

Univerzita Hradec Králové
Přírodovědecká fakulta
Katedra matematiky

Analýza časových řad v programu Gretl

Bakalářská práce

Autor: Andrea Karešová
Studijní program: B1103 Aplikovaná matematika
Studijní obor: Finanční a pojistná matematika
Vedoucí práce: RNDr. Michal Čihák, Ph.D.

Univerzita Hradec Králové

Přírodovědecká fakulta

Zadání bakalářské práce

Autor:	Andrea Karešová
Studijní program:	B1103 Aplikovaná matematika
Studijní obor:	Finanční a pojistná matematika
Název práce:	Analýza časových řad v programu Gretl
Název práce v AJ:	Time series analysis in Gretl
Cíl a metody práce:	Analýzu ekonomických časových řad je možno provádět pomocí komerčních statistických programů (Statistica, Statgraphics, apod.). Pro tyto účely lze však velmi dobře použít i software Gretl (Gnu Regression, Econometrics and Time-series Library), který je k dispozici zdarma. Cílem práce je ukázat na příkladech konkrétních ekonomických časových řad postupy jejich analýzy pomocí programu Gretl.
Garantující pracoviště:	katedra matematiky Přírodovědecké fakulty UHK
Vedoucí práce:	RNDr. Michal Čihák, Ph.D.
Oponent:	Mgr. Jitka Kühnová, Ph.D.
Datum zadání práce:	9. 3. 2014
Datum odevzdání práce:	16. 12. 2015

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a že jsem v seznamu použité literatury uvedla všechny prameny, ze kterých jsem vycházela.

V Hradci Králové dne 16. 12. 2015

Andrea Karešová

Anotace

KAREŠOVÁ, Andrea. *Analýza časových řad v programu Gretl*. Hradec Králové, 2015. Bakalářská práce. Univerzita Hradec Králové, Přírodovědecká fakulta.

Analýzu ekonomických časových řad je možno provádět pomocí komerčních statistických programů (Statistica, Statgraphics, apod.). Pro tyto účely lze však velmi dobře použít i software Gretl (Gnu Regression, Econometrics and Time-series Library), který je k dispozici zdarma. Cílem práce je ukázat na příkladech konkrétních ekonomických časových řad postupy jejich analýzy pomocí programu Gretl.

Klíčová slova

časové řady, trend, metody, Gretl, model

Annotation

KAREŠOVÁ, Andrea. *Time series analysis in Gretl*. Hradec Králové, 2015. Bachelor Thesis. University of Hradec Králové, Faculty of Science.

Analysis of economic time series can be done using commercial statistical systems (Statistica, Statgraphics, etc.). For these purpose, however, free software Gretl (Gnu Regression, Econometrics and Time-series Library) can be used. The aim of this work is to show examples of specific methods of analysis of economic time series using Gretl.

Keywords

time series, trend, methods, Gretl, model

Obsah

Seznam tabulek	7
Seznam obrázků	7
Úvod	8
1 Teoretické základy pro analýzu časových řad	9
1.1 Vymezení termínu časové řady	9
1.2 Druhy časových řad	9
1.2.1 Očištění časové řady o důsledky kalendářních vlivů	11
1.2.1.1 Příklad	11
1.3 Některé specifické problémy časových řad	13
1.4 Základní přístupy k analýze časových řad	14
1.4.1 Dekompozice časové řady	14
1.4.2 Boxova-Jenkinsova metodologie	14
1.4.3 Lineární dynamické modely	15
1.4.4 Spektrální analýza časových řad	16
1.5 Předpovědi v časových řadách	16
1.6 Představení ekonometrického softwaru	17
2 Dekompozice časových řad	18
2.1 Trend	18
2.2 Sezónní složka	18
2.3 Cyklická složka	19
2.4 Náhodná složka	19
3 Modely popisující trendovou složku	20
3.1 Konstantní trend	20
3.2 Lineární trend	21
3.3 Kvadratický trend	22
3.4 Exponenciální trend	22
3.5 Modifikovaný exponenciální trend	23
3.6 Logistický trend	24
3.7 Gompertzova křivka	25
3.8 Příklad	26
3.8.1 Popis dat	26
3.8.2 Úprava dat	26
3.8.3 Uživatelské rozhraní	26

3.8.4	Import dat.....	27
3.8.5	Struktura dat	28
3.8.6	Úprava atributů	29
3.8.7	Sestrojení grafu	29
3.8.8	Odhad parametrů	30
3.8.9	Hledání vhodného modelu.....	33
3.8.9.1	Lineární trend	33
3.8.9.2	Kvadratický trend.....	34
3.8.9.3	Exponenciální trend.....	36
3.8.10	Celkový graf modelů.....	41
3.8.11	Předpověď vývoje budoucích hodnot časové řady.....	42
3.8.12	Chyby v předpovědích	44
3.8.13	Shrnutí.....	45
4	Metoda klouzavých průměrů.....	46
4.1	Konstrukce klouzavých průměrů	46
4.2	Vlastnosti klouzavých průměrů	46
4.3	Volba parametrů klouzavých průměrů	46
4.4	Jednoduché klouzavé průměry.....	47
4.5	Centrované klouzavé průměry	48
4.6	Vážené klouzavé průměry	48
4.7	Exponenciální klouzavé průměry.....	49
4.8	Příklad.....	49
4.8.1	Jednoduché klouzavé průměry	50
4.8.2	Centrované klouzavé průměry.....	52
4.8.3	Exponenciální klouzavé průměry	53
5	Exponenciální vyrovnání	55
5.1	Jednoduché exponenciální vyrovnání.....	55
5.1.1	Volba vyrovnávací konstanty α	56
5.2	Dvojitě exponenciální vyrovnání.....	56
5.3	Trojitě exponenciální vyrovnání.....	58
	Závěr	59
	Seznam použitých zdrojů	60
	Přílohy	62

Seznam tabulek

Tabulka 1- Tabulka výroby cementu roku 1998 a očištění časové řady	12
--	----

Seznam obrázků

Obr. 1.1: Graf původních a očištěných hodnot výroby cementu v ČR.....	12
Obr. 3.1: Modifikovaný exponenciální trend.....	23
Obr. 3.2: Logistický trend a jeho derivace.....	24
Obr. 3.3: Gompertzova křivka, růstová funkce.....	25
Obr. 3.4: Hlavní okno programu Gretl.....	27
Obr. 3.5: Import dat	27
Obr. 3.6: Průvodce strukturou dat 1	28
Obr. 3.7: Průvodce strukturou dat 2	28
Obr. 3.8: Úprava atributů.....	29
Obr. 3.9: Postup vykreslení grafu.....	30
Obr. 3.10: Časová řada „food and beverage stores“ – vývoj dat, leden 1992 – srpen 2015.....	30
Obr. 3.11: Nastavení proměnných k modelu.....	31
Obr. 3.12: Okno s výsledkem regrese.....	32
Obr. 3.13: Odhad parametrů – lineární trend.....	33
Obr. 3.14: Graf - aplikace lineárního trendu.....	34
Obr. 3.15: Odhad parametrů - kvadratický trend.....	35
Obr. 3.16: Graf – aplikace kvadratického trendu.....	36
Obr. 3.17: Odhad parametrů – exponenciální trend	37
Obr. 3.18: Graf – logaritmovaná časová řada	38
Obr. 3.19: Definování předpovědi.....	38
Obr. 3.20: Předpovědi – lineární trend.....	39
Obr. 3.21: Hlavní okno programu Gretl – nadefinované proměnné	40
Obr. 3.22: Popisné statistiky	40
Obr. 3.23: Graf – aplikace exponenciálního trendu	41
Obr. 3.24: Popisné statistiky 2.....	41
Obr. 3.25: Celkový graf trendů	42
Obr. 3.26: Definování předpovědi na následujících 5 let	43
Obr. 3.27: Celkový graf s předpovědí na 5 let.....	44
Obr. 3.28: Graf čtvercových chyb	45
Obr. 4.1: Graf vývoje počtu živě narozených dětí v ČR.....	50
Obr. 4.2: Filtr – jednoduchý klouzavý průměr.....	50
Obr. 4.3: Jednoduché klouzavé průměry délky 3	51
Obr. 4.4: Jednoduché klouzavé průměry délky 9	51
Obr. 4.5: Jednoduché klouzavé průměry délky 15.....	52
Obr. 4.6: Celkový graf centrovaných klouzavých průměrů.....	53
Obr. 4.7: Filtr - exponenciální klouzavý průměr.....	54
Obr. 4.8: Celkový graf exponenciálních klouzavých průměrů	54

Úvod

Analýza ekonomických časových řad se stala v posledních čtyřech desetiletích velice se rozvíjející disciplínou. V posledních letech začalo vznikat mnoho nových efektivních postupů a metod, které modelují časovou řadu. Nyní se nacházíme v době, kdy není možné provádět důležitá ekonomická rozhodnutí bez propracované analýzy ekonomických ukazatelů. Analýza časových řad najde uplatnění v mnoha oborech lidské činnosti jako například v medicíně, v ekonomii, v technice či ve společenských vědách.

Časové řady se v reálném světě vyskytují všude kolem nás. Už sama o sobě nám časová řada udává podstatnou informaci o tom, jak se ukazatel vyvíjel v daném období. Hlavním úkolem matematika tedy je, aby tento vývoj matematicky popsal a případně předpověděl chování ukazatele v budoucnosti.

Cílem analýzy časových řad je konstrukce vhodného modelu. Ve většině případů má formu jedné nebo více stochastických rovnic. Na základě takového modelu můžeme lépe porozumět mechanismu, jehož prostřednictvím jsou vytvářeny sledované údaje.

K modelování časových řad nám slouží celá řada metod, mezi které patří dekompoziční metoda, Boxova-Jenkinsova metodologie, lineární modely nebo spektrální analýza časových řad. V této práci se zaměřím pouze na dekompoziční metodu.

Tématem práce jsou statistické metody, které popisují trendovou složku časových řad. Budu se tedy zabývat klasickými postupy eliminace trendu a tzv. adaptivními metodami, kterými je metoda klouzavých průměrů a exponenciální vyrovnání. Uvedené metody jsou doplněny příklady ekonomických časových řad analyzovaných v softwaru Gretl.

1 Teoretické základy pro analýzu časových řad

V této kapitole vymezím pojem časová řada. Seznámím vás s různými druhy časových řad a s některými specifickými problémy. Dále předkládám stručnou charakteristiku základních přístupů k analýze časových řad a předpovědi v časových řadách. Na závěr se krátce zmíním o programu Gretl, se kterým budu v této bakalářské práci pracovat.

1.1 Vymezení termínu časové řady

Časová řada je chronologicky uspořádaná posloupnost určitého stochastického ukazatele, který je vymezen v čase věcně a prostorově shodně. Z praktického hlediska se jedná o řadu čísel tvořených hodnotami nějaké veličiny, které jsou uspořádány od nejstarších po nejmladší nebo naopak. Časová řada je chápána jako realizace náhodného procesu. [13]

Příkladem časové řady může být například zápis dat srážkoměrné stanice, které jsou zaznamenány po pěti minutách. Časová vzdálenost mezi sousedními prvky této řady je konstantní, proto mluvíme o ekvidistantní časové řadě (tj. jednotlivé hodnoty zahrnují stejně dlouhá období).

Cipra [6] uvádí zápis statistické časové řady, například typu

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad (1.1)$$

kde t označuje čas, β_0 a β_1 jsou parametry tzv. lineárního trendu a ε_t je tzv. bílý šum, tj. nekorelovaná náhodná veličina s nulovou střední hodnotou a s konstantním rozptylem. Chování této řady je zatíženo nejistotou, na rozdíl od deterministické časové řady, například typu $y_t = \cos(2\pi f t)$, kde f je parametr reprezentující tzv. frekvenci, jejíž chování lze striktně popsat matematickým vzorcem, takže lze například zkonstruovat její přesnou předpověď.

1.2 Druhy časových řad

Časové řady můžeme dělit podle různých hledisek, kterými například jsou:

Časové řady intervalové a okamžikové

Podle charakteru dat, jejichž hodnoty tvoří časovou řadu, členíme řady na intervalové a okamžikové. *Intervalovou časovou řadou* rozumíme celou řadu ukazatelů, u kterých data závisí na délce intervalu, který je sledován [15]. U tohoto ukazatele je možné tvořit součty. Příkladem takové řady může být například měsíční výroba cementu v ČR nebo počet rozvodů za rok v ČR. *Okamžiková časová*

řada je řada ukazatelů, u kterých se data vztahují k určitému okamžiku [15]. Součty hodnot této řady nedávají žádný reálný smysl. Příkladem takové řady je počet nezaměstnaných lidí v ČR v jednotlivých měsících nebo počet obyvatel v ČR k 31.12.

V případě, že chceme graficky znázornit časovou řadu, podle které pak poznáme jaká je a jaký bude její další vývoj, je nutné rozlišovat, o jakou časovou řadu se jedná. Okamžikové časové řady se výhradně znázorňují spojnicovými grafy. Zatímco intervalové časové řady lze podle Kropáče [11] graficky znázorňovat třemi způsoby:

- *sloupkovými grafy* – tyto grafy jsou vyobrazeny obdélníky, kde základny se rovnají délkám intervalů a výšky jsou rovny hodnotě časové řady v daném intervalu,
- *hůlkovými grafy* – zde se hodnoty časové řady vynášejí ve středech odpovídajících intervalů jako úsečky,
- *spojnicovými grafy* – zde jsou hodnoty časové řady vynášeny ve středech příslušných intervalů jako body, které se spojují úsečkami.

Časové řady dlouhodobé a krátkodobé

Časové řady lze dále dělit podle periodicity na krátkodobé, které se vztahují zejména ke čtvrtletnímu, měsíčnímu či dennímu časovému úseku a dlouhodobé, které porovnávají roční či delší úseky. Toto dělení je významné při zkoumání sezónních vlivů či trendů. [5]

Časové řady stochastické a deterministické

Zde jsou řady děleny podle možnosti předpovědi jejich budoucího vývoje. Pokud je vývoj řady do budoucna předpověditelný, bez jakékoliv odchylky, mluvíme o deterministické časové řadě. Tyto řady neobsahují prvek náhody. Pro ekonomický obor jsou typičtější stochastické řady, jejichž vývoj je ovlivňován náhodou, a proto jejich předpověď není tak jednoznačná a přesná.

Časové řady sestupné a vzestupné

Z hlediska uspořádání dělíme řady na vzestupné, které se používají nejčastěji a dodržují přirozený tok času. Lze se ale také setkat s využitím obráceného uspořádání.

Časové řady absolutní a odvozené

Podle druhu sledovaných dat se časové řady dělí na absolutní a odvozené. Absolutní neboli primární ukazatel je zjišťován přímo. Příkladem je počet obslužených klientů za měsíc. Odvozené (sekundární) ukazatelé vznikají jako funkce primárních ukazatelů [7]. Příkladem je aktuální počet obslužených klientů od začátku roku.

1.2.1 Očištění časové řady o důsledky kalendářních vlivů

V případě, že chceme porovnávat hodnoty u intervalových krátkodobých časových řad, musí velikost ukazatele záviset na délce intervalu. Důvodem je, že měsíce nejsou stejně dlouhé a také nemají stejný počet pracovních dnů. Údaje je tedy nutné podle Hindlse [10] přepočítat na stejný časový interval a to buď metodou kalendářního očištění, nebo očištění na pracovní dny.

Očištění na kalendářní dny získáme podle vztahu:

$$y_t^{(0)} = y_t \frac{\bar{k}_t}{k_t}, \quad (1.2)$$

kde y_t je hodnota očišťovaného ukazatele, k_t je počet kalendářních dní v daném období a \bar{k}_t je průměrný počet kalendářních dní v dílčím období roku.

Očištění na pracovní dny získáme podle vztahu:

$$y_t^{(0)} = y_t \frac{\bar{p}_t}{p_t}, \quad (1.3)$$

kde y_t je hodnota očišťovaného ukazatele, p_t je počet pracovních dní v příslušném dílčím období roku a \bar{p}_t je průměrný počet pracovních dní v dílčím období roku.

1.2.1.1 Příklad

Měsíční výroba cementu v ČR během roku 1998 tvoří časovou řadu 566, 414, 757, 819, 847, 818, 837, 836, 837, 795, 699, 385 (v tisících tunách). Sestavte časovou řadu produkce pro standardní měsíc o délce 365/12 dnů.

Řešení:

Vypočítáme průměrnou délku měsíce:

$$\bar{k}_t = \frac{365}{12} = 30,42$$

Pro leden je tedy třeba provést výpočet podle vzorečku (1.2):

$$y_1^{(0)} = 566 \cdot \frac{30,42}{31} = 555,4$$

Pro únor provedeme výpočet obdobně:

$$y_2^{(0)} = 414 \cdot \frac{30,42}{28} = 450$$

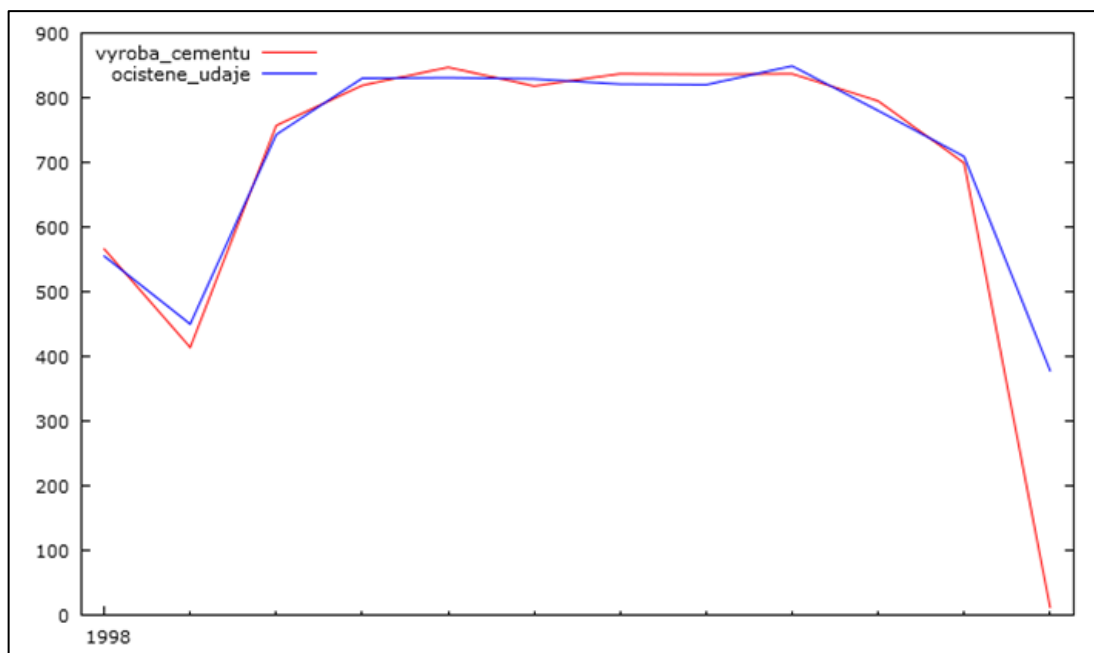
Pro další měsíce provedeme očištění podobně (Tabulka 1).

Měsíc	Původní údaje	Počet dní v měsíci	Očištěné údaje
Leden	566	31	555
Únor	414	28	450
Březen	757	31	743
Duben	819	30	830
Květen	847	31	831
Červen	818	30	829
Červenec	837	31	821
Srpen	836	31	820
Září	837	30	849
Říjen	795	31	780
Listopad	699	30	709
Prosinec	385	31	378

Tabulka 1- Tabulka výroby cementu roku 1998 a očištění časové řady

Závěr:

Z výše uvedených údajů je vidět, že po očištění časové řady připadá nejvyšší výkonnost výroby na září, ačkoliv před očištěním se zdálo, že nejvýkonnější byla výroba v květnu. Na obrázku 1.1 vidíme graf, kdy červená křivka nám udává původní údaje výroby cementu a modrá křivka nám ukazuje časovou řadu, kterou jsme očistili na kalendářní dny.



Obr. 1.1: Graf původních a očištěných hodnot výroby cementu v ČR

1.3 Některé specifické problémy časových řad

V této kapitole se zmíním o problémech, na které můžeme narazit při zpracování dat ve formě časové řady. Jak uvádí Cipra [6], mezi základní problémy patří:

Problémy s volbou časových bodů pozorování

Diskrétní časové řady (tj. řady, které jsou tvořeny určitými pozorováními v daných nespojitéch časových bodech) mohou vznikat třemi způsoby:

- *bud' jsou přímo diskrétní svou povahou* (například úroda obilí za určité roky),
- *akumulací* (neboli agregací), kde se sečtou hodnoty za dané časové období (například denní množství srážek),
- *diskretizací spojitě časové řady* (například teplota ve stanovenou denní dobu a na určitém místě).

Problémy s kalendářem

Tyto problémy jsem již zmínila v kapitole 1.2.1 Očištění časové řady o důsledky kalendářních vlivů.

Problémy s nesrovnalostí jednotlivých měření

Některé ekonomické údaje musíme upravovat pomocí cenových indexů. S technickým vývojem se zvyšuje např. technická vybavenost, a proto není možné srovnávat produkci například v letech 1982 a v roce 2010. Nesrovnalost některých měření také může souviset s tím, že některé podniky např. nedodaly do sestavovaného přehledu některé ukazatele, takže příslušná hodnota za jeden rok se týká např. 85 podniků a za další rok jen 82 podniků [6].

Problémy s délkou časových řad

Délka časové řady souvisí s počtem pozorování a je tedy samozřejmé, že ovlivňuje množství informace pro její analýzu. Nejedná se však o přímou úměru, takže například zdvojnásobení počtu měření, nemusí nutně znamenat zdvojnásobení informací, které jsou obsaženy v těchto měřeních.

Problémy s délkou časových řad jsou dva:

- *časová řada je příliš krátká* – některé metody vyžadují minimální délku (například Boxův-Jenkinsův přístup – doporučuje se používat na řady o délce větší než 50 pozorování),
- *časová řada je příliš dlouhá* – u takové řady hrozí nebezpečí, že se v průběhu času změní charakter modelu.

Délku časové řady je tedy nutné odhadnout s ohledem na určitou situaci a okolnosti k ní připadající.

1.4 Základní přístupy k analýze časových řad

Volba metody analýzy časových řad závisí podle Řezankové aj. [19] na několika faktorech, mezi které patří:

- *účel analýzy* – musíme vědět, o co nám jde, jestli pouze o tvorbu modelu, rozpoznání mechanismu generování hodnot časové řady či o vzájemné vztahy s jinými řadami,
- *typ časové řady* – existuje mnoha různých metod, ale ne každá metoda je vhodná pro všechny řady,
- *zkušenost statistika* – provádí analýzu časové řady a s tím související výpočetní technika a programové vybavení.

Dalšími faktory může být například dostupná databáze nebo softwarové či hardwarové vybavení.

Mezi základní metody pro analýzu časových řad podle Řezankové aj. [19] patří:

- Dekompozice časové řady
- Boxova-Jenkinsova metodologie
- Lineární dynamické modely
- Spektrální analýza časových řad

1.4.1 Dekompozice časové řady

Dekompoziční metodou lze řadu rozložit na součet několika složek, z nichž každá bude značně jednodušší a bude mít jasný výklad. Těmito systematickými složkami jsou: trend, sezónní, cyklická a reziduální (náhodná) složka.

S dekompozicí časové řady vás blíže seznámím v dalších kapitolách, kde vám popíšu jednotlivé složky a dále se zaměřím na nejběžnější trendové křivky, metodu klouzavých průměrů a na exponenciální vyrovnání.

1.4.2 Boxova-Jenkinsova metodologie

Tato metodologie bere v potaz při konstrukci modelu časové řady reziduální (náhodnou) složku, která může být tvořena korelovanými (závislými) náhodnými veličinami. Tato metodologie tedy může nejen zpracovávat časové řady s navzájem závislými pozorováními, ale dokonce těžiště jejich postupů spočívá právě ve vyšetřování těchto závislostí a v tzv. korelační analýze [6]. Boxovy-Jenkinsovy modely jsou zpravidla flexibilnější než modely dekompoziční, což znamená, že se lépe přizpůsobí změnám v průběhu časové řady. Základní zásady této metodologie vytvořili Box a Jenkins.

Jedním z nejjednodušších modelů, s nimiž se Boxova-Jenkinsova metodologie zabývá, je tzv. model klouzavých součtů prvního řádu, který se označuje jako

MA(1). Tento model je vhodný pro časovou řadu, kde tato pozorování jsou nekorelovaná kromě bezprostředně sousedních dvojic. Tento model má tvar typu

$$y_t = \varepsilon_t + k\varepsilon_{t-1}, \quad (1.4)$$

kde y_t je modelovaná řada, k je nějaká reálná konstanta a ε_t zahrnuje bílý šum. Tato metodologie má ve svém arzenálu i jiné typy modelů, kterými jsou tzv. autoregresní modely AR a smíšené modely ARMA. Tento autoregresní model AR 1. řádu je definovaný předpisem

$$y_t = \varepsilon_t + ky_{t-1}. \quad (1.5)$$

Může se zdát, že je zde přehnaná pozornost věnována náhodné složce a není zde možnost modelovat sezónní a trendové řady. Ale i tyto řady je Boxova-Jenkinsova metodologie schopna vyřešit a to na základě tzv. integrovaných modelů ARIMA a tzv. sezónních modelů, v nichž tyto složky mohou být modelovány stochasticky.

1.4.3 Lineární dynamické modely

Data, která se uplatňují v ekonometrii, mají ve většině případů tvar časových řad. Takové modely jsou zpravidla konstruovány tak, že se hodnoty určité časové řady zpravidla vysvětlují pomocí jiných hodnot (tzv. vysvětlujících nebo faktorových časových řad). Podle Cipry [6] může být takovým jednoduchým ekonometrickým modelem například model typu

$$c_t = \alpha + \beta c_{t-1} + \gamma x_t + \delta p_t + \varepsilon_t, \quad (1.6)$$

kde výdaje c_t obyvatelstva na nákup spotřebního zboží v roce t jsou vysvětlovány pomocí výdajů c_{t-1} v bezprostředně předcházejícím roce a navíc pomocí peněžních příjmů x_t obyvatelstva a cenového indexu p_t spotřebního zboží v roce t ($\alpha, \beta, \gamma, \delta$ jsou parametry a ε_t je tzv. bílý šum).

Modely tohoto typu se zpravidla označují jako příčinné (kauzální, faktorové) modely.

Pod názvem lineární dynamické modely budeme chápat Boxovy-Jenkinsovy modely, v nichž vystupují i další vysvětlující časové řady, nejen výše popisované řady a bílý šum. Za velice jednoduchý příklad podle Cipry [6], můžeme uvést model, který zachycuje chování měsíční spotřeby elektrické energie y_t v závislosti na průměrné měsíční teplotě x_t . Takový model má tvar

$$\log y_t = ax_t + u_t, \quad (1.7)$$

kde u_t je stanovený Boxův-Jenkinsov model, který už dále nezahrnuje žádné vysvětlující časové řady.

1.4.4 Spektrální analýza časových řad

Spektrální analýza má od předchozích tří případů odlišný přístup, který tkví v tom, že se považuje zkoumaná časová řada za směs sinusových a kosinusových křivek s různými amplitudami a frekvencemi. Tato analýza se označuje jako analýza časových řad ve spektrální doméně, někdy můžeme slyšet pojem tzv. fourierovská analýza. Pomocí speciálních statistických nástrojů (například periodogram nebo spektrální hustota) lze získat představu o intenzitě zastoupení jednotlivých frekvencí v časové řadě.

1.5 Předpovědi v časových řadách

Konstrukce předpovědí je jeden z důležitých úkolů analýzy časových řad. Tyto předpovědi mají důležitý význam pro národohospodářské plánování. Nyní zmíním některé obecné aspekty, se kterými jsou předpovědi v časových řadách spojeny.

Bodová předpověď a předpovědní interval

Bodová předpověď představuje odhad hodnoty časové řady v určitém budoucím okamžiku. Například bodová předpověď počtu cestujících přepravovaných jistou autobusovou společností je 158 000. Konkrétní číslo, které je zde poskytnuto, je nutno brát s rezervou, neboť bodová předpověď je vždy zatížena jistou chybou. Proto je pro uživatele prospěšnější disponovat s tzv. *předpovědním intervalem*, který je obdobou intervalu spolehlivosti z matematické statistiky. Například 95% interval spolehlivosti udává horní a dolní mez, mezi nimiž bude ležet příslušná sledovaná hodnota s pravděpodobností 0,95.

Kvalitativní a kvantitativní předpovědní modely

Kvalitativní modely (například metoda Delfi, která je založená na postupném dotazování a porovnávání odpovědí od určitých odborníků) jsou založeny na názoru specialistů, a proto mají subjektivní charakter. Oproti tomu *metody kvantitativních předpovědí* vycházejí z objektivního matematicko-statistického pohledu a předpokládá se, že se v budoucím čase charakter dosavadní řady nijak nezmění. Tuto skutečnost je proto nutné mít neustále na paměti.

Výběr předpovědní techniky závisí podle Cipry [6] na mnoha faktorech a to především na požadované formě předpovědi (zda bodová předpověď nebo předpovědní interval), časovém horizontu předpovědi, srozumitelnosti metody, charakteru a dostupnosti dat.

Zmíním se ještě o *chybách v předpovědi*. Chyba v předpovědi \hat{y}_t skutečné hodnoty y_t je stanovena jako

$$e_t = y_t - \hat{y}_t. \quad (1.8)$$

Při hodnocení kvality předpovědi je nutno uvážit všechny zkonstruované předpovědi. V praxi se nejčastěji využívají *míry kvality předpovědí*, které hodnotí vývoj předpovědí v čase. Nejvíce se podle Cipry [6] používá:

- *součet čtvercových chyb* SSE (Sum of Squared Errors) ve tvaru

$$\sum_{t=1}^n (y_t - \hat{y}_t)^2 = \sum_{t=1}^n e_t^2, \quad (1.9)$$

- *střední čtvercová chyba* MSE (Mean Squared Error) ve tvaru

$$\sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{n} = \sum_{t=1}^n \frac{e_t^2}{n}, \quad (1.10)$$

- *střední absolutní odchylka* MAD (Mean Absolute Deviation) ve tvaru

$$\sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{n} = \sum_{t=1}^n \frac{|e_t|}{n}. \quad (1.11)$$

Porovnáme-li všechny uvedené míry, zjistíme, že míry MSE a SSE na rozdíl od MAD posuzují mnohem striktněji větší chyby než ty malé.

1.6 Představení ekonometrického softwaru

Analýza časových řad se v současné době provádí výhradně na počítači pomocí vhodného softwaru, kterým je například Statistica, Statgraphics, Gretl atd.

V této bakalářské práci budu pracovat s programem Gretl. Název tohoto programu je zkratkou Gnu Regression, Econometrics and Time-series Library. Jde o softwarový balíček, který je určen pro ekonometrické analýzy. Autorem tohoto programu je Allin Cottrell z Wake Forest University. Gretl je volně dostupný a díky tomu je možné si ho zdarma stáhnout na internetových stránkách <http://gretl.sourceforge.net>. Na této internetové stránce je také možné najít podrobný manuál k celému softwaru.

2 Dekompozice časových řad

Dekompozice časových řad vychází z domněnky, že náhodný proces, který vytváří časovou řadu, je odkázaný pouze na čas. Dále také předpokládá, že časovou řadu je možné rozčlenit na několik složek. Tento rozklad se dělá proto, že je jednodušší určit chování jednotlivých složek, než chování celé řady najednou.

Časovou řadu lze tedy rozdělit na tzv. systematické složky, kam patří trend, sezónní a cyklická složka a na reziduální (náhodnou, zbytkovou, iregulární) složku. Časovou řadu můžeme vyjádřit součtem

$$y_t = Tr_t + Sz_t + C_t + E_t, \quad (1.12)$$

kde Tr_t je hodnota trendové složky, Sz_t je hodnota sezónní složky, C_t je hodnota cyklické složky a E_t je náhodná složka. Tomuto způsobu rozkladu časové řady se říká *aditivní rozklad*. Aditivní dekompozice se používá v případě, že variabilita hodnot časové řady je přibližně konstantní v čase [9]. Je také zřejmé, že časová řada nemusí obsahovat všechny tyto složky najednou.

Existuje ještě jeden rozklad časové řady, kterému se říká *multiplikativní rozklad*, který má tvar

$$y_t = Tr_t \cdot Sz_t \cdot C_t \cdot E_t. \quad (1.13)$$

Pro tento rozklad je typické, že trendová složka časové řady je ve stejných měrných jednotkách jako ta původní časová řada, ale ostatní tři složky (sezónní, cyklická a náhodná) jsou v relativním vyjádření. Tento způsob se uplatňuje v případě, že variabilita řady roste v čase, nebo se v čase mění.

2.1 Trend

Trend je nejpodstatnější složka, která má významný vliv na dlouhodobý vývoj časové řady (zachycuje tedy dlouhodobý růst či dlouhodobý pokles). Vzniká důsledkem působení stejnoměrných sil [13]. Trend se ve většině případů modeluje pomocí matematických funkcí v celé délce časové řady. Při charakteristice se tedy nejedná o krátkodobý pokles či růst, ale především o zachycení tendence pohybu časové řady.

2.2 Sezónní složka

Sezónní složka charakterizuje pravidelně se opakující změny v časové řadě, které se odehrávají během jednoho kalendářního roku a pravidelně se každý rok opakují. Nejčastěji se sezónnost sleduje u čtvrtletních a měsíčních časových řad. Dalo by se

tedy říci, že sezónnost je způsobena především střídáním ročních období, nebo kulturními zvyky (Vánoce, Velikonoce). Jak již z definice vyplývá, sezónní složka se nemůže vyskytovat u časových řad ročních, protože tato složka může rok od roku měnit svůj charakter.

2.3 Cyklická složka

Cyklická složka je jedna z nejméně problémových složek časové řady. Popisuje dlouhodobou fluktuaci kolem trendu, kde se střídá dlouhodobá fáze růstu s fází poklesu. Tato složka je tedy podobná sezónní složce s výjimkou délky cyklů. Její perioda se může pohybovat až v násobcích let, a proto u krátkodobých časových řad nemusí být cyklická složka vůbec rozpoznatelná.

2.4 Náhodná složka

Náhodná složka je nesystematická složka a je tvořena náhodnými výkyvy v časové řadě. Obsahuje nedefinovatelné jevy, jako jsou například chyby v měření. Pro náhodnou složku se podle Řezankové aj. [19] zavádějí následující požadavky:

1. $E(\varepsilon_i) = 0$ pro všechna $i = 1, 2, 3, \dots, n$
Střední hodnota náhodné složky se rovná nule. Tato podmínka znamená, že náhodná složka nepůsobí systematicky na hodnoty časové řady.
2. $D(\varepsilon_i) = \sigma^2$ pro každé $i = 1, 2, \dots, n$
Rozptyl je konstantní. To znamená, že variabilita náhodné složky nezávisí na systematických hodnotách složek a rovná se neznámé kladné hodnotě.
3. $Cov(\varepsilon_i, \varepsilon_j) = 0$ pro všechna $i \neq j = 1, 2, \dots, n$
Kovariance je nulová, tudíž hodnoty náhodné složky jsou nekorelované.
4. ε_i mají normální rozdělení pro všechna $i = 1, 2, 3, \dots, n$

Pokud jsou splněny první tři požadavky, mluvíme o náhodné veličině jako o *bílém šumu*. V případě, že je splněna i čtvrtá podmínka, mluvíme o tzv. *normálním bílém šumu*.

3 Modely popisující trendovou složku

Popis trendu se uskutečňuje pomocí regresní analýzy. Regresní analýza je metoda, která zkoumá vztah mezi dvěma a více proměnnými. V této analýze existuje určitá závislost mezi nezávisle proměnnou a závisle proměnnou.

Popis vývoje v časových řadách je jedním z nejdůležitějších úkolů. Z celé řady trendových funkcí, které jsou popsány v různých monografiích, se zaměřím pouze na sedm z nich, které se často používají v oblasti analýzy a prognózy časových řad. Jde tedy o konstantní trend, lineární trend, kvadratický trend, exponenciální trend, modifikovaný exponenciální trend, logistický trend a Gompertzovu křivku.

Pro analyzovanou časovou řadu se podle Cipry [6] předpokládá, že má tvar

$$y_t = Tr_t + E_t, \quad (3.1)$$

kde tento součet lze rozdělit na trendovou a reziduální složku.

Pro konstantní, lineární, kvadratický a exponenciální trend platí, že patří z hlediska jejich průběhu mezi funkce jednoduché. Pro ně je typické, že nemají asymptotu, proto jejich růst není ničím omezený. Další tři funkce (modifikovaný exponenciální trend, logistický trend a Gompertzova křivka) nemají už tak jednoduchý průběh. Na rozdíl od prvních čtyř funkcí mají asymptotu, takže jejich průběh je ohraničený buď shora, nebo zdola.

Nejpoužívanější a také nejjednodušší metodou odhadu parametrů trendových funkcí je podle Hindlse [10] *metoda nejmenších čtverců*. Tuto metodu lze využít v případě, že trendová funkce je lineární v parametrech. Její výhodou je, že minimalizuje rozptyl reziduální složky a je poměrně jednoduchá. Z výše uvedených funkcí ji můžeme aplikovat pouze na konstantní, lineární a kvadratický trend. V případě exponenciálního trendu můžeme metodu nejmenších čtverců použít až po provedení linearizující transformace (tzv. exponenciální funkci převedeme na lineární pomocí logaritmu). Tuto metodu nelze použít na modifikovaný exponenciální trend, logistický trend a Gompertzovu křivku, neboť tyto funkce jsou nelineární z hlediska parametrů a nemohou být transformovány na potřebný lineární tvar.

3.1 Konstantní trend

Jeden z nejjednodušších typů polynomiálního trendu je konstantní trend, který se také někdy označuje jako řada bez trendu. Základní vztah pro trendovou složku je podle Cipry [6] tvar

$$Tr_t = \beta_0, \quad t = 1, 2, \dots, n. \quad (3.2)$$

Dostáváme tak jednoduchý odhad b_0 parametru β_0 , který jsme získali po výpočtu normálních rovnic.

Předpis tedy je

$$b_0 = \bar{y} = \sum_{t=1}^n \frac{y_t}{n}. \quad (3.3)$$

Data sledovaného ukazatele u tohoto trendu nerostou ani neklesají, jenom kolísají kolem určité konstanty.

3.2 Lineární trend

Lineární trend je nejvíce používaným typem trendové funkce. Jeho smysl spočívá v tom, že ho můžeme použít kdykoliv, chceme-li přibližně určit směr vývoje analyzované časové řady.

V případě tohoto trendu dostáváme trendovou složku tvaru (viz Hindls [10])

$$Tr_t = \beta_0 + \beta_1 t, \quad (3.4)$$

kde β_0 a β_1 jsou neznámé parametry a $t = 1, 2, \dots, n$ je časová proměnná.

Pro odhady b_0 a b_1 parametrů β_0 a β_1 dostaneme soustavu dvou normálních rovnic:

$$\begin{aligned} b_0 n + b_1 \sum_{t=1}^n t &= \sum_{t=1}^n y_t, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 &= \sum_{t=1}^n t y_t. \end{aligned} \quad (3.5)$$

Řešením této soustavy (3.5) dostaneme pro odhady b_0 a b_1 vzorce

$$\begin{aligned} b_1 &= \frac{\sum t y_t - \bar{t} \sum y_t}{\sum t^2 - n \bar{t}^2}, \\ b_0 &= \bar{y} - b_1 \bar{t}, \end{aligned} \quad (3.6)$$

kde symbolem \sum se rozumí součet pro t od 1 do n .

\bar{y} a \bar{t} jsou výběrové průměry, které získáme pomocí vzorců

$$\bar{y} = \sum_{t=1}^n \frac{y_t}{n}, \quad \bar{t} = \sum_{t=1}^n \frac{t}{n} = \frac{n+1}{2}. \quad (3.7)$$

3.3 Kvadratický trend

Kvadratický trend lze podle Řezankové aj. [19] vyjádřit vztahem

$$Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad (3.8)$$

kde β_0, β_1 a β_2 jsou neznámé parametry a $t = 1, 2, \dots, n$ je časová proměnná.

Tento trend je také velmi často používaný. Z hlediska parametrů jde o lineární trendovou funkci, proto použijeme k odhadu parametrů metodu nejmenších čtverců. Budeme tedy řešit soustavu tří normálních rovnic

$$\begin{aligned} b_0 n + b_1 \sum t + b_2 \sum t^2 &= \sum y_t, \\ b_0 \sum t + b_1 \sum t^2 + b_2 \sum t^3 &= \sum t y_t, \\ b_0 \sum t^2 + b_1 \sum t^3 + b_2 \sum t^4 &= \sum t^2 y_t. \end{aligned} \quad (3.9)$$

Výhodněji se pracuje s vyjádřením trendu (3.8) ve tvaru (viz Cipra [6])

$$Tr_t = \gamma_0 + \gamma_1(t - \bar{t}) + \gamma_2(t - \bar{t})^2, \quad t = 1, 2, \dots, n \quad (3.10)$$

neboť v (3.8) pak platí

$$\sum(t - \bar{t}) = \sum(t - \bar{t})^3 = 0. \quad (3.11)$$

3.4 Exponenciální trend

Podle Hindlse [10] se jedná o dvouparametrický trend tvaru

$$Tr_t = \beta_0 \beta_1^t, \quad t = 1, 2, \dots, n, \quad (\beta_1 > 0) \quad (3.12)$$

kde β_0 a β_1 jsou neznámé parametry. Tento trend se vyznačuje tím, že jeho tzv. *koeficient růstu* a podíly dvou sousedních diferencí

$$\frac{Tr_{t+2} - Tr_{t+1}}{Tr_{t+1} - Tr_t}, \quad (3.13)$$

mají konstantní hodnotu β_1 . Pokud je $\beta_1 > 1$ dochází k růstu, zatímco pro $0 < \beta_1 < 1$ funkce zřejmě klesá.

Exponenciální trend převedeme na obecný zápis lineárního trendu tak, že provedeme zlogaritmování a tím se tento trend převede na lineární

$$\log Tr_t = \log \beta_0 + t \log \beta_1, \quad (3.14)$$

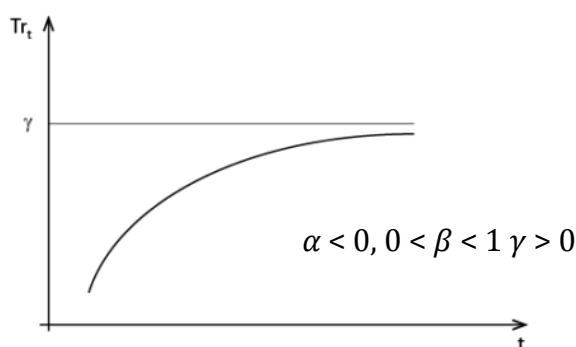
poté odhadneme parametry $\log \beta_0$ a $\log \beta_1$ a zpětně je odlogaritmuje a tím získáme odhady parametrů β_0 a β_1 .

3.5 Modifikovaný exponenciální trend

V případě, že je regresní přímka zdola či shora ohraničena, je dobré použít modifikovaný exponenciální trend, který má podle Cipry [6] podobu

$$Tr_t = \gamma + \alpha \beta^t, \quad t = 1, 2, \dots, n, \beta > 0 \quad (3.15)$$

a jeho graf je znázorněn na obr. 3.1.



Obr. 3.1: Modifikovaný exponenciální trend

Jedná se o nelineární funkci, kterou nelze linearizovat žádnou transformací, proto nelze použít metodu nejmenších čtverců. Je tedy nutné použít pro odhad parametrů funkce jinou metodu, tzv. *metodu částečných součtů*. Rozdělíme celek pozorování na třetiny o délce m a sečteme tato pozorování (sčítance tvoří geometrickou řadu, proto je tak snadné určit jednotlivé částečné součty), takže dostaneme

$$\begin{aligned} S_1 &\sim \sum_{t=1}^m y_t = m\gamma + \frac{\alpha\beta(\beta^m - 1)}{\beta - 1}, \\ S_2 &\sim \sum_{t=m+1}^{2m} y_t = m\gamma + \frac{\alpha\beta^{m+1}(\beta^m - 1)}{\beta - 1}, \\ S_3 &\sim \sum_{t=2m+1}^{3m} y_t = m\gamma + \frac{\alpha\beta^{2m+1}(\beta^m - 1)}{\beta - 1}. \end{aligned} \quad (3.16)$$

Řešením této soustavy dostaneme odhady b , a , c koeficientů β , α , γ ve tvaru

$$b = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{1/m},$$

$$a = \frac{b - 1}{b(b^m - 1)^2} (S_2 - S_1),$$

$$c = \frac{S_1 - ab(b^m - 1)/(b - 1)}{m}.$$

(3.17)

Jinou možností je, že při pevně zvoleném parametru β se model (3.15) stává lineárním. Vypočteme odhady parametrů α a γ pro různé hodnoty β a zvolíme variantu minimalizující MSE (viz vzorec 1.10).

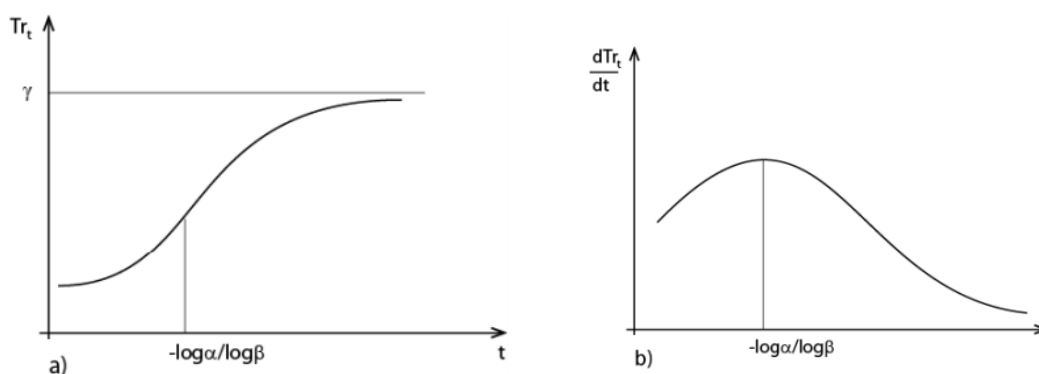
3.6 Logistický trend

Podle Cipry [6] je dán předpisem

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

(3.18)

a jeho graf je znázorněn na obr. 3.2 a). Má inflexní bod ($t = \log \alpha / \log \beta$), je ohraničen shora i zdola a jeho horní asymptota je kladná. Křivka má tvar „S“ a patří mezi tzv. *S-křivky*.



Obr. 3.2: Logistický trend a jeho derivace

Derivací podle proměnné t dostaneme

$$\frac{d Tr_t}{dt} = -\frac{\log \beta}{\gamma} Tr_t (\gamma - Tr_t),$$

(3.19)

což je důležitý ukazatel růstu trendové křivky (někdy se také nazývá jako růstová funkce). Derivace (3.19) je také symetrická kolem inflexního bodu a její graf je zachycen na obr. 3.2 b).

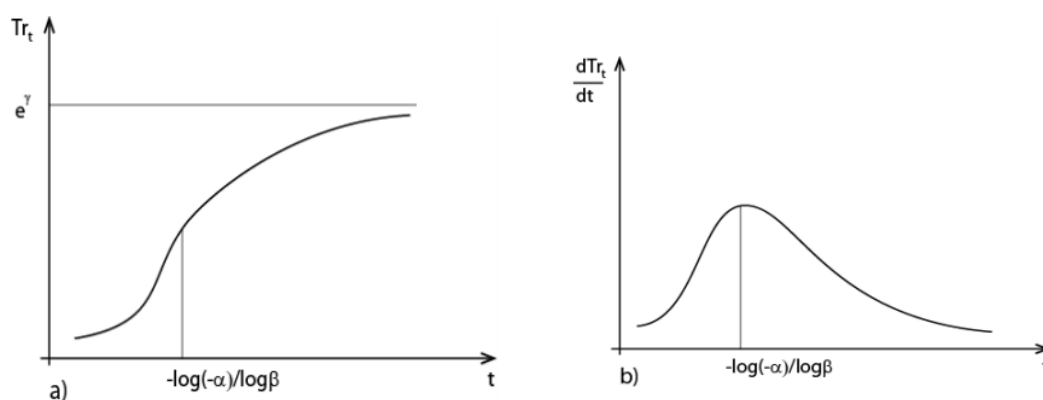
Odhady parametrů logistického trendu můžeme získat například tak, že převrátíme „logistický trend“, čímž získáme modifikovaný exponenciální trend s parametry $\frac{1}{\gamma}, \frac{\alpha}{\gamma}$ a β ty odhadneme a zpětně dopočítáme odhady parametrů α, β a γ . Dalším způsobem je princip tzv. *diferenčních odhadů parametrů*, kdy se pracuje s řadou tzv. *prvních diferencí* $y_{t+1} - y_t$ místo s původní řadou y_t .

3.7 Gompertzova křivka

Patří do skupiny S-křivek a vzniká stejně jako logistický trend transformací modifikovaného exponenciálního trendu. Gompertzova křivka má podle Cipry [6] tvar

$$\log T r_t = \gamma + \alpha \beta^t, \quad t = 1, \dots, n \quad (\beta > 0, \alpha < -1). \quad (3.20)$$

Z obrázku 3.3 a) je vidět, že má inflexi v bodě $t = -\log(-\alpha)/\log \beta$ a je asymptoticky omezena.



Obr. 3.3: Gompertzova křivka, růstová funkce

První derivace, která je zachycená na obr. 3.3 b) není symetrická kolem inflexního bodu, proto Gompertzovu křivku řadíme mezi *S-křivky nesymetrické kolem inflexního bodu*.

Odhady parametrů této křivky se provádí obdobně jako u modifikovaného exponenciálního trendu.

3.8 Příklad

Ukážeme si, jak v programu Gretl modelovat řadu měsíčních maloobchodních tržeb určitého druhu podnikání za období leden 1992 – srpen 2015. Předpokládáme, že se jedná o mnohonásobnou regresi s pomocnými proměnnými dm1 až dm12, které indikují příslušné měsíce.

Vícenásobný regresní model je rozšířením jednoduchého regresního modelu, který má tvar

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \quad (3.21)$$

kde y_t je tzv. závislá proměnná, kterou odhadujeme pomocí parametrů β_1 a β_2 a ε_t je náhodná veličina o které se předpokládá, že má normální rozdělení s nulovou střední hodnotou.

Vícenásobný regresní model spočívá v tom, že budeme pracovat s více než jednou proměnou. Obecný tvar tohoto modelu můžeme psát následovně

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_K x_{iK} + \varepsilon_t, \quad i = 1, 2, \dots, N, \quad (3.22)$$

kde β_0 je tzv. absolutní člen, který se někdy označuje jako úrovněová konstanta, $\beta_1, \beta_2, \dots, \beta_K$ jsou regresní koeficienty, které chceme odhadnout, index i značí jednotlivá pozorování a ε_t je náhodná veličina.

3.8.1 Popis dat

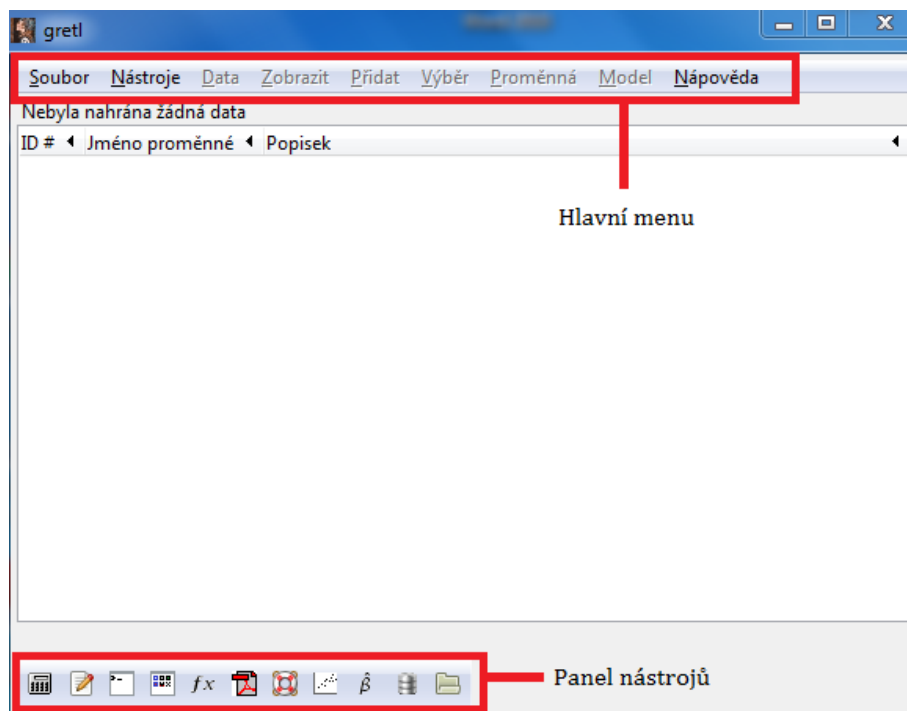
Nejprve si tedy připravíme datový soubor, se kterým budeme pracovat. V našem případě to budou data z internetové stránky [21] pod názvem Food and Beverage Stores (viz příloha č. 1). Data představují měsíční maloobchodní tržby obchodů s jídlem a nápoji od ledna roku 1992 do srpna roku 2015. Údaje v časové řadě jsou v milionech dolarů.

3.8.2 Úprava dat

Data je zapotřebí upravit tak, že hodnoty dáme do jednoho sloupce postupně měsíc za měsícem. Pro naše účely použijeme úpravu dat v poznámkovém bloku, lze ale také použít například Excel, Octave, či jiný typ formátu.

3.8.3 Uživatelské rozhraní

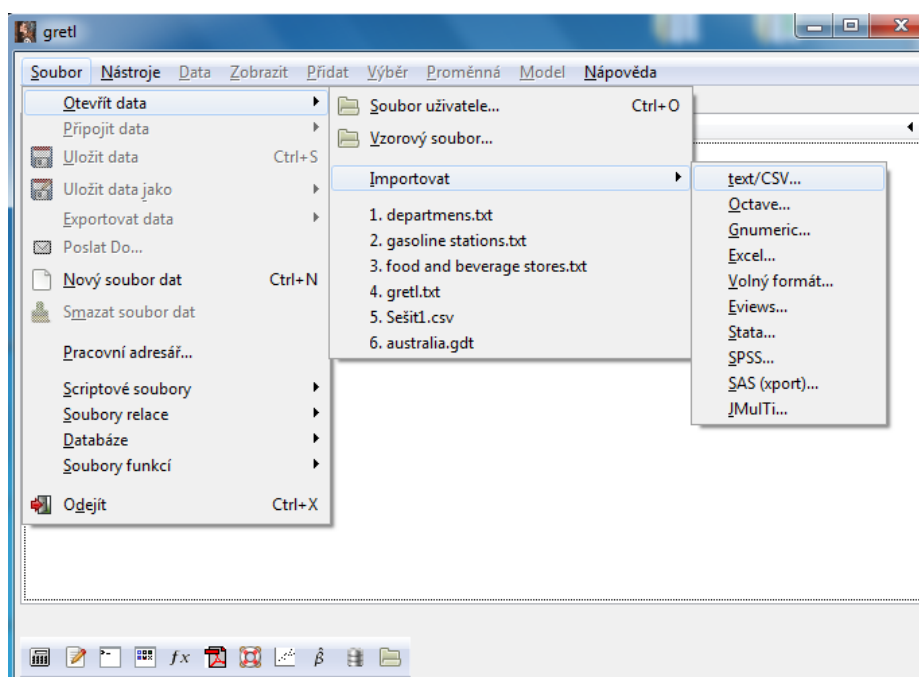
Po spuštění programu se objeví hlavní okno (obr. 3.4), kde v horní části je umístěné *hlavní menu* a ve spodní části se nachází *panel nástrojů*. Hlavní menu obsahuje dvě hlavní nabídky „Soubor“ a „Nástroje“. Ostatní se zpřístupní, až v průběhu dalších kroků. Panel nástrojů obsahuje prakticky odkazy na nejpoužívanější funkce hlavního menu, kalkulačku, návod na používání, seznam příkazů pro práci v konzole atd.



Obr. 3.4: Hlavní okno programu Gretl

3.8.4 Import dat

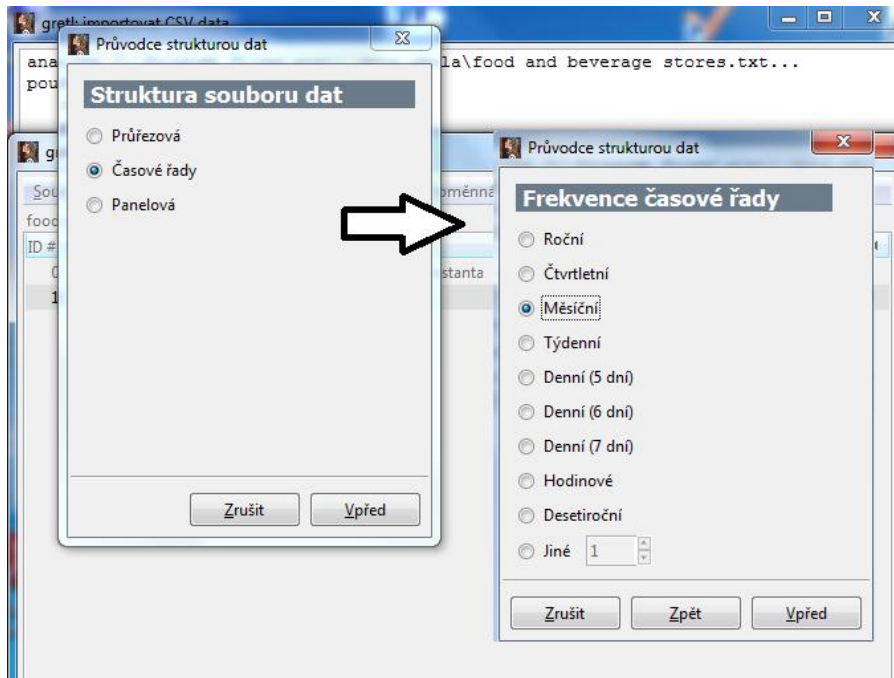
Gretl po nainstalování základního balíčku disponuje s velkým množstvím vzorových údajů. Otevřít se dají z hlavního menu postupem kroků „Soubor“ -> „Otevřít data“ -> „Vzorový soubor“ (obr. 3.5), nebo přímo z panelu nástrojů. Jelikož máme vlastní data, budeme postupovat následovně. Podle obrázku 3.5 klikneme na ikonu „Soubor“ -> „Otevřít data“ -> „Importovat“ -> „text/CSV“.



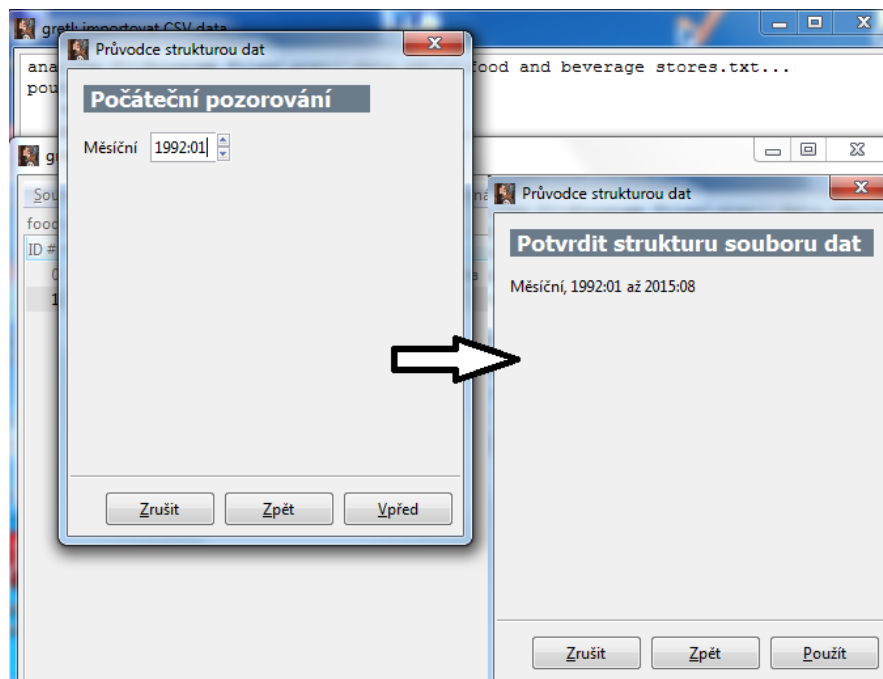
Obr. 3.5: Import dat

3.8.5 Struktura dat

Poté co jsme klikli na „Importovat“, vybereme data, která máme uložena ve svém počítači. Objeví se nám okno nabízející volbu, jestli si přejeme interpretovat data jako časovou řadu nebo panel. Klikneme na tlačítko „ano“. Strukturu souboru dat zvolíme časové řady s měsíční frekvencí (obr. 3.6). Počáteční pozorování v našem případě bude 1992 a poté potvrdíme strukturu souboru dat, kde vidíme, že data jsou opravdu od ledna 1992 do srpna 2015 (viz obr. 3.7).



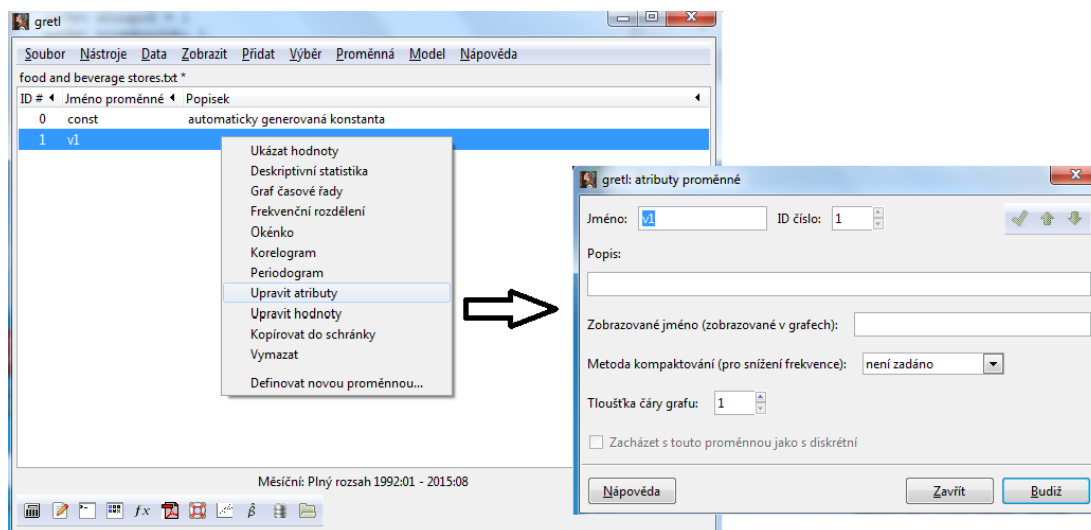
Obr. 3.6: Průvodce strukturou dat 1



Obr. 3.7: Průvodce strukturou dat 2

3.8.6 Úprava atributů

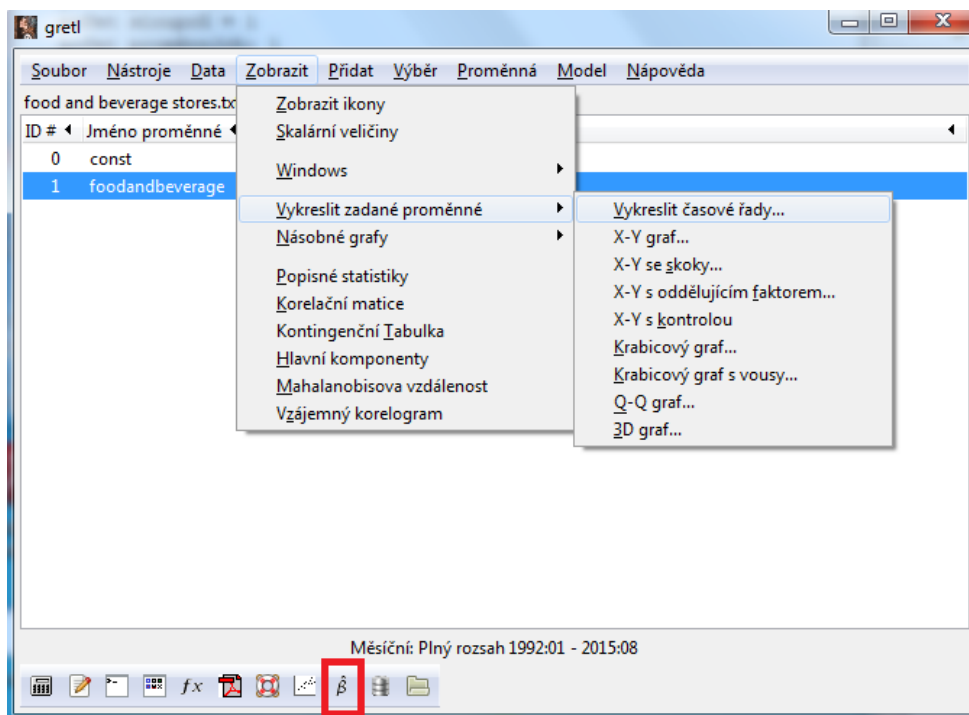
Nyní jsme data nahráli do programu Gretl pod názvem *v1*. Je možné si tento název přepsat a to tak, že pravým tlačítkem myši klikneme na „*Upravit atributy*“ a do pole jméno si vložíme svůj název souboru dat (v našem případě „*foodandbeverage*“) a potvrdíme (obr. 3.8).



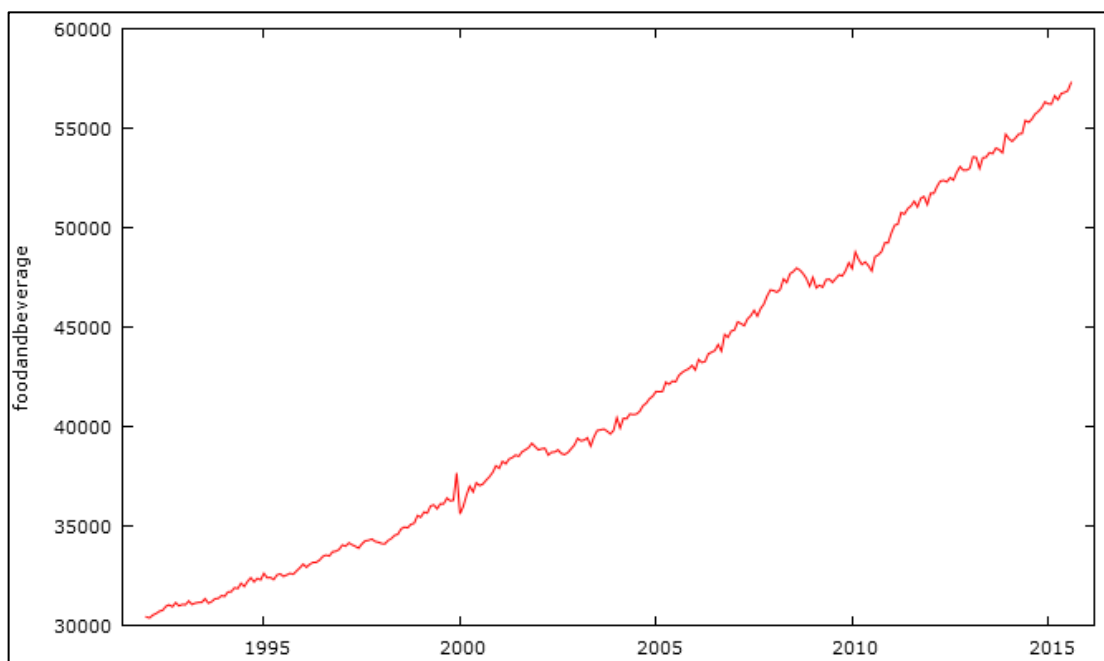
Obr. 3.8: Úprava atributů

3.8.7 Sestrojení grafu

Poté klikneme v hlavním menu na „*Zobrazit*“ -> „*Vykreslit zadané proměnné*“ -> „*Vykreslit časové řady*“ (viz obr. 3.9) a tím se nám vykreslí graf, který vidíme na obrázku 3.10. Ze spojnicového grafu je vidět, že funkce bude rostoucí. Uděláme tedy odhad parametrů regresního modelu, který se provádí metodou nejmenších čtverců.



Obr. 3.9: Postup vykreslení grafu

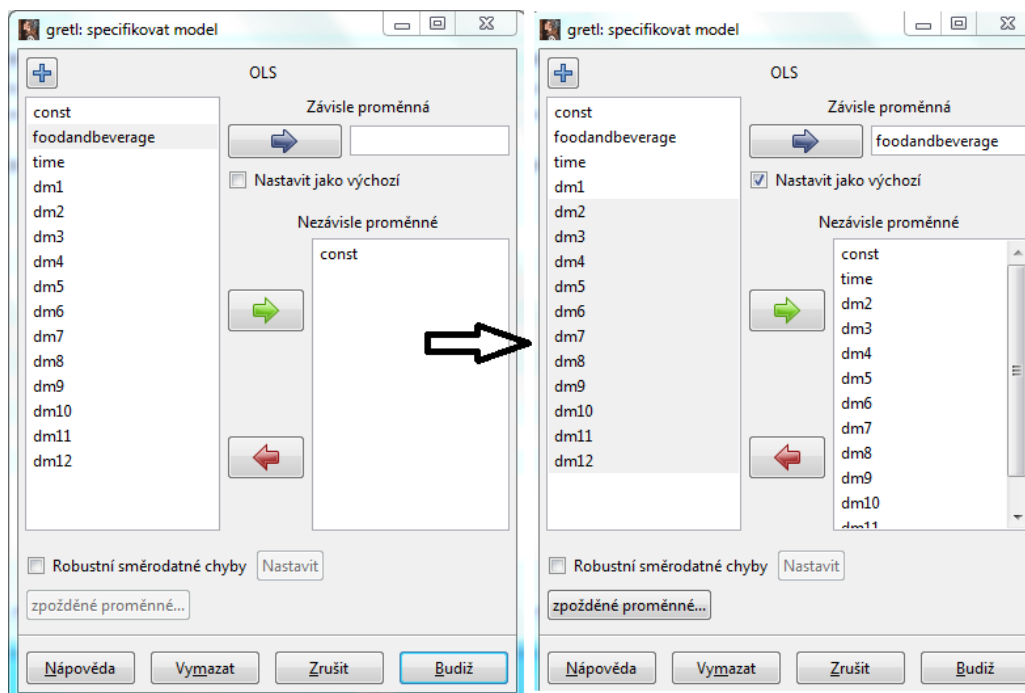


Obr. 3.10: Časová řada „food and beverage stores“ – vývoj dat, leden 1992 – srpen 2015

3.8.8 Odhad parametrů

Abychom mohli udělat odhad modelu s více proměnnými, musíme nejdříve tyto proměnné přidat a to tak, že v hlavním menu klikneme na „Přidat“ -> „Periodické indikátorové proměnné“ a také přidáme trendovou proměnnou tak, že klikneme na „Přidat“ -> „Časový trend“.

Odhad lineárního regresního modelu se v systému Gretl provádí tak, že klikneme na ikonu (viz obr. 3.9, červený rámeček), kde musíme navolit proměnné (konstanta je automaticky předvolená). Závislou proměnnou budou data „food and beverage“ a nezávislé proměnné budou časový trend a pomocné proměnné dm2 až dm12 (obr. 3.11) a následně potvrdíme. Objeví se nám okno s výsledkem regrese (obr. 3.12).



Obr. 3.11: Nastavení proměnných k modelu

Model 1: OLS, za použití pozorování 1992:01-2015:08 (T = 284)
 Závisle proměnná: foodandbeverage

	koeficient	směr. chyba	t-podíl	p-hodnota
const	28078,8	253,888	110,6	1,64e-227 ***
time	95,1747	0,807690	117,8	8,00e-235 ***
dm2	-1,75805	322,041	-0,005459	0,9956
dm3	-10,5578	322,044	-0,03278	0,9739
dm4	-20,6908	322,050	-0,06425	0,9488
dm5	-33,4905	322,057	-0,1040	0,9173
dm6	-7,79024	322,066	-0,02419	0,9807
dm7	-26,1316	322,077	-0,08113	0,9354
dm8	-2,51467	322,090	-0,007807	0,9938
dm9	-104,172	325,526	-0,3200	0,7492
dm10	-82,4336	325,531	-0,2532	0,8003
dm11	-105,217	325,538	-0,3232	0,7468
dm12	-11,6960	325,547	-0,03593	0,9714
Střední hodnota závisle proměnné			41607,92	
Sm. odchylka závisle proměnné			7892,511	
Součet čtverců reziduí			3,37e+08	
Sm. chyba regrese			1115,581	
Koeficient determinace			0,980868	
Adjustovaný koeficient determinace			0,980021	
F(12, 271)			1157,828	
P-hodnota(F)			5,8e-225	
Logaritmus věrohodnosti			-2389,190	
Akaikovo kritérium			4804,380	
Schwarzovo kritérium			4851,817	
Hannan-Quinnovo kritérium			4823,399	
rho (koeficient autokorelace)			0,967063	
Durbin-Watsonova statistika			0,063986	

zde je poznámka o zkratkách statistik modelu

Pomine-li se konstanta, p-hodnota byla nejvyšší pro proměnnou 4 (dm2)

Obr. 3.12: Okno s výsledkem regrese

Jak vidíme na obrázku 3.12, je zde použito 284 pozorování a to od ledna 1992 do srpna 2015. Dále můžeme v prvním sloupci vidět odhady koeficientů β_0, \dots, β_k , ve druhém směrodatnou chybu těchto odhadů, ve třetím realizaci testového kritéria statistiky a ve čtvrtém tzv. *p-hodnotu*. P-hodnota nám vyjadřuje pravděpodobnost, že výsledek testovacího kritéria za platnosti nulové hypotézy padne do kritického oboru. Je-li p-hodnota menší než předem stanovená hladina významnosti α (zpravidla volíme 5%, tj. $\alpha=0,05$), tak nulovou hypotézu zamítáme, v opačném případě nulovou hypotézu na hladině významnosti α nezamítáme. Počet hvězdiček nám znázorňuje, pro jakou nejnižší hladinu významnosti je nulová hypotéza ještě zamítnuta (***) označují 1%, ** označují 5% a * označuje 10%). Dále tu je výčet dalších vypočtených statistických údajů. Za zmínku stojí koeficient determinace, který udává, jak velký podíl celkové variability závislé proměnné se nám podařilo regresí vysvětlit.

Zaměříme se na p-hodnotu, ze které lze vyčíst, že proměnné dm2 až dm12 mají poměrně vysokou p-hodnotu, tudíž by v modelu proměnné nemusely ani figurovat a mohlo by se jednat o jednoduchý regresní model.

3.8.9 Hledání vhodného modelu

Abychom našli vhodný model, je nutné najít nejlépe vyhovující trendovou funkci a proložit ji časovou řadou. Budeme se snažit na data aplikovat lineární trend, kvadratický trend a exponenciální trend. Proměnné dm1 až dm12 uvažovat nebudeme, neboť se v modelu mnohonásobné regrese nejevily jako statisticky významné.

3.8.9.1 Lineární trend

Jako první uděláme odhad parametrů pro *lineární trend* pomocí metody nejmenších čtverců tak, že v panelu nástrojů klikneme na třetí ikonu zprava (viz obr. 3.9, červený rámeček) nebo alternativně vybereme z hlavní nabídky „Model“ -> „Metoda nejmenších čtverců“. Následně se nám otevře okno (podobně jako na obrázku 3.11), ve kterém se model specifikuje. Navolíme proměnné, kde za závislou proměnnou budeme považovat naše původní data („food and beverage“) a nezávislou proměnnou bude časový trend („time“). Po odkliknutí „Budiž“ se nám otevře okno (obr. 3.13). Výsledkem je tabulka s hodnotami.

	koeficient	směr. chyba	t-podíl	p-hodnota
const	28045,4	130,209	215,4	0,0000 ***
time	95,1753	0,792025	120,2	3,03e-244 ***

Střední hodnota závisle proměnné 41607,92
Sm. odchylka závisle proměnné 7892,511
Součet čtverců reziduí 3,38e+08
Sm. chyba regrese 1094,266
Koeficient determinace 0,980845
Adjustovaný koeficient determinace 0,980777
F(1, 282) 14440,13
P-hodnota (F) 3,0e-244
Logaritmus věrohodnosti -2389,361
Akaikovo kritérium 4782,723
Schwarzovo kritérium 4790,021
Hannan-Quinnovo kritérium 4785,649
rho (koeficient autokorelace) 0,967047
Durbin-Watsonova statistika 0,063945
zde je poznámka o zkratkách statistik modelu

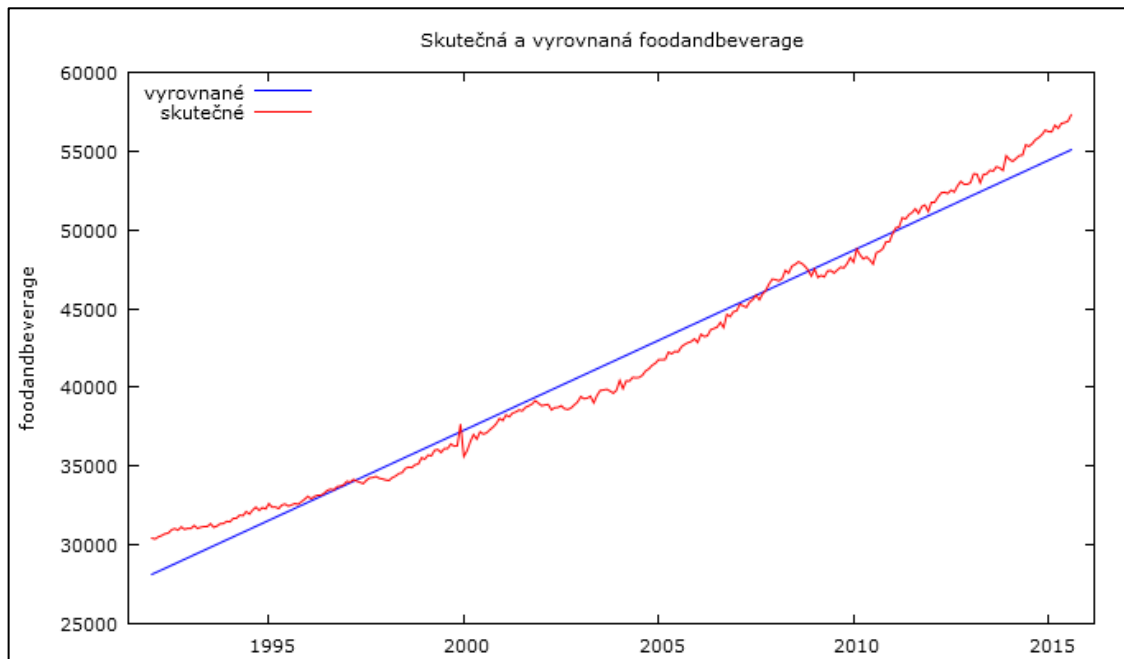
Obr. 3.13: Odhad parametrů – lineární trend

V prvním sloupečku jsou uvedeny odhadnuté koeficienty β_0 a β_1 modelu a tím dostáváme odhad pro lineární trend tvaru

$$Tr_t = 28\,045,4 + 95,1753t.$$

Následně je možné provádět testy či vykreslovat grafy přímo z menu v hlavním okně. Klikneme na „Grafy“ -> „Graf vyrovnaných a skutečných hodnot“ ->

„V závislosti na čase“ a dostaneme následující graf (obr. 3.14), kde jsme proložili časovou řadu přímkou. Červená křivka nám tedy ukazuje skutečné hodnoty a modrá přímkou nám udává vyrovnané hodnoty.



Obr. 3.14: Graf - aplikace lineárního trendu

3.8.9.2 Kvadratický trend

Druhou možností je, že zkusíme data modelovat pomocí *kvadratického trendu*. Najedeme na hlavní okno programu Gretl (obr. 3.4) a v hlavním menu klikneme na „Přidat“ -> „Druhé mocniny vybraných proměnných“, které použijeme v následujícím odhadu parametrů. Odhad parametrů uděláme úplně stejně jako v předchozí situaci, akorát do nezávisle proměnných přidáme druhé mocniny (*sq_time*).

	koeficient	směr. chyba	t-podíl	p-hodnota	
const	30289,0	81,2615	372,7	0,0000	***
time	48,1068	1,31666	36,54	9,50e-109	***
sq_time	0,165153	0,00447415	36,91	8,78e-110	***
Střední hodnota závisle proměnné			41607,92		
Sm. odchylka závisle proměnné			7892,511		
Součet čtverců reziduí			57732524		
Sm. chyba regrese			453,2701		
Koeficient determinace			0,996725		
Adjustovaný koeficient determinace			0,996702		
F(2, 281)			42761,01		
P-hodnota (F)			0,000000		
Logaritmus věrohodnosti			-2138,553		
Akaikovo kritérium			4283,106		
Schwarzovo kritérium			4294,053		
Hannan-Quinnovo kritérium			4287,495		
rho (koeficient autokorelace)			0,814799		
Durbin-Watsonova statistika			0,370236		

zde je poznámka o zkratkách statistik modelu

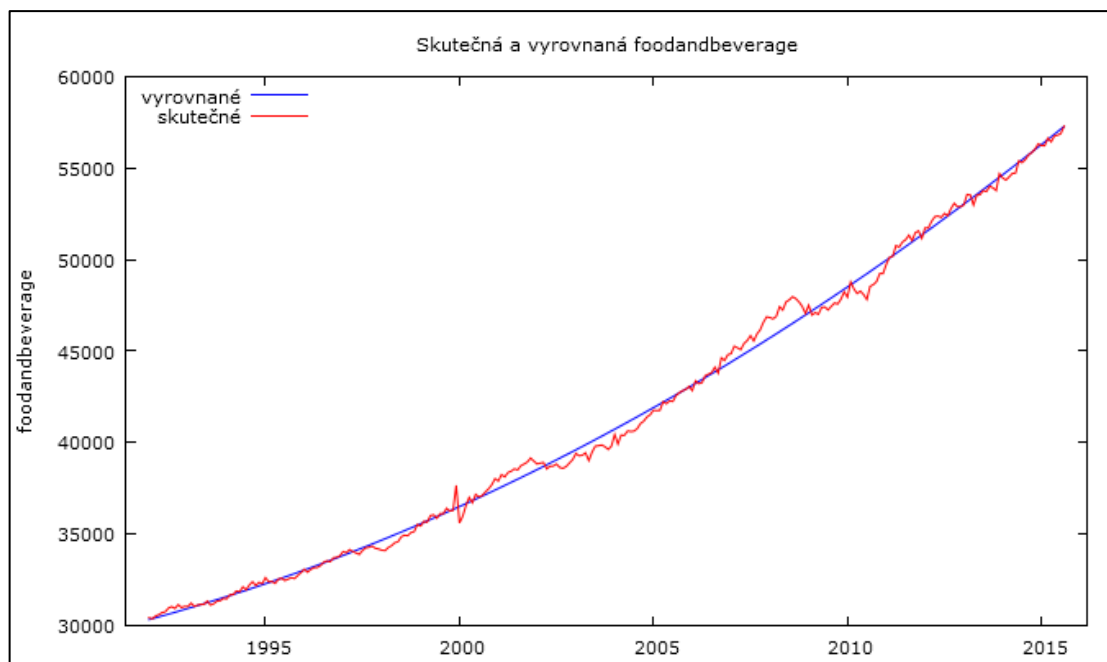
Obr. 3.15: Odhad parametrů - kvadratický trend

Tím jsme dostali odhad parametrů pro kvadratický trend (obr. 3.15) tvaru

$$Tr_t = 30\,289 + 48,1068t + 0,165153t^2.$$

Nyní můžeme porovnat koeficient determinace z obr. 3.13 a obr. 3.15. Je vidět, že koeficient determinace na obr. 3.13 je 0,9808 a na obr. 3.15 je 0,9967. Tudíž s přidáním druhých mocnin trendové proměnné tento koeficient vzrostl. Tento druhý model je tedy přesnější, než model předchozí.

Přejdeme tedy k vykreslení grafu kliknutím na „Grafy“ -> „Graf vyrovnaných a skutečných hodnot“ -> „V závislosti na čase“. V tomto grafu (obr. 3.16), jsme se snažili data proložit parabolou. Z grafu opět vidíme, že červená křivka nám udává skutečná data a modrá křivka nám udává vyrovnané hodnoty pomocí kvadratického trendu.



Obr. 3.16: Graf – aplikace kvadratického trendu

3.8.9.3 Exponenciální trend

Třetí možností je, že na data zkusíme aplikovat *exponenciální trend*. Nejprve najedeme opět na hlavní okno programu a v liště hlavního menu klikneme na „Přidat“ -> „Logaritmy vybraných proměnných“. Odhad parametrů pomocí metody nejmenších čtverců uděláme úplně stejným způsobem jako v předchozích dvou situacích, akorát závislou proměnnou bude „ $l_{\text{foodandbeverage}}$ “ a nezávislou proměnnou bude časový trend („ $time$ “). Tímto způsobem jsme provedli zlogaritmování a trend jsme převedli na obecný zápis lineárního trendu.

	koeficient	směr. chyba	t-podíl	p-hodnota	
const	10,2925	0,00170678	6030	0,0000	***
time	0,00228645	1,03819e-05	220,2	0,0000	***

Střední hodnota závisle proměnné	10,61831
Sm. odchylka závisle proměnné	0,188326
Součet čtverců reziduí	0,058019
Sm. chyba regrese	0,014344
Koeficient determinace	0,994220
Adjustovaný koeficient determinace	0,994199
F(1, 282)	48503,39
P-hodnota(F)	0,000000
Logaritmus věrohodnosti	803,4483
Akaikovo kritérium	-1602,897
Schwarzovo kritérium	-1595,599
Hannan-Quinnovo kritérium	-1599,971
rho (koeficient autokorelace)	0,885078
Durbin-Watsonova statistika	0,218369

zde je poznámka o zkratkách statistik modelu

Obr. 3.17: Odhad parametrů – exponenciální trend

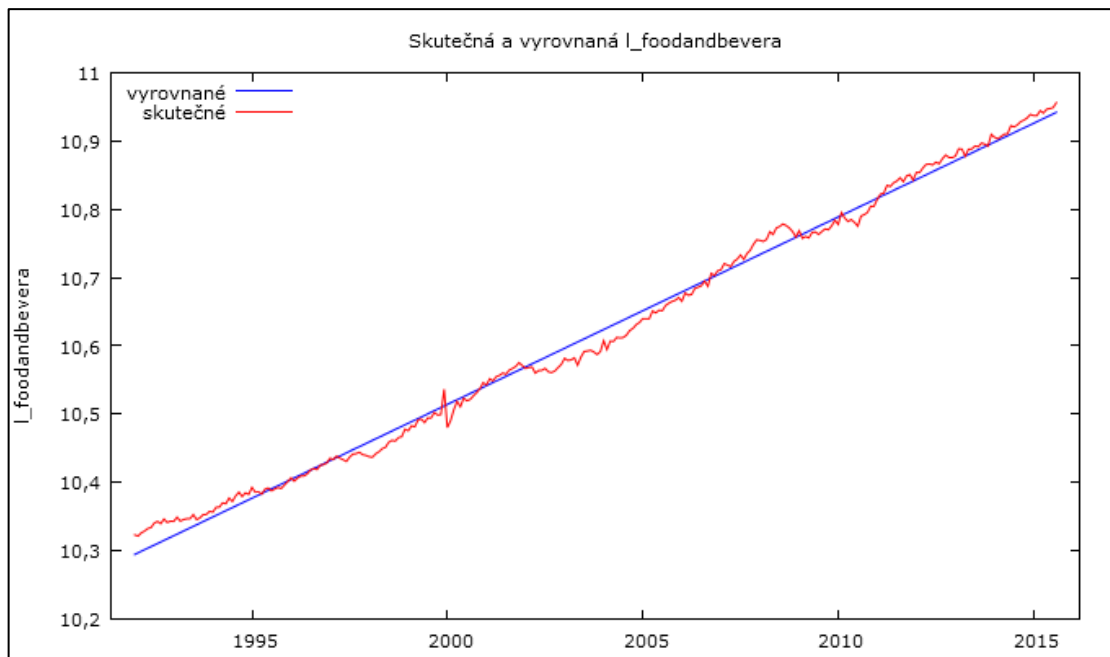
Na obrázku 3.17 vidíme odhad parametrů, který má tvar

$$\log Tr_t = \log 10,2925 + t \log 0,00228645,$$

$$Tr_t = 10,2925 \cdot 0,00228645^t \quad (\text{původní tvar exponenciálního modelu}).$$

Dále vidíme, že koeficient determinace, je o něco menší než u kvadratického trendu.

Přejdeme tedy k vykreslení grafu kliknutím na „Grafy“ -> „Graf vyrovnaných a skutečných hodnot“ -> „V závislosti na čase“. V tomto grafu (obr. 3.18) jsme zobrazili logaritmovanou časovou řadu spolu s příslušným lineárním trendem.



Obr. 3.18: Graf – logaritmovaná časová řada

Následně na hodnoty $\log Tr_t$ aplikujeme exponenciální funkci o základu e . Vrátime se zpátky k prvnímu odhadu parametrů pomocí lineárního trendu (obr. 3.13) a v horní části klikneme na „Analýza“ -> „Předpovědi“. Objeví se nám okno s upozorněním, že nejsou k dispozici žádná pozorování, klikneme na tlačítko „Zavřít“ a objeví se nám tabulka (obr. 3.19) pro zadání kritérií. Vše necháme jak je navolené a dáme „Budiž“.

Obr. 3.19: Definování předpovědi

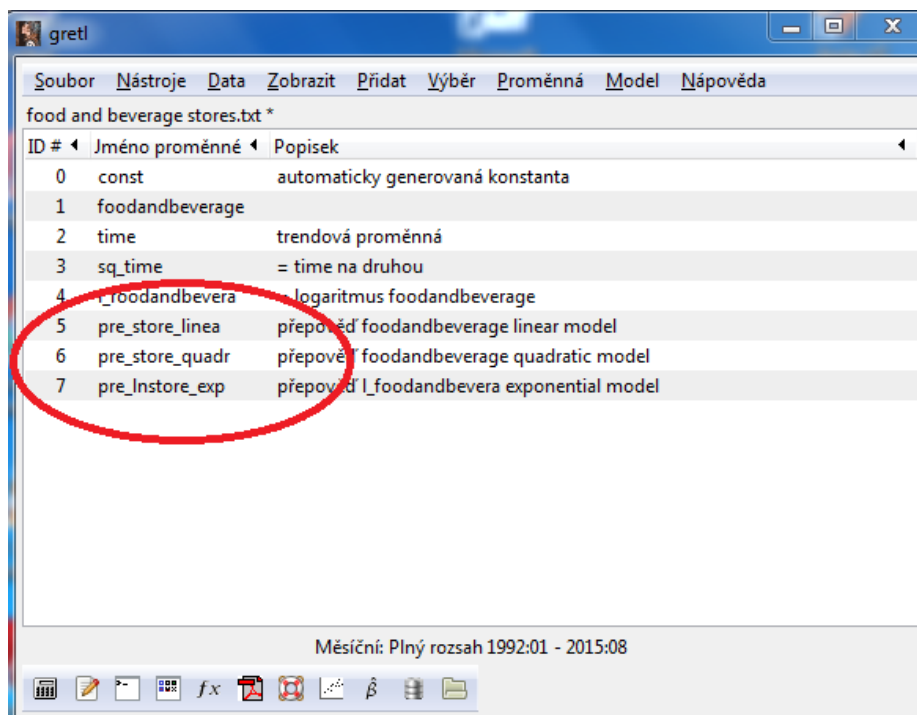
Naskočí nám okno (obr. 3.20), kde první sloupec znázorňuje pozorování, druhý data pozorování, třetí předpověď, čtvrtý směrodatnou chybu a pátý 95% konfidenční interval.

	Pozorování foodandbeverage	předpověď	směr. chyba	95% konfidenční interval
1992:01	30437,00	28140,60	1101,905	25971,60 - 30309,61
1992:02	30382,00	28235,78	1101,825	26066,93 - 30404,63
1992:03	30526,00	28330,95	1101,745	26162,27 - 30499,64
1992:04	30602,00	28426,13	1101,666	26257,60 - 30594,66
1992:05	30718,00	28521,31	1101,588	26352,93 - 30689,68
1992:06	30761,00	28616,48	1101,510	26448,26 - 30784,71
1992:07	30963,00	28711,66	1101,432	26543,58 - 30879,73
1992:08	31041,00	28806,83	1101,355	26638,91 - 30974,75
1992:09	30942,00	28902,01	1101,279	26734,24 - 31069,78
1992:10	31148,00	28997,18	1101,203	26829,56 - 31164,80
1992:11	30992,00	29092,36	1101,128	26924,88 - 31259,83
1992:12	31057,00	29187,53	1101,053	27020,21 - 31354,86
1993:01	31045,00	29282,71	1100,979	27115,53 - 31449,89
1993:02	31233,00	29377,88	1100,906	27210,85 - 31544,92
1993:03	31058,00	29473,06	1100,833	27306,17 - 31639,95
1993:04	31129,00	29568,23	1100,760	27401,48 - 31734,98
1993:05	31164,00	29663,41	1100,689	27496,80 - 31830,02
1993:06	31168,00	29758,59	1100,617	27592,12 - 31925,05
1993:07	31349,00	29853,76	1100,547	27687,43 - 32020,09
1993:08	31126,00	29948,94	1100,477	27782,74 - 32115,13

Obr. 3.20: Předpovědi – lineární trend

Tuto tabulku (obr. 3.20) si uložíme jako novou proměnnou do datového souboru tak, že v horní části klikneme na modrou ikonku „plus“ (viz obr. 3.20, červený rámeček), která značí „Přidat do datového souboru“. Nastavíme atributy (jméno proměnné a popis) a uložíme. To samé uděláme i s odhady parametrů pomocí kvadratického trendu (obr. 3.15) a exponenciálního trendu (obr. 3.17).

Nyní máme nadefinované tyto tři proměnné (obr. 3.21, červený kroužek) a můžeme se podívat na popisné statistiky. Označíme si tyto tři nově nadefinované proměnné, které vidíme na obrázku 3.21 v kroužku a pravým tlačítkem myši klikneme na „Deskriptivní statistika“, kde se nám objeví tabulka (obr. 3.22) s popisnými statistikami.



Obr. 3.21: Hlavní okno programu Gretl – nadefinované proměnné

	Střední hodnota	Medián	Minimum	Maximum
pre_store_linea	41608,	41608,	28141,	55075,
pre_store_quadr	41608,	40498,	30337,	57272,
pre_lstore_exp	10,618	10,618	10,295	10,942

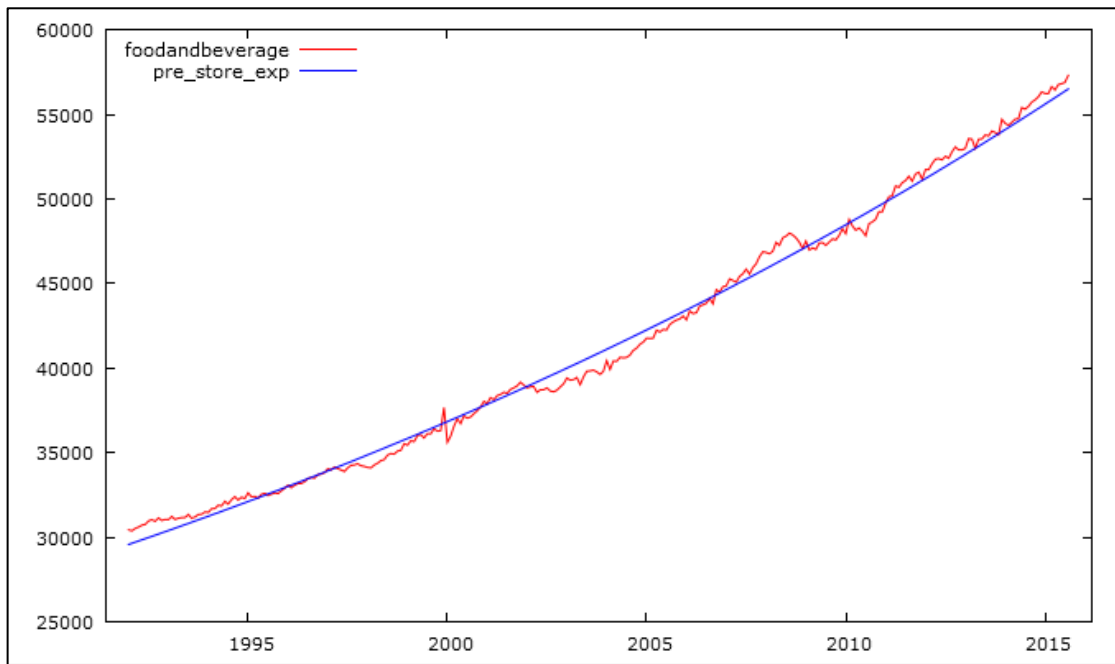
	Směr. odchv. koeficient	Šikmost	Stand. špičatost
pre_store_linea	7816,6	0,18786	-1,4187e-015
pre_store_quadr	7879,6	0,18938	0,33457
pre_lstore_exp	0,18778	0,017685	7,5813e-016

Obr. 3.22: Popisné statistiky

Na obrázku 3.22 vidíme popisné statistiky, kterými je například střední hodnota, která po použití lineárního a kvadratického modelu má hodnotu 41 608, zatímco předpokládané hodnoty pro zlogaritmovaný exponenciální model jsou 10,618. Obdobné nesrovnalosti jsou také u směrodatné odchylky.

Je tedy nutné zpátky odlogaritmovat na exponenciální trend tak, že nadefinujeme novou proměnnou. V hlavním menu klikneme na „Přidat“ -> „Definovat novou proměnnou“. Do otevřeného okna pak zapíšeme vzorec, kterým bude nová proměnná získána. V našem případě to bude: „pre_store_exp = exp(pre_lstore_exp)“. Tento vzorec znamená, že na hodnoty $\log Tr_t$ aplikujeme exponenciální funkci o základu e . Tímto máme nadefinovanou novou proměnnou, která již vyjadřuje exponenciální trend. Můžeme tedy přejít k vykreslení grafu. Klikneme na nově nadefinovanou proměnnou („pre_store_exp“) a na původní data

(„foodandbeverage“) a pravým tlačítkem myši klikneme na „Graf časové řady“, kde chceme řadu vykreslit na jednoduchém grafu. V tomto grafu (obr. 3.23) jsme časovou řadu proložili exponenciálou.



Obr. 3.23: Graf – aplikace exponenciálního trendu

Nyní se můžeme podívat na popisné statistiky. Označíme si proměnné jako v předchozím případě, akorát místo poslední proměnné použijeme proměnnou, kterou jsme teď nadefinovali. Opět klikneme na „Deskriptivní statistika“ a zobrazí se nám následující okno (obr. 3.24). Zde vidíme, že hodnoty u střední hodnoty a směrodatné odchylky jsou přibližně stejné, což je dobře.

gretl: popisné statistiky

Popisné statistiky, za použití pozorování 1992:01 - 2015:08

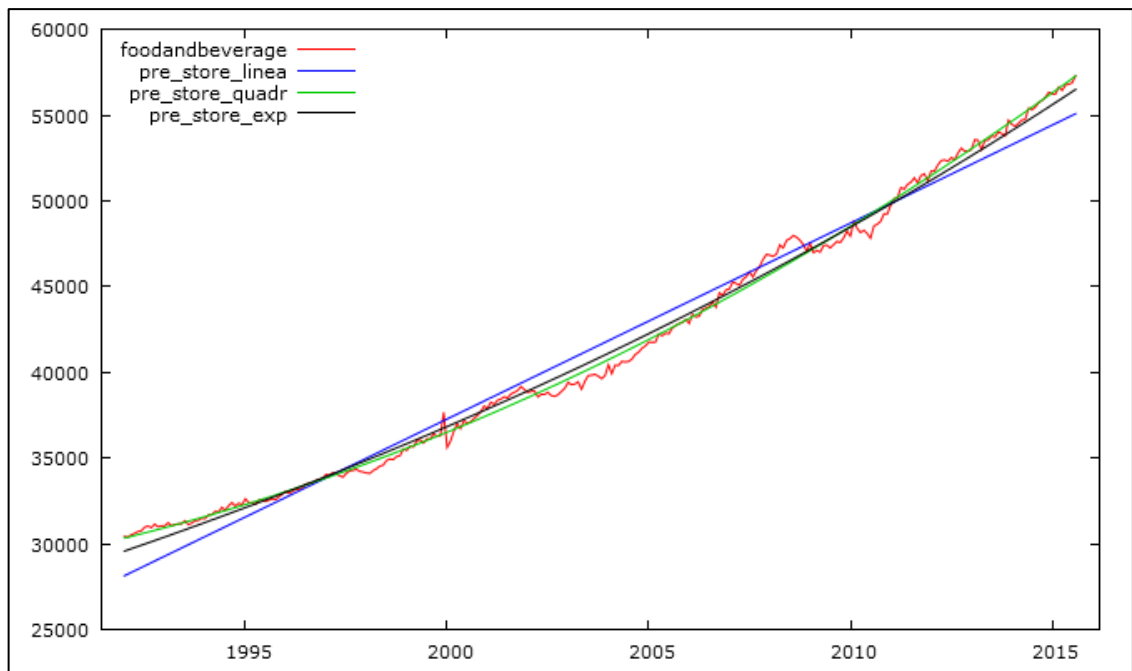
	Střední hodnota	Medián	Minimum	Maximum
pre_store_linea	41608,	41608,	28141,	55075,
pre_store_quadr	41608,	40498,	30337,	57272,
pre_store_exp	41598,	40876,	29578,	56491,

	Směr. odchv.	Směr. odchv. koeficient	Šikmost	Stand. špičatost
pre_store_linea	7816,6	0,18786	-1,4187e-015	-1,2000
pre_store_quadr	7879,6	0,18938	0,33457	-1,1086
pre_store_exp	7784,1	0,18713	0,22430	-1,1449

Obr. 3.24: Popisné statistiky 2

3.8.10 Celkový graf modelů

Snažili jsme se tedy trend namodelovat pomocí lineárního, kvadratického a exponenciálního trendu. Teď si tyto tři grafy dáme do jednoho, abychom viděli, který trend nejlépe vystihuje naši časovou řadu.



Obr. 3.25: Celkový graf trendů

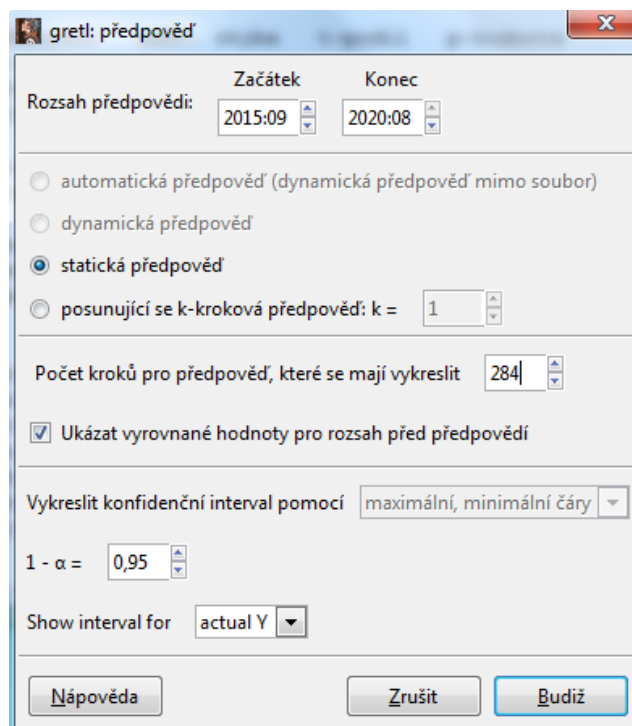
Z obrázku 3.25 vidíme graf skutečných a vyrovnaných hodnot, kde červená křivka znázorňuje naši časovou řadu, kterou je proložena přímka (modrá čára), parabola (zelená čára) a exponenciála (černá čára). Vizuální analýzou grafu je vidět, že nejlépe se datům přizpůsobuje kvadratický trend, neboť odchylky skutečných a odhadnutých hodnot jsou minimální. Tento závěr potvrzuje i nejvyšší index determinace z vytvořených modelů (viz obr. 3.13, obr. 3.15 a obr. 3.17).

3.8.11 Předpověď vývoje budoucích hodnot časové řady

Kromě posouzení toho, jak daný model dobře popisuje data, se musíme také zaměřit na to, jak se tato data mohou vyvíjet v budoucnu. Proto pro naše modely vytvoříme předpovědi trendů pro pět let dopředu.

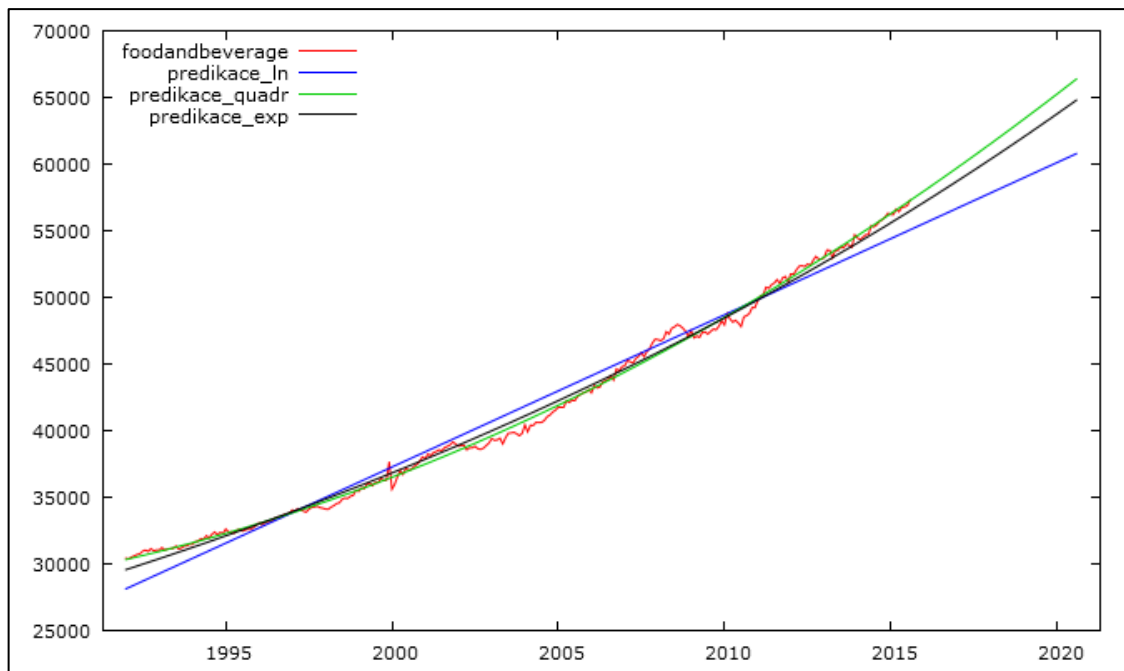
Přidáme tedy počet pozorování tak, že v hlavním okně klikneme na „Data“ -> „Přidat pozorování“. Objeví se nám okno, kde navolíme počet pozorování, které se mají přidat. V našem případě to bude 60 měsíců (5 let). Předpověď tedy bude do srpna roku 2020.

Poté se vrátíme k odhadům parametrů pomocí metody nejmenších čtverců (viz obr. 3.13, obr. 3.15 a obr. 3.17). V horní části těchto odhadů klikneme na „Analýza“ -> „Předpovědi“. Objeví se nám okno (obr. 3.26), kde vidíme rozsah předpovědi do srpna 2020. Počet kroků pro předpověď, které se mají vykreslit, zvolíme 284, neboť to chceme od roku 1992 a potvrdíme. Toto provedeme u všech tří odhadů.



Obr. 3.26: Definování předpovědi na následujících 5 let

Objeví se nám tabulky, které si uložíme tak, že klikneme na modrou ikonku „plus“ (viz obr. 3.20) a do pole jméno si dáme název (v našem případě „*predikace_ln*“ pro lineární trend, „*predikace_quadr*“ pro kvadratický trend a „*predikace_ln_ex*“ pro zlogaritmovaný exponenciální trend). Je nutné na hodnoty $\log Tr$ aplikovat exponenciální funkci o základu e tak, že nadefinujeme novou proměnnou a do pole napíšeme „*predikace_exp = exp(predikace_ln_ex)*“. Nyní máme navolené proměnné pro předpověď na následujících 5 let a můžeme přejít k vykreslení grafu. Klikneme na predikaci pro lineární, kvadratický, exponenciální trend a na původní data a pravým tlačítkem myši najedeme na „*Graf časové řady*“, kde chceme vykreslit řadu na jednoduchém grafu.



Obr. 3.27: Celkový graf s předpovědí na 5 let

Do obrázku 3.27 jsme zakreslili jak dosavadní průběh dat, tak také odhadnuté trendy a jejich předpovědi na následujících pět let. Dále je z grafu vidět, že predikovaný budoucí vývoj má rostoucí trend. Kvadratický model by měl být podle předpokladů, které jsme už zjistili, jeden z nejvhodnějších.

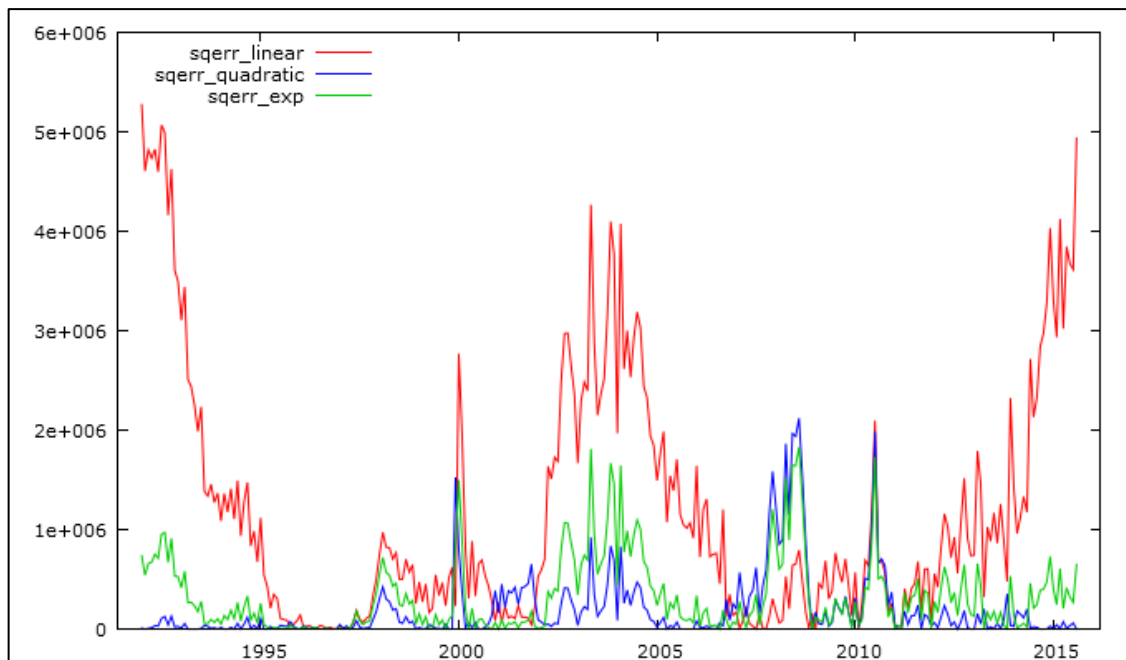
3.8.12 Chyby v předpovědích

Zmíníme se ještě o chybách v předpovědi (viz 1.5 Předpovědi v časových řadách), abychom mohli posoudit, který ze tří modelů má nejnižší čtvercové chyby. Chyba v předpovědi \hat{y}_t skutečné hodnoty y_t je stanovena jako

$$e_t = y_t - \hat{y}_t.$$

Proto si nadefinujeme další proměnné, kde vzorce budou mít tvar „ $sqerr_linear = (foodandbeverage - pre_store_linea)^2$ “ pro čtvercové chyby lineárního trendu, „ $sqerr_quadratic = (foodandbeverage - pre_store_quadr)^2$ “ pro čtvercové chyby kvadratického trendu a „ $sqerr_exp = (foodandbeverage - pre_store_exp)^2$ “ pro čtvercové chyby exponenciálního trendu.

Nyní přejdeme k vykreslení grafu, abychom viděli, který ze tří modelů má nejnižší čtvercové chyby. Označíme si tyto tři nově nadefinované proměnné a pravým tlačítkem myši klikneme na „Graf časové řady“ s vykreslením do jednoho grafu.



Obr. 3.28: Graf čtvercových chyb

Z obrázku 3.28 lze usoudit, že nejnižší čtvercové chyby má kvadratický trend a naopak nejvyšší má lineární trend.

3.8.13 Shrnutí

V tomto příkladu jsem se nejdříve snažila na data aplikovat lineární model s více proměnnými. Poté co jsem udělala odhad parametrů pomocí metody nejmenších čtverců, jsem zjistila, že pomocné proměnné dm1 až dm12 se nejevily jako statisticky významné, tudíž jsem je vyloučila.

Následně jsem se tedy snažila modelovat časovou řadu pomocí metody dekompozice. Byly zde použity tři modely časové řady: model trendové přímky, model trendové paraboly a model trendové exponenciály. Pomocí statistických kritérií a dodatečných kritérií byla porovnána jejich využitelnost. Bylo zjištěno, že nejvhodnějším modelem se jeví model trendové paraboly. Neboť tento model měl nejvyšší koeficient determinace a také nejvíce vystihoval charakter naší časové řady. Co se týče chyb v předpovědích, tak jsem z grafu usoudila, že nejnižší čtvercové chyby má též kvadratický trend.

V celém příkladu byl uveden i postup, jak se s programem Gretl pracuje.

4 Metoda klouzavých průměrů

Metoda klouzavých průměrů stejně jako exponenciální vyrovnání (viz kapitola 5) patří mezi tzv. *adaptivní přístupy*. Obecně můžeme adaptivní přístupy popsat tak, že dokáží pracovat s trendovými složkami, které mění v čase svůj charakter, takže nelze použít žádnou matematickou křivku s neměnnými parametry [6].

Tuto metodu můžeme použít v případě, chceme-li odstranit z časové řady šum, který vzniká působením náhodných vlivů. Spočívá to v tom, že se původní řada pozorování nahradí řadou vypočtených klouzavých průměrů.

Existuje mnoho nejrůznějších druhů klouzavých průměrů. V této práci se zaměřím na jednoduché klouzavé průměry, vážené klouzavé průměry, centrované klouzavé průměry a exponenciální klouzavé průměry.

4.1 Konstrukce klouzavých průměrů

Podle Cipry [6] předpokládáme, že každá „rozumná“ funkce může být aproximována polynomem. Budeme tedy postupovat následovně:

1. Vyrovnáme polynomem prvních $2m + 1$ členů řady.
2. Dále použijeme hodnotu polynomu v bodě $t = m + 1$ jako vyrovnanou hodnotu \hat{y}_{m+1} určité řady v tomto bodě.
3. Pro získání vyrovnaných hodnot v bodě $t = m + 2$, provedeme to samé s pozorováními y_2, \dots, y_{2m+2} .

4.2 Vlastnosti klouzavých průměrů

Nyní si uvedeme některé vlastnosti klouzavých průměrů, které jsou nejdůležitější [17]:

1. Součet vah klouzavého průměru je roven jedné.
2. Váhy jsou symetrické kolem střední hodnoty.
3. Je-li r sudé číslo, potom klouzavé průměry řádu r a $r + 1$ se stejnou délkou $2m + 1$ jsou totožné.

4.3 Volba parametrů klouzavých průměrů

Tato metoda je podle Křivého [12] založena na vyrovnání krátkých úseků časové řady polynomickými funkcemi. Parametry se volí subjektivně podle posouzení charakteru dat tak, že se upřednostňují průměry co nejnižšího řádu a délka je

zvolena podle určitého stupně vyhlazení. Mezi základní dva parametry patří délka a řád klouzavých průměrů.

Délka klouzavých průměrů určuje skutečnou délku vyrovnaných úseků časové řady. Obecně se předpokládá, že se jedná o liché číslo, tj. $2m + 1$. Dále platí, že čím je větší délka klouzavého průměru, tím je větší vyrovnaní časové řady.

Řád klouzavých průměrů (r), představuje stupeň vyrovnávacího polynomu. O řádu lze rozhodnout na základně objektivního kritéria, které se opírá o diferencování dané řady [22]:

$$\begin{aligned}\Delta y_t &= y_t - y_{t-1} \\ \Delta^2 y_t &= \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2} \\ &\cdot \\ &\cdot \\ \Delta^k y_t &= y_t - \binom{k}{1}y_{t-1} + \binom{k}{2}y_{t-2} - \dots + (-1)^k y_{t-k}\end{aligned}\tag{4.1}$$

Označíme-li tedy

$$V_k = \frac{\sum (\Delta^k y_t)^2}{\binom{2k}{k}(n-k)}\tag{4.2}$$

kde symbolem \sum se rozumí součet pro t od $1 + k$ do n a $\Delta^k y_t$ značí k -tou diferenci. Pro $k \geq (r + 1)$ je hodnota kritéria V_k odhad rozptylu bílého šumu. V praxi se počítají hodnoty V_1, V_2, \dots dokud se nezaznamená, že hodnoty začnou konvergovat k nějaké konstantě. Poté, co se hodnoty V_{r+1}, V_{r+2}, \dots blíží ke konstantě, se doporučuje vybrat klouzavé průměry řádu r .

4.4 Jednoduché klouzavé průměry

Nejjednodušší je použít pro výpočet klouzavých průměrů tzv. *jednoduché klouzavé průměry*. Jedná se o prosté aritmetické průměry. Platí pro ně předpoklad, že na jednotlivých klouzavých částech je definovaný lineární trend. Například jednoduché klouzavé průměry délky 5 mají tvar

$$\frac{1}{5}(1,1,1,1,1)y_t = \frac{1}{5}(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}).\tag{4.3}$$

Je tedy pochopitelné, že jednoduchý průměr liché délky $2m + 1$ odpovídá klouzavému průměru řádu 0 nebo 1 téže délky. Pro předpověď budoucí hodnoty $y_{n+\tau}$ ($\tau > 0$), která je konstruovaná v čase t pomocí jednoduchých klouzavých průměrů platí

$$\hat{y}_{n+\tau} = \frac{1}{2m+1} (y_{n-2m} + y_{n-2m+1} + \dots + y_n). \quad (4.4)$$

V uvedeném příkladu byl použit klouzavý průměr délky 5, je ale možné použít i jiné délky jako je například 3, 7, 9 či jiné.

4.5 Centrované klouzavé průměry

Vyrovňovat časovou řadu sudé délky není vhodné, neboť vyrovnaná hodnota neodpovídá žádnému okamžiku měření. V takové situaci je dobré použít tzv. *centrované klouzavé průměry*. Jde o vážené průměry, mající speciální váhy, které jsou zvoleny tak, aby odstranily z časové řady sezónní složku. Délka je vždy o jedničku větší než délka sezóny a váhy jsou vybrány tak, aby krajní pozorování byla poloviční.

4.6 Vážené klouzavé průměry

V případě, že klouzavou část vyrovňované řady můžeme popsat kvadratickou trendovou funkcí, používají se tzv. *vážené klouzavé průměry*. Zde hraje důležitou roli volba vah, ty se dají odvodit na základě metody nejmenších čtverců, jestliže proložíme krátké úseky časové řady polynomem řádu r . Klouzavá část je podle Hindlse [10] označována symbolem $m = 2p + 1$, kde $p = 1, 2, 3, \dots$ a platí, že $m < n$, kde n značí celkový počet pozorování časové řady. Podle Hindlse [10] dostaneme tvar

$$\bar{y}_t = \sum_{i=-p}^p W_i y_{t,i}, \quad t = p + 1, p + 2, \dots, n - p, \quad (4.5)$$

kde

$$W_i = \frac{3}{4m(m^2 - 4)} (3m^2 - 7 - 20i^2), \quad i = -p, \dots, -1, 0, 1, \dots, p. \quad (4.6)$$

Vzorec (4.5) se nazývá vážený klouzavý průměr v čase t a hodnoty W_i , které jsou definované výrazem (4.6) jsou jeho váhy splňující podmínky

$$\sum_{i=-p}^p W_i = 1 \quad \text{a} \quad W_i = W_{-i}, \quad (4.7)$$

tj. váhy jsou symetrické.

4.7 Exponenciální klouzavé průměry

Exponenciální klouzavé průměry (anglicky exponential moving average – EMA) patří mezi vážené klouzavé průměry, kde novějším hodnotám se přiřazují větší váhy a váhy se pohybují exponenciálně.

Existuje zde mnoho způsobů výpočtů exponenciálních klouzavých průměrů. Jeden z nich je následující

$$EMA_t = \alpha \cdot y_t + (1 - \alpha) \cdot EMA_{t-1}, \quad (4.8)$$

kde EMA_t je hodnota exponenciálního klouzavého průměru v čase t , EMA_{t-1} je hodnota exponenciálního klouzavého průměru v čase $t - 1$, y_t je hodnota časové řady v čase t a α (vyrovnávací konstanta) je koeficient. Tento koeficient charakterizuje rychlost snížení vah, hodnotu může mít od 0 do 1. Pro výpočet vyrovnávací konstanty se používá následující vzorec

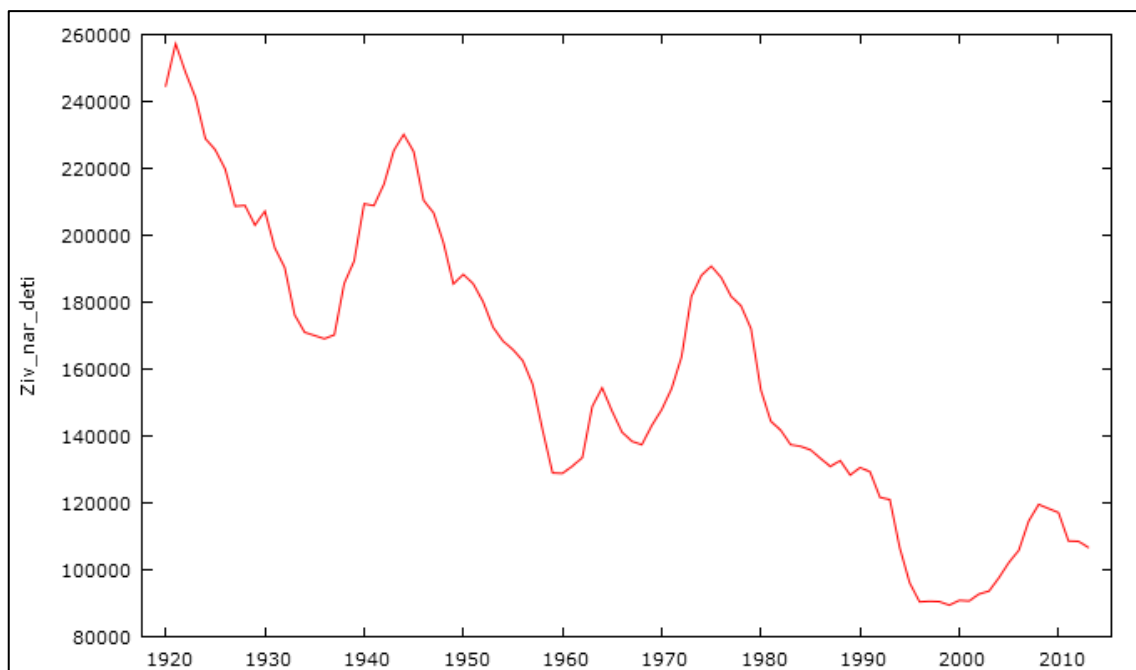
$$\alpha = \frac{2}{n+1}, \quad (4.9)$$

kde n je perioda klouzavého průměru, tj. počet sledovaných časových období.

4.8 Příklad

Máme k dispozici roční časovou řadu počtu živě narozených dětí v České republice v letech 1920 – 2013 (viz příloha č. 2). Ukážeme si, jak v programu Gretl modelovat její průběh pomocí klouzavých průměrů. Data si upravíme a poté importujeme do programu Gretl, jak jsme si ukázali v příkladu 3.8 (viz 3.8.2 Úprava dat a 3.8.4 Import dat). Strukturu souboru dat zvolíme časové řady s roční frekvencí (viz 3.8.5 Struktura dat). Následně se data nahrály do programu, kde si přepíšeme název na „Ziv_nar_deti“.

Nyní můžeme přejít k vykreslení grafu (viz 3.8.7 Sestrojení grafu) této časové řady, kde osa x nám bude udávat jednotlivé roky a osa y bude znázorňovat údaje o počtu živě narozených dětí.

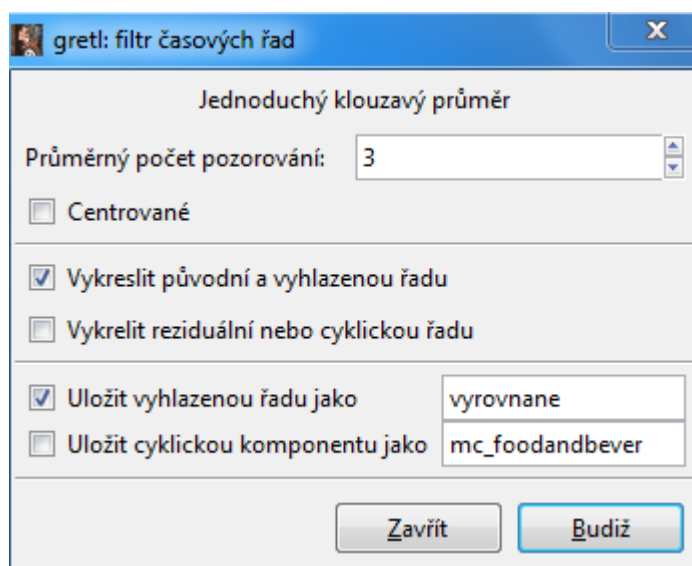


Obr. 4.1: Graf vývoje počtu živě narozených dětí v ČR

Z obr. 4.1 je vidět, že počet živě narozených dětí v ČR má klesající trend za naše sledované období.

4.8.1 Jednoduché klouzavé průměry

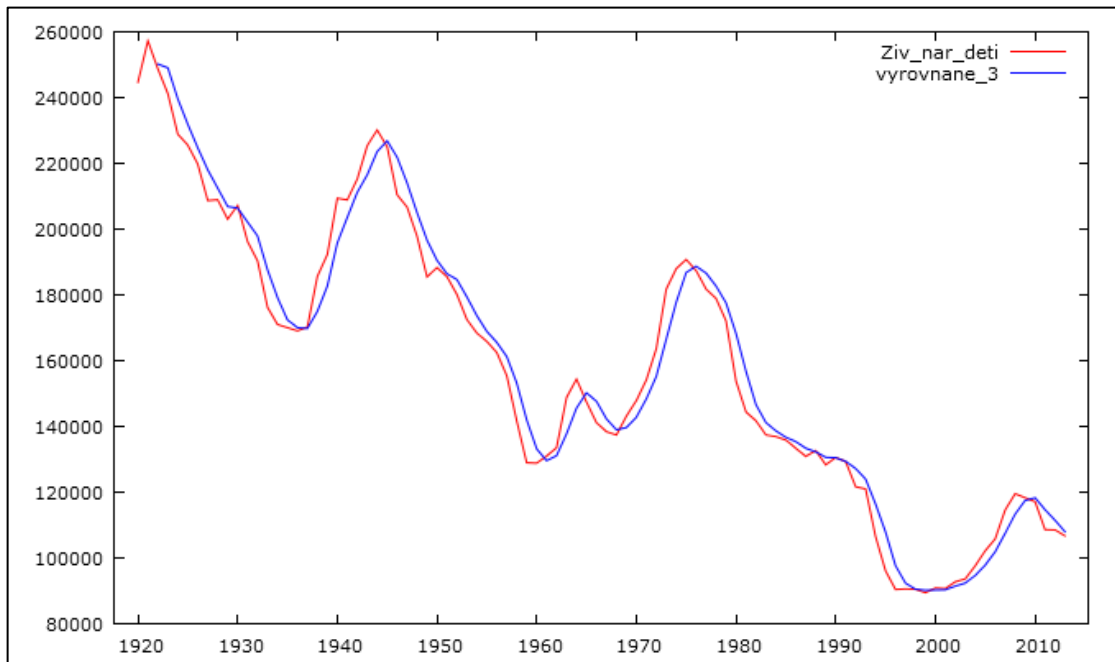
Přejdeme tedy k vyrovnání časové řady jednoduchými klouzavými průměry. V horní části hlavního menu klikneme na „Proměnná“ -> „Filtr“ -> „Jednoduchý klouzavý průměr“. Objeví se nám okno jako na obrázku 4.2, kde si vše navolíme. Průměrný počet pozorování nám udává délku klouzavé části.



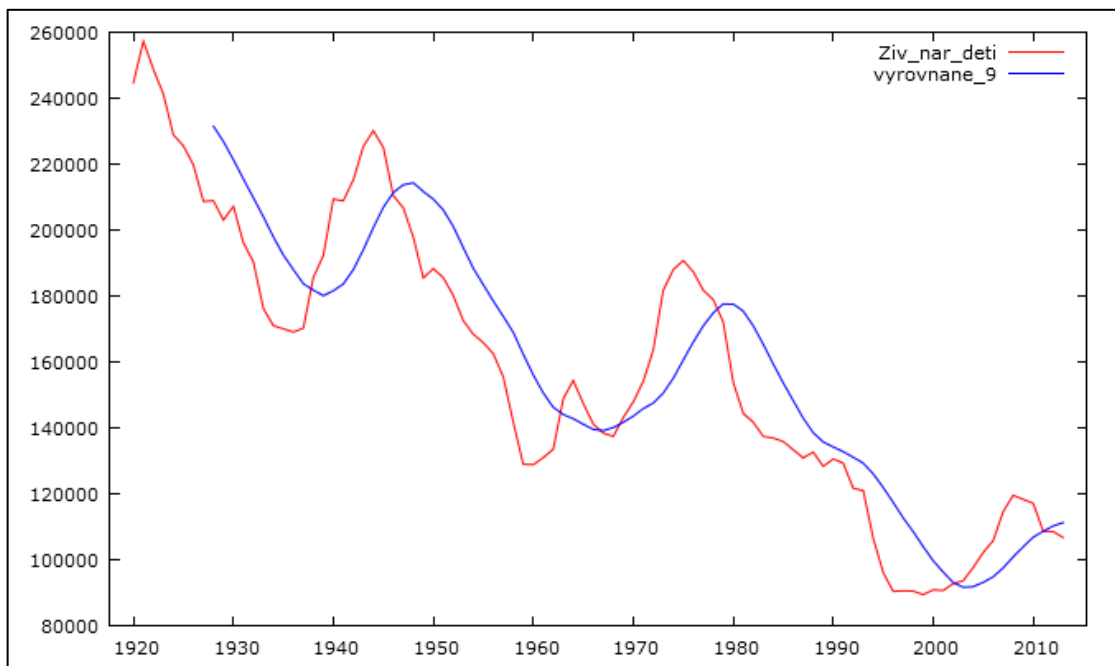
Obr. 4.2: Filtr – jednoduchý klouzavý průměr

Časovou řadu tedy vyrovnáme jednoduchými klouzavými průměry s délkou klouzavé části 3, 9 a 15 a budeme pozorovat, jak tato délka klouzavé části ovlivňuje vyrovnání časové řady.

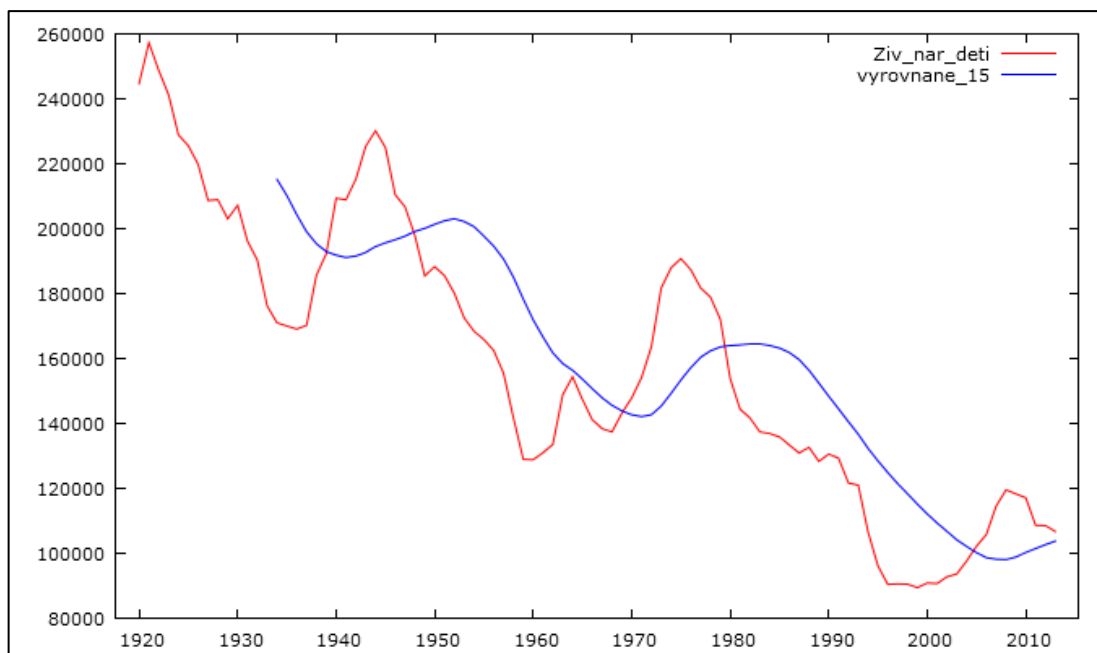
Do pole „průměrný počet pozorování“ postupně zadáme čísla 3, 9 a 15. Vyhlazenou časovou řadu budeme chtít uložit pod názvem „vyrovnane_3“, „vyrovnane_9“ a „vyrovnane_15“. Výsledky vyrovnávání časové řady klouzavými průměry různé délky můžeme pozorovat na obr. 4.3, obr. 4.4 a obr. 4.5.



Obr. 4.3: Jednoduché klouzavé průměry délky 3



Obr. 4.4: Jednoduché klouzavé průměry délky 9

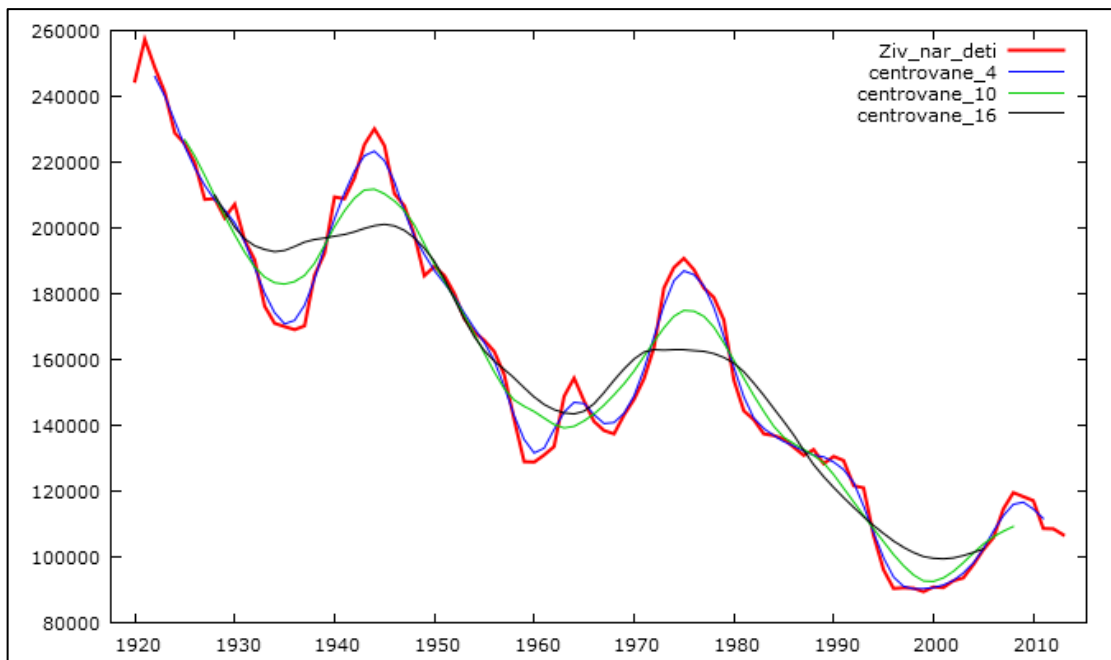


Obr. 4.5: Jednoduché klouzavé průměry délky 15

Z těchto tří grafů je vidět, že pokud je délka klouzavé části malá (3 hodnoty), je vyrovnání podobné původní časové řadě (obr. 4.1). V případě, že je délka klouzavé části větší (15 hodnot), je trendová čára hladší. Čím je tedy délka klouzavé části větší, tím je i větší vyrovnání (vyhlazení) časové řady a naopak.

4.8.2 Centrované klouzavé průměry

V případě, že budeme chtít zadat délku klouzavé části sudé číslo, použijeme centrované klouzavé průměry a to tak, že zaškrtneme v tabulce (obr. 4.2) „Centrované“. Časovou řadu tedy vyrovnáme centrovanými klouzavými průměry s délkou klouzavé části 4, 10 a 16. Pro porovnání uvádím všechny řady najednou v obr. 4.6.



Obr. 4.6: Celkový graf centrovaných klouzavých průměrů

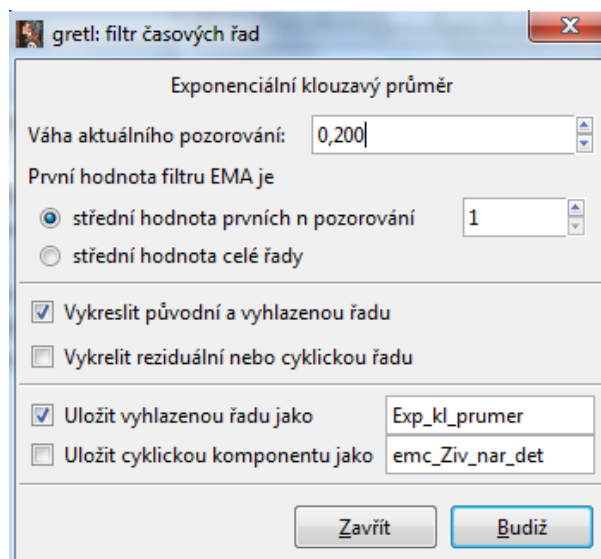
Červená křivka vyjadřuje původní data, modrá křivka vyrovnané hodnoty pomocí centrovaných klouzavých průměrů délky 4, zelená křivka vyrovnané hodnoty pomocí centrovaných klouzavých průměrů délky 10 a černá křivka vyrovnané hodnoty pomocí centrovaných klouzavých průměrů délky 16. Z obrázku je také vidět, že čím větší číslo, tím je trendová čára hladší.

Program Gretl nám také umožňuje upravovat graf, aby byl přehlednější. Pravým tlačítkem myši klikneme na graf -> „Editovat“. Jednou z možností této funkce je navolit si typ, jak se nám má křivka vykreslit, tloušťku čáry nebo také barvu.

4.8.3 Exponenciální klouzavé průměry

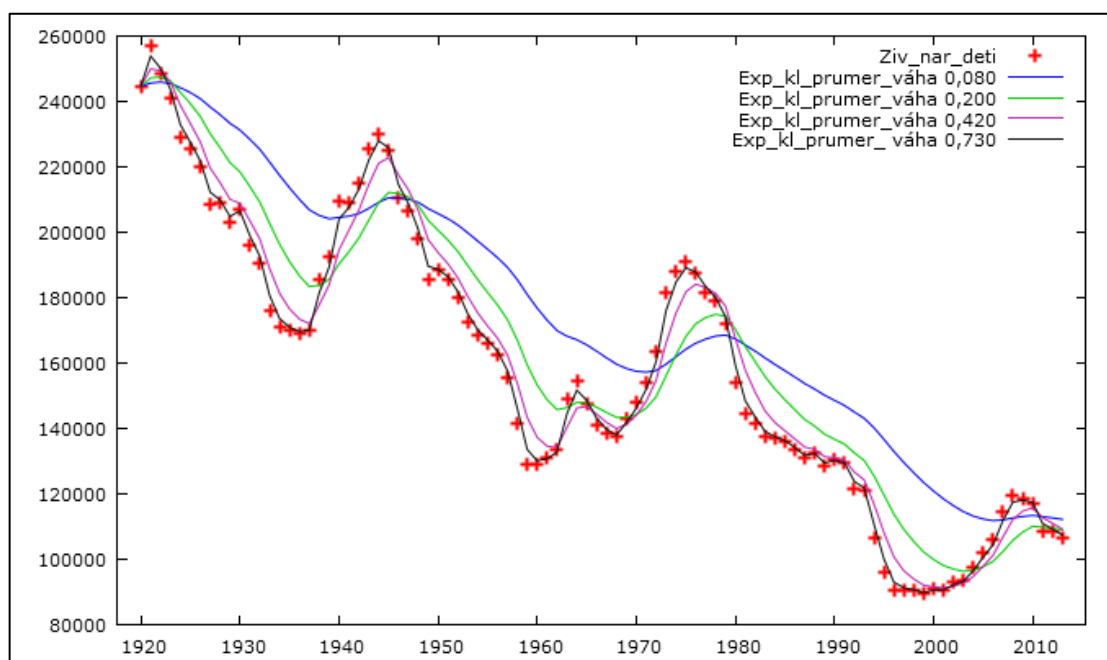
V programu Gretl můžeme kromě jednoduchých klouzavých průměrů využívat i tzv. exponenciální klouzavé průměry.

V hlavním menu klikneme na „Proměnná“ -> „Filtr“ -> „Exponenciální klouzavý průměr“. Objeví se nám okno jako na obr. 4.7, kde si opět vše navolíme.



Obr. 4.7: Filtr - exponenciální klouzavý průměr

Časovou řadu budeme chtít tedy vyrovnat exponenciální klouzavými průměry, kde důležitou roli hraje váha pozorování. Váhy můžeme volit od 0,001 do 0,999. V našem případě zvolíme váhy 0,080, 0,200, 0,420 a 0,730 a budeme pozorovat, jak tato váha ovlivňuje vyrovnání časové řady. Pro porovnání uvádím všechny řady najednou v obr. 4.8.



Obr. 4.8: Celkový graf exponenciálních klouzavých průměrů

Z grafu (obr. 4.8) je vidět, že čím je větší váha aktuálního pozorování, tím je vyrovnání podobné původní časové řadě. V opačném případě, když je váha klouzavé části malá, tím je větší vyhlazení.

5 Exponenciální vyrovnání

Jak jsem již zmínila v kapitole 4, exponenciální vyrovnání je další adaptivní přístup, který se v praxi poměrně často používá. Co se týče metody klouzavých průměrů (viz kapitola 4), ta se snaží vyrovnávat v časové řadě krátké úseky, jejichž délka je předem stanovena. Kdežto u exponenciálního vyrovnání je výpočet založen na všech minulých pozorování časové řady. Pro odhad parametrů se používá metoda nejmenších čtverců, kde váhy pozorování se snižují směrem do minulosti. Podle Řezankové aj. [19] se tedy minimalizuje výraz tvaru

$$(y_t - \hat{y}_t)^2 + (y_{t-1} - \hat{y}_{t-1})^2 \alpha + (y_{t-2} - \hat{y}_{t-2})^2 \alpha^2 + \dots, \quad (5.1)$$

kde α je tzv. *vyrovnávací konstanta* a platí pro ni $0 < \alpha < 1$.

Předpokládá se, že časová řada bude mít tvar

$$y_t = Tr_t + E_t. \quad (5.2)$$

Z tohoto vzorce je vidět, že bude očištěna od sezónní a cyklické složky.

Princip exponenciálního vyrovnání je po výpočetní stránce poměrně jednoduchý a má také malé nároky na potřebný objem uchovávaných dat.

Rozlišují se tři typy exponenciálního vyrovnání a to jednoduché, dvojitě a trojitě exponenciální vyrovnání, se kterými se nyní blíže seznámíme.

5.1 Jednoduché exponenciální vyrovnání

Používá se v případě, kdy trendová složka dané časové řady je v krátkých úsecích konstantní, platí tedy pro ni

$$Tr_t = \beta_0. \quad (5.3)$$

Nyní je nutné nalézt odhad parametru β_0 . Protože exponenciální vyrovnání patří mezi adaptivní přístupy, bude tento odhad závislý na časovém okamžiku, ve kterém byl proveden. Označíme $b_0(t)$ odhad parametru β_0 uskutečněný v čase t . Odhad tedy získáme minimalizací výrazu

$$\sum_{j=0}^{\infty} [y_{t-j} - b_0(t)]^2 \alpha^j, \quad (5.4)$$

kde α je tzv. *vyrovnávací konstanta* a platí pro ni $0 < \alpha < 1$.

Jak vidíme, výraz (5.4) je definovaný jako nekonečný součet, v praxi ale budeme pracovat pouze s konečným počtem hodnot y_1, \dots, y_n .

Položíme-li parciální derivaci výrazu (5.4) podle β_0 rovnou nule, dostaneme odhad jako

$$b_0(t) = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j y_{t-j}, \quad (5.5)$$

nebo jako vyrovnanou hodnotu v čase t

$$\hat{y}_t = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j y_{t-j}. \quad (5.6)$$

Odtud je patrné, že vyrovnaná hodnota řady v čase t je váženým součtem hodnot řady do času t s exponenciálními klesajícími váhami [6]

$$(1 - \alpha), (1 - \alpha)\alpha, (1 - \alpha)\alpha^2, \dots \quad (5.7)$$

Výraz (5.6) můžeme snadno přepsat na tvar

$$\hat{y}_t = (1 - \alpha)y_t + \alpha\hat{y}_{t-1}, \quad (5.8)$$

který symbolizuje rekurentní předpis pro výpočet vyrovnaných hodnot řady.

5.1.1 Volba vyrovnávací konstanty α

Je vhodné volit α z intervalu $0,7 \leq \alpha < 1$. Hodnotu této konstanty lze upřesnit dvěma způsoby:

- a) pomocí vzorce $\alpha = \frac{M-1}{M+1}$, kde M je délka jednoduchého klouzavého průměru,
- b) hodnota α se určuje pomocí simulace, která spočívá v tom, že se vybírají hodnoty $\alpha = 0,70; 0,78; \dots; 0,98$ a vybere se ta hodnota, která má nejlepší předpovědi.

5.2 Dvojitě exponenciální vyrovnání

U tohoto vyrovnání, které se také někdy nazývá jako Brownova metoda, se předpokládá, že trendovou složku lze v krátkých úsecích považovat za lineární, tj.

$$Tr_t = \beta_0 + \beta_1 t. \quad (5.9)$$

Dále budeme odhadovat parametry β_0 a β_1 v čase t , kde jejich odhady označíme $b_0(t)$ a $b_1(t)$ a minimalizací výrazu získáme

$$\sum_{j=0}^{\infty} [y_{t-j} - b_0(t) + b_1(t)(-j)]^2 \alpha^j, \quad (5.10)$$

kde $0 < \alpha < 1$ je opět vyrovnávací konstanta.

Jestliže vypočteme parciální derivace podle β_0 a β_1 a tyto derivace položíme rovno nule, dostaneme soustavu normálních rovnic

$$\begin{aligned} \sum_{j=0}^{\infty} y_{t-j} \alpha^j - b_0(t) \sum_{j=0}^{\infty} \alpha^j + b_1(t) \sum_{j=0}^{\infty} j \alpha^j &= 0, \\ \sum_{j=0}^{\infty} y_{t-j} j \alpha^j - b_0(t) \sum_{j=0}^{\infty} j \alpha^j + b_1(t) \sum_{j=0}^{\infty} j^2 \alpha^j &= 0, \end{aligned} \quad (5.11)$$

kteřou pomocí vzorců můžeme zjednodušit

$$\sum_{j=0}^{\infty} \alpha^j = \frac{1}{1-\alpha}, \quad \sum_{j=0}^{\infty} j \alpha^j = \frac{\alpha}{(1-\alpha)^2}, \quad \sum_{j=0}^{\infty} j^2 \alpha^j = \frac{\alpha(1+\alpha)}{(1-\alpha)^3}, \quad (5.12)$$

na tvar

$$\begin{aligned} b_0(t) - \frac{\alpha}{1-\alpha} b_1(t) &= (1-\alpha) \sum_{j=0}^{\infty} y_{t-j} \alpha^j, \\ \alpha b_0(t) - \frac{\alpha(\alpha+1)}{1-\alpha} b_1(t) &= (1-\alpha)^2 \sum_{j=0}^{\infty} y_{t-j} j \alpha^j. \end{aligned} \quad (5.13)$$

Pro zjednodušení zavedeme dvě veličiny:

- jednoduchá vyrovnávací statistiky S_t

Předpis této statistiky je

$$S_t = (1-\alpha) \sum_{j=0}^{\infty} \alpha^j y_{t-j}, \quad (5.14)$$

a je obdobou vztahu (5.6). Podle vztahu (5.8) platí

$$S_t = (1-\alpha)y_t + \alpha S_{t-1}. \quad (5.15)$$

- dvojité vyrovnávací statistika $S_t^{[2]}$

Ta je definovaná jako

$$S_t^{[2]} = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j S_{t-j}, \quad (5.16)$$

a obdobně jako v předchozím případě platí rekurentní vztah

$$S_t^{[2]} = (1 - \alpha)S_t + \alpha S_{t-1}^{[2]}, \quad (5.17)$$

který vyplývá ze vztahu (5.16).

5.3 Trojité exponenciální vyrovnání

U tohoto typu exponenciálního vyrovnání se předpokládá, že trendovou složku lze v krátkých úsecích popsat kvadratickým polynomem, tj.

$$Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad (5.18)$$

Tento typ je nejsložitější, proto se v praxi používá velice málo.

Postup při odhadu parametrů se počítá obdobně jako u dvojitého exponenciálního vyrovnání. Odvození vztahů je poměrně složité, protože do nich vstupuje navíc trojitá vyrovnávací statistika, která je rekurentně definovaná jako

$$S_t^{[3]} = (1 - \alpha)S_t^{[2]} + \alpha S_{t-1}^{[3]}. \quad (5.19)$$

Závěr

Cílem této bakalářské práce bylo především ukázat možnosti použití některých metod z oblasti analýzy časových řad.

Nejprve jsem vypracovala literární přehled, kde jako první byl vymezen pojem časové řady, její druhy, specifické problémy a základní přístupy k analýze časových řad. V závěru kapitoly jsem se zmínila o předpovědích v časových řadách a představila vám program, se kterým jsem v této práci pracovala. Poté jsem se zaměřila na dekompozici časových řad, kde jsem popsala jednotlivé složky. Dále modely, které popisují trendovou složku a adaptivní přístupy, kam patří metoda klouzavých průměrů a exponenciální vyrovnání.

Praktická část se věnovala modelování časových řad ze získaných dat. Využila jsem možnosti české verze softwaru Gretl a pokusila jsem se v něm analyzovat vybrané časové řady. V tomto programu byly vyzkoušeny klasické postupy eliminace trendu a tzv. adaptivní metody (konkrétně metoda klouzavých průměrů). Uvedené metody byly doplněny i návodem, jak se s programem Gretl pracuje, aby byl čtenář lépe v obrazu.

První příklad analyzoval data pomocí matematických křivek. Jednalo se o časovou řadu měsíčních maloobchodních tržeb obchodů s jídlem a nápoji za období leden 1992 až srpen 2015. Data tak byla popsána pomocí lineárního trendu, kvadratického trendu a exponenciálního trendu. Dále byla provedena predikace na následujících pět let a na závěr jsem se letmo zaměřila na chyby v předpovědích. Podle získaných grafických a statistických výsledků bylo usouzeno, že vybraná data nejlépe vystihují kvadratický trend.

Časová řada udávající počet živě narozených dětí v České republice od roku 1920 do roku 2013 byla použita pro práci s adaptivními přístupy (konkrétně metody klouzavých průměrů), které jsou vhodné pro data, jejichž trend nemůžeme popsat žádnou matematickou křivkou. V programu Gretl máme k dispozici jednoduché klouzavé průměry, centrované klouzavé průměry a exponenciální klouzavé průměry. Výsledky těchto metod nebyly nijak statisticky ověřeny, neboť se jedná především o analýzy subjektivní.

Je nutné poznamenat, že program Gretl nabízí opravdu široké možnosti statistických analýz, ale ne všechny. Při své práci jsem narazila na problém, že v programu Gretl není funkce, která by vyrovnala časovou řadu pomocí metody exponenciálního vyrovnání. Tento problém by se dal snadno vyřešit v Excelu, ve Statgraphicsu nebo například v programu Statistica, kde je přímo funkce na exponenciální vyrovnání.

Pěvně věřím, že čtenáři, kterému se tato práce dostane do ruky, bude aspoň trochu přínosem a v programu Gretl se lépe zorientuje.

Seznam použitých zdrojů

- [1] ADKINS, L. C.: *Using gretl for Principles of Econometrics, 4th Edition* [online]. [cit. 2015-06-28]. Dostupné z: http://www.learneconometrics.com/gretl/using_gretl_for_POE4.pdf
- [2] ANDĚL, Jiří. *Statistická analýza časových řad*. 1. vyd. Praha: Státní nakladatelství technické literatury, 1976, 271 s.
- [3] ARLT, Josef a Markéta ARLTOVÁ. *Ekonomické časové řady: [vlastnosti, metody modelování, příklady a aplikace]*. 1. vyd. Praha: Grada, 2007, 285 s. ISBN 978-80-247-1319-9.
- [4] BIL, J., D. NĚMEC a M. POSPIŠ. *Gretl – uživatelská příručka* [online]. [cit. 2015-09-15]. Dostupné z: http://www.thunova.cz/wp-content/uploads/CZU/Manual_gretl.pdf
- [5] BŘÍZA, Michal. *Analýza výkonnosti firmy Mida, a.s. pomocí časových řad* [online]. [cit. 2015-05-15]. Dostupné z: https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=31207
- [6] CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*. 1. vyd. Praha: Státní nakladatelství technické literatury, 1986, 246 s.
- [7] FISCHEROVÁ, Jana. *Analýza ukazatelů společnosti Bohemia asphalt s.r.o. pomocí časových řad* [online]. [cit. 2015-09-12]. Dostupné z: https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=41909
- [8] *Gnu Regression, Econometrics and Time-series Library – Gretl* [online]. [cit. 2015-09-15]. Dostupné z: <http://gretl.sourceforge.net/>
- [9] HANČLOVÁ, Jana a Lubor TVRDÝ. *Úvod do analýzy časových řad* [online]. [cit. 2015-09-12]. Dostupné z: http://gis.vsb.cz/pan-old/Skoleni_Texty/TextySkoleni/AnalyzaCasRad.pdf
- [10] HINDLS, Richard. *Statistika pro ekonomy*. 8. vyd. Praha: Professional Publishing, 2007, 415 s. ISBN 978-80-86946-43-6.
- [11] KROPÁČ, Jiří. *Statistika B: jednorozměrné a dvourozměrné datové soubory, regresní analýza, časové řady*. 3. vyd. Brno: Akademické nakladatelství CERM, 2012, 145 s. ISBN 978-80-7204-822-9.
- [12] KŘIVÝ, Ivan. *Analýza časových řad* [online]. [cit. 2015-09-18]. Dostupné z: <https://publi.cz/download/publication/20?online=1>

- [13] KUČTOVÁ, Martina. *Tvorba modelu pro přípravu a zpracování experimentálních dat* [online]. [cit. 2015-05-12]. Dostupné z: <http://www.kvhem.cz/wp-content/uploads/2009/02/Kuchtova.pdf>
- [14] LITSCHMANNOVÁ, Martina. *Explorační analýza časových řad (teorie)* [online]. [cit. 2015-05-12]. Dostupné z: http://homel.vsb.cz/~lit40/SMAD/EDA_CR.pdf
- [15] LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad* [online]. [cit. 2015-05-12]. Dostupné z: http://homel.vsb.cz/~lit40/SMAD/Casove_rady.pdf
- [16] LUKÁČIK, Martin a Viktor SLOSIAR. *Základy práce s ekonometrickým programem GRETL* [online]. [cit. 2015-09-15]. Dostupné z: <http://spu.fem.uniag.sk/cvicenia/ksov/obtulovic/EKONOMETRIA/GRETL%20popis.pdf>
- [17] *Metoda klouzavých průměrů* [online]. [cit. 2015-10-18]. Dostupné z: http://www.pf.jcu.cz/stru/katedry/m/petraskova/crek-prednaska_7.pdf
- [18] MÜLLER, Ivo. *Časové řady* [online]. [cit. 2015-09-12]. Dostupné z: http://disk.jabbim.cz/dl/1b769e4e9bc00e6b8920c0e493a1dc1e/565c6fb4/kalisek@jabber.cz/Matematika/Casove_rady/Casovky_projekt_final.pdf
- [19] ŘEZANKOVÁ, H., L. MAREK, a M. VRABEC. *IASTAT – Interaktivní učebnice statistiky* [online]. [cit. 2015-09-18]. Dostupné z: <http://iastat.vse.cz/>
- [20] SEBERA, Martin. *Časové řady v kinantropologickém výzkumu* [online]. [cit. 2015-09-28]. Dostupné z: http://is.muni.cz/do/rect/habilitace/1451/33088294/33088307/Habilitacni_prace_Sebera.pdf
- [21] *Time Series Data – Food and Beverage Stores* [online]. [cit. 2015-10-11]. Dostupné z: <https://www.census.gov/retail/marts/www/timeseries.html>
- [22] *Vyrovnaní počátečních a koncových hodnot řady a předpovědi* [online]. [cit. 2015-10-18]. Dostupné z: http://www.pf.jcu.cz/stru/katedry/m/petraskova/crek-prednaska_8.pdf
- [23] ZDRAŽIL, Tomáš. *Analýza ekonomických časových řad (s využitím systému STATISTICA)* [online]. [cit. 2015-09-12]. Dostupné z: http://is.muni.cz/th/150645/prif_m/diplomka.pdf

Přílohy

Příloha 1 - Data „Food & beverage stores“

FOOD & BEVERAGE STORES												
YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1992	30437	30382	30526	30602	30718	30761	30963	31041	30942	31148	30992	31057
1993	31045	31233	31058	31129	31164	31168	31349	31126	31201	31346	31363	31499
1994	31468	31688	31700	31899	31858	32123	31964	32222	32401	32196	32370	32294
1995	32625	32399	32415	32312	32540	32597	32466	32547	32625	32578	32744	32903
1996	33085	32928	33076	33174	33178	33281	33470	33542	33494	33704	33734	33819
1997	34049	33989	34156	34056	33978	33892	34106	34256	34281	34348	34230	34190
1998	34126	34102	34275	34374	34534	34592	34856	34950	34916	35093	35140	35533
1999	35447	35707	35652	36013	36060	35872	36126	36117	36411	36271	36300	37665
2000	35613	35980	36559	37008	36717	37175	37038	37110	37308	37476	37700	38033
2001	37903	38252	38133	38369	38439	38569	38506	38738	38832	38944	39161	38996
2002	38833	38885	38913	38568	38713	38723	38834	38647	38599	38693	38886	39072
2003	39412	39285	39318	39441	39019	39480	39808	39833	39878	39776	39631	39809
2004	40443	39922	40419	40399	40635	40608	40631	40771	41045	41177	41404	41534
2005	41766	41755	41769	42237	42127	42284	42252	42580	42725	42837	42906	43078
2006	42849	43374	43224	43272	43657	43738	43828	44122	43795	44633	44489	44804
2007	44860	45270	45177	45078	45417	45576	45847	45557	45937	46161	46576	46873
2008	46834	46761	46898	47428	47247	47691	47795	47973	47873	47691	47446	47064
2009	47505	46975	47104	47012	47385	47414	47252	47447	47631	47573	47845	48239
2010	47943	48780	48397	48152	48281	48089	47821	48538	48634	48797	49242	49233
2011	49740	50123	50184	50764	50681	50954	51087	51332	51041	51475	51571	51159
2012	51730	51728	52080	52345	52377	52306	52516	52398	52778	53071	52893	52890
2013	52987	53558	53528	52978	53519	53542	53777	53721	54007	53921	53769	54696
2014	54466	54343	54518	54707	54732	55391	55298	55505	55763	55848	56030	56322
2015	56228	56217	56630	56433	56751	56802	56829	57041	56861			

Příloha 2 – Roční údaje o počtu živě narozených dětí

ROK	Ziv_nar_deti	ROK	Ziv_nar_deti	ROK	Ziv_nar_deti
1920	244668	1953	172547	1986	133356
1921	257281	1954	168402	1987	130921
1922	248728	1955	165874	1988	132667
1923	241230	1956	162509	1989	128356
1924	228894	1957	155429	1990	130564
1925	225555	1958	141762	1991	129354
1926	219802	1959	128982	1992	121705
1927	208711	1960	128879	1993	121025
1928	208942	1961	131019	1994	106579
1929	203064	1962	133557	1995	96097
1930	207224	1963	148840	1996	90446
1931	196214	1964	154420	1997	90657
1932	190397	1965	147438	1998	90535
1933	176201	1966	141162	1999	89471
1934	171042	1967	138448	2000	90910
1935	170052	1968	137437	2001	90715
1936	169124	1969	143165	2002	92786
1937	170251	1970	147865	2003	93685
1938	185623	1971	154180	2004	97664
1939	192344	1972	163661	2005	102211
1940	209432	1973	181750	2006	105831
1941	208913	1974	188015	2007	114632
1942	215259	1975	190776	2008	119570
1943	225379	1976	187378	2009	118348
1944	230183	1977	181763	2010	117153
1945	225025	1978	178901	2011	108673
1946	210454	1979	172112	2012	108576
1947	206745	1980	153801	2013	106751
1948	197837	1981	144438		
1949	185484	1982	141738		
1950	188341	1983	137431		
1951	185570	1984	136941		
1952	180143	1985	135881		