

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

FAKULTA INFORMATIKY A STATISTIKY



BAKALÁŘSKÁ PRÁCE

2017

Vojtěch Sukup

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

FAKULTA INFORMATIKY A STATISTIKY



Název bakalářské práce:

**Ekonometrická analýza závislosti spotřeby piva
na protikuřáckém zákonu**

Autor:	Vojtěch Sukup
Katedra:	Katedra ekonometrie
Obor:	Matematické metody v ekonomii
Vedoucí práce:	Ing. Jan Zouhar, Ph.D.

Prohlášení:

Prohlašuji, že jsem bakalářskou práci na téma „Ekonometrická analýza závislosti spotřeby piva na protikuřáckém zákonu“ zpracoval samostatně. Veškerou použitou literaturu a další podkladové materiály uvádím v seznamu použité literatury.

V Praze dne 29. května 2017

.....

Vojtěch Sukup

Poděkování:

Rád bych na tomto místě poděkoval Ing. Jan Zouhar, Ph.D. za vedení mé bakalářské práce a za podnětné návrhy, které ji obohatily.

Abstrakt

Název práce: Ekonometrická analýza závislosti spotřeby piva na protikuřáckém zákonu
Autor: Vojtěch Sukup
Katedra: Katedra ekonometrie
Vedoucí práce: Ing. Jan Zouhar, Ph.D.

Cílem této bakalářské práce je zjistit, zdali je nějaká závislost mezi spotřebou piva a protikuřáckým zákonem ve 41 zemích Evropy a přilehlých oblastech v časovém období od roku 1998 do roku 2012 pomocí ekonometrické analýzy. Práce je rozdělena na teoretickou část s metodologií a na praktickou část. Empirická část je zpracována pomocí softwarového programu STATA od společnosti StataCorp. Vzhledem k povaze dat je třeba na ně nahlížet jako na data panelová. Ty se liší od dat průřezových i časových řad. V programu STATA jsou k práci s panelovými daty používané modely OLS, GLS, LDSV, GEE a další. V závěrečné fázi práce zde vybírám nejlepší model a interpretuji jeho hodnoty. Předpokládaný cíl práce byl dosažen, protože měl protikuřácký zákon ve všech modelech negativní efekt na spotřebu piva.

Klíčová slova: pivo, protikuřácký zákon, STATA

Abstract

Title: Econometric analysis of dependence beer's consumption on smoke-free law
Author: Vojtěch Sukup
Department: Department of Econometrics
Supervisor: Ing. Jan Zouhar, Ph.D.

The aim of this bachelor thesis is to find out whether there is some dependence between beer consumption and smoke-free law in 41 countries of Europe and adjacent areas in the period from 1998 to 2012 using econometric analysis. The thesis is divided into theoretical part with methodology and practical part. The empirical part is processed by the StataCorp software program STATA. Due to the nature of the database, it is necessary work with them as panel database. These are different from cross-section and time series data. In the STATA program, models OLS, GLS, LDSV, GEE and more, are designed to work with panel data. In the final phase of my work I choose the best model and interpret its values. The intended objective of the work was achieved because the smoke-free law in all models had a negative effect on the consumption of beer.

Keywords: beer, smoke-free law, STATA

Obsah

ÚVOD	- 0 -
1 VYMEZENÍ PROBLÉMU PROTIKUŘÁCKÉHO ZÁKONU V EU A OSTATNÍCH ZEMÍCH	- 1 -
2 DATA A METODY.....	- 3 -
2.1 POUŽITÁ DATA A SOFTWARE STATA	- 3 -
2.1.1 Deskriptivní statistika dat.....	- 3 -
2.2 VYSVĚTLUJÍCÍ A VYSVĚTLOVANÉ PROMĚNNÉ.....	- 6 -
2.3 METODOLOGIE.....	- 7 -
2.3.1 Ekonometrické pojmy	- 7 -
2.3.2 Teoretické okénko	- 8 -
3 VÝSLEDKY	- 10 -
3.1 MODEL Y ODHADNUTÉ MNČ A ZNČ	- 10 -
3.1.1 Odhady pořízené MNČ pomocí funkce regress	- 10 -
3.1.2 Modely GLS a GEE	- 12 -
3.1.3 Modely s fixními a náhodnými efekty.....	- 15 -
3.1.4 Ostatní modely	- 18 -
3.1.5 Porovnání vybraných modelů.....	- 20 -
3.2 MODEL Y S TRENDEM.....	- 23 -
4 DISKUZE.....	- 26 -
ZÁVĚR.....	- 27 -
LITERATURA	28

Seznam grafů, obrázků a tabulek

Tabulka 2-1: Vysvětlovaná a vysvětlující proměnné v datovém souboru.....	- 3 -
Obrázek 3-1: Deskriptivní statistika dat.....	- 4 -
Obrázek 3-2: Spotřeba piva v Evropě	- 5 -
Obrázek 3-3 Korelační matice.....	- 6 -
Obrázek 3-4 Model OLS pomocí funkce regress bez VCE	- 10 -
Obrázek 3-5 Model OLS pomocí funkce regress s VCE	- 11 -
Obrázek 3-6 Model OLS pomocí funkce regress s prvními diferencemi.....	- 11 -
Obrázek 3-7 Model GLS pomocí funkce xtreg	- 12 -
Obrázek 3-8 Model GLS pomocí funkce xtreg se zpožděními a VCE (robust).....	- 13 -
Obrázek 3-9 Model GEE pomocí funkce xtgee s robustní VCE.....	- 15 -

Obrázek 3-10 Hausmanův test na fixní nebo náhodné efekty u modelu GLS	- 15 -
Obrázek 3-11 Hausmanův robustní test na fixní nebo náhodné efekty	- 16 -
Obrázek 3-12 Model GLS pomocí funkce xtreg – fixní vs náhodné efekty.....	- 17 -
Obrázek 3-13 Model funkce areg.....	- 18 -
Obrázek 3-14 Metoda Bootstrap (10000 opakování) – funkce xtreg s fixními efekty.....	- 19 -
Obrázek 3-15 Tabulka s výsledky	- 20 -
(3.1)	- 22 -
Tabulka 3-2 Výsledky modelů s trendy	- 24 -

Úvod

Při výběru vhodné tematiky pro mou bakalářskou práci jsem se zamýšlel nad zajímavými problémy dnešní doby, až jsem došel k protikuřáckému zákonu a jeho vlivu na spotřebu piva. Je to možná i tím, že pivo je pro mě, stejně jako pro většinu populace v České republice, lék a neodmyslitelný nápoj pro dlouhé letní večery, ale i mrazivé zimní noci. Jak se říká mezi štamgasty: „V hospodě neprší, ale ani slunce nesvítí.“ Pít pivo v hospodách a restauračních zařízeních se dá proto za každého počasí. Problém s protikuřáckým zákonem řeší ne jeden hospodský, a proto jsem se zaměřil na analýzu jeho závislosti na spotřebě piva.

Tento zákon o úplném zákazu kouření v restauračních zařízeních, na veřejných místech a ve veřejných budovách je v diskuzi už od roku 2015. Velký podíl na zavedení protikuřáckého zákonu má Evropská Unie a Světová zdravotnická organizace se svým programem FCTC (Frame Convention on Tobacco Control), který jsme jako ČR ratifikovali [24].

Cílem mé práce je tedy analýza panelových dat. Pro zpracování jsem si vybral softwarový program STATA od společnosti StataCorp. V empirické části je mým hlavním úkolem najít nejvhodnější model, který by aspoň částečně vysvětloval závislost spotřeby piva na taxaci cigaret a zavedení částečného nebo úplného zákazu kouření. Tento model a jeho výsledky se poté pokusím interpretovat.

Pro tuto analýzu jsem si zvolil celkem 41 států Evropy a přilehlých oblastí v časovém rozmezí od roku 1998 až do roku 2012. Jako vysvětlující proměnné jsem zvolil daňové zatížení cigaret a roky zavedení částečného a úplného zákazu kouření. Na tento datový soubor nelze tedy pohlížet jako na průřezová data ani jako na časovou řadu. Tyto data se nazývají panelová. Jelikož se téma panelových dat probírá na bakalářském studiu pouze velmi omezeně, bylo potřeba před začátkem práce nastudovat danou problematiku.

V první části se zaměřím na vymezení problému protikuřáckého zákonu napříč Evropou. Protikuřácký zákon jako takový si každá země určuje legislativně sama, proto jsou zde popsány výjimky a důsledky protikuřáckého zákonu. Druhá část se zabývá datovým souborem, jeho tvorbou a základními statistikami. Součástí je i popis vysvětlovaných a vysvětlujících proměnných, které se budou v modelech vyskytovat. Na konci této části jsou uvedeny ekonometrické modely z pohledu teorie.

V třetí nejrozsáhlejší části jsou popsány výsledky mého úsilí. Modely jsou rozřazeny do dvou významných skupin, a to s ohledem na vývoj trendu a bez něj. Popsány jsou zde rozdíly mezi fixními a náhodnými efekty, které mohou zcela zásadně zasahovat do ekonometrických modelů. Závěrem je uvedeno, jaký dopad má na spotřebu piva protikuřácký zákon, který je postupně zaváděn do všech zemí Evropské Unie.

1 Vymezení problému protikuřáckého zákona v EU a ostatních zemích

V dnešní době je velkým tématem zdraví populace a snižování rizik tvorby smrtelných onemocnění. Velmi podrobně se problému kouření z druhé ruky zabývá [8], ve kterém je uvedeno, že zavedení protikuřáckého zákona snižuje velmi významně riziko tvorby smrtelných nemocí u nekuřáků. Je to také jeden z hlavních důvodů v ČR, proč byl tento zákon s platností od 31.5.2017 schválen. Další studie [9] se zabývá, zdali má zavedení protikuřáckého zákona vliv na tržby v restauracích a pohostinstvích v Norsku. V tomto článku byla použita metoda ARIMA pro časové řady a z modelu vyplynulo, že v dlouhodobém časovém horizontu nemá zavedení protikuřáckého zákona žádný vliv na tržby.

V USA byl proveden průzkum na univerzitách v Indianě a Purdue mezi studenty [10]. Jednalo se o rozsáhlý průzkum zaměřující se na počet kuřáků mezi studenty a efekt zavedení protikuřáckého zákona omezující kouření na veřejných místech jako jsou autobusové zastávky, koncerty, divadla apod. I zde byl nalezen pozitivní efekt na snížení počtu kuřáků po zavedení protikuřáckého zákona.

Během hledání podkladových materiálů k mé práci jsem narazil také na studii [6], která se zabývá vlivem protikuřáckého zákona v New Yorku a Kalifornii na fatální automobilové nehody s následkem smrti, které proběhly v noci pod vlivem alkoholu. Výsledky této studie jsou celkem zřejmé a to, že protikuřácký zákon nemá žádný vliv na smrtelné nehody na silnicích, které byly zaviněny pod vlivem alkoholu.

Se zajímavými fakty jsem se seznámil ve článku [11], který popisuje situaci v Brazílii v letech 1989 až 2008. V této zemi je zákaz kouření podpořen velmi silně pomocí policejní služby. V průběhu těchto let došlo k průměrnému ke snížení počtu kuřáků o 45 %. V tomto modelu je stejně jako v mých modelech zahrnuta proměnná popisující daňové zatížení tabáku. Daňové zatížení zde vzrostlo z 60 % na 75 %, čímž se Brazílie velmi podobá mnohým zemím Evropy.

Ve studii popisující situaci v Mexico City [13] jsou použity fixní efekty na daný model popisující výši tržeb v restauracích a zaměstnanost. Výsledky této studie se dají interpretovat dvěma způsoby. První je nalezení pozitivního efektu zavedení protikuřáckého zákona na výši tržeb a zaměstnanosti. Bohužel zde v tomto modelu nebyly koeficienty vysvětlujících proměnných statisticky významné, a proto se nedá přesně určit jakýkoliv efekt.

Situaci na Africkém kontinentu popisuje článek autora Jacqueline Tumwine [14]. Ten ve svém článku uvádí, že Africké země jsou teprve na počátku tabákové epidemie, která se projevuje zvyšujícím se počtem kuřáků mužského pohlaví při skoro nulovém počtu žen

kuřáček. Bohužel i země Afriky budou muset časem zavádět protikuřácké zákony, akorát se neočekává tak významných výsledků jako je tomu v zemích Evropy. Jedním z hlavních důvodů, proč tomu tak nebude, je nízký stupeň bezpečnostních složek a také to, že to jsou země rozvojové.

Efekty protikuřáckého zákona v evropských podmínkách velmi dobře popisují články [12][5]. Je zde nahlíženo na krátké časové období po zavedení protikuřáckého zákona. Ve článku autorů M. Kvasnicka a T. Siedler a S. Anger je výsledkem malý pokles počtu kuřáků v Německu a také omezení spotřeby cigaret u denních kuřáků. Naproti tomu článek popisující situaci v Itálii popisuje nulový efekt zavedení protikuřáckého zákona na většině území Itálie. Dále je zde testována oboustranná závislost kouření a spotřeby piva a byl zde nalezena pozitivní závislost. Tohoto předpokladu jsem i já využíval v mé práci.

Protikuřácký zákon je zaváděn i napříč celou Evropou, avšak v různých formách. Některé země mají povoleno kouření v klubech a barech pod určitými podmínkami. V Belgii například bylo povoleno mezi lety 2007-2011 kouření v restauracích, které měli pouze studená jídla a jejich podíl na tržbách nebyl vyšší než 30 % [19]. Ve Finsku je zavedena výjimka, pro bary s uzavřeným prostorem, kde se nesmí konzumovat jídlo a neobsluhuje zde obsluha [19]. V Lucembursku například dohlíží nad bary, kde je povoleno kouřit, Ministerstvo zdravotnictví a uděluje výjimky [19]. Další zemí s určitými specializace je Rakousko. Zde je do roku 2018 povoleno pro malé bary a kavárny do 50 m² vybrat si, zda budou kuřácké nebo nekuřácké, větší podniky musí mít stavebně oddělené prostory [19]. Poslední stojí za zmínku fakt, že ve Španělsku je maximální hranice kuřáckých pokojů v hotelu rovna 30 % z celkového počtu pokojů [19].

U nás v ČR byl zaveden částečný protikuřácký zákon už v roce 2006. Bohužel se jednalo o velmi mírný zákaz, který se vztahoval pouze na veřejná místa jako jsou nemocnice, úřady, školy a také zastávky hromadné dopravy. Podle mého názoru nám i v tomto bodě ujel pomyslný vlak se Západní Evropou, kdy ve stejném období už některé země měly úplný zákaz. Díky politickému zdržování zavedení zákona o úplném zákazu bude u nás tento zákon zaveden až od 31.5.2017.

2 Data a metody

2.1 Použitá data a software STATA

Datový soubor ke zpracování jsem vytvářel svépomocí z velkého množství různých zdrojů. Soubor popisuje celkem 41 zemí převážně z Evropské Unie (EU) a států ležících v Evropě nebo sousedících. Časová řada je vymezena roky 1998 a 2012 a jedná se o roční data. V průběhu sbírání dat jsem se setkal i s problémy, že některá data byla ve formě PDF, a proto jsem je musel opisovat jednotlivě za každý rok. Dále jsem také narazil na neexistenci dat z menších zemí nebo zemí, které se připojili k EU později. Důsledek těchto problémů je poté vidět níže v daných analýzách.

Celkem je v datovém souboru použito 5 proměnných, které vysvětluje následující tabulka:

Vysvětlovaná proměnná	Jednotky	Zdroje
Spotřeba piva	Litr čistého alkoholu na 1 obyvatele	[15]
Vysvětlující proměnné	Jednotky	Zdroje
Protikuřácký zákon - částečný	Rok zavedení	[18][19][20]
Protikuřácký zákon - úplný	Rok zavedení	[18][19][20]
Taxace cigaret	Procentuální část z celkové ceny 1 krabičky cigaret	[17]
HDP (GDP)	US dolar na 1 obyvatele (stálé ceny 2010)	[16]

Tabulka 2-1: Vysvětlovaná a vysvětlující proměnné v datovém souboru

2.1.1 Deskriptivní statistika dat

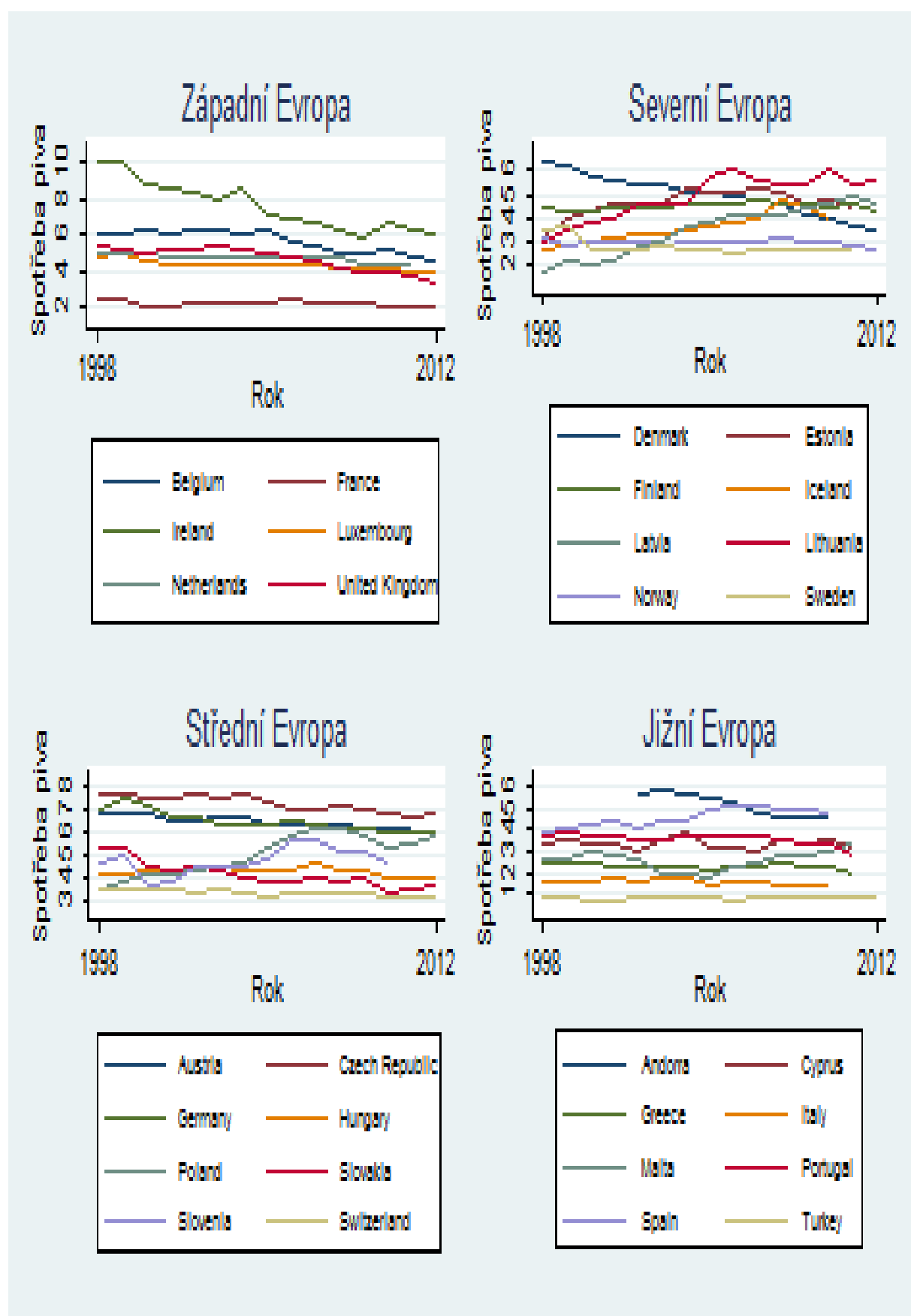
Nejdříve je potřeba si uvědomit, jak nahlížet na data jako celek. Jde o data panelová, která jsou opakovanými měřeními v různých časových bodech u daných proměnných. Nejde tedy na data nahlížet ani jako na časovou řadu ani jako na průřezová data. Popisné statistiky uvedené níže jsou průměr, směrodatná odchylka, minimální a maximální hodnota. Také jsou v obrázku uvedeny hodnoty pro meziskupinový a vnitroskupinový rozptyl. Lepší pohled na věc nám poskytne tento výstup:

Variable	Mean	Std. Dev.	Min	Max	Observations
spotre~h overall	3.837254	1.746758	.17	9.94	N = 579
between		1.647007	.2742857	7.562667	n = 41
within		.5976897	1.868587	6.214587	T-bar = 14.122
tax_ra~s overall	75.8266	5.429656	55.59	97.62	N = 332
between		3.15678	68.36	81.98778	n = 27
within		4.512522	57.40215	91.45882	T-bar = 12.2963
gdp overall	23300.84	19830.05	570.1318	87772.69	N = 615
between		19932.69	805.948	77948.19	n = 41
within		2231.114	9374.042	33125.34	T = 15
smoke_~t overall	2006.706	2.273557	2003	2010	N = 255
between		2.338929	2003	2010	n = 17
within		0	2006.706	2006.706	T = 15
smoke_~l overall	2007.897	2.400242	2004	2012	N = 435
between		2.439918	2004	2012	n = 29
within		0	2007.897	2007.897	T = 15

Obrázek 2-1: Deskriptivní statistika dat

Jak je vidět v obrázku výše uvedeném je maximální počet pozorování roven 615 při délce časové řady 15 let a počet států je 41. Bohužel počet pozorování není u každé proměnné stejný a poukazujete to na data, které neexistují nebo nebyla mnou dohledána.

Průměrná spotřeba piva napříč všemi zeměmi je 3,84 litru čistého alkoholu na jednoho člověka za rok. Toto číslo samo o sobě mnoho nevypovídá, prostým výpočtem jde dosáhnout více srozumitelnému výsledku a to 96 litrů piva ročně. Pro výpočet byl použit můj osobní poznatek, že průměrné pivo má v sobě 4 % alkoholu. Ač se to nezdá jako nikterak velké číslo, tak se jedná stále o počet piv na jednoho člověka, to zahrnuje i mladistvé od 15, kterým je alkohol odepřen. Největší spotřebu piva na člověka mělo Irsko v roce 1999 – 9,94 litrů. Pro lepší představu přikládám grafy vývoje průměrné spotřeby piva za jednotlivé regiony:



Obrázek 2-2: Spotřeba piva v Evropě

Dále je zajímavým číslem průměrné daňové zatížení na jednu krabičku cigaret, tj. 75,83 %. Poukazuje to na vysokou aktivitu států proti kouření jako takovému. K zemím s nejvyšším daňovým zatížením se řadí Portugalsko, Francie, Slovensko, Estonsko a Velká Británie. Nejmenší zdanění v mých datech má Litva v roce 2006 – 55,59 %.

První zemí, která zavedla částečný protikuřácký zákon je Polsko v roce 2003, které bylo následováno v roce 2004 Kypr, Andorou, Norskem a Irskem. Na opačném konci se nachází země, které nemají žádnou formu protikuřáckého zákona, jako je Rusko, Rumunsko a Moldávie. V České republice existuje od roku 2009 také částečný protikuřácký zákon, který by měl být podpořen úplným zákazem kouření ve veřejných prostorách a restauracích dne 31. května 2017, na který připadá Mezinárodní den bez tabáku.

Poslední statistikou, kterou uvádím je korelační matice. Tato matice může odhalit prvotní vztahy mezi jednotlivými proměnnými, které by mohly negativně zasáhnout do tvorby modelů. V mé korelační matici Obrázek 2-3 je vše v pořádku, nenašel jsem ani jednu dvojici proměnných, které by měly silnou závislost.

	spotre~h	tax_ra~s	gdp	law1	law2
spotreba_p~h	1.0000				
tax_rate_c~s	0.0154	1.0000			
gdp	0.0927	-0.1392	1.0000		
law1	0.0280	0.0590	0.0297	1.0000	
law2	-0.0838	0.2051	0.0727	-0.2720	1.0000

Obrázek 2-3 Korelační matice

V této práci byl použit softwarový program STATA, který byl vytvořen v roce 1985 společností StataCorp. Tento program byl přednostně vyvinut pro analýzu dat, jejich zpracování a vyobrazení výsledků do grafů. K používání programu slouží grafické uživatelské rozhraní a příkazový řádek. Stejně jako u dalších softwarových programů jako je např. R, Matlab, Eviews je k programu STATA mnoho rozšiřujících balíčků. Za jednu z největších výhod programu považuji jeho uživatelský manuál a celkovou podporu předními experty ze statisticko-ekonometrického okruhu.

2.2 Vysvětlující a vysvětlované proměnné

Základní vysvětlovanou proměnnou v mé práci je proměnná **spotřeba piva**. Jedná se o proměnnou, která popisuje, jaká je spotřeba litrů čistého alkoholu piva na jednu osobu za rok v dané zemi. V odborných textech je tato jednotka udávána jako *liters per capita*. Jelikož se jedná o uměle vytvořenou jednotku, tak zde může být malá chyba v zaokrouhlování.

Další proměnné už jsou vysvětlující. První z nich je **částečný zákaz**. Tato veličina uvádí, v jakém roce byl v dané zemi zavedený částečný zákaz kouření na veřejných místech. V mém datovém souboru jde o binární (dichotomickou) proměnnou. Nabývá tedy pouze hodnot 0, pokud není zaveden zákaz a 1, pokud je zaveden zákaz. Tuto vlastnost má i

proměnná **úplný zákaz**. V tomto případě se jedná o striktní zákaz kouření ve všech restauračních zařízeních a na veřejných místech.

Proměnnou, která by mohla zasáhnout do modelu je i **tax rate cigarettes**. Ta udává, jaké je procentuální daňové zatížení 1 krabičky cigaret. Toto daňové zatížení se skládá z více částí a podmínek a upravuje si to každá země, tak aby splňovala potřebné limity dané Evropskou Unií. Pro představu zde cituji část článku [25] o tvorbě taxace u nás v ČR:

„Nastavení spotřební daně na cigarety v České republice:

- podle evropských předpisů se v České republice uplatňuje smíšený model zdanění, Česko patří mezi země, které zavedly minimální spotřební daň.
- celková spotřební daň tvoří v Česku 62 % vážené průměrné maloobchodní ceny cigaret propuštěných ke spotřebě.
- celková spotřební daň činí 92,41 eur za 1.000 kusů cigaret.
- pevná část daně činí 1,29 Kč za jednu cigaretu, to je 44 % celkového daňového zatížení.
- procentní část daně je stanovena na 27 % z maloobchodní ceny.
- minimální spotřební daň je stanovena na 2,37 Kč za jednu cigaretu.“

Poslední proměnnou, která se objevuje jen na začátku práce s modely, je **gdp**. Jde o proměnnou, která popisuje hrubý domácí produkt na člověka v dolarech při stálých cenách 2010.

2.3 Metodologie

V této části nastíním důležité pojmy, se kterými se budu v průběhu mé práce setkávat a dále už nebudou popisovány. Tyto pojmy se týkají hlavně ekonometrických modelů a jejich rozšiřujícími vlastnostmi.

2.3.1 Ekonometrické pojmy

V této krátké kapitole uvedu přehled všech používaných ekonometrických zkratk v praktické části, které by bez vysvětlení nedávaly moc smysl. Jelikož je většina literatury k tomuto tématu v angličtině, tak je lepší používat buďto originální anglické znění nebo počeštěné verze těchto slov:

- **R² (R-squared)** – koeficient determinace
- **Standard errors (Std. Err.)** – směrodatná odchylka
- **VCE (variance-covariance matrix of an estimator)** – variační-kovariační matice odhadu, může nabývat vlastností *cluster* (sendvičová forma VCE) a *robust* (forma VCE s robustními směrodatnými odchylkami)

- **GLS** model (*Generalized least squares*) – metoda zobecněných nejmenších čtverců (ZNČ)
- **GEE** model (*Generalized estimating equation*) – rovnice sloužící k odhadu parametrů ze zobecněného lineárního modelu
- **LSDV** model (*Least square dummy variable*) – metoda nejmenších čtverců s *dummy* proměnnými
- **OLS** model (*Ordinary least squares*) – základní metoda nejmenších čtverců (MNČ)
- **FDL** model (*Finite distributed lag*) – metoda používající zpoždění efektu v čase
- **ML** model (*maximum likelihood*) metoda maximální věrohodnosti
- **Hausmanův** test – test k určení vhodnosti fixních nebo náhodných efektů v modelu

2.3.2 Teoretické okénko

V mé práci se objevují ekonometrické modely, které jsou většinou odhadnuty pomocí MNČ nebo zobecněných nejmenších čtverců. Teorie k těmto odhadům je tisíckrát použita ve skoro každé bakalářské práci, a proto nevidím význam je uvádět v mé. Rád bych se pozastavil nad rozšířeními běžných modelů o fixní a náhodné efekty a také jaké jsou rozdíly v meziskupinovém, vnitroskupinovém a celkovém koeficientu determinace. Na závěr zde uvádím základní *Hausmanův* test.

Základní metoda OLS s fixními efekty má odhad pomocí funkce *xtreg*, *fe* daný:

$$\left(y_{it} - \bar{y}_i + \bar{y}\right) = \alpha + \left(x_{it} - \bar{x}_i + \bar{x}\right)\beta + \left(\varepsilon_{it} - \bar{\varepsilon}_i + \bar{\varepsilon}\right) + \bar{\varepsilon} \quad (2.3.1)$$

, kde jsou odhadnuty hodnoty α a β .

Interpretace R^2 vnitroskupinový je taková, že R^2 dostaneme jako průměrnou odchylku regrese

Interpretace R^2 meziskupinový je taková, že R^2 dostaneme jako $\text{corr}(x_i\hat{\beta}, y_i)^2$.

Interpretace R^2 celkový je taková, že R^2 dostaneme jako $\text{corr}(x_{it}\hat{\beta}, y_{it})^2$.

Avšak pokud chceme použít náhodné efekty v našem modelu, tak musíme použít metodu GLS a je dána pomocí funkce *xtreg*, *re*. Klíčové je pro tento model transformace z GLS na OLS, kterou provedeme pomocí:

$$z_{it}^* = z_{it} - \hat{\theta}_i \bar{z}_i,$$

$$\text{kde } z_i = \frac{1}{T_i} \sum_t^{T_i} z_{it} \text{ a } \hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{T_i \hat{\sigma}_u^2 + \hat{\sigma}_e^2}}. \quad (2.3.2)$$

Interpretace R^2 vnitroskupinový je taková, že R^2 dostaneme jako $\text{corr}\left\{(x_{it} - \bar{x}_i)\hat{\beta}, y_{it} - \bar{y}_i\right\}^2$.

Interpretace R^2 meziskupinový je taková, že R^2 dostaneme jako $\text{corr}\left(x_i \hat{\beta}, y_i\right)^2$.

Interpretace R^2 celkový je taková, že R^2 dostaneme jako $\text{corr}\left(x_{it} \hat{\beta}, y_{it}\right)^2$.

Hausmanův test a jeho robustní verze slouží k určení vhodnosti fixních nebo náhodných efektů. Bohužel tento test je i předmětem velké diskuze [1], zda se jeho výsledky řídit vždy, a nebo přihlídnout k nedokonalosti tohoto testu a použít fixní efekty i přesto, že nebyla zamítnuta hypotéza o vhodnosti náhodných efektů.

Hausmanova statistika má rozdělení χ^2 a je spočítána:

$$H = (\beta_c - \beta_e)' (V_c - V_e)^{-1} (\beta_c - \beta_e) \quad (2.3.3)$$

, kde β_c je koeficient vektoru z konzistentního odhadu

β_e je koeficient vektoru z vydatného odhadu

V_c je kovariační matice z konzistentního odhadu

V_e je kovariační matice z vydatného odhadu.

Problematika robustní verze *Hausmanova* testu je důkladněji probrána v literatuře [1].

3 Výsledky

3.1 Modely odhadnuté MNČ a ZNČ

Všechny modely a jejich výstupy byly zpracovány v programu STATA. Pojmenování důležitých hodnot se může lišit ve formě zápisu od jiných softwarových programů, postupně na tyto hodnoty budu upozorňovat.

3.1.1 Odhady pořízené MNČ pomocí funkce *regress*

V tomto modelu jsou použity všechny proměnné a jeho výstup je následující:

Source	SS	df	MS	Number of obs = 316		
Model	15.3858454	4	3.84646134	F(4, 311) = 1.53		
Residual	780.238607	311	2.50880581	Prob > F = 0.1923		
Total	795.624452	315	2.52579191	R-squared = 0.0193		
				Adj R-squared = 0.0067		
				Root MSE = 1.5839		

spotreba_piva_v_l~h	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tax_rate_cigarettes	.0152072	.0170869	0.89	0.374	-.0184134	.0488278
gdp	.00001	5.32e-06	1.88	0.060	-4.43e-07	.0000205
law1	-.0294831	.2628643	-0.11	0.911	-.5467005	.4877343
law2	-.3615683	.2096785	-1.72	0.086	-.7741361	.0509995
_cons	3.173152	1.312838	2.42	0.016	.5899837	5.756319

Obrázek 3-1 Model OLS pomocí funkce *regress* bez VCE

Při prvním pohledu na model je zjištěno, že R^2 je 0,0193. Z toho vyplývá, že se variabilita vysvětlované proměnné *spotřeba_piva* nebyla dostatečně vysvětlena pomocí vysvětlujících proměnných. Bohužel tento model nebyl správně namodelován, protože hodnota F-testu je 0,1923 a tím se nezamítá nulová hypotéza, pokud si stanovíme hladinu statistické významnosti na hodnotu 0,05. Dalším indikátorem problému v modelu jsou jednotlivé P hodnoty proměnných dílčích t-testů, které jsou uvedeny ve čtvrtém sloupci. Posledním problémem v tomto modelu je fakt, že jsme při tvorbě zapomněli na VCE. V dalším výstupu už jsem VCE zahrnul do modelu:

Linear regression				Number of obs = 316		
				F(4, 26) = 0.47		
				Prob > F = 0.7547		
				R-squared = 0.0193		
				Root MSE = 1.5839		
(Std. Err. adjusted for 27 clusters in id)						
spotreba_piva_v_l~h	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
tax_rate_cigarettes	.0152072	.0343158	0.44	0.661	-.05533	.0857443
gdp	.00001	.0000137	0.73	0.472	-.0000182	.0000382
law1	-.0294831	.4765331	-0.06	0.951	-1.009011	.9500448
law2	-.3615683	.3140006	-1.15	0.260	-1.007006	.2838693
_cons	3.173152	2.717788	1.17	0.254	-2.413341	8.759644

Obrázek 3-2 Model OLS pomocí funkce *regress* s VCE

V obou modelech jsou stejné hodnoty koeficientů a R^2 , ale směrodatné odchylky jsou počítány se zahrnutím VCE. Proto dosahují už vyšších hodnot, které už nejsou abnormálně malé. Špatným důsledkem je to, že se nám zvýšila hodnota F-testu na 0,7547 i dílčí hodnoty t-testů. Jako následující možnost se jeví nejlépe podle [4] model s prvními diferencemi. K modelování byla použita znovu funkce *regress* s robustní VCE:

Linear regression

Number of obs = 288
F(3, 26) = 1.11
Prob > F = 0.3623
R-squared = 0.0062
Root MSE = .27924

(Std. Err. adjusted for 27 clusters in id)

D. spotreba_piva_v_l~h	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
tax_rate_cigarettes D1.	-.0013892	.0030261	-0.46	0.650	-.0076095	.0048312
law1 D1.	.0154611	.0545087	0.28	0.779	-.0965833	.1275054
law2 D1.	-.0791325	.0634143	-1.25	0.223	-.2094824	.0512174

Obrázek 3-3 Model OLS pomocí funkce *regress* s prvními diferencemi

Bohužel tento model mě k ničemu nepomohl, protože všechny hypotézy testových kritérií nebyly zamítnuty. Tímto způsobem jsem dokázal, že funkce *regress* se v programu STATA nehodí k modelům panelových dat. Musím tedy přistoupit ke skupině funkcí, které začínají *xt* a jsou vhodné k použití na panelová data. Jsou to například funkce *xtreg*, *xtgee* a jejich verze s fixními nebo náhodnými efekty.

3.1.2 Modely GLS a GEE

V této části byla použita funkce *xtreg*, která má i různé varianty. Struktura proměnných zůstala neměnná, a proto je mohu porovnávat s předchozími modely. Více prozradí obrázek výstup z programu STATA:

Random-effects GLS regression		Number of obs	=	316
Group variable: id		Number of groups	=	27
R-sq: within	= 0.2760	Obs per group: min	=	7
between	= 0.0068	avg	=	11.7
overall	= 0.0027	max	=	15
corr(u_i, X) = 0 (assumed)		Wald chi2(4)	=	102.47
		Prob > chi2	=	0.0000

spotreba_piva_v_l~h	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
tax_rate_cigarettes	.0174252	.0058032	3.00	0.003	.006051	.0287993
gdp	-.0000343	.0000104	-3.30	0.001	-.0000547	-.0000139
law						
castecny zakaz	-.1652407	.0990935	-1.67	0.095	-.3594604	.0289791
uplny zakaz	-.5235035	.0745642	-7.02	0.000	-.6696466	-.3773604
_cons	4.283763	.5970797	7.17	0.000	3.113509	5.454018
sigma_u	1.5909654					
sigma_e	.43099111					
rho	.93163102	(fraction of variance due to u_i)				

Obrázek 3-4 Model GLS pomocí funkce *xtreg*

V tomto výstupu je vidět trochu více informací, a tak by bylo dobré nastínit, co mohou vypovídat o modelu. V prvním řádku je vidět název použité regrese. V základním nastavení pracuje program STATA s náhodnými efekty, pokud bych chtěl použít fixní efekty, musím přidat tuto vlastnost do modelu. Veličina R^2 má nyní 3 položky: vnitroskupinový (*within*), meziskupinový (*between*) a celkový (*overall*) koeficient determinace. Při počítání s panelovými daty si program sám nastavuje jednotlivé skupiny a podle toho počítá jejich hodnoty koeficientů. S mém modelem vytvořil program STATA celkem 27 skupin při celkovém počtu použitých pozorování 316. Toto číslo je mnohem menší než celkový počet pozorování, který je 615. Je to důsledek chybějících dat.

Další změnou ve výstupu je použití *Waldova* chí-kvadrát testu, při daném stupni volnosti, v mém případě 4. Prostřední část výstupu zůstala stejná, ale na konci přibýly hodnoty *sigma_u*, *sigma_e* a *rho*. Tyto hodnoty budou použity v další části mé práce v *Hausmannově* robustním testu.

Výsledky tohoto modelu lze interpretovat jako uspokojivé, protože většina testových kritérií nepřekročila hladinu významnosti 5 %. Jediná výjimka je proměnná *castecny zakaz*, u které je to dáno tím, že se jedná o předstupeň proměnné *uplny zakaz* a je zde přítomna slabá korelace. Pro další postupy vyřadím z modelu proměnnou *gdp*, jelikož koeficienty jsou tak nízké, že se nevyplatí tuto proměnnou v modelu držet. V dalším výstupu se zaměřím, jestli by pomohla zpoždění u proměnných *castecny zakaz* a *uplny zakaz* ke zlepšení mého modelu:

Random-effects GLS regression		Number of obs		=	301	
Group variable: id		Number of groups		=	27	
R-sq: within = 0.2615		Obs per group: min		=	7	
between = 0.0023		avg		=	11.1	
overall = 0.0066		max		=	14	
		Wald chi2(5)		=	15.36	
corr(u_i, X) = 0 (assumed)		Prob > chi2		=	0.0089	
(Std. Err. adjusted for 27 clusters in id)						
spotreba_piva_v_l~h	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
tax_rate_cigarettes	.017305	.0069437	2.49	0.013	.0036956	.0309145
law1						
--.	-.1163831	.1123577	-1.04	0.300	-.3366001	.1038339
L1.	-.2129934	.1045016	-2.04	0.042	-.4178127	-.0081741
law2						
--.	-.2441482	.1160845	-2.10	0.035	-.4716696	-.0166268
L1.	-.4230368	.1832642	-2.31	0.021	-.782228	-.0638455
_cons	3.406551	.5531694	6.16	0.000	2.322359	4.490743
sigma_u	1.6085129					
sigma_e	.41579258					
rho	.93736554	(fraction of variance due to u_i)				

Obrázek 3-5 Model GLS pomocí funkce *xtreg* se zpožděními a VCE (robust)

Výsledky této verze se zpožděními nejsou pro další postup v práci potřebné, protože hodnota vnitroskupinového koeficientu determinace R^2 je 0,2615 a je menší než u verze bez zpoždění. Tudíž má teze, že zavedení protikuřáckého zákona se projeví ve spotřebě piva až s ročním zpožděním, mě nepomohla.

Doposud jsem neinterpretovat žádné výsledky koeficient a pouze se zaměřoval na hodnocení statistických testů, zdali má tento model význam a uvedené proměnné mají pro model statistickou významnost. Ohlédnutím na předešlé modely se zjistilo, že proměnná *tax_rate_cigarettes* má ve všech modelech kladný koeficient. Tohle jsem při tvorbě modelu vůbec nečekal, protože můj předpoklad byl opačný. Interpretace toho koeficientu je následující: Pokud bych vzal dvě naprosto stejné země ve stejném roce, kde by hodnoty všech

ostatních proměnných byly stejné a pouze u první země by bylo daňové zatížení cigaret o 1 % vyšší, tak by tato země měla spotřebu piva o cca 0,0174 litrů čistého alkoholu vyšší za každého člověka za jeden rok. V přepočtu na litry piva je to o 0,435 litrů piva na člověka za rok. Pokud bych zvážil, že rozdíly mezi jednotlivými zeměmi v taxaci cigaret činí v určité roky 10-15 procentních bodů, tak by mohla být hodnota spotřeby piva o 4-6 litrů vyšší než u zemí s nižší taxací, což už není zanedbatelná hodnota.

Naproti tomu proměnné, které popisují zavedení částečného nebo úplného zákazu kouření, potvrzují můj předpoklad o negativním vlivu na spotřebu piva. Průběžné výsledky z modelu funkce *xtreg* nám ukazují velké pokles spotřeby piva po zavedení úplného zákazu. Při stejné interpretaci by to znamenalo pokles spotřeby piva u země o -0,523 litru čistého alkoholu na osobu za rok, tj. 13,075 litrů piva na osobu za rok.

Tyto průběžné výsledky jsou celkem znepokojující pro většinu majitelů hospod a pivovarnických společností v České republice, která se chystá na zavedení zákona o úplném zákazu kouření na veřejných místech a pohostinstvích. V médiích se hovoří pouze o zavírání hospod a pohostinství, ale jen málokdo se zabývá hlubším pojetím této změny. Nejdůkladněji to popisují autoři M. Kvasnicka a H. Tauchman v jejich článku 28[7]. Výsledkem jejich práce je zjištění, že po zavedení zákazu v Německu v roce 2007 a 2008, byl zjištěn pokles průměrných reálných tržeb o 2 %.

Po malém odbočení k průběžnému hodnocení výsledků se vracím zpět k dalšímu modelu a to *xtgee*. Tento model by se dal do češtiny přeložit jako Zobecněná odhadovací rovnice (*Generalized estimating equation*) a se liší výpočtem, ke kterému se využívá Gaussovo rozdělení a iterace. Iterace v mém modelu jsou celkem 3. Koeficienty jednotlivých proměnných jsou odlišné od modelu s funkcí *xtreg*. Tyto získané výsledky se zatím jeví nadějně, ale bohužel nemáme zde uvedenou srovnávací hodnotu koeficientu determinace R^2 . Také zde není možnost přidat do modelu fixní efekty, protože se jedná o zobecněný model. Dále se bude model porovnávat s konečné tabulce. Tyto informace nám doloží výstup:

```
Iteration 1: tolerance = .28190613
Iteration 2: tolerance = .00029062
Iteration 3: tolerance = 1.076e-08
```

```
GEE population-averaged model      Number of obs      =      316
Group variable:                    id      Number of groups   =      27
Link:                              identity Obs per group: min =      7
Family:                            Gaussian      avg      =     11.7
Correlation:                       exchangeable      max      =     15
                                      Wald chi2(3)      =     12.99
Scale parameter:                   2.519916      Prob > chi2      =     0.0047
```

(Std. Err. adjusted for clustering on id)

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
spotreba_piva_v_1~h						
tax_rate_cigarettes	.0180903	.0072753	2.49	0.013	.0038311	.0323496
law						
castecny zakaz	-.2777208	.1491969	-1.86	0.063	-.5701414	.0146998
uplny zakaz	-.6359333	.1897831	-3.35	0.001	-1.007901	-.2639651
_cons	3.369878	.5900012	5.71	0.000	2.213497	4.52626

Obrázek 3-6 Model GEE pomocí funkce *xtgee* s robustní VCE

3.1.3 Modely s fixními a náhodnými efekty

Panelová data jsou zajímavá také tím, že mohou obsahovat některé proměnné, které se v čase nijak nevyvíjí nebo se také nemusí vyvíjet napříč jednotlivými kategoriemi, tyto proměnné mohou být zafixovány. Z tohoto důvodu je potřeba také v modelech rozlišovat, zdali mohou použít fixní efekty nebo náhodné. K otestování vhodného typu efektů v mých modelech slouží *Hausmanův* test a jeho varianta s robustními směrodatnými odchylkami. Testová hypotéza je vyřčena, zdali náhodné efekty (RE) jsou konzistentními odhady pro náš model:

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) FE	(B) RE		
tax_rate_c~s	.0181167	.0179629	.0001538	.0003834
lbn.law	-.2782192	-.2753062	-.002913	.0072522
2.law	-.63637	-.6338135	-.0025565	.0038481

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(3) = (b-B)' [(V_b-V_B)^(-1)] (b-B)
= 0.61
Prob>chi2 = 0.8931

Obrázek 3-7 *Hausmanův* test na fixní nebo náhodné efekty u modelu GLS

Výsledkem toho testu je hodnoty $P = 0,8931$, tudíž nemůžeme zamítnout testovou hypotézu. Vhodnější by byly pro moje modely náhodné efekty. V tomto základním testu však nejsou zahrnuty robustní směrodatné odchylky. S tímto problémem nám pomůže rozšířený *Hausmanův* test, který zde uvádím jako náhled do programu STATA:

```
. * Hausman robust
. quietly xtreg $xlist, re

. keep if e(sample)
(299 observations deleted)

. bysort id: egen T_i = count(id)

. scalar s_u = e(sigma_u)

. scalar s_e = e(sigma_e)

. generate theta = 1 - sqrt( s_e^2 / (T_i*s_u^2 + s_e^2)) // Z manualu k xtreg.

. global xforhausman tax_rate_cigarettes law1 law2 gdp

. sort id

. * Zavisle promenna
. global yvar spotreba_piva_v_litrech

. by id : egen mean_$yvar = mean($yvar)

. generate y_md_$yvar = $yvar - mean_$yvar

. generate y_red_$yvar = $yvar - theta*mean_$yvar

. * Vysvetlujici promenne
. foreach x of varlist $xforhausman {
2.      by id : egen mean_`x' = mean(`x')
3.      generate md_`x' = `x' - mean_`x'
4.      generate red_`x' = `x' - theta*mean_`x'
5. }

. * Vlastni
. quietly regress y_red_* red_* md_*

. testparm md_*

( 1) md_tax_rate_cigarettes = 0
( 2) md_law1 = 0
( 3) md_law2 = 0
( 4) md_gdp = 0

      F( 4, 307) = 1.33
      Prob > F = 0.2604
```

Obrázek 3-8 *Hausmanův* robustní test na fixní nebo náhodné efekty

Na konci výstupy je uveden F-test, který má stanovit, zda se hypotéza o vhodnosti náhodných efektů zamítne nebo nezamítne. Hodnota testu je 0,2604, a proto při statistické hladině významnosti 0,05 nezamítám hypotézu. Toto je důsledek malého počtu vysvětlujících proměnných v mém modelu a menšího rozsahu počtu pozorování. Bohužel zde zase narážím na problém s daty, které neexistují nebo nejsou dohledatelné. Proto jsem se rozhodl dále počítat i s fixními efekty jak časovými, tak individuálními.

Pro názornou ukázkou rozdílu mezi fixními a náhodnými efekty jsou si vybral model GLS pomocí funkce *xtreg*, která obě dvě varianty umožňuje:

Regresní tabulka

	FE	RE
<i>tax_rate_cigarettes</i>	0.0181** (0.0059)	0.0180** (0.0059)
<i>zadny zakaz</i>	0.0000 (.)	0.0000 (.)
<i>castecny zakaz</i>	-0.2782** (0.0957)	-0.2753** (0.0950)
<i>uplny zakaz</i>	-0.6364*** (0.0682)	-0.6338*** (0.0679)
Constant	3.3745*** (0.4421)	3.3785*** (0.5316)
N	316.0000	316.0000
r2	0.2351	
r2_o	0.0071	0.0071
r2_b	0.0015	0.0015
r2_w	0.2351	0.2351

Standard errors in parentheses
 * p<0.05, ** p<0.01, *** p<0.001

Obrázek 3-9 Model GLS pomocí funkce *xtreg* – fixní vs náhodné efekty

Jak je možné z výsledků vyčíst, tak oba dva modely mají stejné hodnoty koeficientů determinace. V čem se liší jednotlivé modely jsou koeficienty proměnných a jejich směrodatné odchylky. Ani v jednom modelu nejsou zahrnuty robustní VCE, a proto lze říct, že všechny vysvětlující proměnné jsou statisticky významné na hladině významnosti 0,05. Rozdíly mezi koeficienty modelů jsou velmi malé, čímž se potvrzuje výsledek *Hausmanova* testu a to, že fixní efekty nehrají v mých datech tak velkou roli.

3.1.4 Ostatní modely

V této části nastíním i další modely, které jde na panelová data použít. Základní literatura [4][2] se těmito modely zabývá okrajově, a proto je zde uvádím jen pro srovnání s ostatními výsledky.

Prvním model je LSDV (*Least square dummy variable*), pro který je důležitá podmínka $T \rightarrow \infty$. V mých datech je N rovno 15, takže bych mohl použít tento model, který je v programu STATA zastoupen funkcí *areg*:

Linear regression, absorbing indicators

Number of obs = 316

F(3, 26) = 3.94

Prob > F = 0.0192

R-squared = 0.9292

Adj R-squared = 0.9220

Root MSE = 0.4439

(Std. Err. adjusted for 27 clusters in id)

spotreba_piva_v_l~h	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
tax_rate_cigarettes	.0181167	.0076276	2.38	0.025	.0024381	.0337954
law						
castecny zakaz	-.2782192	.1566832	-1.78	0.087	-.6002861	.0438476
uplny zakaz	-.63637	.1993618	-3.19	0.004	-1.046164	-.2265759
_cons	3.374452	.5546606	6.08	0.000	2.234331	4.514573
id	absorbed	(27 categories)				

Obrázek 3-10 Model funkce *areg*

Zajímavý je fakt, že model GLS funkce *xtreg* s fixními efekty má stejné koeficienty u vysvětlujících proměnných jako tento model: Obrázek 3-10. Model LSDV je vlastně totéž, co vnitroskupinový odhad (*within estimator*). Slouží nám tedy pouze k odhadu koeficientů proměnných a další hodnoty v modelu nejsou relevantní pro vysvětlení.

Poslední model, který zde uvedu, není až tak model, ale spíš metoda k odhadu teoretických hodnot koeficientů a směrodatných odchylek a je to *Bootstrap*. Tato metoda má základ v náhodném losování vzorků z dat a vytváření odhadu ke každému z nich a následném zprůměrování. Pokud je tato metoda opakována s velkou četností – pro představu 1000krát a více, tak slouží k získání teoretických hodnot pro celou populaci v mém případě pro všechny země. Výstup z této metody uvádím níže a počet kroků jsem zvolil 10000:

Fixed-effects (within) regression		Number of obs	=	316
Group variable: id		Number of groups	=	27
R-sq: within	= 0.2351	Obs per group: min	=	7
between	= 0.0015	avg	=	11.7
overall	= 0.0071	max	=	15
corr(u_i, Xb) = -0.0924		Wald chi2(3)	=	62.34
		Prob > chi2	=	0.0000
(Replications based on 27 clusters in id)				

	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
spotreba_piva_v_l~h						
tax_rate_cigarettes	.0181167	.0061415	2.95	0.003	.0060796	.0301539
law						
castecny zakaz	-.2782192	.0732965	-3.80	0.000	-.4218778	-.1345607
uplny zakaz	-.63637	.0828767	-7.68	0.000	-.7988055	-.4739346
_cons	3.374452	.4653347	7.25	0.000	2.462413	4.286491
sigma_u	1.4929856					
sigma_e	.44391571					
rho	.91877341	(fraction of variance due to u_i)				

Obrázek 3-11 Metoda *Bootstrap* (10000 opakování) – funkce *xtreg* s fixními efekty

Koeficienty vysvětlujících proměnných jsou stejné jako u modelu funkce *xtreg* s fixními efekty. Důležitější jsou zde hodnoty směrodatných odchylek, které jsou nižší než v normálním modelu OLS funkce *xtreg*. Komentovat hodnoty testových kritérií nemá význam, jelikož jsou je opakováním dosaženo minimálních hodnot.

Další složitější modely v mé práci nebudu uvádět, protože teorie k nim je rozsáhlá a převyšující meze bakalářské studia. Takovéto modely se formulují v programu STATA pomocí funkcí *xtabond*, *xtdpdys* nebo *xtmixed*. V poslední části mé analýzy vyberu nejnadějnější modely a porovnáám jejich hodnoty mezi sebou.

3.1.5 Porovnání vybraných modelů

V průběhu mé práce jsem vybral celkem 6 modelů, které podle mého názoru nejlépe popisují závislost spotřeby piva na vysvětlujících proměnných. Nejdůležitější hledisko pro výběr těchto modelů bylo splnění testových kritérií, resp. zamítnutí nulových hypotéz na hladině statistické významnosti 0,05. Dalším kritériem byl vnitroskupinový koeficient determinace R^2 (*within*). Pro ulehčení práce jsem si nadefinoval v programu *\$xlist*, který zahrnuje proměnné *tax_rate_cigarettes*, *castecny zakaz*, *uplny zakaz*. Do mého výběru jsem vybral tyto modely:

- xtreg *\$xlist*, fe vce(robust) – model s fixní efekty a robustní VCE,
- xtgee *\$xlist*, vce(robust) – model GEE s robustní VCE,
- bootstrap, rep (1000): xtreg *\$xlist*, fe vce(robust) – metoda Bootstrap,
- areg *\$xlist*, absorb(id) vce(cluster id) – model LSDV s robustními VCE,
- xtreg *\$xlist*, re vce(cluster id) – model s náhodnými efekty a robustní VCE,
- regress *\$xlist*, vce(cluster id) – základní OLS regresní model s robustní VCE.

Ekonometrická analýza závislosti spotřeby piva na protikuřáckém zákonu

	FE_rob	GEE_rob	Boot	Areg	RE_rob	OLS
tax_rate_cigarettes	0.0181* (0.0073)	0.0181* (0.0073)	0.0181** (0.0060)	0.0181* (0.0076)	0.0180* (0.0074)	0.0099 (0.0303)
zadny zakaz	0.0000 (.)	0.0000 (.)	0.0000 (0.0000)	0.0000 (.)	0.0000 (.)	0.0000 (.)
castecny zakaz	-0.2782 (0.1500)	-0.2777 (0.1492)	-0.2782*** (0.0725)	-0.2782 (0.1567)	-0.2753 (0.1495)	0.0063 (0.4474)
uplny zakaz	-0.6364** (0.1909)	-0.6359*** (0.1898)	-0.6364*** (0.0818)	-0.6364** (0.1994)	-0.6338*** (0.1898)	-0.3138 (0.3130)
Constant	3.3745*** (0.5310)	3.3699*** (0.5900)	3.3745*** (0.4531)	3.3745*** (0.5547)	3.3785*** (0.5949)	3.8596 (2.2324)
N	316.0000	316.0000	316.0000	316.0000	316.0000	316.0000
r2	0.2351		0.2351	0.9292		0.0081
r2_o	0.0071		0.0071		0.0071	
r2_b	0.0015		0.0015		0.0015	
r2_w	0.2351		0.2351		0.2351	

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Obrázek 3-12 Tabulka s výsledky

Při prvním pohledu na výstup lze vidět, že některé modely nemají všechny srovnávací kritéria. Postupně budu tedy vyřazovat jednotlivé modely, abych vyseletoval model, který se bude blížit *BLUE* (*Best linear unbiased estimator*).

Model OLS se jeví jako nejhorší varianta. Ani jedna z vysvětlujících proměnných nesplňuje dílčí t-test při hladině statistické významnosti 0,05. Dále tento model má koeficient determinace R^2 roven pouze 0,0081. Jak už bylo řečeno, funkce *regress* se nehodí na panelová data, ale na data průřezová.

Model GEE je v této tabulce uveden jako ukázka zobecněné odhadové rovnice a bohužel nemá koeficient determinace. Ani tento model se tedy nehodí k interpretaci koeficientů proměnných.

Metoda Bootstrap zde představuje pouze přirovnání, k jakým teoretickým hodnotám by měly směřovat směrodatné odchylky. Jelikož vychází z modelu *xtreg* s fixními efekty, tak má také stejné koeficienty proměnných.

Model LSDV, který je zastoupen funkcí *areg*, slouží pouze k odhadnutí koeficientů vysvětlujících proměnných. Jak je uvedeno výše u Obrázek 3-10, nedá se dále použít, kvůli nemožnosti interpretace koeficientu determinace.

Zbývají nám tedy modely dva, které zastupuje v programu STATA funkce *xtreg*. Jeden model s fixními efekty a druhý model s náhodnými efekty. Jelikož mají oba dva modely stejné vnitroskupinové koeficienty determinace R^2 , tak se musíme podívat na hodnoty dílčích t-testů. Tyto hodnoty se liší pouze minimálně, a proto ani toto kritérium nedokáže rozhodnout, který model je na tom lépe. Je to také způsobeno tím, že nebyla zamítnuta hypotéza *Hausmanova* testu, která by potvrdila lepší vhodnost modelu s fixními efekty. Je nutné se tedy podívat na přesná čísla koeficientů proměnných a jejich směrodatných odchylek.

Hodnoty směrodatných odchylek v modelu s fixními efekty jsou menší než v modelu s náhodnými efekty u proměnné *tax_rate_cigarettes* a u konstanty. Na druhou stranu jsou v FE modelu vyšší koeficient proměnných *castecny zakaz* a *uplny zakaz*. Po uvážení se přikláním k modelu funkce *xtreg* s fixními efekty.

Pro interpretaci výsledků byl tedy vybrán model FE s robustní VCE. V tomto modelu je hodnota koeficientu vnitroskupinové determinace R^2 rovna 0,2351, což lze vysvětlit následovně. Model vysvětluje 23 % vnitroskupinové variability vysvětlované proměnné *spotreba_piva*. Toto číslo není nikterak vysoké, ale to jsem v začátku své práce ani neočekával. Toto číslo je ovlivněno hlavně nekompletními daty, se kterými jsem měl problém s dohledáním, protože některé země je neuvádějí. Koeficient celkové determinace R^2 je tak malý, protože mám v modelu zahrnutý pouze 3 proměnné a bylo by neskutečným štěstím, pokud bych něco našel. V článku autorů Kvasnicka a Tauchman [7] jsou například uvedeny další vysvětlující proměnné *rainfall* a *unemployment*, které popisují množství srážek a míru nezaměstnanosti a jejich vliv na příjem barů a pohostinství.

Konstanta v mém modelu má hodnotu 3,375 a odpovídá zhruba tedy i reálnému průměru spotřeby piva v litrech s hodnotou 3,837. Pokud bych měl všechna data, tak by se tyto dvě čísla měla lišit jen minimálně. Vzniklý rozdíl nabývá hodnoty 0,462 a je to důsledkem, že k popsání modelu bylo použito pouze 316 pozorování, ale celkový počet pozorování spotřeby piva je 579. Nicméně hypotéza dílčího t-testu byla zamítnuta a mohu tedy uvést konstantu v mém výsledném modelu.

Proměnné *castecny zakaz* a *uplny zakaz* mají hodnoty koeficientů se záporným znaménkem, tudíž mají negativní vliv na vysvětlovanou proměnnou. U proměnné *castecny zakaz* je hodnota dílčího t-testu $P > 0,05$. Z tohoto důvodu by finální model neměl obsahovat tuto proměnnou. Z praktického hlediska bych ale ztratil další vysvětlující proměnnou, a to nechci dopustit. Vyšší P hodnotu u této proměnné dávám za vinu slabé korelace mezi částečným zákazem a úplným zákazem. Je to pochopitelné, protože pokud přejde částečný zákaz plynule do úplného zákazu kouření, tak je mezi nimi vztah.

Největším překvapením je proměnná *tax_rate_cigarettes*, která popisuje v mém modelu procentuální daňové zatížení 1 krabičky cigaret. Ta má v modelu pozitivní vliv na spotřebu piva. Dílčí t-test je v pořádku, a tak tedy mohu vytvořit rovnici mého odhadu:

$$spotreba_piva = 3,3745 + 0,0181 \times tax_rate_cigarettes - 0,2782 \times castecny_zakaz - 0,6364 \times uplny_zakaz \quad (3.1)$$

Hodnota 3,375 udává průměrnou spotřebu piva v litrech na osobu za rok v časovém rozmezí roků 1998-2012 v zemích, které byly zahrnuty do modelu. Koeficient 0,018 u proměnné *tax_rate_cigarettes* lze interpretovat jako pozitivní vliv na vysvětlovanou proměnnou. Pokud bych měl dvě stejné země, které by ve stejném roce měly stejné ostatní proměnné a v první zemi by byla míra daňového zatížení krabičky cigaret o 1 % vyšší, tak bude mít tato první zem i spotřebu piva vyšší, a to o cca 0,018 litru na osobu na rok.

Stejným způsobem lze interpretovat zbývající dvě vysvětlující proměnné. Pokud bych měl dvě naprosto stejné země, ve stejném roce, se stejnou taxací cigaret a ani jedna by neměla zavedený jakýkoliv zákaz kouření a první země by zavedla částečný zákaz kouření, tak by v této zemi poklesla průměrná spotřeba piva o 0,278 litru na osobu na rok. V případě zavedení úplného zákazu kouření v první zemi, by poklesla průměrná spotřeba piva o 0,636 litru čistého alkoholu na osobu za rok. Pokud bych použil přepočít jako v datové sekci, kdy průměrné pivo má 4 % alkoholu, tak by to znamenalo pokles spotřeby o 15,9 litrů piva. Pokud bych toto číslo konkrétně vztáhl na Českou republiku, ve které nyní žije podle [21] 10 578 820 obyvatel, tak ročně se sníží spotřeba piva za ČR o cca 168 203 238 litrů. Toto číslo mě přijde jako příliš velká daň za zavedení protikuřáckého zákona. Podle posledních zdrojů [22][23] je Česká republika nejsvobodnější zemí Evropy, a právě zavedení protikuřáckého zákona by nám mohlo tuto první příčku vzít. Bohužel s touto politicko-

ekonomickou strategií přiblížit se, co nejvíce Západní a Severní Evropě přichází i velká ztráta nezaměnitelnosti České republiky a jejího působivého charakteru, pro který jsme známí ve světě.

3.2 Modely s trendem

V této části se zaměřím na možnost, že mi mé modely ovlivňuje trendová část, kterou jsem zatím nezahrnoval do modelu. Dále jsem poupravil seznam vysvětlovaných proměnných pro tyto modely a zahrnul jsem zde `tax_rate_cigarettes`, zpoždění 1 rok proměnné `castecny_zakaz`, zpoždění 1 rok proměnné `uplny_zakaz` a také proměnnou `year`, která zde je z důvodu trendu.

Tyto metody se označují jako FDL modely a je v nich nejdůležitější složkou zpoždění (*Lag*) po zavedení daného zákazu. S těmito modely se pojí více problémů jako je například multikolinearita mezi více zpožděními. Celkově jsou tyto modely vhodnější k použití na datové soubory s větší časovou osou. Jelikož mám v mých modelech časovou osu pouze 15 let, tak nepředpokládám vynikající výsledky. Nově je přidán model odhadnutý metodou ML, který je v programu STATA zastoupen funkcí *mixed*. Výstup je následující:

Výsledky					
	FE_rob	GEE_rob	RE_rob	OLS	MIXED
tax_rate_cigarettes	0.0232** (0.0079)	0.0232** (0.0078)	0.0230** (0.0079)	0.0128 (0.0337)	0.0230*** (0.0055)
L.law1	-0.0820 (0.1162)	-0.0816 (0.1152)	-0.0801 (0.1150)	0.1867 (0.3766)	-0.0798 (0.1014)
L.law2	-0.3119* (0.1323)	-0.3119* (0.1306)	-0.3120* (0.1286)	-0.1458 (0.5725)	- 0.3120*** (0.0926)
year	-0.0510** (0.0172)	-0.0509** (0.0171)	-0.0508** (0.0175)	-0.0361 (0.0735)	- 0.0507*** (0.0109)
Constant	105.1100** (34.1593)	105.0187** (34.1896)	104.6845** (34.8524)	75.9515 (145.8647)	104.6235*** (21.6451)
lns1_1_1 Constant					- 24.3103*** (3.9534)
lns1_1_2 Constant					0.3714** (0.1370)
lnsig_e Constant					- 0.9161*** (0.0427)
N	301.0000	301.0000	301.0000	301.0000	301.0000
r2	0.3035			0.0121	
r2_o	0.0098		0.0099		
r2_b	0.0126		0.0125		
r2_w	0.3035		0.3035		

Tabulka s výsledky
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabulka 3-2 Výsledky modelů s trendy

Prvním pohledem na tabulku zjistíme, že přidání trendu změnilo koeficienty u konstanty. Proto zde není možné z těchto modelů vhodně interpretovat hodnoty koeficientů proměnných. Statistickou významnost splňují všechny proměnné až na proměnnou *L.law1*, která popisuje zpoždění 1 rok po zavedení částečného zákazu. Toto připisuji malému vlivu a částečné propojenosti zavádění částečného a úplného zákazu kouření. V principu je jasné, že země nejdříve zavede částečný zákaz a poté plynule přejde na úplný zákaz. Zajímavá je trendová složka, která vychází ve všech modelech se záporným znaménkem. Z tohoto jde vyvodit fakt, že průměrná spotřeba piva napříč Evropou každým rokem postupně klesá.

Ostatní proměnné jsou se stejnými matematickými znaménky jako v modelech bez trendové složky. Zavedení trendu nám i částečně vzrostla hodnota vnitroskupinového koeficientu determinace R^2 na 0,3035. Což říká, že pomocí modelu se zpožděními jsem schopen popsat 30 % vnitroskupinové variability vysvětlované proměnné spotřeba piva. Tento výsledek už je celkem zajímavý vzhledem k faktu, že do hodnoty průměrné spotřeby piva jsou započtena všechna piva vypitá daným národem tedy i lahvová. Jelikož se v hospodách prodává hlavně pivo čepované, tak nedokážu popsat svým modelem piva, která jsou vypita mimo restaurační zařízení, na které se nevztahuje protikuřácký zákon. Nicméně i v těchto modelech se potvrdila má teze, že protikuřácký zákon má negativní vliv na spotřebu piva.

4 Diskuze

V mé práci byly sledovány 41 zemí Evropy a přilehlých států za období 15 let mezi roky 1998 až 2012. Výsledkem mého zkoumání je poznatek, že protikuřácký zákon negativně působí na spotřebu piva. V modelech byly použity fixní a náhodné efekty a robustní VCE, dále jsem také nahlížel na možnost promítnutí trendové složky do mých modelů. Stejných výsledků dosáhli ve své studii [7] i M. Kvasnicka s H. Tauchman, kteří dokázali negativní vliv protikuřáckého zákona na výši tržeb v restauračních zařízeních v Německu. Použili ve svých modelech proměnné *Rainfall* a *Unemployment*, které měli zlepšit model, aby co nejlépe popisoval skutečnost.

Naproti tomu se staví studie [13] provedená v Mexiko City, kde autorům vyšel pozitivní vliv protikuřáckého zákona na výši tržeb a zaměstnanost. Při této studii, ale nebyla zjištěna statistická významnost u proměnných, takže jsou tyto výsledky pouze k zamyšlení. Podle mého názoru je celkem limitující fakt při těchto studiích, že nejsou dostupné informace za jednotlivé restaurační zařízení z celé země. Výběrový vzorek může být často zkreslený a nepoužitelný pro aplikování na celou společnost

Obecně téma protikuřáckého zákona je celkem aktuální a předmětem zkoumání mnoha lidí z prostředí ekonometrie a statistiky. Bohužel výsledky jsou zatím nevalné, ale to bych přisouval celkem krátkým zkoumaným časovým obdobím. Vzhledem k tomu že Světová zdravotnická organizace WHO bojuje s tabákovou epidemií už několik dekad a až v posledních pár letech se začíná postupně všude po světě zavádět protikuřácký zákon, tak nás teprve výsledky jejich snažení čekají. Na druhou stranu záleží na každém člověku, jestli chce kouřit nebo ne, je to jeho svobodná volba, jak naloží se svým zdravím. Nejhorším důsledkem této tabákové epidemie je razantní nárůst smrtelných onemocnění lidí, kteří kouří tzv. z druhé ruky. Proto bych i já chtěl podpořit tento protikuřácký zákon, aby lidé, kteří nekouří, nemuseli trpět svobodnou volbou druhých, kteří kouřit chtějí.

Závěr

Cílem této práce bylo zjistit pomocí ekonometrické analýzy závislost spotřeby piva na protikuřáckém zákonu. Na začátku tvorby datového souboru jsem narazil na vícero překážek. Jednalo se především o neexistující nebo nedohledatelná data. Celkově tvorba datového souboru a přípravy pro vhodné zpracování v programu STATA mi zabrala velkou část času, a to i z důvodu, že data musela být převedena do *Long* formátu. Po dokončení datové souboru jsem se pustil do deskriptivních statistik. Z výše uvedeného grafu Obrázek 2-2 lze celkem jasně vyčíst, že Česká republika je na nejvyšších pozicích, co se spotřeby piva týče. Zajímavým číslem je i průměrná spotřeba piva v litrech čistého alkoholu každého obyvatele pozorovaných zemí za rok a to 3,837 litru.

V nejrozsáhlejší praktické části jsem se zaměřil na hledání nejlepšího modelu, který by splňoval testovací kritéria a byl vhodný k interpretaci. Při hledání jsem používal v programu STATA funkce jako např. *regress*, *xtreg*, *xtgee*, *areg* a využíval jejich rozšíření, abych zamezil vlivu robustní statistické indukce. Tyto rozšíření jsou zastoupeny v programu jako VCE (robust) a také VCE (cluster). Dále jsem se zabýval fixními a náhodnými efekty v modelech. Fixní efekty mohou být dvojího typu buď se zafixovaným časem nebo zafixovanou vlastností. K porovnání vhodnosti použití fixní nebo náhodných efektů slouží *Hausmanův* robustní test. Ten nezamítnul hypotézu, a tudíž není v mém datovém souboru velký vliv fixních efektů. Je to dáno také tím, že mám v datovém souboru malý počet vysvětlujících proměnných.

Po pečlivém zkoumání jsem vybral jako nejlepší model funkce *xtreg* s fixními efekty a robustní VCE. Hodnota vnitroskupinového koeficientu determinace R^2 vyšla 0,2351 respektive 0,3035 v modelu s trendem, což lze interpretovat, že model popsal 23 % (30 %) vnitroskupinové variability u vysvětlované proměnné spotřeba piva. Toto číslo není velké, ale s tím jsem předpokládal už v úvodu své práce. Protože naleznout ekonometrický model, který významně popisuje variabilitu vysvětlované proměnné, není jednoduché. Navíc se zde promítnul problém s neúplností datového souboru, protože bylo do modelu vybráno pouze 316 z celkových 579 pozorováními spotřeby piva.

Hlavní účel, proč jsem tuto bakalářskou práci vypracoval, byl dosažen. Ve všech modelech byly koeficienty u vysvětlujících proměnných částečný zákaz a úplný zákaz se zápornými znaménky, a tudíž i negativním efektem na spotřebu piva. Důsledkem zavedení úplného protikuřáckého zákazu bude skoro s jistotou i to, že neobhájíme v příštím roce 2018 první pozici mezi Evropskými státy v *Nanny State Indexu*.

Literatura

- [1] WOOLDRIDGE, J M. *Econometric analysis of cross section and panel data*. Cambridge: MIT Press, 2010. ISBN 978-0-262-23258-6.
- [2] WOOLDRIDGE, J M. *Introductory econometrics : a modern approach*. Boston: Cengage Learning, 2016. ISBN 978-1-305-27010-7.
- [3] CAMERON, A C. -- TRIVEDI, P K. *Microeconometrics : methods and applications*. New York: Cambridge University Press, 2005. ISBN 0-521-84805-9.
- [4] CAMERON, A C. -- TRIVEDI, P K. *Microeconometrics using Stata*. College Station: Stata Press, 2010. ISBN 978-1-59718-073-3.
- [5] Anger, Silke; Kvasnicka, Michael; Siedler, Thomas (2011) : One Last Puff? Public Smoking Bans and Smoking Behavior, *Journal of Health Economics*, ISSN 0167-6296, Vol. 30, Iss. 3, pp. 591-601, <http://dx.doi.org/10.1016/j.jhealeco.2011.03.003>
- [6] Bernat, Debra H., Mildred Maldonado-Molina, Andrew Hyland, and Alexander C. Wagenaar (2013): Effects of smoke-free laws on alcohol-related car crashes in California and New York: time series analyses from 1982 to 2008. *American journal of public health* 103, no. 2: 214-220.
- [7] Michael Kvasnicka, Harald Tauchmann. Much Ado About Nothing? Smoking Bans and Germany's Hospitality Industry. *Applied Economics*, Taylor & Francis (Routledge), 2011, pp.1. <10.1080/00036846.2011.591739>
- [8] Joaquin Barnoya, MD, MPH and Ana Navas-Acien, MD, PhD : Protecting the World From Secondhand Tobacco Smoke Exposure: Where Do We Stand and Where Do We Go From Here?, *Nicotine Tob Res*: 789–804, 10.1093/ntr/nts200
- [9] Hans Olav Melberg, Karl E. Lund: Do smoke-free laws affect revenues in pubs and restaurants?, *Eur J Health Econ* (2012) 13:93–99, DOI 10.1007/s10198-010-0287-6
- [10] Dong-Chul Seo, Jonathan T. Macy, Mohammad R. Torabi, Susan E. Middlestadt: The effect of a smoke-free campus policy on college students' smoking behaviors and attitudes, D.-C. Seo et al. / *Preventive Medicine* 53 (2011) 347–352
- [11] Levy D, de Almeida LM, Szklo A (2012) The Brazil SimSmoke Policy Simulation Model: The Effect of Strong Tobacco Control Policies on Smoking Prevalence and Smoking-Attributable Deaths in a Middle Income Nation. *PLoS Med* 9(11): e1001336. doi:10.1371/journal.pmed.1001336
- [12] Luca Pieroni , Manuela Chiavarini , Liliana Minelli , Luca Salmasi: The role of anti-smoking legislation on cigarette and alcohol consumption habits in Italy, *Health Policy* 111 (2013) 116–126, <http://dx.doi.org/10.1016/j.healthpol.2013.04.001>

- [13] Carlos Manuel Guerrero Lo'pez, Jorge Alberto Jime'nez Ruiz, Luz Myriam Reynales Shigematsu, Hugh R Waters: The economic impact of Mexico City's smoke-free law, *Tobacco Control* (2011). doi:10.1136/tc.2010.036467
- [14] Jacqueline Tumwine: Implementation of the Framework Convention on Tobacco Control in Africa: Current Status of Legislation, *Int. J. Environ. Res. Public Health* **2011**, 8, 4312-4331; doi:10.3390/ijerph8114312, ISSN 1660-4601

Internetové zdroje:

- [15] <http://data.euro.who.int/hfadb/>: Data o spotřebě piva
- [16] <http://databank.worldbank.org/data/reports.aspx?source=2&series=NY.GDP.PCAP.KD&country> : Data o GDP
- [17] <https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp> : Data o taxaci cigaret
- [18] https://en.wikipedia.org/wiki/List_of_smoking_bans : Zdroj k vytvoření proměnné zavedení protikuřáckého zákona v čase
- [19] https://cs.wikipedia.org/wiki/Z%C3%A1kaz_kou%C5%99en%C3%AD_podle_zem%C3%AD#cite_note-14 : Zdroj k vytvoření proměnné zavedení protikuřáckého zákona v čase
- [20] https://ec.europa.eu/health/sites/health/files/tobacco/docs/smoke-free_legislation_overview_en.pdf : Zdroj k vytvoření proměnné zavedení protikuřáckého zákona v čase
- [21] <http://www.ceskenoviny.cz/zpravy/pocet-obyvatel-cr-loni-vzrostl-na-10-58-milionu-hlavne-diky-migraci/1463711> : Počet obyvatel ČR, zobrazeno 22.5.2017
- [22] <https://www.novinky.cz/ekonomika/437372-cesko-je-nejsvobodnejsi-zemi-eu-tvrdi-ekonomicky-institut-protikuracky-zakon-to-muze-zmenit.html> : Nanny State Index, zobrazeno 21.5.2017
- [23] <http://nannystateindex.org/index-2017/> : Nanny State Index, zobrazeno 22.5.2017
- [24] https://www.mzcr.cz/obsah/svetova-zdravotnicka-organizace_3115_8.html : FCTC, zobrazeno 14.1.2017
- [25] <http://euractiv.cz/factsheet/ekonomika-a-euro/spotrebni-dan-na-cigarety-v-cesku-a-v-eu-000114/> : Daňové zatížení cigarette, zobrazeno 19.5.2017