

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Trendy ve spotřebě alkoholu v České republice
a okolních zemích



Katedra matematické analýzy a aplikací matematiky

Vedoucí diplomové práce: **Mgr. Ondřej Vencálek, Ph.D.**

Vypracoval(a): **Václav Číkl**

Studijní program: N1103 Aplikovaná matematika

Studijní obor Aplikace matematiky v ekonomii

Forma studia: prezenční

Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Václav Číkl

Název práce: Trendy ve spotřebě alkoholu v České republice a okolních zemích

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: Mgr. Ondřej Vencálek, Ph.D.

Rok obhajoby práce: 2017

Abstrakt: Cílem práce je analýza dat o spotřebě alkoholu v české republice a okolních zemích. Pokusit se najít vhodné modely pro vyrovnání dat, provést analýzu strukturálních změn v souboru a zjistit možné závislosti jak mezi konzumací jednotlivých druhů alkoholu, tak na úrovni jednotlivých států.

Klíčová slova: Trendy, časové řady, joinpoint regrese, strukturální změny

Počet stran: 76

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Václav Číkl

Title: Trends in alcohol consumption in the Czech Republic and neighbouring countries

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: Mgr. Ondřej Vencálek, Ph.D.

The year of presentation: 2017

Abstract: The aim of this thesis is to analyze alcohol consumption data in the Czech Republic and neighbouring countries. Try to find appropriate models for data, analyze structural changes in datasets and identify potential dependency between alcohol consumption.

Key words: Trends, Time series, joinpoint regression, structural changes

Number of pages: 76

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně pod vedením pana Mgr. Ondřeje Vencálka, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	7
I Teoretická část	9
1 Časové řady	9
1.1 Dekompozice časových řad	9
1.1.1 Speciální případy trendových funkcí	11
1.1.2 Detekce změn v časových řadách	21
1.2 Míry vhodnosti modelu	25
2 Joinpoint regrese	29
3 Korelační analýza	31
3.1 Korelační koeficient	31
3.2 Korelační matice	33
4 Lokální regrese	35
II Praktická část	38
5 Použitá data	38
5.1 Metodologie dat	39
5.2 Náhled na data	39
6 Analýza dat	44
6.1 Číselné charakteristiky	44
6.2 Proložení dat trendovými funkcemi	45
6.2.1 Lineární trend	46
6.2.2 Kvadratický trend	47
6.2.3 Exponenciální trend	51
6.3 Proložení dat joinpoint regresí	52
6.4 Korelační analýza	63
6.4.1 Korelační mapy	63
6.4.2 Proložení bodových grafů lokální regresí	66
Literatura	72
Seznam obrázků	73
Seznam tabulek	76

Poděkování

Na tomto místě bych chtěl poděkovat především svému vedoucímu diplomové práce panu Mgr. Ondřeji Vencálkovi, Ph.D., že měl se mnou dostatek trpělivosti, aby mi pomohl dovést tuto práci ke zdárnému konci. Také bych rád poděkoval své rodině, přátelům a všem, kteří mě po celou dobu studia podporovali.

Úvod

K sepsání diplomové práce na téma „Trendy ve spotřebě alkoholu v České republice a okolních zemích“ mě „hnalo“ více motivů. Tím prvním bylo jistě mé okolí, které se mě vždy tázalo, k čemu je studium matematiky vlastně dobré a jak se dá využít v praxi. Dalo by se tedy říci, že jsem zvolil takové téma, které zajme již při přečtení nadpisu a pevně doufám, že přiláká potenciální čtenáře alespoň k nahlédnutí. Dalším motivem byly téměř pravidelné články v tiskovinách o nadměrné konzumaci alkoholu v České republice, ve kterých byly využity vždy jen částečné informace, čímž docházelo k prohloubení mé zvědavosti ohledně toho, jak to tedy ve skutečnosti je a zda je na tom Česká republika, co se konzumace alkoholu týče, skutečně tak špatně. Nelze také opomenout, že jedním z motivů bylo sepsání práce v rámci završení mého dvouletého navazujícího magisterského studia aplikované matematiky v ekonomii.

Práce je rozdělena na dvě části, teoretickou a praktickou. Jednotlivé části jsou pak děleny do podkapitol. První část se věnuje základním teoretickým aspektům, které je dobré znát při části praktické. V první kapitole jsou postupně vysvětleny základy dekompozičního přístupu při analýze časových řad, míry pro posouzení vhodnosti modelu a detekce změn v časových řadách. Druhá kapitola je věnována joinpoint regresi. V následující kapitole jsou popsány základy korelační analýzy a poslední kapitola z teoretické části se zabývá lokální regresi. Na závěr jednotlivých kapitol v teoretické části budou přidány některé možnosti, jakým způsobem lze problematiku řešit pomocí softwaru R.

Praktická část se věnuje analýze dat o spotřebě alkoholu v České republice a okolních zemích v letech 1961 - 2014, které jsou veřejně dostupné v databázi Světové zdravotnické organizace. Podstatná část analýzy se bude zabývat nalezením vhodného modelu, který by se vhodně dokázal přizpůsobit datům a zároveň si zachoval určitou jednoduchost. Vhodnost modelu bude posouzena dle charakteristik uvedených v teoretické části a v případě nalezení více vhodných modelů budou tyto modely srovnány. Dále budou analyzovány strukturální změny v datech

o spotřebě alkoholu. Poslední částí práce s daty je analýza závislostí na základě korelace. Předmětem zájmu budou především závislosti úrovní konzumace alkoholu mezi jednotlivými státy nebo závislosti úrovní konzumace jednotlivých složek celkové spotřeby alkoholu v rámci jednotlivých států. Na závěr analýzy bude rozebráno, jak se sledované závislosti mění v čase.

K provedení analýzy dat bude využit freewarový software R, ve kterém budou zpracovány všechny grafy a výpočty. Práce je sázena v programu L^AT_EX.

Část I

Teoretická část

1. Časové řady

Vzhledem k velice širokému spektru využitelnosti analýzy časových řad můžeme říct, že jde o jedno z nejdůležitějších oblastí statistiky. Dnes se s časovými řadami setkáváme téměř na každém kroku, v téměř všech oblastech lidského bádání. Může se jednat například o vývoj HDP v jednotlivých zemích, o EEG záznamy, o extrémny denních teplot nebo také o sledování úrovně konzumace alkoholu. Základní myšlenkou časových řad je konstrukce modelu, který co nejvíce odpovídá reálnému chování časové řady. Zároveň však dbáme na to, aby zkontruovaný model nebyl příliš složitý a nebylo obtížné s ním pracovat a porozumět mu. Aby data mohla tvořit časovou řadu je nutné, aby byla chronologicky uspořádána, viz následující definice.

Definice 1 ([3]). *Časovou řadou rozumíme řadu věcně a prostorově srovnatelných hodnot příslušného statistického znaku, uspořádanou z hlediska času ve směru od minulosti do přítomnosti.*

Časovou řadu lze nejjednodušším způsobem vyjádřit jako soubor pozorování y_1, y_2, \dots, y_n , kde y_t je hodnota sledovaného znaku Y , $t = 1, \dots, n$ značí časovou proměnnou a n je počet pozorování.

1.1. Dekompozice časových řad

K modelování časových řad existuje více přístupů. Mezi základní přístupy můžeme zařadit dekompoziční přístup, Box - Jenkinsovu metodologii nebo například spektrální analýzu. Každý z těchto přístupů má své klady a zápory. Box - Jenkinsova metoda je například limitována délkou časových řad, kdy se nedoporučuje ji využívat pro řady, kde $n < 50$. Naopak je tato metoda daleko fle-

xibilnější než dekompoziční přístup, protože dokáže podstatně rychleji reagovat na změny charakteru časové řady. V další části textu se budeme zabývat právě dekompozičním přístupem.

V dekompozičním přístupu, jak již název vypovídá, pracujeme s rozložením časové řady na jednotlivé složky. Těmito složkami jsou:

1. Trend - T_t
2. Sezónní složka - S_t
3. Cyklická složka - C_t
4. Reziduální (náhodná) složka - E_t

Každá z těchto složek odpovídá určitému chování časové řady. Základní složkou je trend, který odráží dlouhodobé chování časové řady. Sezónní složka popisuje opakující se změny v časové řadě, které se dějí během jednoho kalendářního roku. Faktorem, který působí na sezónní složku, tak může být například střídání ročního období. Cyklická složka odráží dlouhodobější periodické změny. Příkladem může být hospodářský cyklus nebo volební období. Poslední složka svou povahou odbočuje od předchozích a nepočítá se mezi tzv. systematické složky časových řad. Náhodná složka pokrývá nejen náhodné pohyby, ale také chyby při samotném měření a zpracování dat.

Existují dva typy dekompozice:

$$y_t = T_t + S_t + C_t + E_t$$

a

$$y_t = T_t S_t C_t E_t$$

Rozdíl mezi tvary je v tom, že při aditivním tvaru dekompozice uvažujeme všechny složky ve svých absolutních hodnotách a stejných jednotkách, v nichž je měřena veličina y_t . Při multiplikativním tvaru je v absolutních hodnotách a stejných jednotkách jako y_t uvažována pouze trendová složka T_t . Ostatní složky jsou bezrozměrné a v relativních hodnotách vůči trendu. Mezi jednotlivými složkami lze přecházet pomocí logaritmické transformace, avšak ta má za následek změnu vlastností reziduální složky.

1.1.1. Speciální případy trendových funkcí

Jak už jsme zmínili, je nutné, aby zkonstruovaný model byl co nejjednodušší, ale zároveň aby vystihoval do určité míry realitu. V otázce jednoduchosti nám může pomoci proložení časové řady některou z jednoduchých křivek (přímka, parabola, exponenciální křivka atp.). Pro takto modelované řady je jednoduché odhadnout přibližně budoucí hodnoty. Samozřejmě za předpokladu, že trend se v čase drasticky nemění.

Pro proložení časové řady trendovou křivkou uvažujeme řadu ve tvaru

$$y_t = T_t + E_t$$

Běžně využíváme některý z následujících trendových funkcí:

1. Konstantní trend:

$$T_t = \beta_0, \quad t = 1, \dots, n$$

Jde o nejjednodušší možnou trendovou funkci. Vyrovnané hodnoty jsou konstantní. Odhad $\hat{\beta}_0$ parametru β_0 získáme metodou nejmenších čtverců (dále jen MNC) jako výběrový průměr časové řady.

$$\hat{y}_t = \hat{\beta}_0 = \frac{1}{n} \sum_{t=1}^n y_t = \bar{y}$$

Stejně jako vyrovnané hodnoty, tak i bodová předpověď je konstantní a je rovna odhadu $\hat{\beta}_0$. Intervalová předpověď je dána následovně:

$$\hat{y}_{n+h}^P = \bar{y} \pm \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n}\right)} t_{n-1, \alpha},$$

kde $\hat{\sigma}^2$ je výběrový rozptyl, viz 1.1 a $t_{n-1, \alpha}$ je kritická hodnota t rozdělení s $n - 1$ stupni volnosti na hladině významnosti α . Konstantní trend volíme v případě, kdy první diference časové řady $(y_t - y_{t-1})$ kolísají okolo 0.

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n (y_t - \hat{y})^2}{n - 1} \quad (1.1)$$

2. Lineární trend:

$$T_t = \beta_0 + \beta_1 t, \quad t = 1, \dots, n$$

Trend ve tvaru přímky obvykle volíme v případě, že první diference časové řady jsou přibližně konstantní a nenulové. Parametry β_0 a β_1 opět odhadujeme pomocí MNČ, jejíž postup si tentokrát alespoň nastíníme.

U MNČ nám jde v zásadě o to, abychom minimalizovali odchylky vyrovnaných hodnot od těch skutečných. Toho docílíme minimalizováním výrazu

$$\min_{\beta_0, \beta_1} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)^2. \quad (1.2)$$

Výraz 1.2 zderivujeme podle β_0 a β_1

$$\frac{\partial}{\partial \beta_0} = -2 \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t),$$

$$\frac{\partial}{\partial \beta_1} = -2 \sum_{t=1}^n t(y_t - \beta_0 - \beta_1 t)$$

a položíme rovno nule a vydělíme -2 :

$$\sum_{t=1}^n (y_t - \beta_0 - \beta_1 t) = 0,$$

$$\sum_{t=1}^n t(y_t - \beta_0 - \beta_1 t) = 0.$$

Vyřešením soustavy tzv. normálních rovnic (např. Cramerovým pravidlem) dostáváme odhady parametrů β_0 a β_1

$$\hat{\beta}_0 = \frac{\sum_{t=1}^n y_t - \hat{\beta}_1 \sum_{t=1}^n t}{n} = \bar{y} - \hat{\beta}_1 \bar{t},$$

$$\hat{\beta}_1 = \frac{n \sum_{t=1}^n t y_t - \sum_{t=1}^n t \sum_{t=1}^n y_t}{n \sum_{t=1}^n t^2 - (\sum_{t=1}^n t)^2}.$$

Vyrovnané hodnoty dostáváme dosazením odhadů parametrů jako

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t.$$

Bodová předpověď lineárního trendu je dána

$$\hat{y}_{n+h}^P = \hat{\beta}_0 + \hat{\beta}_1(n+h)$$

a intervalová předpověď jako

$$\hat{y}_{n+h}^P \pm \sqrt{\hat{\sigma}^2 f_{n+h} t_{n-2, \alpha}},$$

kde

$$f_{n+h} = 1 + (\mathbf{1}, \mathbf{n} + \mathbf{h})(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{1}, \mathbf{n} + \mathbf{h})^T,$$

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n-p}.$$

Matrice \mathbf{X} zde značí tzv. designovou matici, která je tvořena n (počet pozorování) řádky a k (počet vlastních regresorů) sloupce. V případě, že pracujeme s modelem, ve kterém pracujeme s absolutním členem máme $k+1$ sloupců, kde první sloupec je tvořen 1. Přesnou podobu naší matice \mathbf{X} můžeme vidět v rovnici 1.3. p značí počet parametrů. V našem případě je $p=2$.

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & n \end{pmatrix} \quad (1.3)$$

Pro zjednodušení odhadu parametrů se dá pracovat s vyjádřením ve tvaru

$$T_t = \gamma_0 + \gamma_1(t - \bar{t}), \quad t = 1, \dots, n. \quad (1.4)$$

Zjednodušení spočívá v tom, že

$$\sum_{t=1}^n (t - \bar{t}) = 0,$$

kde

$$\bar{t} = \sum_{t=1}^n \frac{t}{n},$$

čímž se zjednoduší soustava normálních rovnic na tvar

$$nc_0 = \sum_{t=1}^n y_t,$$

$$c_1 \sum_{t=1}^n (t - \bar{t})^2 = \sum_{t=1}^n (t - \bar{t})y_t,$$

kde

$$c_0 = \bar{y}$$

a

$$c_1 = \frac{\sum_{t=1}^n (t - \bar{t})y_t}{\sum_{t=1}^n (t - \bar{t})^2}.$$

3. Kvadratický trend:

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, \dots, n.$$

Parabolu volíme obvykle v případě, kdy řada prvních diferencí má cca lineární trend a řada druhých diferencí ($y_t - 2y_{t-1} + y_{t-2}$) je blízká konstantnímu trendu. Parametry β_0, β_1 a β_2 opět odhadujeme pomocí MNČ. I zde často pracujeme s vyjádřením trendu v podobě 1.5.

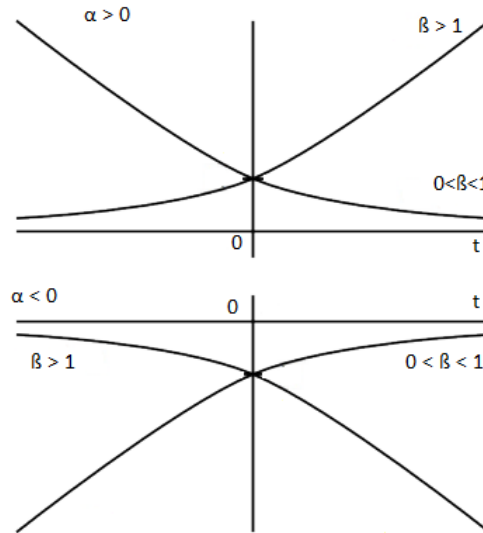
$$T_t = \gamma_0 + \gamma_1(t - \bar{t}) + \gamma_2(t - \bar{t})^2, \quad t = 1, \dots, n. \quad (1.5)$$

V soustavě normálních rovnic pak lze využít toho, že

$$\sum_{t=1}^n (t - \bar{t}) = \sum_{t=1}^n (t - \bar{t})^3 = 0.$$

Vyrovnané hodnoty mají tvar

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2.$$



Obrázek 1: Exponenciální trend

Bodová a intervalová předpověď pak má tvar

$$\hat{y}_{n+h}^P = \hat{\beta}_0 + \hat{\beta}_1(n+h) + \hat{\beta}_2(n+h)^2,$$

$$\hat{y}_{n+h}^P \pm \sqrt{\hat{\sigma}^2 f'_{n+h} t_{n-2, \alpha}},$$

kde

$$f'_{n+h} = 1 + (\mathbf{1}, \mathbf{n} + \mathbf{h}, (\mathbf{n} + \mathbf{h})^2)(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{1}, \mathbf{n} + \mathbf{h}, (\mathbf{n} + \mathbf{h})^2)^T$$

a

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n-3}.$$

Designová matice \mathbf{X} má podobu:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1^2 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{pmatrix}$$

4. Exponenciální trend:

$$T_t = \alpha \beta^t, \quad t = 1, \dots, n, \quad \beta > 0 \quad (1.6)$$

Jednoduchý exponenciální trend je vhodný, pokud řada koeficientů růstu y_t/y_{t-1} nebo řada prvních diferencí $(\ln y_t - \ln y_{t-1})$ kolísá okolo konstantní hodnoty různé od nuly. Exponenciální trend není lineární v parametrech, proto se v této podobě nedá pro odhad parametrů využít metoda MNČ. Proto je na trendovou funkci nutné uplatnit logaritmickou transformaci 1.7.

$$\ln T_t = \ln \alpha + (\ln \beta)t = \gamma_0 + \gamma_1 t. \quad (1.7)$$

Pomocí této transformace dostáváme klasický lineární trend. Můžeme tedy využít MNČ a odhadnout parametry γ_0 a γ_1 . Inverzní transformací (exponenciální) dostáváme odhady parametrů pro původní trendovou funkci 1.6:

$$\hat{\alpha} = e^{\hat{\gamma}_0},$$

$$\hat{\beta} = e^{\hat{\gamma}_1}.$$

Vyrovnané hodnoty jsou ve tvaru

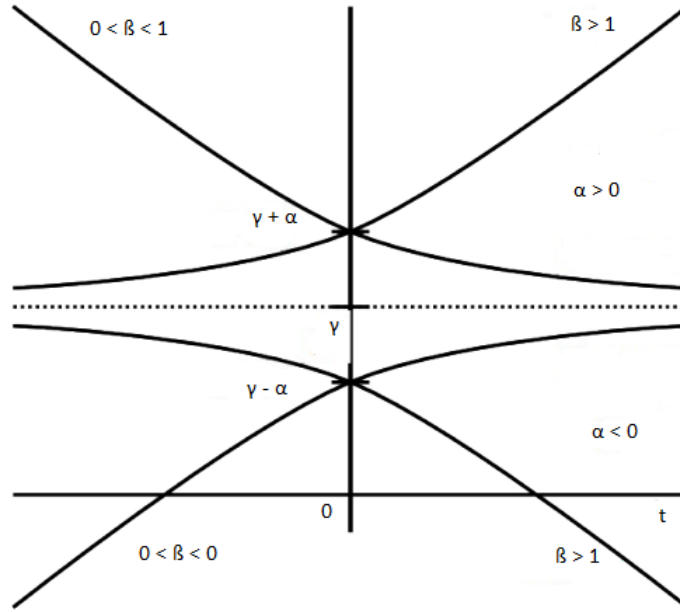
$$\hat{y}_t = \hat{\alpha}\hat{\beta}^t.$$

5. Modifikovaný exponenciální trend:

$$T_t = \gamma + \alpha\beta^t, \quad t = 1, \dots, n, \quad \beta > 0$$

Modifikovaný exponenciální trend je zobecněním jednoduchého exponenciálního trendu a je vhodný pro případy, kdy podíl sousedních diferencí kolísá okolo nenulové konstanty. Vzhledem ke svému posunu nelze pomocí logaritmické transformace převést na lineární trend. Z toho důvodu se pro odhad parametrů γ , α a β využívá metoda částečných součtů.

Princip této metody spočívá v tom, že rozdělíme soubor na tři stejně velké části o délce m a uděláme součty jednotlivých pozorování v rámci každé části. Vzhledem k tomu, že ne vždy musíme mít soubor, který lze rozdělit na tři stejně velké části, tak v případě, že $n \neq 3m$ odřízneme některá pozorování ze začátku řady.



Obrázek 2: Modifikovaný exponenciální trend

Dostáváme tak soustavu 3 rovnic o třech neznámých.

$$\sum_1 y_t \sim \sum_1 T_t = m\gamma + \frac{\alpha\beta(\beta^m - 1)}{\beta - 1}$$

$$\sum_2 y_t \sim \sum_2 T_t = m\gamma + \frac{\alpha\beta^{m+1}(\beta^m - 1)}{\beta - 1}$$

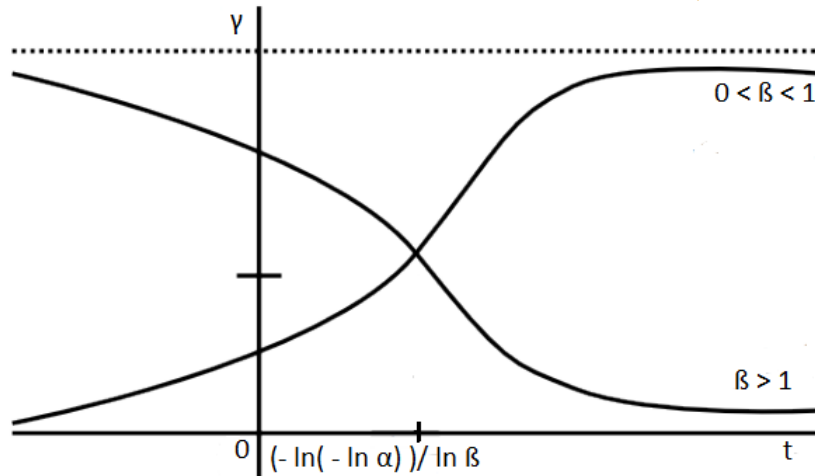
$$\sum_3 y_t \sim \sum_3 T_t = m\gamma + \frac{\alpha\beta^{2m+1}(\beta^m - 1)}{\beta - 1}.$$

Vyřešením této soustavy dostáváme odhady $\hat{\gamma}$, $\hat{\alpha}$ a $\hat{\beta}$

$$\hat{\beta} = \left(\frac{\sum_3 y_t - \sum_2 y_t}{\sum_2 y_t - \sum_1 y_t} \right)^{1/m},$$

$$\hat{\alpha} = \frac{b-1}{b(b^m-1)^2} \left(\sum_2 y_t - \sum_1 y_t \right),$$

$$\hat{\gamma} = \frac{\sum_1 y_t - \frac{ab(b^m-1)}{(b-1)}}{m}.$$



Obrázek 3: Gompertzova křivka

Další přístup k odhadu parametrů modifikovaného exponenciálního parametru je založen na volbě pevné hodnoty parametru β . Dostáváme

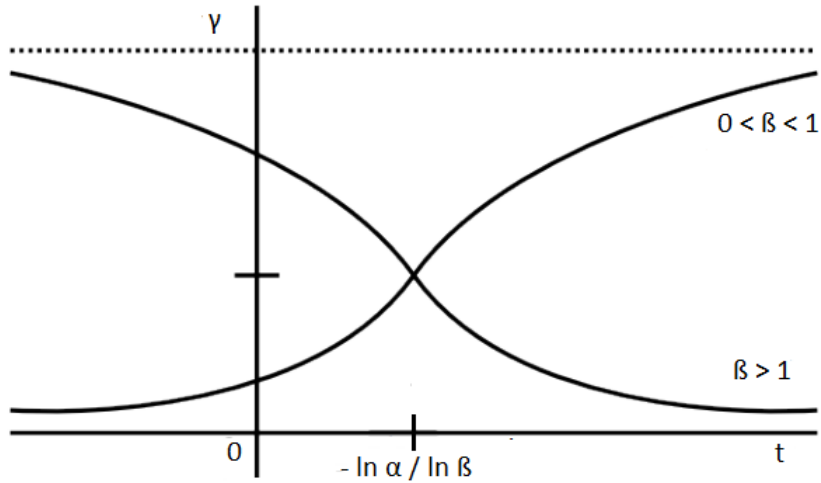
$$T_t = \gamma + \alpha\beta^t = \gamma + \alpha + f_\beta(t),$$

kde funkci $f_\beta(t)$ známe. Tím pádem přecházíme k lineárnímu trendu a můžeme odhadnout zbylé parametry. Problémem však je samotná volba hodnoty parametru β . Proto volíme více hodnot a následně pak vybíráme trojici $(\hat{\gamma}, \hat{\alpha}, \hat{\beta})$ s nejmenším residuálním součtem čtverců (dále jen RSČ).

6. Gompertzova křivka:

$$T_t = \gamma\alpha^{\beta^t}, \quad t = 1, \dots, n, \quad \alpha, \beta > 0$$

Gompertzova křivka vzniká transformací modifikovaného exponenciálního trendu. Signálem pro použití tohoto trendu je kolísání hodnot podílů sousedních diferencí $(\ln y_t - \ln y_{t-1})/(\ln y_{t-1} - \ln y_{t-2})$ okolo nenulové konstanty. Využití najdeme především v ekonomii. Příkladem může být odbyt nového produktu na trhu. Na začátku uvedení roste pomalu, až dosáhne do bodu, kdy si na trhu najde své místo a odběratele. Poté odbyt roste až do svého maxima. Na obrázku 14 můžeme vidět, že Gompertzova křivka je ve tvaru nesymetrické S křivky (kolem bodu inflexe). Odhady parametrů Gompert-



Obrázek 4: Logistický trend

zovy křivky provádíme pomocí logaritmické transformace, čímž docházíme k modifikovanému exponenciálnímu trendu.

$$\ln T_t = \ln \gamma + \ln \alpha \beta^t = \delta + \omega \varphi^t$$

7. Logistický trend:

$$T_t = \frac{\gamma}{1 + \alpha \beta^t}, \quad t = 1, \dots, n, \quad \beta, \gamma > 0$$

Posledním základním trendem, který si zde uvedeme, je logistický trend. Tento trend ve tvaru symetrické S křivky (podle bodu inflexe) je vhodný použít v případě, že řada přirozených logaritmů hodnot y_t má přibližně hyperbolický průběh. Parametry logistického trendu můžeme odhadnout inverzní transformací, tedy převedením na modifikovaný exponenciální trend nebo lze využít princip tzv. diferenčních odhadů parametrů, kdy neppracujeme s hodnotami řady y_t ale s řadou $y_t - y_{t-1}$. Inverzní transformace vypadá následovně:

$$\frac{1}{T_t} = \frac{1 + \alpha \beta^t}{\gamma} = \frac{1}{\gamma} + \frac{\alpha}{\gamma} \beta^t = \gamma^* + \alpha^* \beta^t.$$

Pro samotný výpočet odhadů parametrů se obecně spíše využívá vzorec 1.8, který

vychází ze samotné metody nejmenších čtverců a využívá designovou matici \mathbf{X} .

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (1.8)$$

Právě využití tohoto vzorce je v praxi často jednodušší, než dosazovat do vztahů pro jednotlivé parametry, případně vůbec tyto vztahy odvozovat, jak ukážeme na příkladu níže.

Příklad 1. Uvažujme časovou řadu úrovně konzumace alkoholu v Polsku v letech 1961 - 2014, kterou budeme chtít proložit lineárním trendem. Časová řada je k nahlédnutí v tabulce na přiloženém CD. Víme, že máme 54 pozorování a vzhledem k tomu, že budeme uvažovat lineární trend včetně absolutního členu β_0 , máme celkově 2 parametry. Designová matice tedy vypadá takto:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 54 \end{pmatrix}$$

Součin matic $\mathbf{X}^T \mathbf{X}$ má podobu čtvercovou matici 2x2:

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 54 & 1485 \\ 1485 & 53955 \end{pmatrix}$$

Inverzí $\mathbf{X}^T \mathbf{X}$ dostaneme:

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} 0.076170 & -0.002096 \\ -0.002096 & 0.000762 \end{pmatrix} \quad (1.9)$$

Součinem transponované designové matice a vektoru pozorování \mathbf{y} dostáváme:

$$\mathbf{X}^T \mathbf{y} = \begin{pmatrix} 168.01 \\ 5754.44 \end{pmatrix} \quad (1.10)$$

Na závěr již stačí jen vynásobit mezi sebou matice 1.9 a 1.10 a dostáváme odhady parametrů $\hat{\boldsymbol{\beta}}$.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \begin{pmatrix} 0.73359 \\ 0.08646 \end{pmatrix}.$$

Dá se říci, že při práci s jednoduchými trendy si uživatel v softwaru R vystačí pouze s funkcemi pro lineární regresi `lm`, případně pro zobecněnou lineární regresi `glm`. Před samotnou prací je však vhodné data na časovou řadu převést pomocí funkce `ts`. Pro samotné predikce pak může posloužit balíček `forecast`, ze kterého lze pro predikci v lineárních modelech využít funkci `forecast.lm`.

1.1.2. Detekce změn v časových řadách

V řadě praktických aplikací analýzy časových řad si během analýzy nemusíme vystačit pouze s jedním trendem nebo modelem. Ba naopak je daleko pravděpodobnější, že se v čase parametry trendových funkcí mění nebo dochází ke kompletní změně trendu. Bylo by proto chybou zvolit na začátku analýzy některou z trendových funkcí, případně sestavit odpovídající model a předpokládat, že tento model musí odpovídat realitě za desítky dalších pozorování. Pro detekci časových změn v časové řadě máme více způsobů. Můžeme využít poznatků z intervenční analýzy, flukтуаční testy (především CUSUM a MOSUM testy), významové testy nebo F testy založené na Chowových testech. My si zde představíme základy flukтуаčních a F testů.

1. **Flukтуаční testy:** Základní myšlenka u obou testů je naprosto identická a vychází z modelu lineární regrese 1.11.

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + e_i \quad i = 1, \dots, n, \quad (1.11)$$

kde $\mathbf{x}_i = (1, x_{i_2}, \dots, x_{i_k})^T$ je vektor regresorů s rozměrem $k \times 1$ a $\boldsymbol{\beta}_i = (\beta_0, \dots, \beta_k)$ je vektor regresních koeficientů s rozměrem $k \times 1$. Rezidua, která jsou nezávislá a stejně rozdělená, zde značíme jako e_i .

Nulovou hypotézou je, že máme soubor bez strukturálních změn oproti alternativě, že v časové řadě dochází ke strukturálním změnám. Což můžeme napsat ve tvaru 1.12.

$$H_0 : \beta_i = \beta_0 \quad \forall i = 1, \dots, n \quad H_A : \exists i, \quad i = 1, \dots, n : \beta_i \neq \beta_0 \quad (1.12)$$

- (a) **CUSUM:** Proces CUSUM je založen na kumulativních součtech standardizovaných reziduí. Označme:

$$W_n(t) = \frac{1}{\hat{\sigma}\sqrt{\eta}} \sum_{i=k+1}^{k+\lfloor t\eta \rfloor} \tilde{u}_i \quad (0 \leq t \leq 1),$$

kde \tilde{u} značí rekurzivní rezidua 1.13, $\eta = n - k$ značí jejich počet a $\lfloor t\eta \rfloor$ značí celou část $t\eta$.

$$\tilde{u}_i = \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{(i-1)}}{\sqrt{1 + \mathbf{x}_i^T (\mathbf{X}^{(i-1)T} \mathbf{X}^{(i-1)})^{-1} \mathbf{x}_i}} \quad (i = k + 1, \dots, n), \quad (1.13)$$

kde vektor $\boldsymbol{\beta}^{(i-1)}$ vyjadřuje odhad regresních parametrů z modelu, do kterého zahrnujeme $i - 1$ prvních pozorování. Podobně matice $\mathbf{X}^{(i-1)}$ je maticí regresorů založených na všech pozorováních až do i -tého pozorování. V případě platnosti nulové hypotézy mají tato rezidua nulovou střední hodnotu a rozptyl σ^2 , jehož odhad je dán následovně:

$$\hat{\sigma}^2 = \frac{1}{n - k} \sum_{i=k+1}^n (\tilde{u}_i - \bar{\tilde{u}})^2,$$

kde n je počet pozorování a k značí počet regresorů.

Za platnosti nulové hypotézy $W_n(t)$ konverguje ke standardnímu Brownovu pohybu W_t .

$$W_n(t) \Rightarrow W_t$$

pro $n \rightarrow \infty$. Za předpokladu platnosti alternativní hypotézy, tedy že existuje strukturální změna v nějakém bodě t_0 , tak mají rekurzivní rezidua nulovou střední hodnotu až do tohoto bodu a hodnoty CUSUM procesu kolísají okolo nuly. Po dosažení tohoto bodu se střední hodnota rekurzivních reziduí mění a hodnota CUSUM procesu se vzdaluje od nuly.

- (b) **MOSUM:** Dalším způsobem jak detekovat změnu struktury je založen na sledování změny sumy reziduí. Výsledný empirický proces pak neobsahuje součet všech reziduí až do určitého zlomu t_0 , ale pouze součet

zbytků daný pevným počtem v datovém okně, jehož velikost je dána parametrem šířky pásma $h \in (0, 1)$. MOSUM proces je definován jako:

$$M_n(t|h) = \frac{1}{\hat{\sigma}\sqrt{\eta}} \sum_{i=k+\lfloor N_\eta t \rfloor + 1}^{k+\lfloor N_\eta t \rfloor + \lfloor \eta h \rfloor} \bar{u}_i \quad (0 \leq t \leq 1 - h),$$

kde $N_\eta = (\eta - \lfloor \eta h \rfloor)/(1 - h)$.

MOSUM proces lze vyjádřit také pomocí Brownova pohybu jako:

$$M_n(t|h) = W_n \left(\frac{\lfloor N_\eta t \rfloor + \lfloor \eta h \rfloor}{\eta} \right) - W_n \left(\frac{\lfloor N_\eta t \rfloor}{\eta} \right). \quad (1.14)$$

Z 1.14 lze vidět, že limitujícím procesem pro proces MOSUM jsou přírůstky Brownova pohybu. Stejně jako v případě CUSUM procesu MOSUM proces za platnosti alternativní hypotézy MOSUM proces fluktuuje okolo nuly a v případě dosažení bodu t_0 , ve kterém dochází ke strukturální změně, se hodnoty MOSUM procesu od nuly odcházejí.

2. F testy: Rozdílný přístup oproti zmíněným fluktuacním testům je využití testů založených na F statistikách. Tím nejpodstatnějším rozdílem je, že oproti fluktuacním testům, které reagují v podstatě na jakoukoliv strukturální změnu v souboru, jsou F testy konstruovány pouze k odhalení jednoho bodu zlomu. Myšlenku lze na základě modelu 1.11 zapsat následovně

$$\beta_i = \begin{cases} \beta_A & (1 \leq i \leq i_0), \\ \beta_B & (i_0 < i < n) \end{cases}$$

kde i_0 je bod změny v intervalu $(k, n - k)$. Označení n a k koresponduje s předchozím, n vyjadřuje celkový počet pozorování a k označuje počet vlastních regresorů. První test na základě F statistiky navrhl Chow, a to tak, že v bodě i_0 dojde k rozdělení modelu na dva podmodely. F statistika pak vypadá následovně

$$F_{i_0} = \frac{\hat{u}^T \hat{u} - \hat{e}^T \hat{e}}{\hat{e}^T \hat{e} / (n - 2k)},$$

kde $\hat{e} = (\hat{u}_A, \hat{u}_B)^T$ jsou residua z celkového modelu a \hat{u} jsou rezidua z restriktivního modelu. Regresní koeficienty z restriktivních modelů jsou odhadovány zvlášť. F statistika má za platnosti nulové hypotézy asymptoticky χ_k^2 rozdělení. F_{i_0}/k má pak přesné F rozdělení s k a $n - 2k$ stupni volnosti. Nulovou hypotézu zamítáme při vysokých hodnotách F_{i_0} . Velkou nevýhodou „Chow - testu“ je fakt, že předem musíme znát bod zlomu i_0 .

Přirozeným rozšířením myšlenky „Chow- testu“ je vypočítat F statistiky pro všechny potenciální body zlomu nebo pro všechny potenciální body zlomu v nějakém intervalu (\underline{i}, \bar{i}) . Výpočtem hodnot jednotlivých F_i statistik v intervalu $k < \underline{i} \leq i \leq \bar{i} < n - k$ dostáváme časovou řadu F - statistik.

Získáním časové řady F statistik jsme vlastně dostali proces, který porovnáváme s určitými hranicemi a v případě překročení těchto hranic zamítáme nulovou hypotézu. Hranice můžeme za platnosti nulové hypotézy konstruovat tak, že asymptotická pravděpodobnost toho, že suprema (případně průměry) statistiky F_i pro $(\underline{i} \leq i \leq \bar{i})$ překročí tuto hranici, je α .

Možnosti agregací jednotlivých statistik F_i do jedné je více. Níže jsou uvedeny tři nejpoužívanější, jež byly zkonstruovány D. W. K. Andrewsem a W. Plobergerem.

$$\begin{aligned} \text{sup}F &= \sup_{(\underline{i} \leq i \leq \bar{i})} F_i, \\ \text{ave}F &= \frac{1}{\bar{i} - \underline{i} + 1} \sum_{i=\underline{i}}^{\bar{i}} F_i, \\ \text{exp}F &= \log \left(\frac{1}{\bar{i} - \underline{i} + 1} \sum_{i=\underline{i}}^{\bar{i}} \exp(0.5 \cdot F_i) \right). \end{aligned}$$

Nulovou hypotézu zamítáme při velkých hodnotách $\text{sup}F$, $\text{ave}F$ nebo $\text{exp}F$.

Všechny z uvedených testů jsou v softwaru R implementovány do balíčku **strucchange**, kde flukтуаční testy můžeme provést pomocí funkce **efp**. Pomocí

parametru „type“ volíme, pomocí kterého procesu chceme test provést. Defaultně je „type“ nastaven na „Rec-CUSUMb“, což je proces CUSUM, který jsme popsali výše.

K provedení testu založeného na F statistikách slouží funkce **Fstats**, ve které mimo modelu, který chceme testovat, můžeme pomocí parametrů „from“ a „to“ zadat, která pozorování chceme do F testu zahrnout. Ruční zadání je vhodné v případě, kdy tušíme, v jakém bodě by ke strukturálním změnám mohlo docházet. V případě, že tyto parametry nezadáme, vezme R automaticky celou časovou řadu. Pokud zadáváme parametry „from“ a „to“ ručně a použijeme data ze začátku/konce časové řady, tak se příkaz neprovede s tím, že parametry jsou příliš blízko začátku/konci časové řady. Je tedy nutné několik prvních a posledních dat z testu vyloučit.

1.2. Míry vhodnosti modelu

Po sestavení samotného modelu, pomocí kterého máme například predikovat budoucí hodnoty, musíme ještě ověřit, zda je tento model vůbec vhodný. U jednotlivých trendů jsme uvedli případy, kdy je na místě tyto trendy využít. V případě, kdy však máme daleko složitější modely (což obvykle daleko lépe odpovídá realitě), tak vhodnost modelu ověřujeme až po odhadnutí parametrů pomocí různých interpolačních kritérií. Zde si uvedeme základní kritéria, která pak budou použita při samotné analýze dat.

1. Residuální součet čtverců

Jednou z nejzákladnějších charakteristik pro určení, zda jde o vhodný model či nikoliv, je residuální součet čtverců (RSČ). Obvykle jej označujeme jako S_e^2 a je dán vztahem

$$S_e^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2,$$

kde y_t značí pozorování v čase t a \hat{y}_t značí vyrovnané hodnoty. Obecně se dá říci, že čím je menší hodnota residuálního součtu čtverců, tím lépe. Po vydělení RSČ počtem pozorování dostáváme charakteristiku MSE - mean square error neboli střední čtvercovou chybu.

$$MSE = \frac{S_e^2}{n}$$

2. Index determinace

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})} = \frac{S_M^2}{S_T^2}$$

Pomocí indexu determinace lze vyjádřit, kolik % variability vysvětlované proměnné se nám podařilo modelem objasnit. Může nabývat hodnot od 0 do 1, přičemž čím vyšší, tím lepší. Problémem u obyčejného indexu determinace je to, že pokud přidáme regresory navíc, které mohou být nadbytečné, tak automaticky zvyšují hodnotu R^2 . To jaká hodnota R^2 je dostačující, je vždy předmětem diskuze. Některé zdroje uvádí, že v technických a přírodních vědách je postačující hodnota 0,6 a ve společenských vědách stačí i méně. Já jsem ve své analýze bral za uspokojivou hodnotu bral hodnotu větší než 0,7.

V softwaru R lze index determinace k příslušnému regresnímu modelu zjistit jednoduše příkazem `summary(model lin. regrese)$r.squared`.

3. Modifikovaný index determinace

Celkový počet parametrů zohledňuje modifikovaný index determinace, který dává „spravedlivější“ výsledky, protože „penalizuje“ nadbytečné parametry. Je definován jako

$$R_{adj}^2 = R^2 - \frac{(1 - R^2)(p - 1)}{n - p},$$

kde n vyjadřuje celkový počet pozorování a p značí počet parametrů v modelu.

V softwaru R lze, stejně jako klasický index determinace, i modifikovaný index determinace jednoduše „vytáhnout“ z příslušné lineární regrese aplikací příkazu `summary`. Pouze místo `$r.squared` připojíme `$adj.r.squared`. V samotné analýze budu pracovat s oběma variantami. S modifikovaným indexem determinace pak především při srovnání výsledků s `joinpoint` regresí, kde v důsledku „lomení“ přímky v bodech zlomu dochází k nárůstu parametru.

4. F test kvality regrese

Uvažujme lineární model s absolutním členem

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + e_t, \quad t = 1, \dots, n, \quad (1.15)$$

případně bez absolutního členu

$$y_t = \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + e_t, \quad t = 1, \dots, n. \quad (1.16)$$

Další možností, jak ověřit že zvolený model 1.15, případně 1.16 je vhodný, je využít F test založený na testové statistice

$$F = \frac{\frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{p - 1}}{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - p}},$$

kde p značí celkový počet parametrů v modelu. V případě, že uvažujeme model 1.15 je $p = k + 1$, a v případě 1.16 je $p = k$. F statistika je dána jako

podíl vysvětlené variability ku nevysvětlené variabilitě. Za platnosti nulové hypotézy 1.17 má toto kritérium F rozdělení s $p - 1$ a $n - p$ stupni volnosti.

$$H_0 : \beta_1 = \dots = \beta_k = 0 \quad \text{vs.} \quad H_A : \exists j : \beta_j \neq 0 \quad (1.17)$$

Nulová hypotéza nám vlastně říká, že regresory na vysvětlovanou proměnnou nemají vliv a proto je model nevyhovující. Hypotézu zamítám, pokud je $F > F_{p-1, n-p, \alpha}$. Testové kritérium se dá využít i pro srovnání dvou a více modelů, kdy vhodnější je vybrat ten, pro který je F statistika největší.

5. Test nadbytečnosti regresorů

Uvažujme model

$$y = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{e}_1.$$

V případě, že chceme otestovat, zda posledních q regresorů je nevýznamných, využíváme následující testovou statistiku:

$$F = \frac{R^2 - R_r^2}{1 - R^2} \frac{n - p}{q},$$

kde R_r^2 značí index determinace modelu bez q posledních regresorů. F statistika má za platnosti nulové hypotézy 1.18 F rozdělení s q a $n - p$ stupni volnosti. Nulovou hypotézu zamítáme při vysokých hodnotách F statistiky.

$$H_0 : \beta_{k-q+1} = \dots = \beta_k = 0 \quad (1.18)$$

Tento test lze využít pro jednoduché porovnání dvou modelů, kde jeden z nich je podmodelem druhého. Například tedy k porovnání lineárního a kvadratického trendu.

Čerpáno z [1],[2], [3], [5] a [8].

2. Joinpoint regrese

Joinpoint regrese nebo také breakpoint regrese či segmented regression je model, ve kterém závisle proměnnou vysvětlujeme za pomoci lineární lomené funkce. Tato lineární lomená funkce mění svou směrnici v předem neznámých bodech zlomu - tzv. breakpointech. Myšlenkou celé metody je najít právě tyto body zvratu. Jak se uvádí v [4] značnou výhodou joinpoint regrese je jednoduchost a snadná interpretace parametrů. Právě díky snadné interpretaci parametrů lze joinpoint regresi uplatnit na popis změn ve sledovaném trendu. Navíc lze tuto regresi využít i v případě zobecněných lineárních modelů jako např. loglineární regrese nebo logistická regrese. V případě, že chceme odhadnout podobu joinpoint modelu, který má bod zlomu v nějakém předem neznámém bodu ψ , tedy bodu, ve kterém se mění směrnice z β_1 na $\beta_2 = \beta_1 + \delta_1$, můžeme model zapsat pomocí 2.1. U zobecněných lineárních modelů je model ve tvaru 2.2.

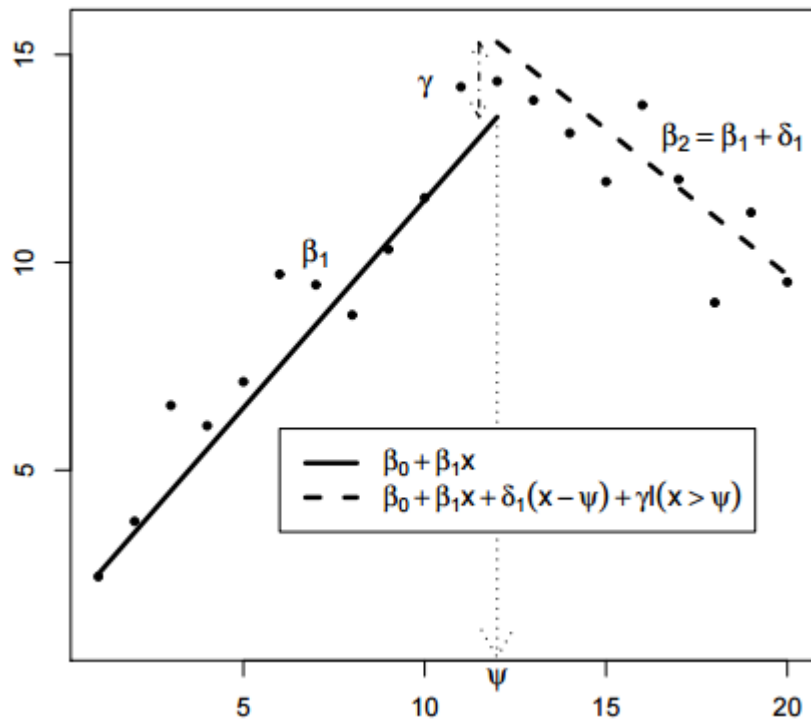
$$E(Y) = \beta_0 + \beta_1 x + \delta_1(x - \psi)^+ \quad (2.1)$$

$$g(E(Y)) = \beta_0 + \beta_1 x + \delta_1(x - \psi)^+ \quad (2.2)$$

Znaménko $+$ značí kladnou část a $g()$ u zobecněných modelů je tzv. linková funkce. β_0, β_1 a δ_1 jsou neznámé regresní koeficienty. Tyto koeficienty odhadujeme pomocí metody maximální věrohodnosti. Odhad parametrů modelu je založen na iteračním postupu, ve kterém předpokládáme, že máme k dispozici nějaký počáteční odhad bodu zvratu, označme jej $\hat{\psi}_0$. Parametry modelu 2.1 pak můžeme odhadnout pomocí iteračních odhadů lineárního modelu 2.3

$$E(Y) = \beta_0 + \beta_1 x + \delta_1(x - \hat{\psi}_0) + \gamma I(x > \hat{\psi}_0), \quad (2.3)$$

kde $I(\cdot)$ je identifikační funkce a γ je tzv. parametr nespojitosti, který měří nespojitost v bodu zlomu. Následně je pomocí γ aktualizován odhad bodu zlomu $\hat{\psi}_0$. V případě, že algoritmus konverguje, tak výsledná regresní funkce by měla být spojitá, tj. $\gamma \approx 0$. Situaci si můžeme ilustrovat na obrázku 5. Uvedený algoritmus má však i své nedostatky. Vzhledem k tomu, že věrohodnostní funkce v joinpoint modelu není obecně konkávní, algoritmus nemusí nalézt její globální



Obrázek 5: Model, pomocí kterého je iteračně odhadována joinpoint regrese.
Zdroj: [4]

minimum. Další nevýhodou je, že metoda v každém kroce pouze aproximuje základní model 2.1. To bohužel způsobuje nesrovnalosti a problémy při odhadování parametrů. Může nastat i situace, kdy algoritmus vůbec nekonverguje. Obvykle zmíněná situace nastává v případě, kdy algoritmus dojde do fáze, kdy alternuje mezi dvěma hodnotami. Je proto vhodné zkusit algoritmus spustit pro více počátečních odhadů $\hat{\psi}_0$ a pokud obdržíme různé odhady parametrů, pak pouze vybrat dle vlastního uvážení tu nejlepší možnou alternativu.

V softwaru R se této problematice věnuje knihovna **segmented**. A právě stejnojmenná funkce v rámci této knihovny pracuje s výše uvedeným algoritmem. Knihovna však neobsahuje funkce pro odhad bodů zlomu. Pro tuto problematiku můžeme zvolit knihovnu **strucchange** a funkci **breakpoints**.

Čerpáno z [4]

3. Korelační analýza

V praktických příkladech se mnohdy setkáváme se situací, kdy jedna z proměnných je závislá na jiných proměnných. V takovém případě je dobré nějakým způsobem tuto závislost číselně vyjádřit. Jedním z nejzákladnějších přístupů, který se zabývá lineárním vztahem mezi dvěma náhodnými veličinami, je korelační analýza.

3.1. Korelační koeficient

Pojem korelace v nejšířším možném slova smyslu označuje vztah neboli asociaci mezi dvěma proměnnými. Přitom jde o vztah vzájemný, kdy ani o jedné z veličin nemůžeme říct, že je závislá a druhá nezávislá. V praxi může jít například o závislost mezi dobou využívání elektrospotřebiče a náklady na spotřebu elektřiny při jeho používání.

Korelační koeficient, který vyjadřuje sílu lineárního vztahu mezi náhodnými veličinami X a Y definujeme jako

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}. \quad (3.1)$$

$\rho_{X,Y}$ může nabývat hodnot od -1 do 1 . Pokud $\rho > 0$, tak jde o tzv. *pozitivní korelaci*, a pokud je $\rho < 0$, tak jde o *negativní korelaci*. Pokud je $\rho = 0$, tak jsou veličiny nekorelované. Pro korelační koeficient mimo jiné platí, že

$$\rho_{X,Y} = \rho_{Y,X},$$

$$\rho_{X,X} = 1.$$

Při práci s daty však musíme využít výběrovou obdobu korelačního koeficientu, který je protějškem toho teoretického 3.1. Výběrový korelační koeficient má následující vyjádření:

$$R_{X,Y} = \frac{S_{X,Y}}{\sqrt{S_X^2 S_Y^2}}, \quad (3.2)$$

kde

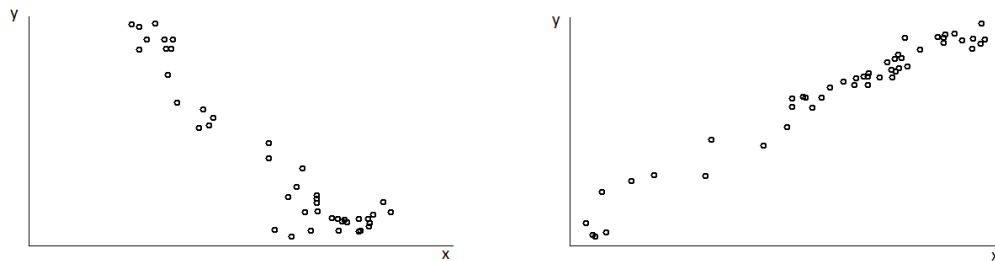
$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

je výběrový rozptyl a

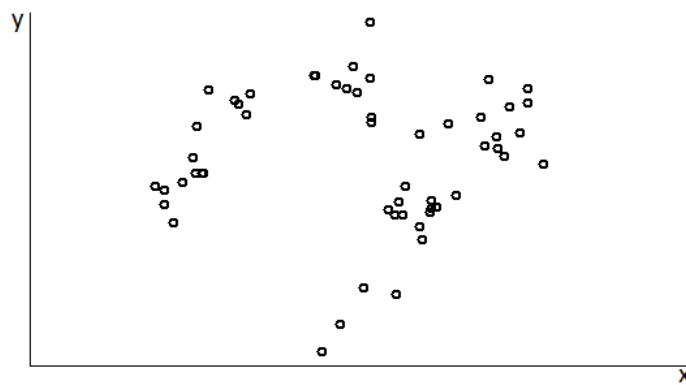
$$S_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

je výběrová kovariance. Vyobrazení jednotlivých typů korelací můžeme vidět na obrázku č. 6 a 7.

Korelační koeficient má však svá úskalí. Jednak slouží pouze pro popis lineárního typu závislosti, což může být mnohdy omezující a v případě jiného



Obrázek 6: Negativně a pozitivně korelovaná data



Obrázek 7: Nekorelovaná data

typu závislosti může dávat hodnoty, které nevystihují sílu této závislosti. Navíc v případě jeho použití mnohdy nebere ohled na to, že proměnné, mezi kterými korelační koeficient počítáme, mohou být ovlivněny nějakým třetím, externím faktorem. Další nepříjemností, kterou se budu dále zabývat více v praktické části, je velmi nutná opatrnost na to, s jakými daty pracujeme. Problém totiž mohou způsobit také odlehle hodnoty nebo situace, kdy jsou data složena z nějakých podskupin nebo shluků.

3.2. Korelační matice

V případě, že máme k dispozici náhodný vektor $\mathbf{X} = (X_1, \dots, X_p)$, tak jednotlivé korelační koeficienty mezi dvěma proměnnými lze seskupit do tzv. korelační matice $cor(\mathbf{X})$.

$$cor(\mathbf{X}) = (\rho_{X_i, X_j})_{i,j=1}^p.$$

V případě práce s daty opět využíváme výběrovou korelační matici \mathbf{R}_x , která vychází z výběrového korelačního koeficientu.

$$\mathbf{R}_x = (R_{ij})_{i,j=1}^p = \left(\frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \right)_{i,j=1}^2,$$

kde indexy i odpovídají řádkům matice a indexy j jejích sloupcům. Celkový počet náhodných veličin je označen jako p . Z vlastnosti korelačního koeficientu je jasné, že jde o symetrickou matici, která má na diagonále 1 a mimo diagonálu jednotlivé korelační koeficienty.

Grafické vyobrazení korelační matice pak nazýváme korelační mapou. Využití korelační mapy je vhodné při větších souborech, kdy lze lehce barvami odlišit hodnoty korelačních koeficientů a na první pohled je pak zřejmé, jakých hodnot přibližně korelační koeficient dosahuje.

Práce s korelacemi v softwaru R je opět poměrně jednoduchá záležitost. Korelační koeficient, potažmo korelační matici, lze získat aplikací příkazu `cor` na data, ze kterých chceme tyto charakteristiky počítat. Pro vizualizaci korelačních

matic lze v softwaru využít balíček **corrplot**. K dispozici je spousta typů vizualizací, které lze využít a uživatel tak má možnost přizpůsobit si korelační mapu i z hlediska designu. Pro proložení dat přímkou lze využít bodových grafů, které můžeme získat pomocí funkce **plot** nebo **qplot** v rámci knihovny **ggplot2**.

Čerpáno z [9].

4. Lokální regrese

Lokální regrese, někdy taky lokální polynomičká regrese, je jedním z neparametrických přístupů k vyrovnání dat, tedy k vysvětlení závislosti střední hodnoty vysvětlované proměnné na hodnotách vysvětlující proměnné (na regresech). Neparametrické přístupy dopředu nepředpokládají žádný předpis regresní funkce jako je tomu u klasické regrese. Proto výsledkem nejsou odhady jednotlivých parametrů a taktéž nelze využít testování parametrických hypotéz. Na druhou stranu vzhledem k absenci předem známého parametrického modelu je možné vyhnout se jeho špatné specifikaci a následným odchylkám modelu od reality. Lokální regrese je tak velice flexibilním přístupem k vyrovnávání dat.

Lokální regrese vychází z metody vážených nejmenších čtverců, přičemž váhy jsou specifikovány pomocí tzv. jádrové funkce.

Zvolme si bod x_0 . Normovaná vzdálenost bodu x_0 od bodu x je dána jako:

$$d = \frac{x - x_0}{h},$$

kde parametr $h > 0$ je šířka okna (v angl. terminologii se používá výraz *bandwidth*) a jeho hodnota je dána naší volbou. $K(d)$ označuje hodnotu jádrové funkce v bodě d . Požadujeme, aby jádrová funkce splňovala tyto vlastnosti:

1. $K(d) \geq 0$, pro $\forall d \in \mathbb{R}$,
2. $K(-d) = K(d)$, pro $\forall d \in \mathbb{R}$,
3. $K(d)$ je nerostoucí pro $d \geq 0$.

Jádrová funkce zajišťuje, že nejvyšší váha je přiřazena bodům, které jsou bližší bodu x_0 a naopak vzdálenějším bodům přiřazuje váhu nižší. Výsledný odhad regresní funkce však není příliš ovlivněn volbou jádrové funkce. Nejčastěji využívané jádrové funkce jsou:

1. Gaussova jádrová funkce:

$$K(d) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

2. Trikubická jádrová funkce:

$$K(d) = \begin{cases} (1 - |z|^3)^3 & (|d| < 1), \\ 0 & (|d| \geq 1). \end{cases}$$

3. Obdélníková jádrová funkce:

$$K(d) = \begin{cases} 1 & (|d| < 1), \\ 0 & (|d| \geq 1). \end{cases}$$

Parametr h můžeme volit dvojím způsobem. Buďto jej volíme dle vzdálenosti bodu x od bodu x_0 a to tak, že pokud $|x - x_0| \geq h$, tak bude bodu x přiřazena nulová váha. Nebo lze parametr h volit nezávisle na hodnotě x_0 tak, aby se v okolí bodu x_0 vyskytoval pevný počet bodů l z celkového počtu pozorování n .

Předpokládejme, že pro náhodný vektor $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ platí:

$$Y_i = f(x_i) + \epsilon_i,$$

kde $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$, $\epsilon \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ a funkce $f : \mathbb{R} \rightarrow \mathbb{R}$ je hladká funkce regresorů. Díky předpokladu o hladkosti regresní funkce f lze pro body x_i z okolí bodu x_0 využít aproximaci funkce f polynomem:

$$Y_i = f(x_i) + \epsilon_i = \alpha_0 + \alpha_1(x_i - x_0) + \dots + \alpha_k(x_i - x_0)^k + E_i = Q_k(x_i - x_0) + E_i,$$

kde v obecném případě nemusí platit, že $\epsilon_i = E_i$. Polynom Q_k , se kterým pracujeme na okolí bodu x_0 , nazýváme lokálním polynomem.

Vektor parametrů lokálního polynomu $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_k)^T \in \mathbb{R}^{k+1}$ odhadujeme pomocí vážené metody nejmenších čtverců. Necht' $\mathbf{a} = (a_0, \dots, a_k)^T$ je odhad vektoru $\boldsymbol{\alpha}$ a hodnota šířky okna h je nezávislá na x_0 . Potom

$$\mathbf{a} = \min_{\boldsymbol{\alpha} \in \mathbb{R}^{k+1}} \sum_{i=1}^n K\left(\frac{x_i - x_0}{h}\right) E_i^2,$$

kde váhy $K(\cdot)$ udává jádrová funkce.

Ukažme si odhad parametrů pro nejjednodušší lokální polynom, tedy lokální přímkou. Počítáme tedy s $k = 1$. Vektor odhadnutých parametrů bude mít podobu:

$$\mathbf{a} = (a_1, a_2)^T.$$

Matice vah \mathbf{K}_{x_0} , kde index x_0 znamená příslušnost dané hodnotě x_0 , je diagonální o rozměru $n \times n$ s prvky:

$$k_{ii} = K \left(\frac{x_i - x_0}{h} \right).$$

A matice regresorů je dána jako:

$$\mathbf{X} = \begin{pmatrix} 1 & (x_1 - x_0) \\ \vdots & \vdots \\ 1 & (x_n - x_0) \end{pmatrix}.$$

Odhad vektoru \mathbf{a} je získán pomocí vážené metody nejmenších čtverců, můžeme tedy psát:

$$\mathbf{X}^T \mathbf{K}_{x_0} \mathbf{X} \mathbf{a} = \mathbf{X}^T \mathbf{K}_{x_0} \mathbf{Y}.$$

Za předpokladu regularity a invertovatelnosti matice $\mathbf{X}^T \mathbf{K}_{x_0} \mathbf{X}$ platí:

$$\mathbf{a} = (\mathbf{X}^T \mathbf{K}_{x_0} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K}_{x_0} \mathbf{Y}.$$

Vyrovnanou hodnotu v bodě x_0 dostaneme následnou lineární kombinací

$$\hat{y}(x_0) = \mathbf{u}_1^T (\mathbf{X}^T \mathbf{K}_{x_0} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K}_{x_0} \mathbf{Y},$$

kde $\mathbf{u}_1^T = (1, 0) \in \mathbb{R}^2$. Zopakováním postupu pro různé hodnoty x_0 dostáváme vektor vyrovnaných hodnot. Za hodnoty x_0 dosazujeme hodnoty regresorů (x_i).

Algoritmus počítá vyrovnané hodnoty v bodech x_i . Pro zakreslení výsledné regresní křivky tyto body spojíme úsečkami. V literatuře lze takto získanou křivku najít pod termínem *loess*. Termín také odpovídá funkci v softwaru R, pomocí kterého lze lokální regresi provést.

Čerpáno z [7] a [10].

Část II

Praktická část

V praktické části se zabývám, jak již samotný název práce napovídá, analýzou konzumace alkoholu v České republice a sousedních zemích, tedy na Slovensku, v Rakousku, v Německu a v Polsku. V analýze jsem se zaměřil jak na trendy konzumace v jednotlivých zemích nebo změny úrovní konzumace v průběhu let, tak na závislosti konzumace alkoholu mezi jednotlivými státy a změnu těchto závislostí v čase.

5. Použitá data

Data, která budeme analyzovat, pocházejí z úložiště dat Světové zdravotnické organizace (dále jen WHO). Každý z datasetů se skládá ze 4 sloupců - úrovně konzumace piva, vína, tvrdého alkoholu a celkové spotřeby alkoholických nápojů. Spotřeba je uváděna v litrech čistého alkoholu na osobu starší 15 let za rok. V litrech čistého alkoholu na osobu starší 15 let jsou data uváděna především kvůli srovnatelnosti jednotlivých složek. Pokud bychom nebrali ohled na přepočtení % alkoholu na čistý alkohol, pak bychom se u vína a piva dostali ke značnému nepochopitelnému poměru oproti tvrdému alkoholu. Uvažujme příklad, kdy 10° pivo má v průměru asi 4 % alkoholu a běžný tvrdý alkohol 37,5%. Při konzumaci 0,5l 10° piva tak spotřebujeme 20 ml čistého alkoholu, kdežto u 0,5l čiré vodky se dostáváme na 187,5 ml. Na stejnou úroveň spotřeby 100% alkoholu při vypití takové vodky se dostaneme až po vypití 9,375 „půl-litrů“ piva. Časové řady konzumace alkoholu začínají v roce 1961 a aktuálně končí rokem 2014. Máme tedy k dispozici celkem 54 pozorování v každé složce z jednotlivých zemí.

Nejednoho čtenáře by mohlo překvapit, že data v letech 1961 - 1992 jsou rozdělena na Českou republiku a Slovensko i přes fakt, že v letech 1945 - 1992 tvořily tyto dva státy jeden. Právě z důvodu vývoje politických hranic jsou data konstruována spíše geograficky.

5.1. Metodologie dat

Ačkoliv by si zřejmě každý pod úrovní konzumace alkoholických látek představil množství alkoholu, které obyvatelstvo dané země skutečně spotřebuje, tak v případě našich dat tomu tak není. Bylo by velmi obtížné sledovat, zda se zakoupený sud piva skutečně vypil nebo je někde skladován. Proto zde úroveň konzumace (UK) alkoholu nabírá spíše ekonomického charakteru a je definována následovně:

$$UK = Výroba + (Export - Import) - změna stavu zásob u výrobců$$

Takto vypočtená celková úroveň konzumace se pak přepočítá na osoby starší 15 let pomocí tzv. středního stavu obyvatelstva¹.

5.2. Náhled na data

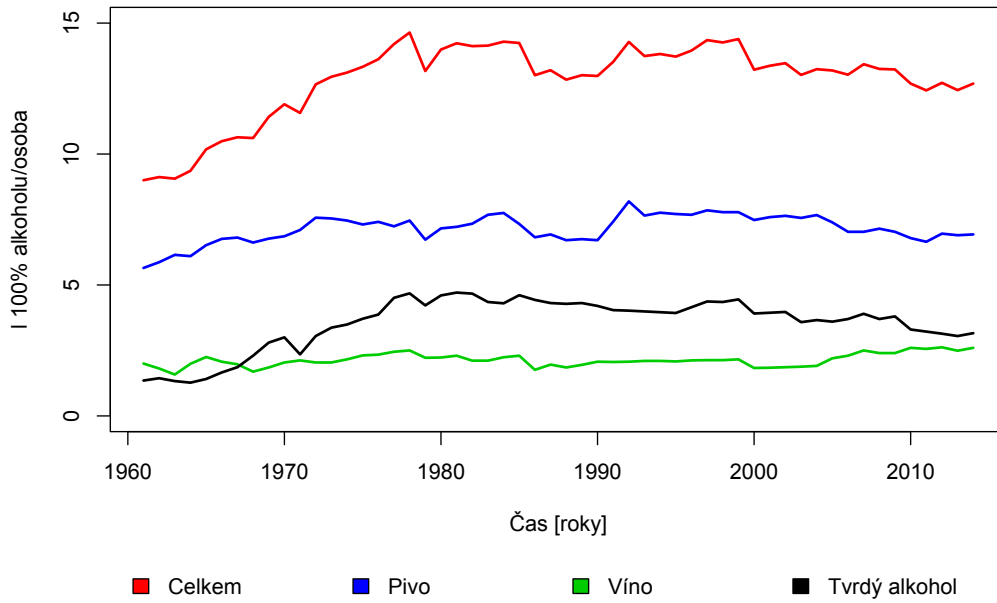
Vzhledem k tomu, že tabulky s daty pro jednotlivé státy jsou poměrně velké, ukážeme zde pouze data pro Českou republiku (viz tabulka 1). Tabulky s daty pro ostatní státy jsou vloženy k nahlédnutí v přílohách.

Často je mnohem efektivnější nahlížet na data v grafu místo tabulky. Můžeme tak pouhým okem bez větší znalosti problematiky zjistit, jakým směrem se časová řada ubírá, jaké jsou maximální/minimální hodnoty nebo jaké jsou zlomové okamžiky. Níže jsou postupně v grafech vyobrazeny všechny časové řady, se kterými jsem pracoval. Z důvodu úspornosti jsou všechny složky konzumace vyobrazeny v jednom grafu. Grafy pro jednotlivé spotřeby jsou k nahlédnutí v přílohách. Pro zajímavost pak uvádím obrázek 13, kde můžeme vidět kolik % celkové spotřeby tvoří spotřeba jednotlivých složek.

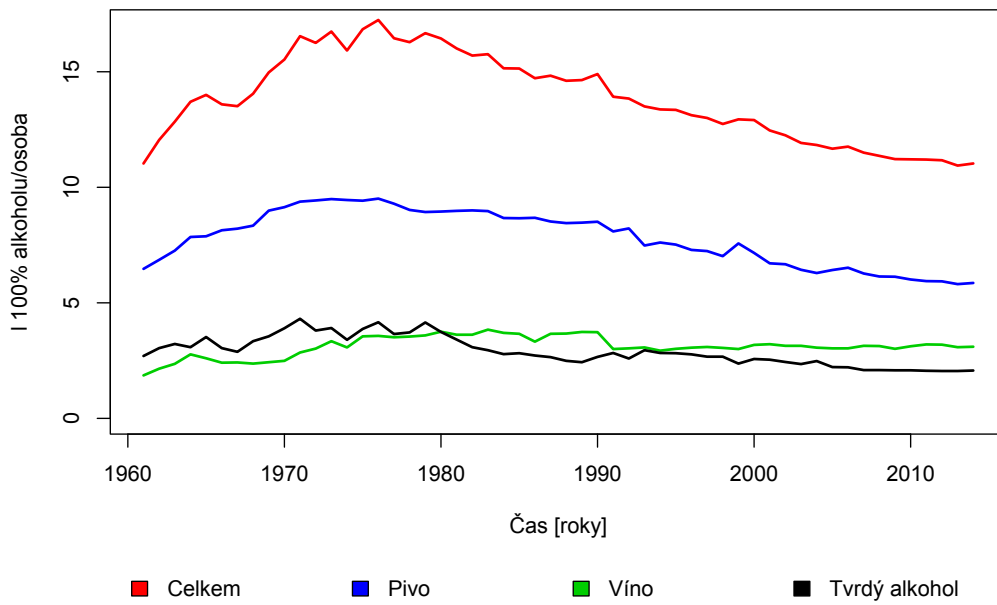
¹viz https://www.czso.cz/csu/czso/pocet_obyvatel_m

Rok	Pivo	Víno	Tvrký	Celkem	Rok	Pivo	Víno	Tvrký	Celkem
1961	5.65	2.00	1.35	9.00	1988	6.71	1.85	4.28	12.84
1962	5.87	1.81	1.44	9.12	1989	6.75	1.95	4.31	13.01
1963	6.15	1.58	1.33	9.06	1990	6.71	2.07	4.20	12.98
1964	6.10	1.99	1.27	9.36	1991	7.42	2.06	4.04	13.52
1965	6.52	2.25	1.41	10.18	1992	8.19	2.07	4.02	14.28
1966	6.76	2.07	1.66	10.49	1993	7.65	2.10	3.99	13.74
1967	6.81	1.97	1.86	10.64	1994	7.76	2.10	3.96	13.82
1968	6.62	1.69	2.30	10.61	1995	7.71	2.08	3.93	13.72
1969	6.77	1.85	2.80	11.42	1996	7.68	2.12	4.15	13.95
1970	6.86	2.04	3.00	11.90	1997	7.85	2.13	4.37	14.35
1971	7.10	2.12	2.35	11.57	1998	7.78	2.13	4.35	14.26
1972	7.57	2.04	3.05	12.66	1999	7.78	2.16	4.45	14.39
1973	7.54	2.04	3.37	12.95	2000	7.48	1.83	3.91	13.22
1974	7.46	2.16	3.49	13.11	2001	7.59	1.84	3.94	13.37
1975	7.31	2.31	3.71	13.33	2002	7.64	1.86	3.97	13.47
1976	7.41	2.34	3.87	13.62	2003	7.56	1.88	3.58	13.02
1977	7.24	2.45	4.51	14.20	2004	7.67	1.91	3.66	13.24
1978	7.46	2.50	4.68	14.64	2005	7.39	2.20	3.60	13.19
1979	6.73	2.22	4.22	13.17	2006	7.03	2.30	3.70	13.03
1980	7.16	2.23	4.60	13.99	2007	7.03	2.50	3.90	13.43
1981	7.22	2.30	4.71	14.23	2008	7.15	2.40	3.70	13.25
1982	7.34	2.11	4.67	14.12	2009	7.03	2.40	3.80	13.23
1983	7.68	2.11	4.35	14.14	2010	6.79	2.60	3.30	12.69
1984	7.75	2.24	4.30	14.29	2011	6.65	2.56	3.22	12.43
1985	7.33	2.30	4.61	14.24	2012	6.96	2.62	3.14	12.72
1986	6.82	1.76	4.43	13.01	2013	6.90	2.49	3.05	12.44
1987	6.93	1.96	4.31	13.20	2014	6.93	2.60	3.16	12.69

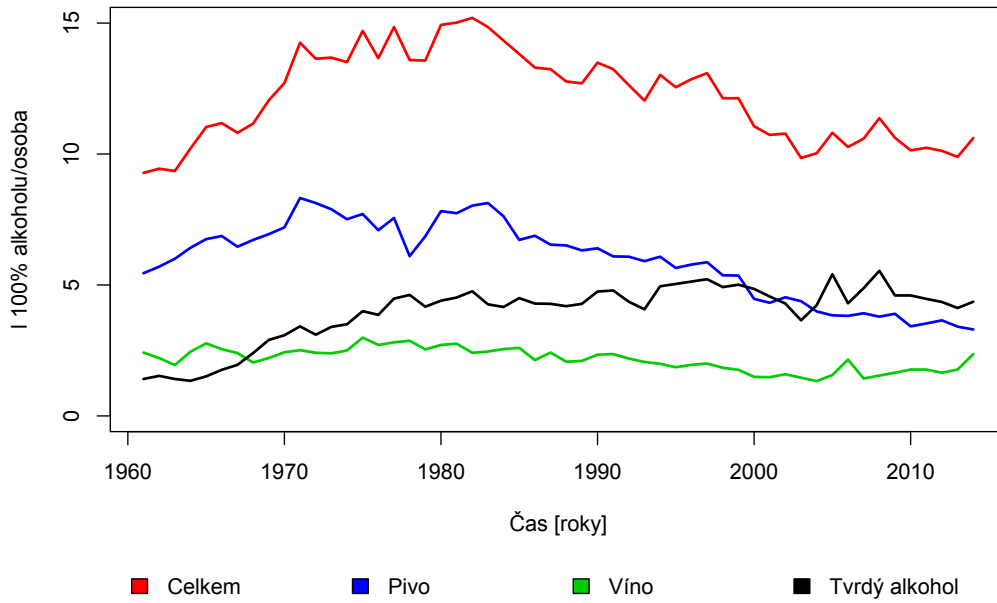
Tabulka 1: konzumace alkoholu v litrech čistého alkoholu na osobu starší 15 let v ČR v letech 1961 - 2014



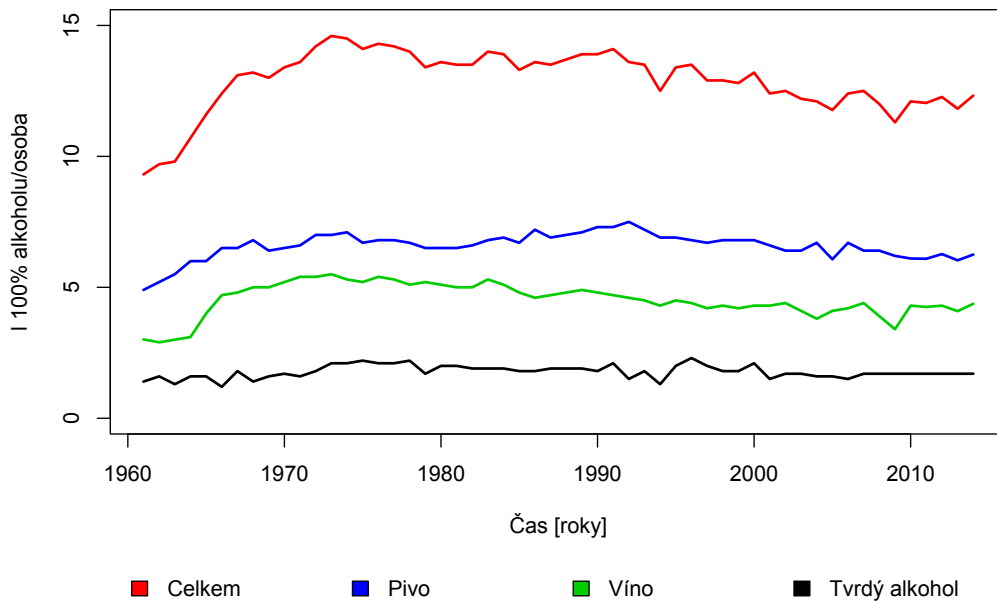
Obrázek 8: Konzumace alkoholu v ČR v letech 1961 - 2014



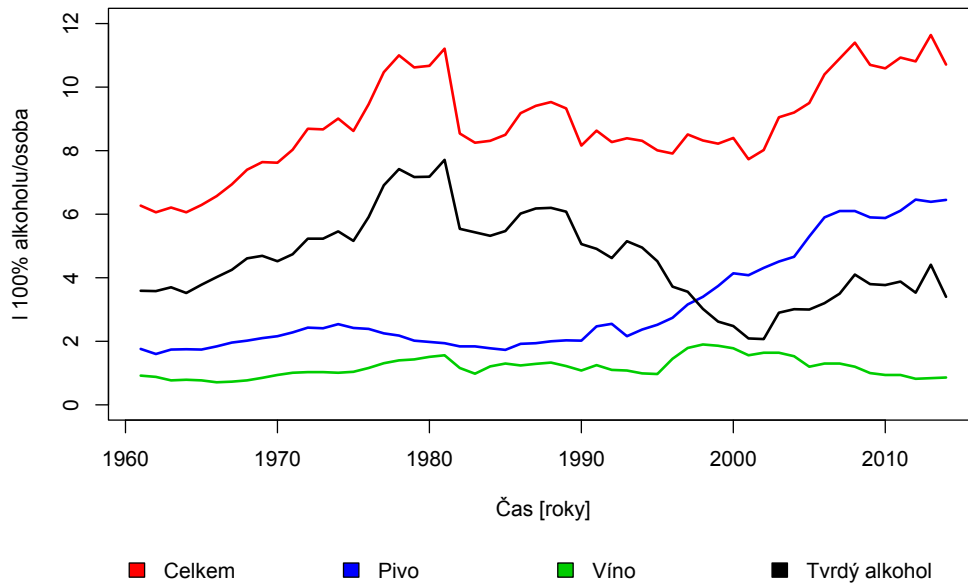
Obrázek 9: Konzumace alkoholu v Německu v letech 1961 - 2014



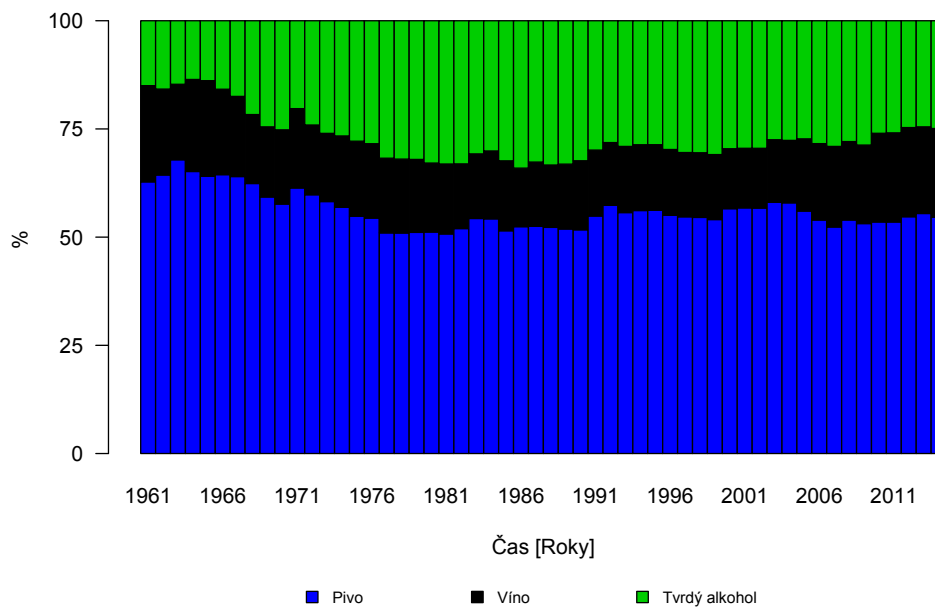
Obrázek 10: Konzumace alkoholu v SR v letech 1961 - 2014



Obrázek 11: Konzumace alkoholu v Rakousku v letech 1961 - 2014



Obrázek 12: Konzumace alkoholu v Polsku v letech 1961 - 2014



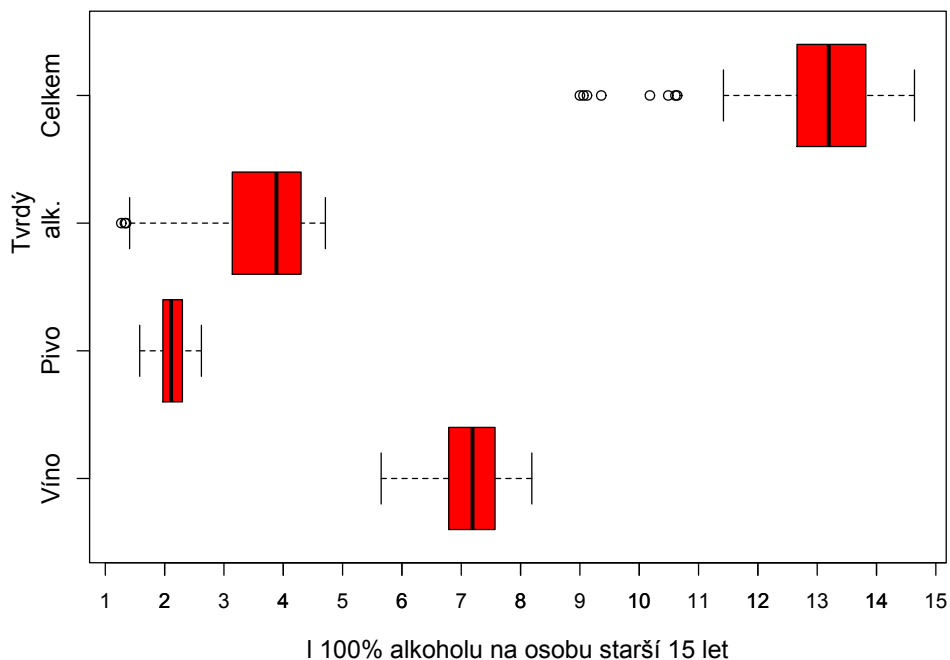
Obrázek 13: Procentuální zastoupení konzumace piva, vína a tvrdého alkoholu v České republice v letech 1961 - 2014

6. Analýza dat

V této kapitole se již zaměříme na samotnou analýzu dat. Vzhledem k četnosti některých grafů budu do textu vybírat pouze nejzajímavější reprezentativní vzorky. Ostatní grafy jsou umístěny na příloženém CD k nahlédnutí.

6.1. Číselné charakteristiky

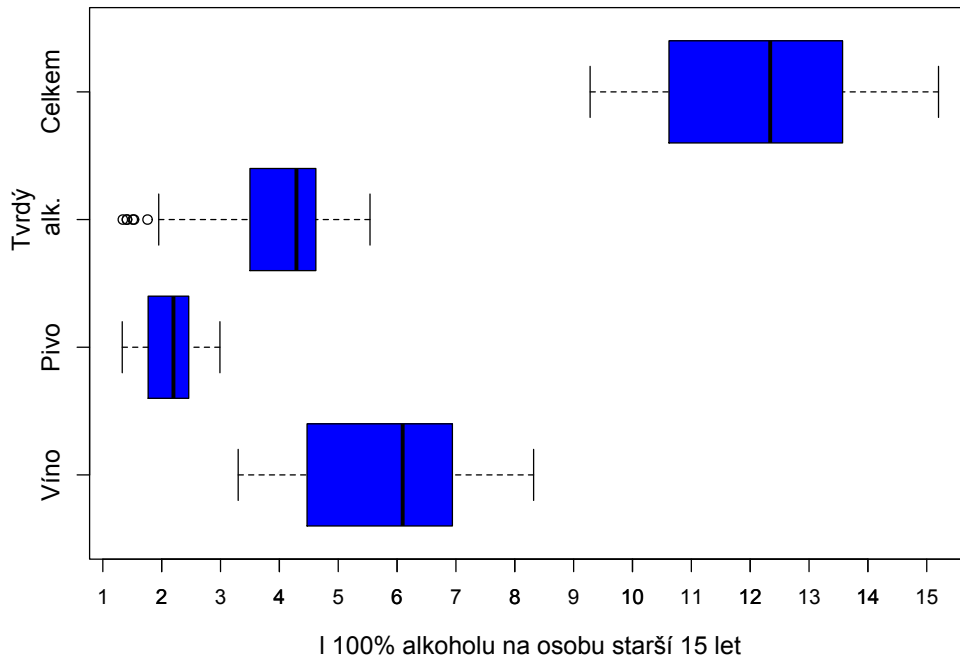
Pro samotnou analýzu časových řad jsou podstatné i její číselné charakteristiky. Mezi základní číselné charakteristiky můžeme zařadit minimum, maximum, medián, horní a dolní kvartil nebo třeba rozptyl a směrodatnou odchylku. Místo toho, abychom všechny tyto údaje vypisovali do zvláštních tabulek, zvolil jsem přístup pomocí krabicových grafů - boxplotů, ve kterých jsou tyto údaje zřetelné na první pohled. Alespoň tedy orientačně. Níže uvádím tabulky a boxploty pro konzumaci alkoholu v ČR a na Slovensku. Ostatní boxploty a tabulky s číselnými charakteristikami jsou dostupné na příloženém CD.



Obrázek 14: Boxploty konzumace alkoholu v ČR v letech 1961 - 2014

Česká republika				
	Pivo	Víno	Tvrký alkohol	Celkem
Minimum	5.65	1.58	1.27	9.00
Maximum	8.19	2.62	4.74	14.64
Průměr	7.15	2.13	3.54	12.82
Medián	7.19	2.11	3.86	13.20
Horní kvartil	7.57	2.30	4.29	12.66
Dolní kvartil	6.79	1.97	3.14	13.82
Rozptyl	0.28	0.06	0.97	2.15

Tabulka 2: Tabulka základních číselných charakteristik dat konzumace alkoholu v ČR v letech 1961 - 2014



Obrázek 15: Boxploty konzumace alkoholu na Slovensku v letech 1961 - 2014

6.2. Proložení dat trendovými funkcemi

Jak již bylo uvedeno v teoretické části, základním přístupem v analýze časových řad je snaha odhadnout její trendovou složku pomocí matematických křivek. Ze základních křivek, které jsme si uvedli, jsem daty zkusil proložit lineární, kvadratický a jednoduchý exponenciální trend. Na základě charakteristik pro vhodnost modelu jsem zde vybral ty nejvhodnější modely jednotlivých spotřeb

Slovenská republika				
	Pivo	Víno	Tvrký alkohol	Celkem
Minimum	3.30	1.33	1.34	9.28
Maximum	8.32	2.99	5.54	15.20
Průměr	5.94	2.16	3.96	12.17
Medián	6.09	2.20	4.29	12.34
Horní kvartil	6.93	2.46	4.62	13.55
Dolní kvartil	4.49	1.79	3.54	10.65
Rozptyl	2.22	0.18	1.25	2.96

Tabulka 3: Tabulka základních číselných charakteristik dat konzumace alkoholu na Slovensku v letech 1961 - 2014

v jednotlivých zemích. Grafy a tabulky k proložení dat jednotlivými trendy, které nejsou uvedeny přímo v textu práce jsou dostupné na příloženém CD.

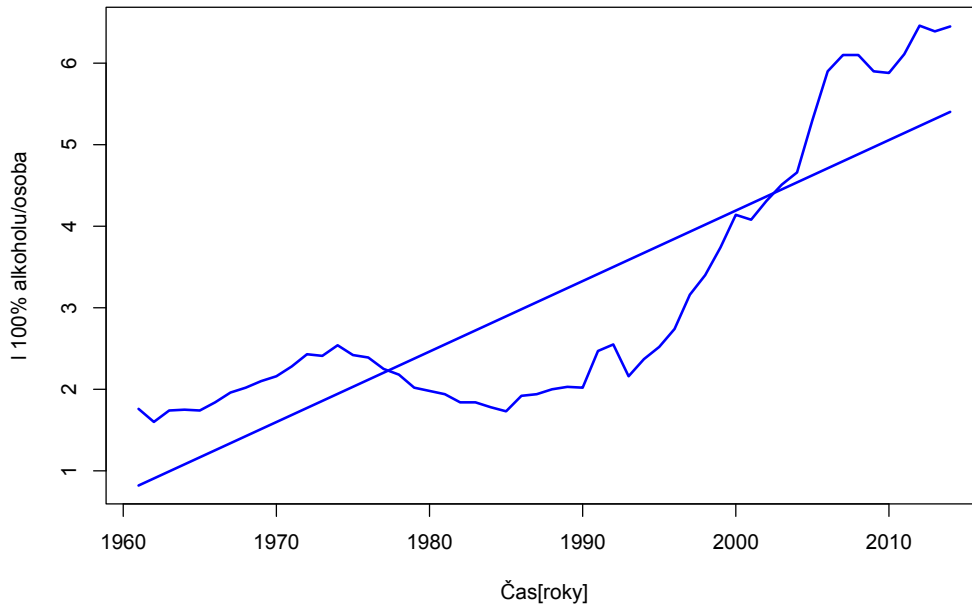
6.2.1. Lineární trend

Už na začátku samotné analýzy se dalo očekávat, že lineární trend reálně příliš odpovídat nebude. Přesto se dá říci, že vzhledem ke své jednoduchosti v určitých případech dosáhl hezkých výsledků. Dle jednotlivých měř vhodnosti modelu lze lineární trend akceptovat u vývoje konzumace piva v Polsku. Proložení dat lineárním trendem spolu s mírami vhodnosti modelu můžeme vidět na obrázku 16 a v tabulce 4.

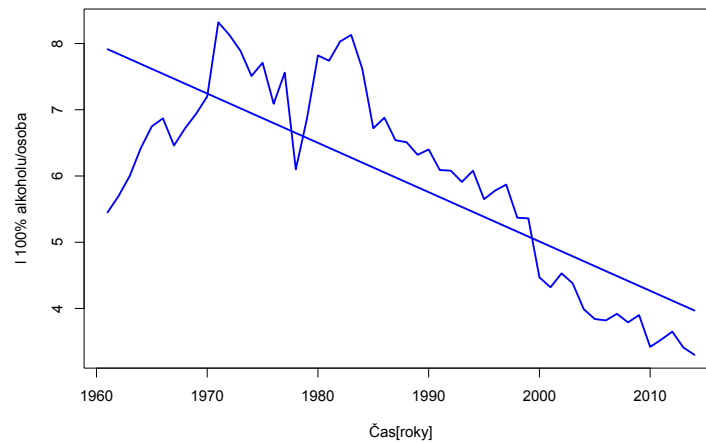
Míry vhodnosti lineárního trendu při proložení dat spotřeby piva v Polsku			
RSC	R ²	R _{adj} ²	Hodnota F statistiky
38.36	0.7188	0.7134	132.9 on 1 and 52 DF

Tabulka 4: Míry vhodnosti lineárního trendu při proložení dat spotřeby piva v Polsku

Za zmínku u lineárního trendu ještě stojí výsledky proložení dat spotřeby piva na Slovensku a tvrdého alkoholu v Německu, kdy index determinace dosahuje hodnot 0.6167 respektive 0.6103. Zejména u spotřeby tvrdého alkoholu v Německu by se při odříznutí starších dat dalo o užití lineárního trendu uvažovat. Proložení dat je k vidění na obrázku 17 a 18. Souhrnné tabulky a grafy jsou k dispozici na příloženém CD.



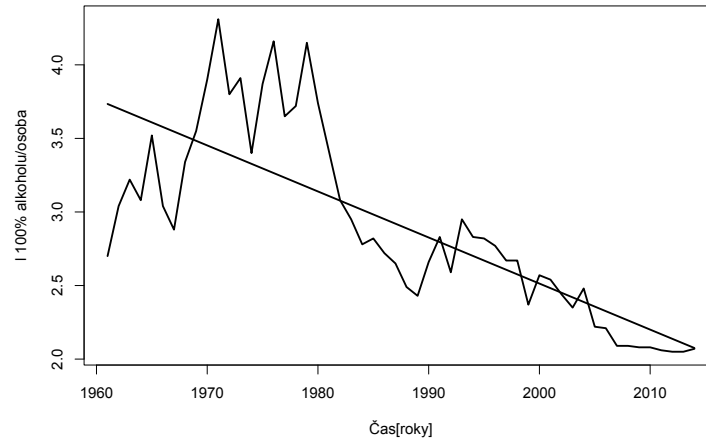
Obrázek 16: Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 lineárním trendem



Obrázek 17: Proložení dat spotřeby piva na Slovensku v letech 1961 - 2014 lineárním trendem

6.2.2. Kvadratický trend

V případě kvadratického trendu se dá říci, že předčil má očekávání. I přes to, že jde stále o poměrně jednoduchý model, tak index determinace dosáhl v 6 případech nad hodnotu 0,8. Mimo proložení dat celkové spotřeby v České re-



Obrázek 18: Proložení dat spotřeby tvrdého alkoholu v Německu v letech 1961 - 2014 lineárním trendem

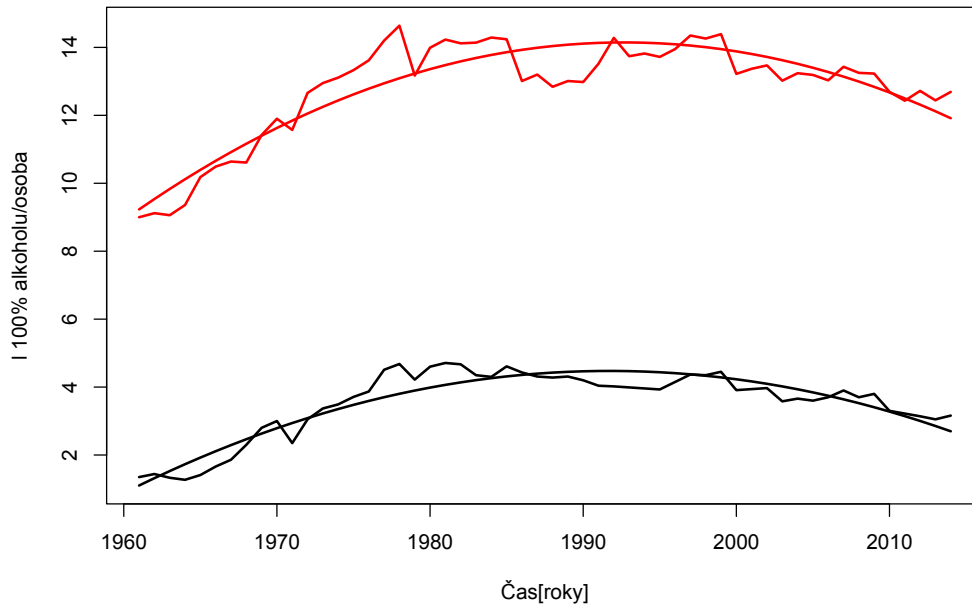
publice (obr. 19 + tab. 5) a Německu (obr. 20 + tab. 6) a konzumace tvrdého alkoholu v České republice (obr. 19 + tab. 5) a Slovensku (obr. 21 + tab. 7), dosahuje index determinace vysokých hodnot při proložení dat spotřeby piva na Slovensku (obr. 21 + tab. 7), Německu (obr. 20 + tab. 6) a především pak v Polsku, kde již má smysl vzhledem k výsledkům z proložení dat lineárním trendem přistoupit k porovnání obou modelů.

Míry vhodnosti kvadratického trendu ČR				
	RSC	R ²	R ² _{adj}	Hodnota F statistiky
Celkem	19.28	0.8311	0.8244	125.4 on 2 and 51 DF
Tvrdý alkohol	6.6918	0.8692	0.8641	169.5 on 2 and 51 DF

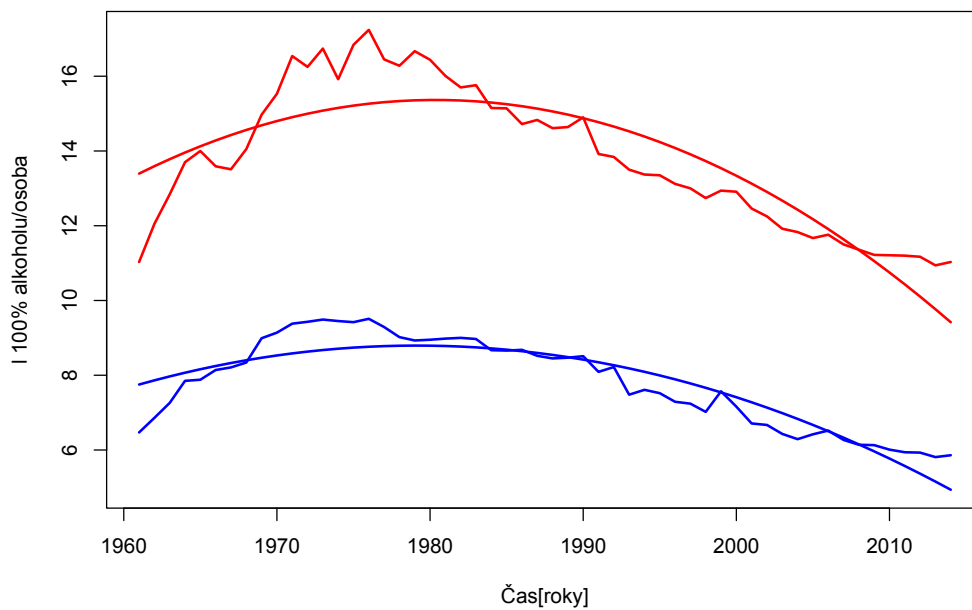
Tabulka 5: Míry vhodnosti kvadratického trendu při proložení dat celkové konzumace a konzumace tvrdého alkoholu v ČR v letech 1961 - 2014

Míry vhodnosti kvadratického trendu Německo				
	RSC	R ²	R ² _{adj}	Hodnota F statistiky
Celkem	47.31	0.7529	0.7432	77.7 on 2 and 51 DF
Pivo	13.16	0.8263	0.8194	121.3 on 2 and 51 DF

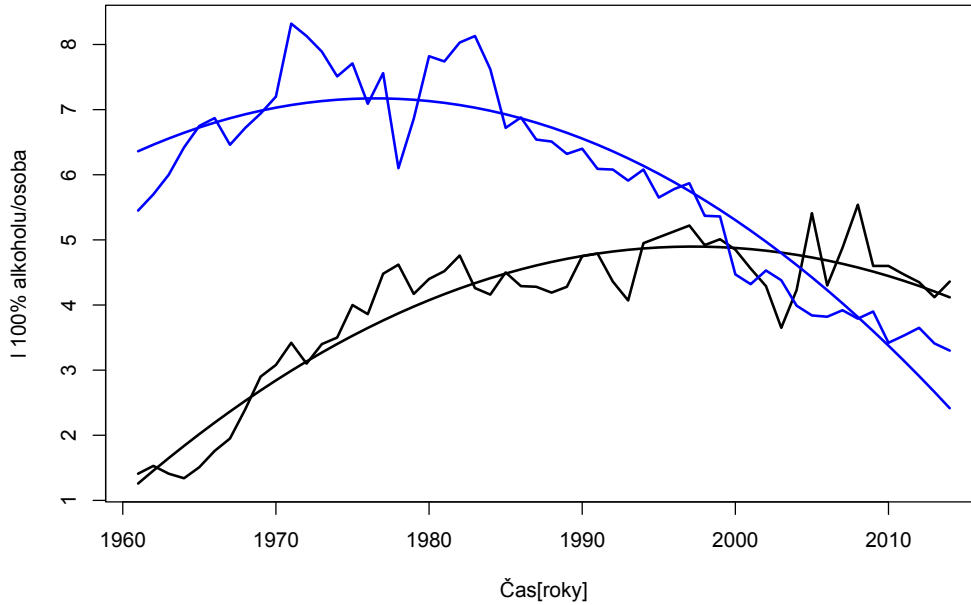
Tabulka 6: Míry vhodnosti kvadratického trendu při proložení dat konzumace piva a celkové konzumace v Německu v letech 1961 - 2014



Obrázek 19: Proložení dat celkové konzumace alkoholu a konzumace tvrdého alkoholu v ČR v letech 1961 - 2014 kvadratickým trendem



Obrázek 20: Proložení dat celkové konzumace alkoholu a konzumace piva v Německu v letech 1961 - 2014 kvadratickým trendem



Obrázek 21: Proložení dat celkové konzumace piva a tvrdého alkoholu na Slovensku v letech 1961 - 2014 kvadratickým trendem

Míry vhodnosti kvadratického trendu Slovensko				
	RSC	R ²	R _{adj} ²	Hodnota F statistiky
Tvrký alkohol	8.54	0.8711	0.866	172.3 on 2 and 51 DF
Pivo	16.05	0.8639	0.8585	161.8 on 2 and 51 DF

Tabulka 7: Míry vhodnosti kvadratického trendu při proložení dat konzumace piva a tvrdého alkoholu na Slovensku v letech 1961 - 2014

Porovnání lineárního a kvadratického trendu při proložení dat spotřeby piva v Polsku v letech 1961 - 2014:

Porovnáváme následující modely:

$$y_t = 0.733592 + 0.086462x_t,$$

$$y_t = 2.454132 - 0.097882x_t + 0.003352x_t^2.$$

Již při prvním pohledu na tabulky s jednotlivými mírami pro vhodnost modelu (tedy tabulky 4 a 8) je jasné, že kvadratický trend se daleko lépe přizpůsobil datům, čímž dosahuje i nižší hodnoty residuálního součtu čtverců. Vzhledem

k tomu, že kvadratický trend má více parametrů, tedy přesněji o jeden parametr víc než trend lineární, je třeba uvážit spíše hodnoty modifikovaného indexu determinace. I přes penalizaci za vyšší počet parametrů je však hodnota modifikovaného indexu determinace u kvadratického trendu vyšší.

Ještě než prohlásíme kvadratický trend za lepší, tak můžeme provést test nadbytečnosti regresorů. Předpokládejme tedy následující nulovou hypotézu pro model s kvadratickým trendem:

$$H_0 : \beta_2 = 0 \quad vs. \quad H_A : \beta_2 \neq 0. \quad (6.1)$$

Dosažením jednotlivých indexů determinace do testovací statistiky 5 dostáváme:

$$F = \frac{(0.9285 - 0.7188) 54 - 3}{(1 - 0.9285)} \frac{54 - 3}{1} = 51 \cdot \frac{0.2097}{0.0715} = 149.5762.$$

Vzhledem k tomu, že:

$$F_{q,n-p,1-\alpha} = F_{1,51,0.95} = 4.030393,$$

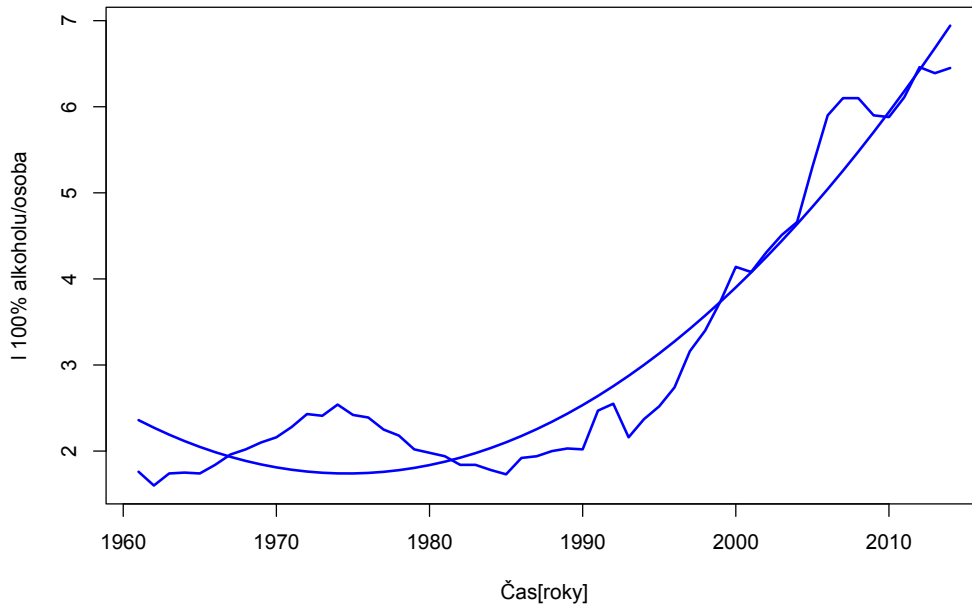
tak při porovnání s hodnotou F statistiky docházíme k závěru, že nulovou hypotézu zamítáme. Testem jsme potvrdili, že kvadratický trend nemá žádné nadbytečné regresory a můžeme jej tedy prohlásit za kvalitnější než trend lineární. Samozřejmě pouze v rámci konzumace piva v Polsku v letech 1961 - 2014.

Míry vhodnosti kvadratického trendu při proložení dat spotřeby piva v Polsku v letech 1961 - 2014			
RSC	R²	R_{adj}²	Hodnota F statistiky
9.75	0.9285	0.9257	331.2 on 2 and 51 DF

Tabulka 8: Míry vhodnosti kvadratického trendu při dat spotřeby piva v Polsku v letech 1961 - 2014

6.2.3. Exponenciální trend

Stejně jako v případě lineárního trendu se u proložení dat exponenciálním trendem nedá hovořit o modelech s dobrými výsledky. Jediným světlým bodem může být proložení dat o spotřebě piva na Slovensku, které můžeme vidět na



Obrázek 22: Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 kvadratickým trendem

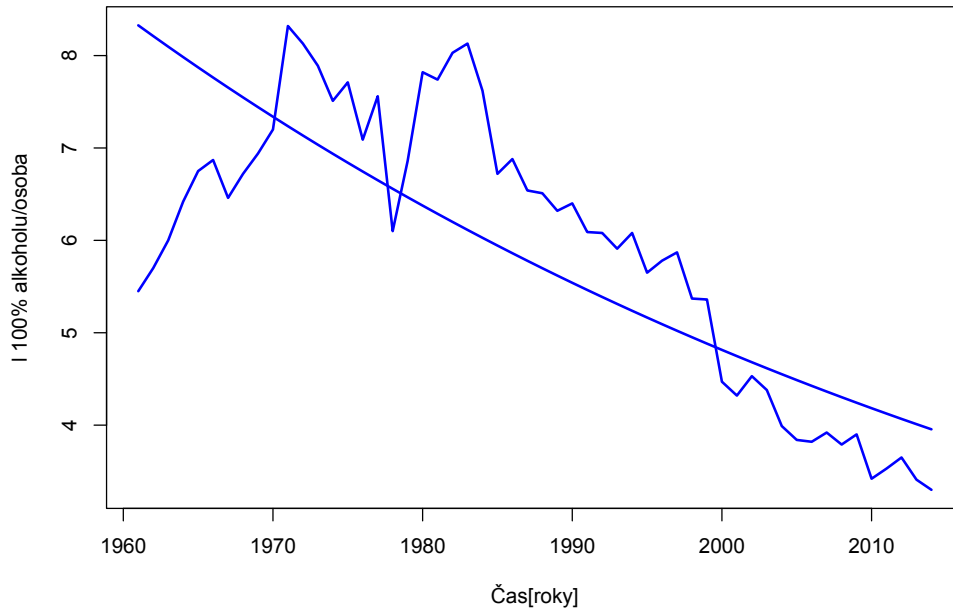
obrázku 23. Pomocí exponenciálního trendu se nám podařilo vysvětlit 75,33 % variability v datech, což je lepší výsledek než u trendu lineárního. Ostatní proložení dat exponenciální křivkou jsou opět spolu s tabulkami charakteristik dostupné na příloženém CD.

Míry vhodnosti exponenciálního trendu při proložení dat spotřeby piva na Slovensku v letech 1961 - 2014			
RSC	R^2	R_{adj}^2	Hodnota F statistiky
57.135	0.7533	0.7486	158.8547 on 1 and 52 DF

Tabulka 9: Míry vhodnosti exponenciálního trendu při dat spotřeby piva na Slovensku v letech 1961 - 2014

6.3. Proložení dat joinpoint regresí

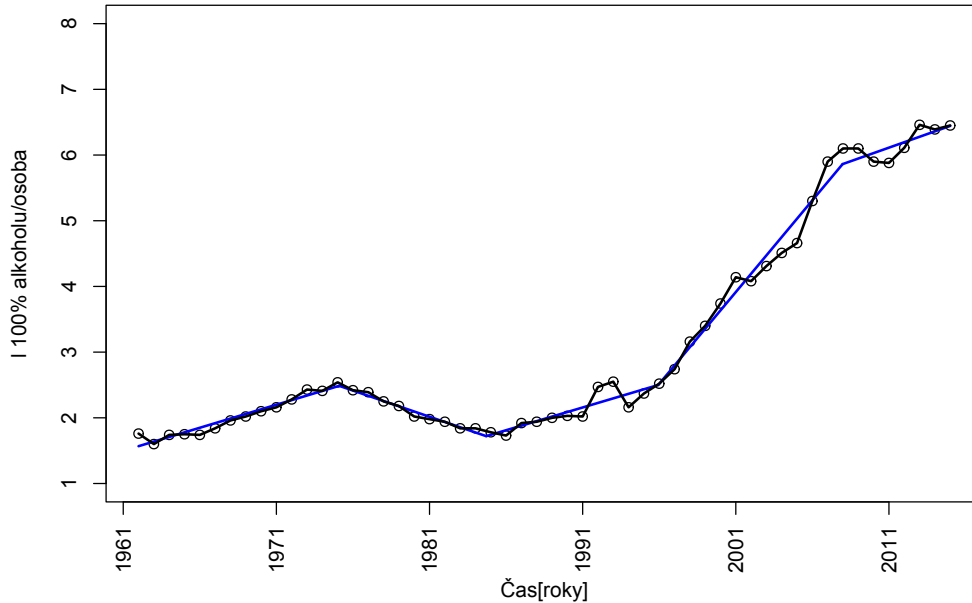
Joinpoint regrese je velice flexibilní metodou, proto není divu, že výsledky proložení dat joinpoint regresí dávaly zdaleka nejlepší výsledky co se našich charakteristik týče. Problematickým krokem u joinpoint regrese může být určení vhodných bodů zlomu a případně jejich počtu.



Obrázek 23: Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 exponenciálním trendem

V textu samotné práce jsem se rozhodl zahrnout proložení dat joinpoint regresí pro konzumaci piva v Polsku vzhledem k tomu, jaké výsledky mělo proložení dat jednotlivými trendovými funkcemi. Dále jsem zahrnul proložení dat spotřeby tvrdého alkoholu v České republice. Proložení ostatních dat je spolu s tabulkami charakteristik k nahlédnutí na příloženém CD.

Jak již bylo zmíněno na konci teoretické kapitoly o joinpoint regresi, tak počet a odhad bodů zlomu lze v softwaru R získat pomocí funkce **breakpoints**. Aplikací funkce na lineární model dostáváme odhady bodů zlomu. Aplikoval jsem tedy funkci na lineární trendovou funkci, kterou jsem prokládal data o spotřebě piva v Polsku. Výsledkem byly odhady bodů zlomu v 13., 24., 32., a 45. pozorování, což odpovídá postupně rokům 1973, 1984, 1992 a 2005. Výsledné proložení dat můžeme vidět na obrázku 24 a jednotlivé charakteristiky vhodnosti modelu v tabulce 10.



Obrázek 24: Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 joinpoint regresí se 4 body zlomu

Míry vhodnosti modelu joinpoint regrese se 4 body zlomu při proložení dat konzumace piva v Polsku v letech 1961 - 2014				
	RSC	R²	R²_{adj}	Hodnota F statistiky
Pivo	0.8891	0.9935	0.9921	747.2479 on 9 and 44 DF

Tabulka 10: Míry vhodnosti modelu joinpoint regrese se 4 body zlomu při proložení dat konzumace piva v Polsku v letech 1961 - 2014

Jednotlivé odhady regresních koeficientů pak můžeme vidět na obrázku 25, který je výstupem ze softwaru R. Veličiny U1.VektorT - U4.VektorT znázorňují změnu směrnice oproti předchozímu stavu.

Výsledný model má celkově 10 parametrů. To je i přes tak výborné charakteristiky vhodnosti modelu docela veliký počet. Zkusil jsem proto tento počet parametrů omezit tím, že jsem zvolil pouze 2 odhady bodů zlomu, čímž se počet parametrů zredukuje na 6. Nejdříve jsem vyšel z odhadů prvních dvou bodů zlomů, tedy zlomy v pozorování č. 13 a 24 a poté jsem zkusil vyjít pouze s posledních dvou odhadů, tedy v pozorování č. 32 a 45. Výsledky vidíme na obrázcích

```

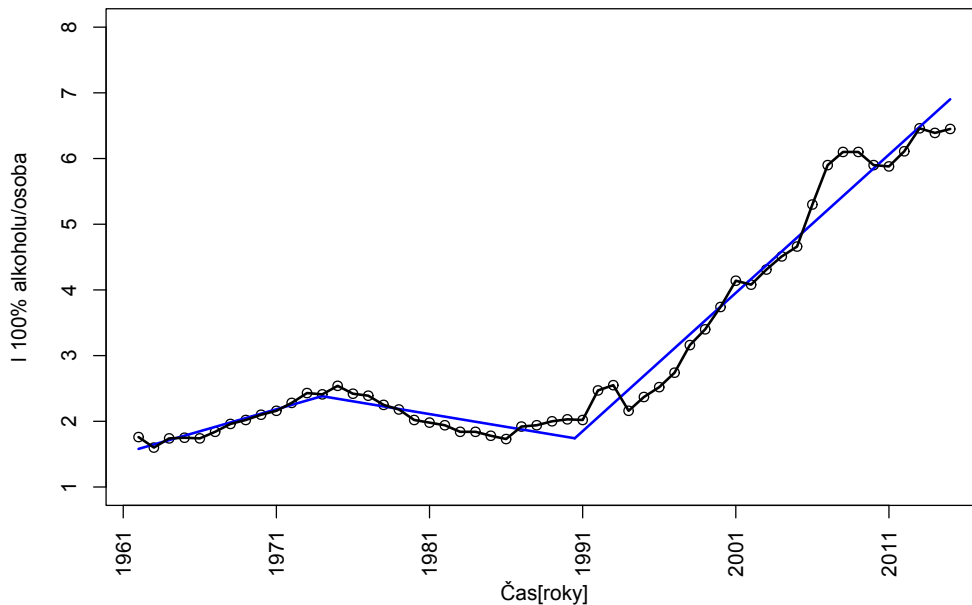
Estimated Break-Point(s):
      Est. St.Err
psi1.vektorT 14.102 0.835
psi2.vektorT 23.701 0.868
psi3.vektorT 34.935 0.570
psi4.vektorT 46.974 0.645

Meaningful coefficients of the linear terms:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.497695   0.080247  18.664 < 2e-16 ***
vektorT      0.070110   0.009425   7.439 2.61e-09 ***
U1.vektorT  -0.149840   0.020630  -7.263    NA
U2.vektorT   0.148833   0.022814   6.524    NA
U3.vektorT   0.210353   0.018028  11.668    NA
U4.vektorT  -0.196684   0.024949  -7.884    NA
---

```

Obrázek 25: Odhady regresních koeficientů joinpoint regrese se 4 body zlomu při proložení dat o spotřeba piva v Polsku v letech 1961 - 2014

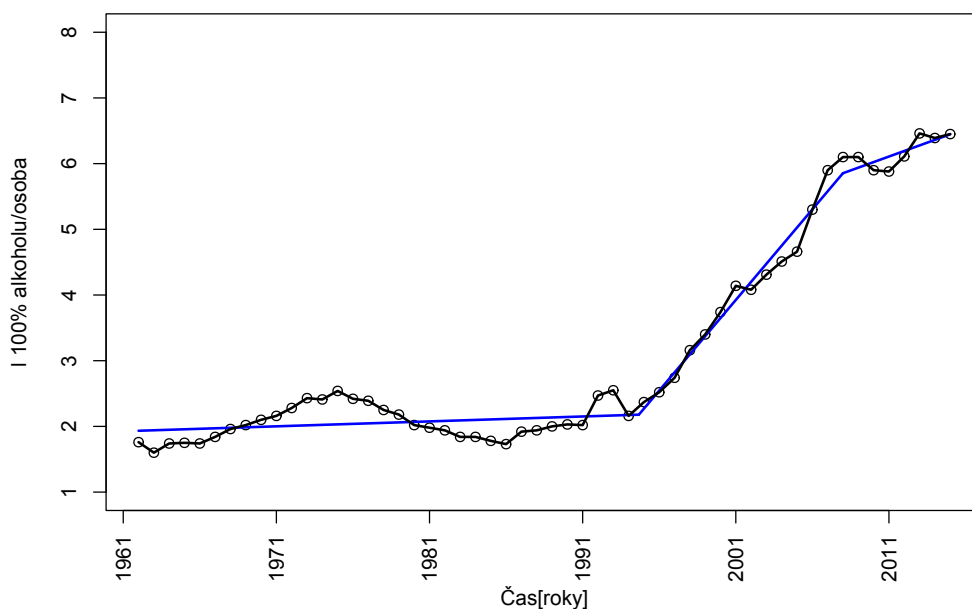
26 a 27 a v tabulce 11.



Obrázek 26: Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 joinpoint regresí se 2 body zlomu na začátku řady

Míry vhodnosti modelu joinpoint regrese se 2 body zlomu při proložení dat konzumace piva v Polsku v letech 1961 - 2014				
	RSC	R ²	R ² _{adj}	Hodnota F statistiky
Body zlomu - konec řady	2.8925	0.9788	0.9766	672.5231 on 5 and 48 DF
Body zlomu - začátek řady	2.7080	0.9801	0.9781	472.8121 on 5 and 48 DF

Tabulka 11: Míry vhodnosti modelu joinpoint regrese se 2 body zlomu při proložení dat konzumace piva v Polsku v letech 1961 - 2014

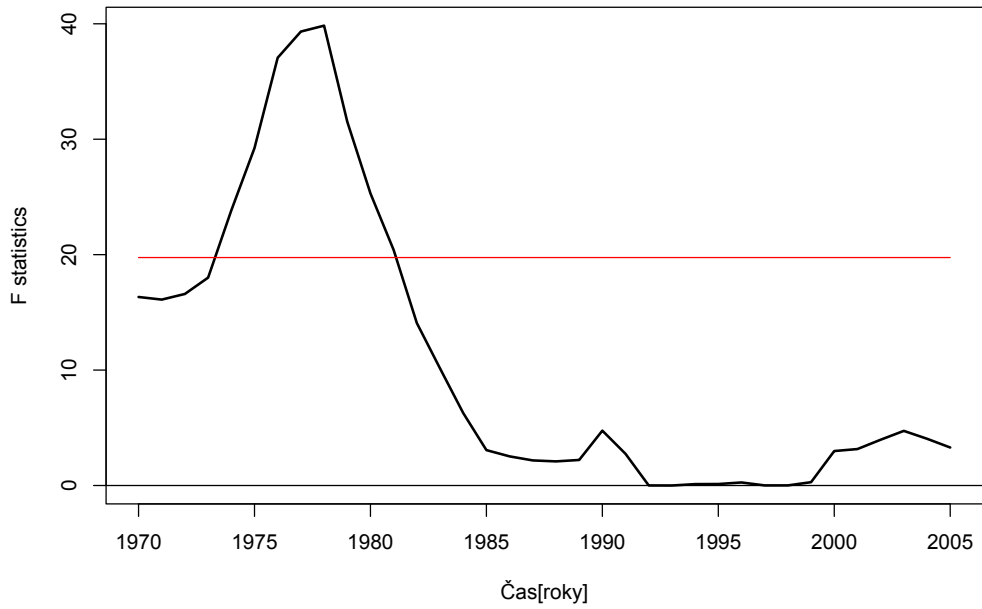


Obrázek 27: Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 joinpoint regresí se 2 body zlomu v druhé polovině řady

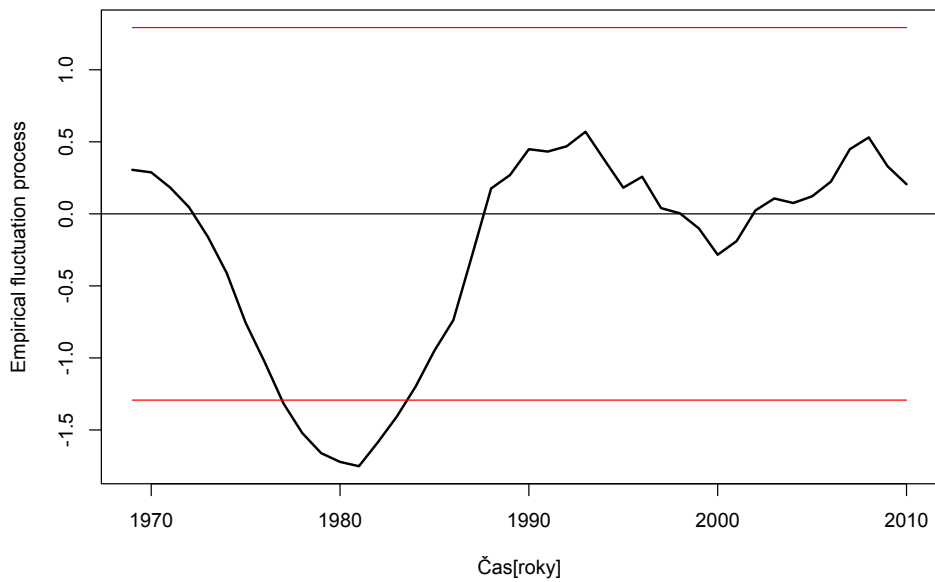
Dá se říci, že z hlediska charakteristik jsou na tom oba modely s dvěma body zlomu obdobně. Pokud se však zaměříme na posouzení modelu z hlediska možných predikcí, tak i z grafu lze vidět, že pokud volíme body zlomu na začátku časové řady, tak konec již není tak flexibilní a uchyluje se jiným směrem než samotný vývoj dat. Z tohoto hlediska je proto adekvátnější volit model, který lépe přibližuje data „aktuální“ než ten, který dobře vyrovná data stará 50 let.

Výsledky odhadů bodů zlomu můžeme porovnat s výsledky flukтуаčních a F testů strukturálních změn. Je nutné podotknout, že výstupem funkce **breakpoints** nejsou body strukturálních změn, ale pouze odhady bodů, které jsou

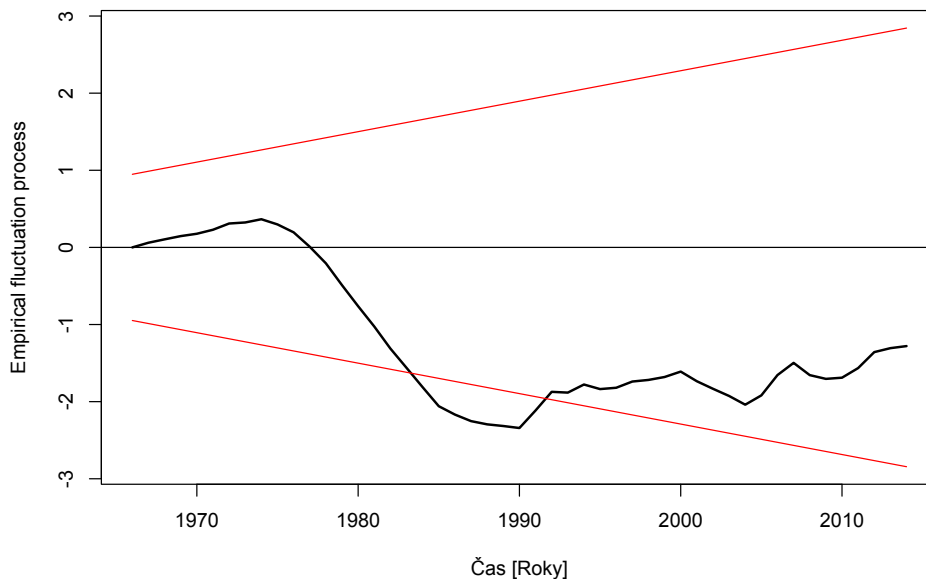
dobré jako vstupní odhady bodů zlomu pro joinpoint regresi. Funkce **breakpoints** funguje na principu dynamického programování a Bellmanově přístupu optimality, kdy vstupní model rozdělíme na $m + 1$ segmentů a algoritmus funkce poté minimalizuje RSC těchto segmentů, proto se výsledky funkce **breakpoints** a testů strukturálních změn mohou lišit. Na obrázcích 28, 29 a 30 můžeme vidět postupně výsledky F testu, MOSUM procesu a CUSUM procesu. Dle F testu dochází k významným strukturálním změnám v souboru mezi lety 1973 - 1982, přičemž jako nejpravděpodobnější bod strukturální změny se jeví 1978. Podobné výsledky jako F test dává i flukтуаční test založený na MOSUM procesu. Podle něj dochází k významným strukturálním změnám v letech 1977 - 1984, kde jako nejpravděpodobnější je změna v roce 1982. V roce 1992 a 2008 pak dochází k mírnému odchýlení procesu, nicméně ne natolik, abychom nulovou hypotézu mohli na hladině 5 % zamítnout. Mírnou změnu zaznamenáváme u flukтуаčního testu založeném na CUSUM procesu. Tam můžeme vidět, že významná strukturální změna nastává v letech 1983 - 1992. Vzhledem k tomu, že na výsledky všech testů ukazovaly na významnou strukturální změnu na začátku 80. let, tak lze prohlásit, že v tomto období opravdu dochází ke strukturální změně v datech o spotřebě piva v Polsku v letech 1961 - 2014.



Obrázek 28: F test strukturálních změn v datech o spotřebě piva v Polsku v letech 1970 - 2005



Obrázek 29: Proces MOSUM fluktuálního testu strukturálních změn v datech o spotřebě piva v Polsku v letech 1969 - 2011



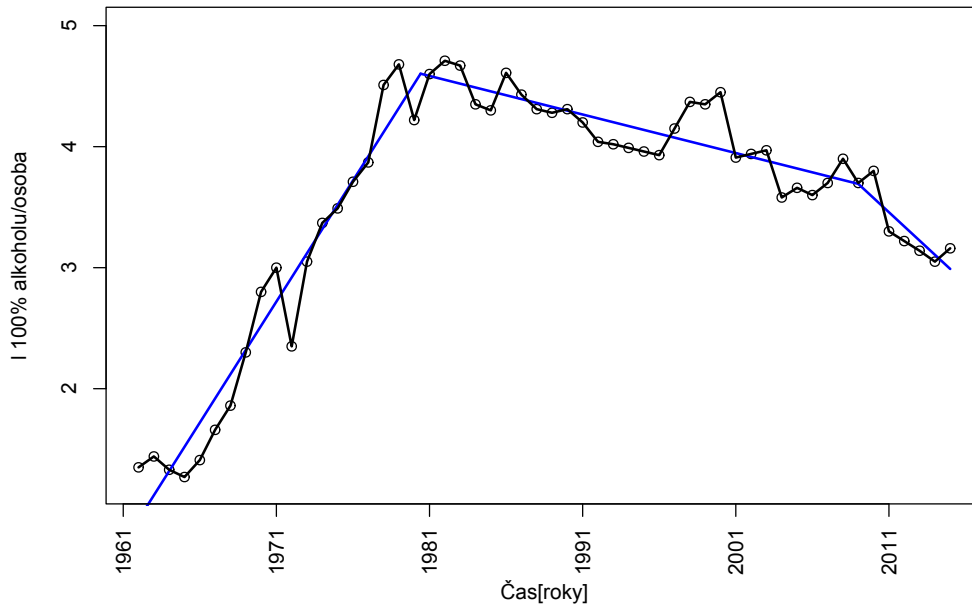
Obrázek 30: Proces CUSUM flukтуаčního testu strukturálních změn v datech o spotřebě piva v Polsku v letech 1967 - 2012

V případě aplikace funkce **breakpoints** na lineární regresní přímku, kterou jsem prokládal data o spotřebě tvrdého alkoholu v České republice v letech 1961 - 2014, jsem dostal odhady bodů zlomů v letech 1976 a 1996. To odpovídá zlomu v 16. a 36. pozorování. Proložení je možné vidět na obrázku 31 a míry vhodnosti pak v tabulce 12.

Míry vhodnosti modelu joinpoint regrese se 2 body zlomu při proložení dat konzumace tvrdého alkoholu v České republice v letech 1961 - 2014				
	RSC	R²	R²_{adj}	Hodnota F statistiky
Pivo	2.5055	0.9510	0.9459	186.3184 on 5 and 48 DF

Tabulka 12: Míry vhodnosti modelu joinpoint regrese se 2 body zlomu při proložení dat konzumace tvrdého alkoholu v České republice v letech 1961 - 2014

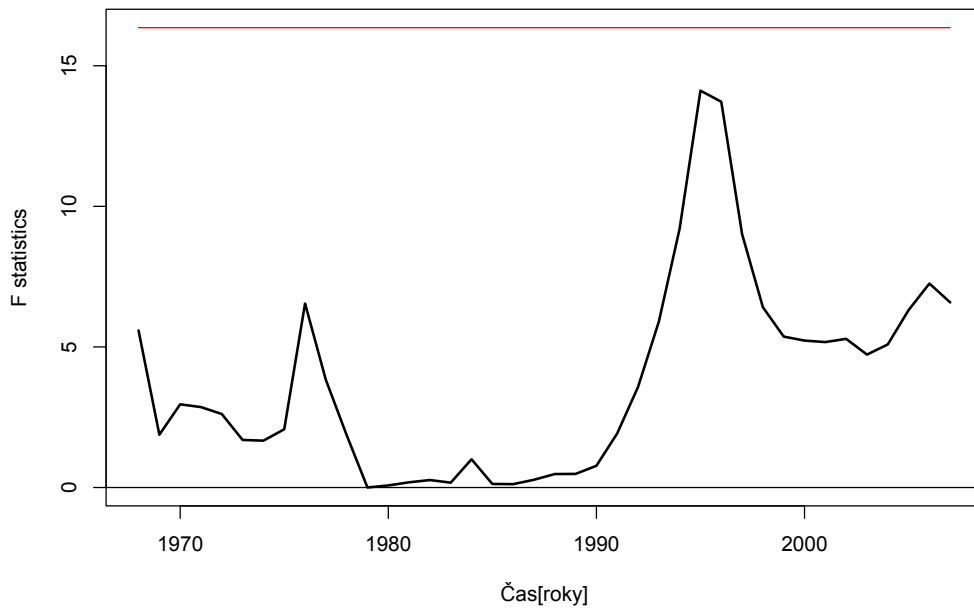
Samotné proložení dat joinpoint regrese se zdá být v pořádku. Avšak když se podíváme na samotný konec řady, tak vidíme, že poslední zlom křivky by mohl být zbytečný a v případě predikcí se zbytečně upíná ke čtyřem posledním pozorováním. Pokud se zase podíváme na strukturální změny z hlediska flukтуаčních



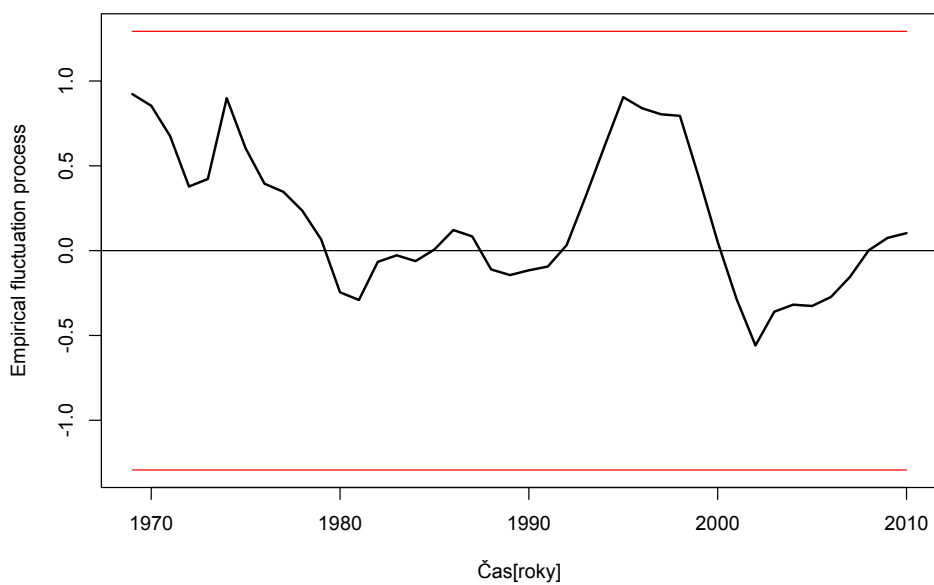
Obrázek 31: Proložení dat spotřeby piva v České republice v letech 1961 - 2014 joinpoint regresí se 2 body zlomu

a F testů, tak vidíme, že jako jediný potvrdil významnou strukturální změnu test založený na procesu CUSUM a to mezi lety 1977 - 1978. O zamítnutí nulové hypotézy o absenci strukturálních změn by se dalo na základě CUSUM procesu uvažovat i mezi lety 1999 - 2000. Tam je to ale velmi hraniční. Naopak proces MOSUM nepotvrdil žádné strukturální změny. K vychýlení u něj došlo okolo let 1975, 1994 a 1999, ale nikoliv k tak velikému, aby mohlo dojít k zamítnutí nulové hypotézy na hladině 5 %. Podobný je případ F testu, kdy největší odchýlení je v letech 1995 - 1997. Jednotlivé procesy můžeme vidět na obrázcích 32, 33 a 34.

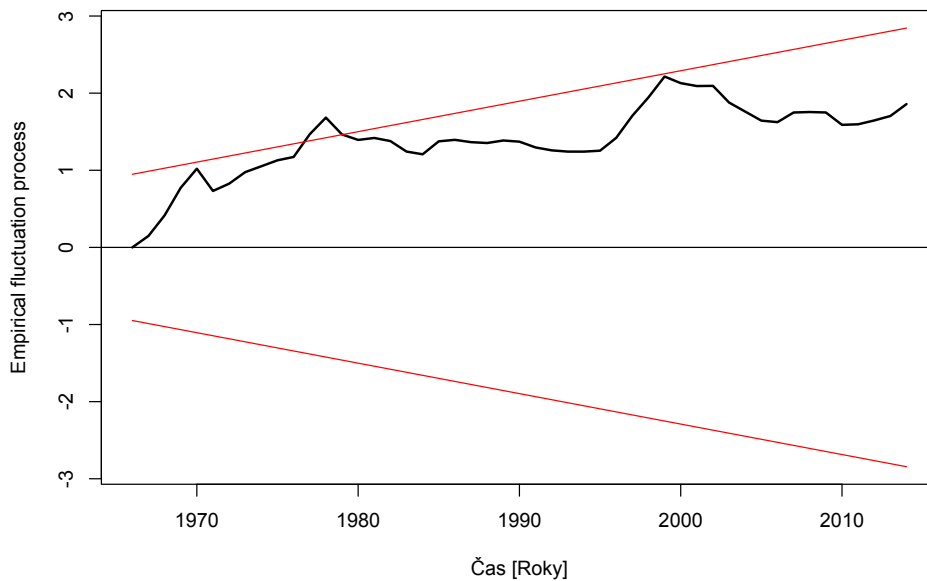
Na základě výsledků testů jsem proto došel k závěru, že vhodnějším modelem bude joinpoint regrese s jedním bodem zlomu. A to i za cenu nižších indexů determinace a většího RSC. Jistou odměnou je na druhou stranu snížení počtu parametrů ze 6 na 4. Výsledný model a mír vhodnosti můžeme vidět na obrázku 35 a v tabulce 13.



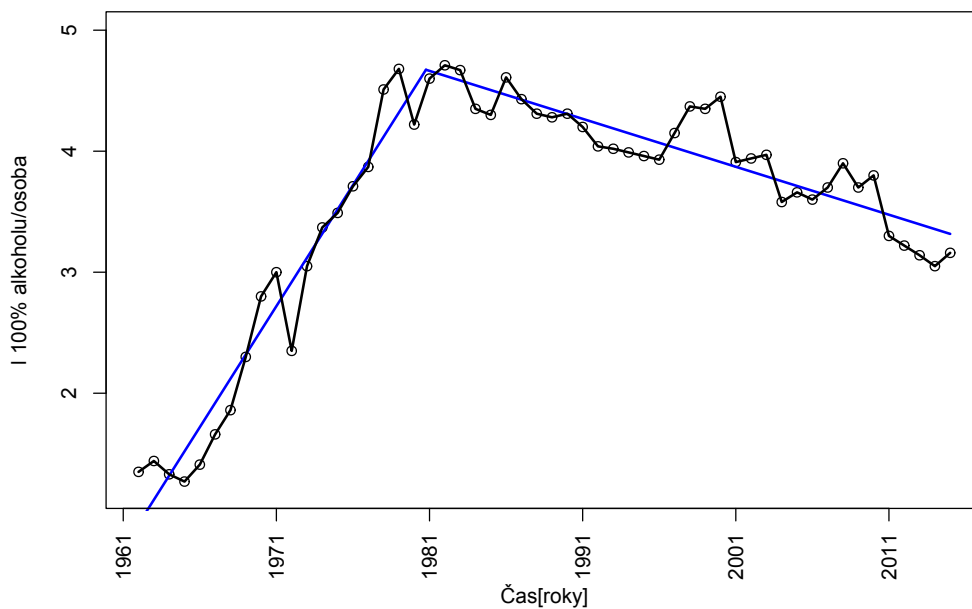
Obrázek 32: F test strukturálních změn v datech o spotřebě tvrdého alkoholu v České republice v letech 1969 - 2007



Obrázek 33: Proces MOSUM flukтуаčního testu strukturálních změn v datech o spotřebě tvrdého alkoholu v České republice v letech 1969 - 2011



Obrázek 34: Proces CUSUM flukтуаčního testu strukturálních změn v datech o spotřebě tvrdého alkoholu v České republice v letech 1967 - 2012



Obrázek 35: Proložení dat spotřeby piva v České republice v letech 1961 - 2014 joinpoint regresí s jedním bodem zlomu

Míry vhodnosti modelu joinpoint regrese s 1 bodem zlomu při proložení dat konzumace tvrdého alkoholu v České republice v letech 1961 - 2014				
	RSC	R²	R_{adj}²	Hodnota F statistiky
Pivo	2.864089	0.9440	0.9407	280.9524 on 3 and 50 DF

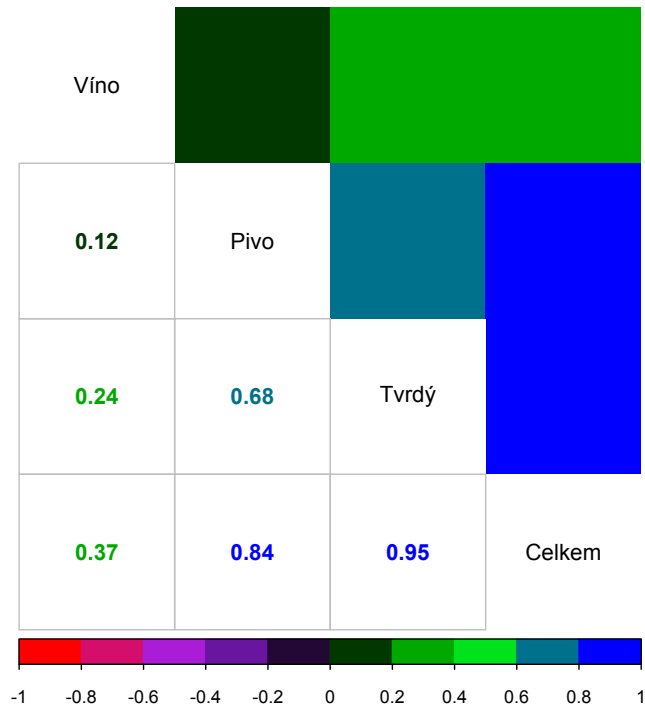
Tabulka 13: Míry vhodnosti modelu joinpoint regrese se 1 bodem zlomu při proložení dat konzumace tvrdého alkoholu v České republice v letech 1961 - 2014

6.4. Korelační analýza

Velice zajímavé vzhledem ke zkoumané oblasti jsou výsledky korelační analýzy. Zde si ukážeme, jak jsou závislé konzumace jednotlivých složek v rámci jedné země, ale také jak se ovlivňují konzumace složek v rámci dvou zemí. Díky možnosti rozdělit data na více částí (v mém případě nejčastěji na 2) se také podíváme na to, jak se mění závislosti v čase a můžeme pak tyto období porovnávat.

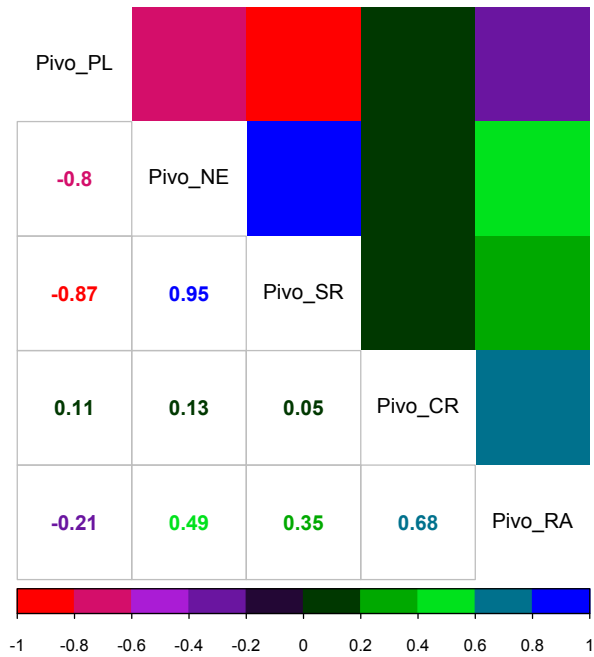
6.4.1. Korelační mapy

Co se týče korelací mezi jednotlivými složkami celkové spotřeby v rámci jednoho státu, tak se zřejmě logicky dala očekávat silná pozitivní korelace mezi složkami a celkovou spotřebou. V rámci jednoho státu jsem naopak neočekával přítomnost negativní korelace mezi dvojicí. Tomu nasvědčují i korelační koeficienty na obrázku 36, který zobrazuje jednotlivé korelační koeficienty mezi daty o spotřebě jednotlivých složek, případně celkovou spotřebou v České republice. Jistým překvapením je pro mě tak nízký korelační koeficient mezi daty o celkové spotřebě a spotřebě vína.

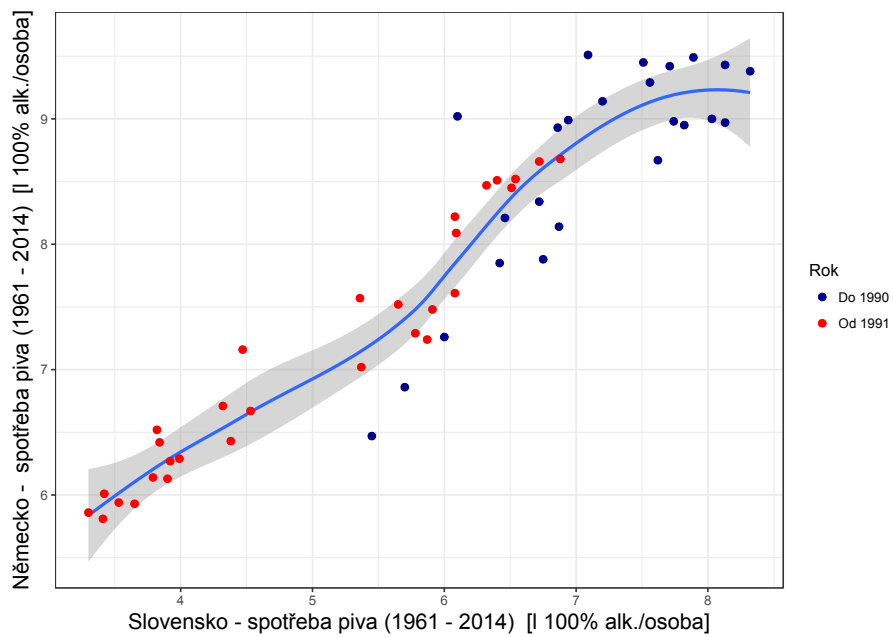


Obrázek 36: Korelační mapa dat o spotřebě alkoholu v České republice v letech 1961 - 2014

Daleko pestřejší je korelační mapa zobrazující korelační koeficienty mezi daty o spotřebě piva v jednotlivých státech. Silně negativně korelovaná jsou data o spotřebě piva mezi Polskem a Slovenskem nebo mezi Polskem a Německem. Naopak mezi Slovenskem a Německem sledujeme velmi silnou pozitivní korelaci. Pokud se podíváme na proložení dat lokální regresí - obr 38, tak vidíme, že křivka má blízko k přímce a při poklesu konzumace piva v Německu v posledních 24 letech dochází k poklesu konzumace piva i na Slovensku.



Obrázek 37: Korelační mapa dat o spotřebě piva v České republice a sousedních zemích v letech 1961 - 2014

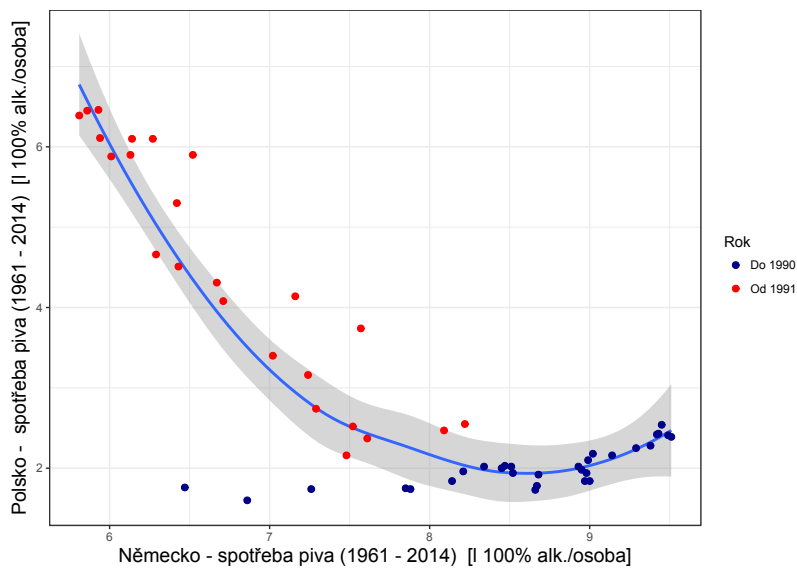


Obrázek 38: Proložení bodového grafu grafu konzumace piva v Německu a na Slovensku v letech 1961 - 2014 lokální regresi s parametrem 2/3

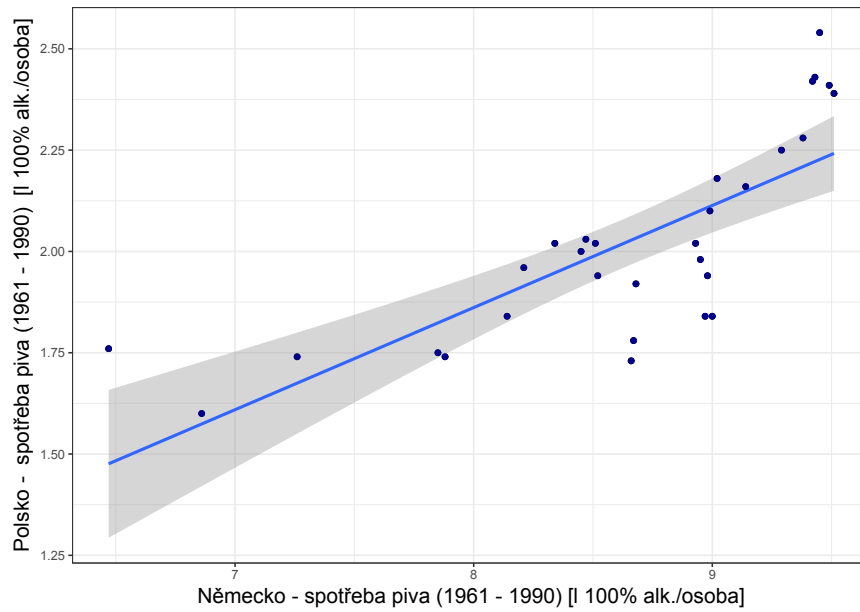
6.4.2. Proložení bodových grafů lokální regresí

V rámci analýzy pomocí bodových grafů jsem se zaměřil především na to, jakým způsobem se mění závislost v čase mezi jednotlivými složkami. Na začátku jsem si rozdělil časové řady na dvě části - před rokem 1990 a po tomto roce. Pomocí barevného odlišení v grafu jsem pak zjišťoval, jak se nejen posunula konzumace jednotlivých složek, ale jak se mění proložení dat v čase. V mnoha případech jde o velice zajímavé výsledky.

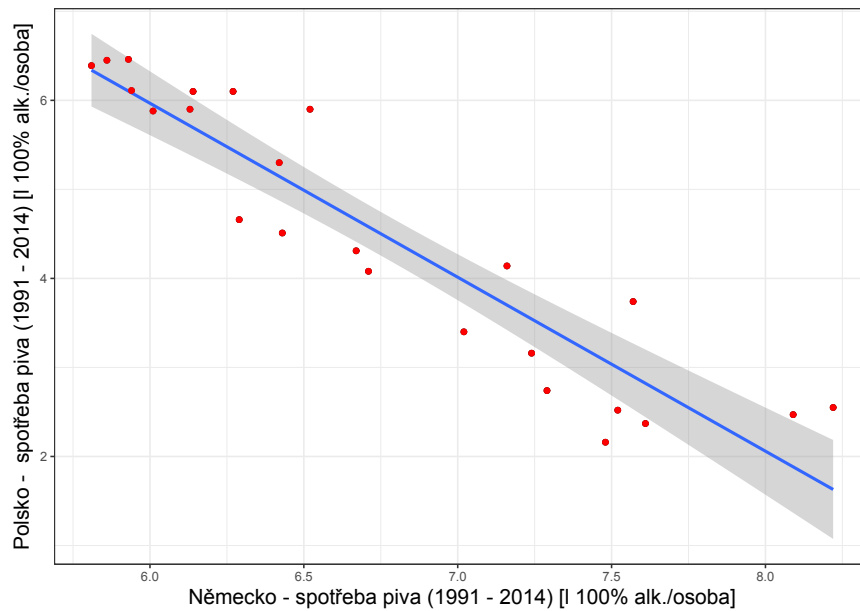
Například na obrázku č. 39 můžeme vidět bodový graf konzumace piva v Polsku a Německu. Už na první pohled je jasné, že v časové řadě došlo nejen ke změně úrovně konzumací piva v jednotlivých státech, kde Polsko oproti počátečním pozorováním svou spotřebu zvedlo až o 4 litry čistého alkoholu na osobu a naopak v Německu došlo ke snížení této úrovně, ale také to, že pokud proložíme data do roku 1990, tedy body znázorněné modrou barvou, tak můžeme mluvit o pozitivní korelaci. Proložíme-li data od roku 1991, tak naopak můžeme mluvit o silně negativní korelaci. Proložení rozdělenými časovými řadami je pak možné porovnat na obrázcích č. 40 a 41.



Obrázek 39: Proložení bodového grafu grafu konzumace piva v Polsku a Německu v letech 1961 - 2014 lokální regresí s parametrem 2/3

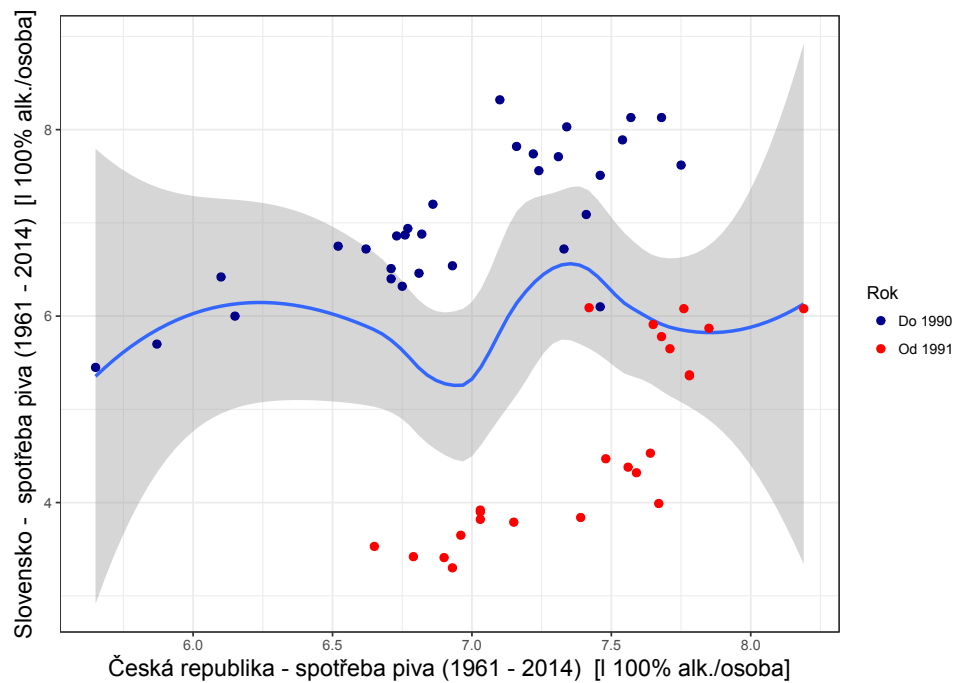


Obrázek 40: Proložení bodového grafu konzumace piva v Polsku a Německu v letech 1961 - 1990 regresní přímkou



Obrázek 41: Proložení bodového grafu konzumace piva v Polsku a Německu v letech 1991 - 2014 regresní přímkou

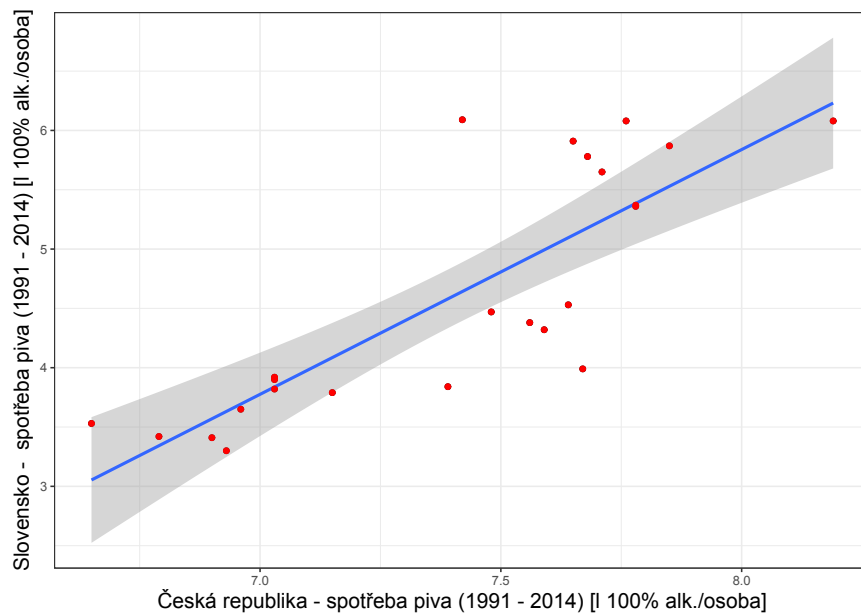
Další ze zajímavých příkladů ilustruje situaci, kdy jsou v souboru pozorování tvořeny shluky. Jedná se o bodový graf spotřeby piva v České republice a na Slovensku v letech 1961 - 2014. Pokud vezmeme v úvahu celé časové řady, tak můžeme na obrázku 42 vidět, že proložení dat nedává úplně adekvátní výsledek. Tím, že na Slovensku došlo v průběhu let ke snížení spotřeby piva a u nás se konzumace ve své podstatě nijak dramaticky neměnila, tak data vytvořila dva shluky. Pokud shluky vezmeme jednotlivě a křivku proložíme nimi, tak zjistíme, že v samotné korelaci k žádným dramatickým změnám nedošlo, což dokazují i korelační koeficienty jednotlivých souborů. Korelační koeficient pro soubor do roku 1990 je 0.7786 a korelační koeficient pro soubor v letech 1991 - 2014 je 0.8253. Proložení rozdělených souborů je pak k vidění na obrázcích č. 43 a 44.



Obrázek 42: Proložení bodového grafu konzumace piva v České republice a na Slovensku v letech 1961 - 2014 lokální regresí s parametrem 2/3



Obrázek 43: Proložení bodového grafu konzumace piva v České republice a na Slovensku v letech 1961 - 1990 regresní přímkou



Obrázek 44: Proložení bodového grafu konzumace piva v České republice a na Slovensku v letech 1991 - 2014 regresní přímkou

Závěr

Cílem diplomové práce bylo analyzovat data o spotřebě alkoholu v České republice a okolních zemích, najít vhodné modely pro vyrovnání dat, zjistit strukturální změny v datech a provést analýzu závislostí na základě korelace. Snad se také dá říci, že díky zvolenému tématu bylo druhotným cílem popularizovat studium matematiky a matematiku jako takovou.

V první části diplomové práce byly vysvětleny základní teoretické aspekty potřebné k samotné analýze. Postupně se jim věnovaly kapitoly o časových řadách, joinpoint regresi, korelační analýze a lokální regresi. Na konec každé kapitoly byly přiřazeny komentáře s možným provedením v softwaru R.

Samotná analýza dat byla zahrnuta do druhé části, tedy části praktické. Na začátku analýzy byla představena data, se kterými se v analýze pracovalo. Následně byly představeny základní číselné charakteristiky vybraných časových řad a to ve formě boxplotů a souhrnných tabulek. Při hledání vhodného modelu se práce zaměřila jak na proložení dat trendovými funkcemi, ze kterých byly použity trendy lineární, kvadratické a exponenciální, tak na proložení dat joinpoint regresi. Jak se ukázalo, nebyly tyto jednoduché trendové funkce příliš vhodné k modelování námi analyzovaných řad. Avšak našly se i světlé momenty, kdy především trend kvadratický, i přes svou jednoduchost, proložil data s uspokojivými mírami vhodnosti modelu. Jiný případ byl využití joinpoint regrese. Ta se díky své flexibilitě přizpůsobila datům a docházelo tak k minimálním odchylkám vyrovnaných hodnot od reality. Navíc se pomocí testů strukturálních změn, především flukтуаčních testů a testů založených na F statistice, podařilo prokázat přítomnost strukturální změny v bodech zlomu, ve kterých při joinpoint regresi křivka mění svou směrnici. V poslední části analýzy pak bylo ukázáno, jak se mění závislost konzumace alkoholu mezi státy v čase a také změna závislosti mezi konzumací jednotlivých složek v rámci jednoho státu.

Diplomovou práci jsem vypracovával se zájmem a zaujetím. V závěru mohu říci, že mi psaní této práce mnohé dalo. Od jiného úhlu pohledu na analýzu dat, než jaký nám byl poskytován v základních kurzech, po prohloubení svých znalostí

o časových řadách a možných přístupech k jejich modelování. Při zpracování samotné analýzy jsem se také sblížil se softwarem R, kde jsem si své znalosti posunul dále.

Literatura

- [1] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*, 1. vydání. SNTL, Praha, 1986.
- [2] Kočenda, E.: *Element of time series econometrics an applied approach*, 2. vydání. Karolinum Press, Praha, 2014.
- [3] Sebera, M., Klárová, R., Zháněl, J.: *Časové řady*, 1. vydání. Masarykova Univerzita, Brno, 2014.
- [4] Informační Bulletin České statistické společnosti [online], dostupné z: <http://www.statspol.cz/bulletiny/ib-2011-3-4-web.pdf>, [cit. 2017-03-23]
- [5] strucchange: An R Package for Testing for Structural change in Linear Regression Models [online], dostupné z: <https://cran.r-project.org/web/packages/strucchange/vignettes/strucchange-intro.pdf>, [cit. 2017-]
- [6] Regression Models with Breakpoints/Changepoints Estimation [online], dostupné z <https://cran.r-project.org/web/packages/segmented/segmented.pdf>, [cit. 2017-]
- [7] Venables, W. N., Ripley, B. D.: *Modern applied statistic with S*, 4. vydání. Springer, New York, 2002.
- [8] Artl, J., Artlová, M., Rublíková, E.: *Analýza ekonomických časových řad s příklady*, 1. vydání. VŠE ,Praha ,2002.
- [9] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky*, 1. vydání, Univerzita Palackého v Olomouci ,Olomouc, 2013.
- [10] Fox, J.: *Applied Regression Analysis and Generalized Linear Models*, 2. vydání, Sage, Los Angeles ,2008.

Seznam obrázků

1	Exponenciální trend	15
2	Modifikovaný exponenciální trend	17
3	Gompertzova křivka	18
4	Logistický trend	19
5	Model, pomocí kterého je iteračně odhadována joinpoint regrese. .	30
6	Negativně a pozitivně korelovaná data	32
7	Nekorelovaná data	32
8	Konzumace alkoholu v ČR v letech 1961 - 2014	41
9	Konzumace alkoholu v Německu v letech 1961 - 2014	41
10	Konzumace alkoholu v SR v letech 1961 - 2014	42
11	Konzumace alkoholu v Rakousku v letech 1961 - 2014	42
12	Konzumace alkoholu v Polsku v letech 1961 - 2014	43
13	Procentuální zastoupení konzumace piva, vína a tvrdého alkoholu v České republice v letech 1961 - 2014	43
14	Boxploty konzumace alkoholu v ČR v letech 1961 - 2014	44
15	Boxploty konzumace alkoholu na Slovensku v letech 1961 - 2014 .	45
16	Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 lineárním trendem	47
17	Proložení dat spotřeby piva na Slovensku v letech 1961 - 2014 lineárním trendem	47
18	Proložení dat spotřeby tvrdého alkoholu v Německu v letech 1961 - 2014 lineárním trendem	48
19	Proložení dat celkové konzumace alkoholu a konzumace tvrdého alkoholu v ČR v letech 1961 - 2014 kvadratickým trendem	49
20	Proložení dat celkové konzumace alkoholu a konzumace piva v Německu v letech 1961 - 2014 kvadratickým trendem	49
21	Proložení dat celkové konzumace piva a tvrdého alkoholu na Slo- vensku v letech 1961 - 2014 kvadratickým trendem	50

22	Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 kvadratickým trendem	52
23	Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 exponenciálním trendem	53
24	Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 joinpoint regresí se 4 body zlomu	54
25	Odhady regresních koeficientů joinpoint regrese se 4 body zlomu při proložení dat o spotřeba piva v Polsku v letech 1961 - 2014 . .	55
26	Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 joinpoint regresí se 2 body zlomu na začátku řady	55
27	Proložení dat spotřeby piva v Polsku v letech 1961 - 2014 joinpoint regresí se 2 body zlomu v druhé polovině řady	56
28	F test strukturálních změn v datech o spotřebě piva v Polsku v letech 1970 - 2005	58
29	Proces MOSUM flukтуаčního testu strukturálních změn v datech o spotřebě piva v Polsku v letech 1969 - 2011	58
30	Proces CUSUM flukтуаčního testu strukturálních změn v datech o spotřebě piva v Polsku v letech 1967 - 2012	59
31	Proložení dat spotřeby piva v České republice v letech 1961 - 2014 joinpoint regresí se 2 body zlomu	60
32	F test strukturálních změn v datech o spotřebě tvrdého alkoholu v České republice v letech 1969 - 2007	61
33	Proces MOSUM flukтуаčního testu strukturálních změn v datech o spotřebě tvrdého alkoholu v České republice v letech 1969 - 2011	61
34	Proces CUSUM flukтуаčního testu strukturálních změn v datech o spotřebě tvrdého alkoholu v České republice v letech 1967 - 2012	62
35	Proložení dat spotřeby piva v České republice v letech 1961 - 2014 joinpoint regresí s jedním bodem zlomu	62
36	Korelační mapa dat o spotřebě alkoholu v České republice v letech 1961 - 2014	64

37	Korelační mapa dat o spotřebě piva v České republice a sousedních zemích v letech 1961 - 2014	65
38	Proložení bodového grafu grafu konzumace piva v Německu a na Slovensku v letech 1961 - 2014 lokální regresí s parametrem $2/3$.	65
39	Proložení bodového grafu grafu konzumace piva v Polsku a Německu v letech 1961 - 2014 lokální regresí s parametrem $2/3$	66
40	Proložení bodového grafu konzumace piva v Polsku a Německu v letech 1961 - 1990 regresní přímkou	67
41	Proložení bodového grafu konzumace piva v Polsku a Německu v letech 1991 - 2014 regresní přímkou	67
42	Proložení bodového grafu konzumace piva v České republice a na Slovensku v letech 1961 - 2014 lokální regresí s parametrem $2/3$.	68
43	Proložení bodového grafu konzumace piva v České republice a na Slovensku v letech 1961 - 1990 regresní přímkou	69
44	Proložení bodového grafu konzumace piva v České republice a na Slovensku v letech 1991 - 2014 regresní přímkou	69

Seznam tabulek

1	konzumace alkoholu v litrech čistého alkoholu na osobu starší 15 let v ČR v letech 1961 - 2014	40
2	Tabulka základních číselných charakteristik dat konzumace alkoholu v ČR v letech 1961 - 2014	45
3	Tabulka základních číselných charakteristik dat konzumace alkoholu na Slovensku v letech 1961 - 2014	46
4	Míry vhodnosti lineárního trendu při proložení dat spotřeby piva v Polsku	46
5	Míry vhodnosti kvadratického trendu při proložení dat celkové konzumace a konzumace tvrdého alkoholu v ČR v letech 1961 - 2014	48
6	Míry vhodnosti kvadratického trendu při proložení dat konzumace piva a celkové konzumace v Německu v letech 1961 - 2014	48
7	Míry vhodnosti kvadratického trendu při proložení dat konzumace piva a tvrdého alkoholu na Slovensku v letech 1961 - 2014	50
8	Míry vhodnosti kvadratického trendu při dat spotřeby piva v Polsku v letech 1961 - 2014	51
9	Míry vhodnosti exponenciálního trendu při dat spotřeby piva na Slovensku v letech 1961 - 2014	52
10	Míry vhodnosti modelu joinpoint regrese se 4 body zlomu při proložení dat konzumace piva v Polsku v letech 1961 - 2014	54
11	Míry vhodnosti modelu joinpoint regrese se 2 body zlomu při proložení dat konzumace piva v Polsku v letech 1961 - 2014	56
12	Míry vhodnosti modelu joinpoint regrese se 2 body zlomu při proložení dat konzumace tvrdého alkoholu v České republice v letech 1961 - 2014	59
13	Míry vhodnosti modelu joinpoint regrese se 1 bodem zlomu při proložení dat konzumace tvrdého alkoholu v České republice v letech 1961 - 2014	63