekty musí být ve všech časových okamžicích stejné.

Cílem diplomové práce bylo zodpovědět otázku, zda je daňová politika u cigaret v kontextu zemí EU účinná. Účinnost daňových politik členských států se na základě analýzy provedené v mé diplomové práci nepodařila prokázat. Z tohoto důvodu se domnívám, že daňová politika zemí Evropské unie nedostatečně motivuje chování spotřebitele.

Zdroje

- CELNÍ SPRÁVA, 2016. Tabákové výrobky - přehled sazeb spotřební daně [online]. [cit. 2016-08-25]. Dostupné z: https://www.celnisprava.cz/cz/dane/spotrebni-dane/tabak/Informace/Informace_16_10797_priloha.pdf
- CROISSANT, Yves, a MILLO, Giovanni, 2008. Panel Data Econometrics in R: The plm Package. In: Journal of Statistical Software [online]. 27(2), s. 43 [cit. 2016-08-16]. DOI: 10.18637/jss.v027.i02. ISSN 1548-7660. Dostupné z: https://www.jstatsoft.org/article/view/v027i02
- CZUBEK, Magdalena a JOHAL, Surjinder, 2010. Econometric Analysis of Cigarette Consumption in the UK. HM Revenue & Customs [online]. (HMRC Working Paper Number 9), 48 [cit. 2016-07-21]. Dostupné z: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331580/cig-consumption-uk.pdf
- DECKER, S.,L., a SCHWARTZ, E., A, 2000. Cigarettes and alcohol: substitutes or complements? [online] NBER working paper n. 7535,[cit. 2016-07-23] Dostupné z: http://www.nber.org/papers/w7535
- ERIKSEN, Michael P., Judith MACKAY, Neil W. SCHLUGER, Farhad ISLAMI a Jeffrey DROPE, 2015. The Tobacco Atlas [online]. Fifth edition. Atlanta, Georgia, 30303, USA: Published by the American Cancer Society, [cit. 2016-07-14]. ISBN 16-044-3235-7. Dostupné z: http://3pk43x313ggr4cy0lh3tctjh.wpengine.netdna-cdn.com/wp-content/uploads/2015/03/TA5_2015_WEB.pdf
- EU cigarette prices: Cigarette prices across Europe, 2015. tma [online]. [cit. 2016-08-25]. Dostupné z: http://www.the-tma.org.uk/tma-publications-research/facts-figures/eu-cigarette-prices/
- EUROPEAN COMMISSION, 2013. Report on the implementation of the Council Recommendation of 30 November 2009 on Smoke-free Environments (2009/C 296/02). In: COMMISSION STAFF WORKING DOCUMENT [online]. Brussels, s. 21 [cit. 2016-07-23]. Dostupné z: http://ec.europa.eu/health/tobacco/docs/smoke-free_implementation_report_en.pdf
- EUROPEAN COMMISSION, 2016a. Kouření: Politika. Ec.europa.eu [online]. [cit. 2016-08-24]. Dostupné z: http://ec.europa.eu/health/tobacco/policy/index_cs.htm

- EUROPEAN COMMISSION, 2016b. Kouření: Nekuřácké prostředí. Ec.europa.eu [online]. [cit. 2016-08-24]. Dostupné z: http://ec.europa.eu/health/tobacco/smoke-free_environments/index_cs.htm
- HEISS, Florian, c2016. Using R for Introductory Econometrics. Germany: CreateSpace. [cit. 2016-08-17]. ISBN 978-1-523-28513-6. ISBN-10: 1-523-28513-3. Dostupné z: http://www.urfie.net/index.html
- How to interpret a QQ plot, 2014. In: Cross Validated [online]. [cit. 2016-08-20]. Dostupné z: http://stats.stackexchange.com/questions/101274/how-to-interpret-a-qq-plot
- KATCHOVA, Ani L., 2013. Panel Data Models in R [Software Program and YouTube Video]. In : Econometrics Academy Website: https://sites.google.com/site/econometricsacademy/, [online], [cit. 2016-08-16]. Dostupné z: https://sites.google.com/site/econometricsacademy/econometrics-models/panel-data-models
- KRÁLÍKOVÁ, Eva, 2011. Kouření škodí nejen zdraví, ale i ekonomice (státu i občanů). TEMPUS MEDICORUM: časopis české lékařské komory. 2011(9) [online]. [cit. 2016-08-24], str. 15. ISSN 1214-7524. Dostupné z: file:///C:/Users/kudla/Downloads/tempus_11_09.pdf
- KURZYCZ, 2015. GBP britská libra, od 31.12.2015 do 2.1.2015, historie kurzů měn. Kurzy.cz [online]. [cit. 2016-08-25]. Dostupné z: http://www.kurzy.cz/kurzy-men/kurzy.asp?A=H&KM=GBP&D1=2.1.2015&D2=31.12.2015&I=1
- MLČOCH, Tomáš, DOLEŽAL, Tomáš, 2014. Jaký je vztah mezi cenou cigaret a jejich spotřebou? Institut pro zdravotní ekonomiku a technology assessment [online] [cit. 2016-07-20]. Dostupné z: http://www.iheta.org/ext/publication/files/vzath%20mezi%20cenou%20c.%20a%20jejich%20spot%C5%99ebou_final.pdf
- SHLEIFER, Andrei, TREISMAN Daniel, 2014. Normal Countries: The East 25 Years After Communism [online]. [cit. 2016-08-09]. Dostupné z: https://scholar.harvard.edu/files/shleifer/files/normal_countries_draft_sept_12_annotated.pdf. Vědecký článek pro Foreign Affairs, plná verze. Harvardova universita.
- STATISTICA, c2015. Proportional consumption of cigarettes by world region as of 2013. Statistica.com [online]. [cit. 2015-10-03]. Dostupné z: http://www.statista.com/statistics/279570/cigarette-consumption-around-the-world/
- TORRES-REYNA, Oscar, 2010. Getting Started in Fixed/Random Effects Models using R. In: Princeton university: Data & Statistical services [online]. [cit. 2016-08-20]. Dostupné z: http://www.princeton.edu/~otorres/Panel101R.pdf

- vlastní poznámky k předmětu Regrese LS 2014/2015, ident: 4ST416, garant a přednášející: Mgr. Milan Bašta, Ph.D., přednášky na VŠE
- vlastní poznámky k předmětu Úvod do Úvod do ekonometrie LS 2014/2015, ident: 4EK214, garant a přednášející: doc. Ing. Tomáš Cahlík, CSc., přednáky na VŠE
- WOOLDRIDGE, Jeffrey M, 2013. Introductory econometrics: a modern approach. 5th ed. Mason, OH: South-Western Cengage Learning, c2013. ISBN 978-111-1531-041

Smoking bans:

- EUROPEAN COMMISSION, 2013. Report on the implementation of the Council Recommendation of 30 November 2009 on Smoke-free Environments (2009/C 296/02). In: COMMISSION STAFF WORKING DOCUMENT [online]. Brussels, s. 21 [cit. 2016-07-23]. Dostupné z: http://ec.europa.eu/health/tobacco/docs/smoke-free_implementation_report_en.pdf
- List of smoking bans wordwide, 2009. Smoking Ban Worldwide - In which countries is smoking prohibited? [online]. Germany, [cit. 2016-07-23]. Dostupné z: http://en.rauchverbotweltweit.de/smokingbans-worldwide.php
- List of smoking bans, 2001-a. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, [cit. 2016-08-08]. Dostupné z: https://en.wikipedia.org/wiki/List_of_smoking_bans
- Zákaz kouření podle zemí, 2001-b. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, [cit. 2016-08-08]. Dostupné z: https://cs.wikipedia.org/wiki/Z%C3%A1kaz_kou%C5%99en%C3%AD_podle_zem%C3%AD

Databáze Passport:

- EUROMONITOR INTERNATIONAL, c2016a. Market Sizes, Retail Volume, sticks Per Capita [statistická data]. In: Passport [online]. [cit. 2016-07-20]. Dostupné z: http://www.portal.euromonitor.com
- EUROMONITOR INTERNATIONAL, c2016b. Calculation Variables [statistická data]. In: Passport [online]. [cit. 2016-07-20]. Dostupné z: http://www.portal.euromonitor.com
- EUROMONITOR INTERNATIONAL, c2016c. Smoking prevalence - % of adult population [statistická data]. In: Passport [online]. [vid. 2016-07-20]. Dostupné z: http://www.portal.euromonitor.com
- EUROMONITOR INTERNATIONAL, c2016d. GDP € Per Capita, Constant 2015 Prices, Fixed 2015 Exchange Rates [statistická data]. In: Passport [online]. [cit. 2016-08-09]. Dostupné z: http://www.portal.euromonitor.com

- EUROMONITOR INTERNATIONAL, c2016e. % of Population Aged 15+ with Higher Education [statistická data]. In: Passport [online]. [cit. 2016-08-09]. Dostupné z: http://www.portal.euromonitor.com
- EUROMONITOR INTERNATIONAL, c2016f. Unemployment Rate, % of economically active population [statistická data]. In: Passport [online]. [cit. 2016-08-09]. Dostupné z: http://www.portal.euromonitor.com

Excise duties tables:

- EUROPEAN COMMISSION, 2003a. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], July 2003, [cit. 2016-05-08]. Dostupné z: http://dielebensmittel.at/Dokumente/steuern/uebersicht-moel.pdf
- EUROPEAN COMMISSION, 2003b. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], April 2003, [cit. 2016-05-08]. Dostupné z: http://aei.pitt.edu/38716/1/A3645.pdf
- EUROPEAN COMMISSION, 2005. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], January 2005, [cit. 2016-05-08]. Dostupné z: http://ec.europa.eu/taxation_customs/resources/documents/excise_duties-part_III_tobacco-en.pdf
- EUROPEAN COMMISSION, 2006. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], July 2006, [cit. 2016-05-08]. Dostupné z: http://apps.who.int/fctc/reporting/fiscal_policy.pdf
- EUROPEAN COMMISSION, 2007. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], January 2007, [cit. 2016-05-08]. Dostupné z: http://www.pedz.uni-mannheim.de/daten/edz-h/gdz/07/excise_duties-part_II_energy_products-en.pdf
- EUROPEAN COMMISSION, 2011. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], July 2011, [cit. 2016-05-08]. Dostupné z: http://www.alliantienederlandrookvrij.nl/wp-content/uploads/2012/docs/accijns/2011-06%20excise_duties-part_iii_tobacco_en.pdf
- EUROPEAN COMMISSION, 2014. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], July 2014, [cit. 2016-05-08]. Dostupné z: https://circabc.europa.eu/sd/a/7daee41d-f33d-420e-96ac-e01020693795/III-Tobacco_July2014%20final.pdf
- EUROPEAN COMMISSION, 2015. Excise Duty Tables, Part III – Manufactured Tobbaco, [online], January 2015, [cit. 2016-05-08]. Dostupné z: http://ec.europa.eu/taxation_customs/resources/documents/excise_duties-part_III_tobacco-en.pdf
- General Excise Duties: Tobacco and Alcohol Product, 2009. In: Tax Strategy Group Papers [online]. Tax Policy Unit, s. 11 [cit. 2016-08-25]. Dostupné z: http://taxpolicy.gov.ie/wp-content/uploads/2011/03/09.19b.pdf

- General Excise Duties: Tobacco and Alcohol Product, 2010. In: Tax Strategy Group Papers [online]. Tax Policy Unit, s. 11 [cit. 2016-08-25]. Dostupné z: http://www.finance.gov.ie/sites/default/files/10.21-General-Excise-Duties.pdf
- General Excise Duties: Tobacco and Alcohol Product, 2012. In: Tax Strategy Group Papers [online]. Tax Policy Unit, s. 13 [cit. 2016-08-25]. Dostupné z: http://taxpolicy.gov.ie/wp-content/uploads/2013/07/12-20-General-Excise-Duties.pdf
- General Excise Duties: Tobacco and Alcohol Product, 2013. In: Tax Strategy Group Papers [online]. Tax Policy Unit, s. 14 [cit. 2016-08-25]. Dostupné z: http://www.finance.gov.ie/sites/default/files/TSG%201302.pdf

Seznam obrázků

Seznam tabulek

Seznam příloh