

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra ekonomiky, manažerství a humanitních věd

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Leoš Klik

Studijní program: elektrotechnika, energetika a management
Obor: ekonomika a řízení elektrotechniky

Název tématu: Analýza telekomunikačního odvětví ČR z pohledu investic

Pokyny pro vypracování:

- vývoj telekomunikací a HDP v ČR
- statistická analýza časové řady investic sektoru telekomunikací
- krátkodobé prognózy vývoje investic do odvětví
- ekonometrický model souvislostí HDP a investic
- závěry, přínosy a hodnocení

Seznam odborné literatury:

Kaňok, J.: Statistické metody v managementu, Praha, Vydavatelství ČVUT, 2005
Hušek, R.: Ekonometrická analýza, vyd. Oeconomica, 2007, ISBN 8024513005

Vedoucí diplomové práce: Ing. Miroslav Vitek, CSc. – ČVUT FEL, K 13116

Platnost zadání: do konce letního semestru akademického roku 2015/2016
L.S.

Doc.Ing. Jaroslav Knápek, CSc.
vedoucí katedry

Prof.Ing. Pavel Ripka, CSc.
děkan

V Praze dne 28.1. 2015

České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra ekonomiky, manažerství a humanitních věd



Diplomová práce

**Analýza telekomunikačního odvětví České republiky z pohledu
investic**

Ing. Leoš Klik

Vedoucí práce: Ing. Miroslav Vitek, CSc.

Studijní program: Elektrotechnika, energetika a management, Navazující magisterský

Obor: Ekonomika a řízení elektrotechniky

10. května 2015

Poděkování

Především děkuji vedoucímu své diplomové práce panu Ing. Miroslavu Vítkovi, CSc. za cenné rady, připomínky a za čas strávený konzultacemi k této práci. Rovněž děkuji Mgr. Marii Klikové za jazykovou korekturu práce.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Lubné dne 9. 5. 2015

.....

Abstract

The thesis deals with an analysis of the telecommunications industry in the Czech Republic from the perspective of investment. It describes the general characteristics and the amount of investment in the telecommunications industry in the last fifteen years. It also deals with the development of gross domestic product as a macroeconomic indicator of the country progression. It tries to find a suitable model of the behavior of these economic indicators and their mutual relationship by using time series analysis and econometric models construction. The last part of the work is focused on testing the suitability of this model and forecasting future values of these economic time series.

Abstrakt

Cílem této diplomové práce je analyzovat telekomunikační sektor České republiky z pohledu investic. Práce popisuje obecné charakteristiky a výši investic do telekomunikačního sektoru v posledních patnácti letech. Dále se zabývá vývojem hrubého domácího produktu, jakožto makroekonomického ukazatele vývoje země, a snaží se nalézt za pomoci nástrojů pro analýzu časových řad vhodný model chování těchto ekonomických ukazatelů a jejich vzájemného vztahu. V další části pak testuje vhodnost tohoto modelu a zabývá se kvalitou prognózování budoucích hodnot těchto ekonomických časových řad.

Obsah

1	Úvod	1
2	Telekomunikační sektor v České republice	3
2.1	Charakteristika odvětví telekomunikací	3
2.1.1	Účastník veřejné telefonní sítě	3
2.1.2	Typy pevných telefonních linek	4
2.1.3	Vývoj počtu pevných telefonních linek a VoIP	4
2.1.4	Aktivní SIM karty pro mobilní telefonní síť	4
2.1.5	Vývoj počtu aktivních SIM karet	5
2.1.6	Srovnání vývoje počtu pevných telefonních linek a aktivních SIM karet	6
2.2	Ekonomická struktura telekomunikačního sektoru	7
2.2.1	Základní pojmy	7
2.2.2	Struktura sektoru z pohledu subjektů	8
2.2.3	Investice a výdaje ICT v posledních 15 letech	9
2.2.3.1	Výdaje za ICT	9
2.2.3.2	Investice v ICT	9
3	Hrubý domácí produkt	11
3.1	Charakteristika HDP	11
3.1.1	Tržní ceny	11
3.1.2	Stínová ekonomika	11
3.1.3	Období výpočtu HDP	12
3.1.4	Území daného státu	12
3.1.5	Reálný a nominální HDP	12
3.2	Metodiky výpočtu HDP	13
3.2.1	Produkční metoda	13
3.2.2	Výdajová metoda	13
3.2.3	Důchodová metoda	13
3.3	Vývoj HDP v České republice za posledních 15 let	14
3.3.1	HDP na obyvatele	15
4	Analýza časových řad	16
4.1	Časové řady obecně	16
4.1.1	Časová proměnná řady	16
4.2	Metody analýzy časových řad	17

4.2.1	Grafické zobrazení časové řady	17
4.2.2	Popisné charakteristiky časových řad	18
4.2.2.1	Charakteristiky polohy	18
4.2.2.2	Charakteristiky variability	19
4.2.2.3	Charakteristiky dynamiky	19
4.2.2.4	Korelace	20
4.2.3	Dekompozice časových řad	20
4.2.3.1	Trend časové řady	20
4.2.3.2	Sezónní složka	20
4.2.3.3	Cyklická složka	20
4.2.3.4	Náhodná složka	21
4.2.3.5	Typy dekompozice časové řady	21
4.3	Analýza složek časové řady	21
4.3.1	Analýza trendové složky	21
4.3.1.1	Modelování trendu analytickými funkcemi	22
4.3.1.2	Metoda klouzavých průměrů	24
4.3.1.3	Exponenciální vyrovnání	25
4.3.2	Analýza sezónní složky	27
4.3.2.1	Očištění od sezónní složky metodou klouzavých průměrů	28
4.3.3	Analýza cyklické složky	29
4.4	Kvalita vyrovnání a vhodnost modelu	30
4.4.1	Kvalita prognózy	31
5	Ekonometrické modely	33
5.1	Ekonometrický model obecně	33
5.1.1	Formulace a specifikace modelu	33
5.1.2	Kvantifikace modelu	35
5.1.3	Verifikace modelu	36
5.1.4	Aplikace modelu	37
5.1.5	Data vstupující do ekonometrického modelu	37
5.2	Konstrukce jednorovnicového modelu	37
5.2.1	Metoda nejmenších čtverců	40
6	Analýza časových řad prakticky	42
6.1	Analýza časové řady HDP	42
6.1.1	Získání vstupních dat	42
6.1.2	Základní parametry časové řady	42
6.1.3	Dekompozice časové řady HDP	43
6.1.3.1	Trendová složka	43
6.1.3.2	Sezónní složka	43
6.1.3.3	Náhodná a cyklická složka	44
6.1.4	Tvorba modelu	46
6.1.4.1	Modelování trendu přímkou	46
6.1.4.2	Modelování trendu exponenciálou	48
6.1.4.3	Modelování trendu dvojitým exponenciálním vyrovnáním	50
6.1.4.4	Porovnání kvality modelů	51

6.1.5	Dekompozice v softwaru R	51
6.1.6	Tvorba modelu v R	52
6.1.6.1	Modelování trendu dvojitým exponenciálním vyrováním	53
6.1.6.2	Modelování trendu trojitým exponenciálním vyrováním	54
6.1.6.3	Porovnání kvality modelů	55
6.2	Analýza časové řady investic	56
6.2.1	Získání vstupních dat	56
6.2.2	Základní parametry časové řady	56
6.2.3	Tvorba modelu	57
6.2.3.1	Modelování trendu přímkou	57
6.2.3.2	Modelování trendu exponenciálou	59
6.2.3.3	Modelování trendu dvojitým exponenciálním vyrováním	60
6.2.3.4	Porovnání kvality modelů	62
6.2.4	Tvorba modelu v R	62
6.2.4.1	Modelování trendu dvojitým exponenciálním vyrováním	62
7	Krátkodobé prognózy budoucího vývoje	65
7.1	Krátkodobé prognózy vývoje HDP	65
7.1.1	Prognóza na základě ručního výpočtu v Excelu	65
7.1.2	Kvalita prognózy	65
7.1.3	Prognóza v R	67
7.1.4	Kvalita prognózy	67
7.2	Krátkodobé prognózy vývoje investic	68
7.2.1	Prognóza na základě ručního výpočtu v Excelu	68
7.2.2	Kvalita prognózy	68
7.2.3	Prognóza v R	69
7.2.4	Kvalita prognózy	70
8	Ekonometrický model souvislostí HDP a investic	71
8.1	Specifikace modelu	71
8.2	Kvantifikace modelu	72
8.3	Verifikace modelu	73
8.3.1	Test heteroskedasticity	73
8.3.2	Test autokorelace	74
8.3.3	Test multikolinearity	74
8.3.4	Výsledek verifikace modelu	75
8.4	Aplikace modelu	75
9	Závěr	78
A	Seznam zkratk	82
B	Vstupní data - HDP	83
C	Vstupní data - výše investic	85
D	Vstupní data - model	86

OBSAH

ix

E Obsah přiloženého CD

87

Seznam obrázků

2.1	Vývoj počtu pevných linek a VoIP [21]	5
2.2	Vývoj počtu aktivních SIM karet [22]	6
2.3	Srovnání počtu pevných linek a SIM karet	6
2.4	Struktura telekomunikačního sektoru	8
2.5	Výdaje za ICT [19]	9
2.6	Výše investic v ICT sektoru [19]	10
3.1	Vývoj HDP v ČR v cenách roku 2000 [20]	14
3.2	Struktura HDP v ČR v cenách roku 2000 [20]	15
3.3	Porovnání HDP na obyvatele v ČR s okolními státy [8]	15
4.1	Vývoj počtu aktivních SIM karet na 100 obyvatel [22]	18
4.2	Průběhy trendových funkcí	23
6.1	Časová řada kvartálního vývoje HDP [?]	43
6.2	Trendová složka časové řady HDP	44
6.3	Sezónní složka časové řady HDP	45
6.4	Časová řada HDP očištěná o sezónní složku	45
6.5	Cyklická a náhodná složka časové řady HDP	46
6.6	Model HDP vytvořený regresní přímkou	47
6.7	Model HDP vytvořený regresní exponenciálou	49
6.8	Model HDP vytvořený dvojitým exponenciálním vyrovnáním	51
6.9	Časová řada HDP v R	52
6.10	Dekompozice časové řady HDP v R	53
6.11	Model HDP - dvojitě exponenciální vyrovnání v R	54
6.12	Model HDP - trojitě exponenciální vyrovnání v R	55
6.13	Časová řada výše investic do telekomunikačního vybavení[?]	56
6.14	Model investic vytvořený regresní přímkou	58
6.15	Model investic vytvořený regresní exponenciálou	60
6.16	Model investic vytvořený dvojitým exponenciálním vyrovnáním	61
6.17	Výše investic do telekomunikačního vybavení	63
6.18	Model investic - dvojitě exponenciální vyrovnání	63
7.1	Prognóza budoucí hodnoty HDP	66
7.2	Prognóza budoucí hodnoty HDP v R	68
7.3	Predikce budoucí výše investic	69

7.4	Prognóza budoucí výše investic v R	70
8.1	Model HDP a investic 1 - významnost proměnných	72
8.2	Model HDP a investic 2 - významnost proměnných	73
8.3	Model HDP a investic 3 - významnost proměnných	73
8.4	Aplikace modelu - předpověď	76
E.1	Obsah příloženého CD	87

Seznam tabulek

4.1	Vývoj počtu aktivních SIM karet na 100 obyvatel [22]	18
6.1	Popisná statistika časové řady HDP	44
6.2	Vstupní data pro regresní analýzu - přímka	47
6.3	Vstupní data pro regresní analýzu - exponenciála	48
6.4	Porovnání kvality modelů HDP	51
6.5	Porovnání kvality modelů HDP - R	55
6.6	Popisná statistika časové řady investic	57
6.7	Vstupní data pro regresní analýzu - přímka	58
6.8	Vstupní data pro regresní analýzu - exponenciála	59
6.9	Porovnání kvality modelů investic	62
7.1	Prognóza vývoje HDP	66
7.2	Prognóza budoucí hodnoty HDP v R	67
7.3	Prognóza vývoje investic	68
7.4	Prognóza budoucí hodnoty investic v R	70
8.1	Proměnné modelu	71
8.2	Hodnoty vypočtené modelem	77
8.3	Předpověď na základě modelu	77
B.1	Neočištěné hodnoty HDP	84
C.1	Výše investic do telekomunikačního vybavení	85
D.1	Vstupní data ekonometrického modelu	86

Kapitola 1

Úvod

Komunikace je v dnešní době jedním z nejdůležitějších prvků, na kterém je založena existence lidské společnosti. Jak již napovídá latinský původ slova *communicare* (sdílet, radit se), podstatou komunikace je v první řadě přenos a sdílení informací. Bez ohledu na komunikační prostředky, které mohou být velmi různorodé, půjde při komunikaci vždy právě o přenášenou informaci. Historický vývoj komunikace souvisí rozhodně se strukturou a rozvojem společnosti. Prvotní potřeby dorozumívání byly uspokojeny různými formami neverbální komunikace, postupným vývojem pak komunikací verbální, tedy slovem nebo písmem.

Tento druh předávání informací má samozřejmě své výhody i nevýhody. V mnoha situacích jistě není nad osobní kontakt a přímý rozhovor. Problémem pak může být nesnadnost realizovat právě onen osobní kontakt. V případě komunikace písmem nebývá největším problémem samotná realizace přenosu informace, ale jeho rychlost, která zejména v minulosti velmi omezovala tento způsob komunikace.

V dnešní době, kdy patří informace k nejcennějšímu a také nejžádanějšímu zboží napříč všemi odvětvími obchodu, ekonomiky i společnosti, je požadavek na její velice rychlý přenos a široké možnosti šíření kritický. Technologická vyspělost nynější společnosti také rozhodně není překážkou pro realizaci takového způsobu komunikace. Logickým vyústěním těchto faktorů je vznik odvětví telekomunikací. Z řeckého slova *tele* (vzdálený) lze odvodit, že telekomunikace je zaměřena především na předávání informací, tedy komunikaci, na dálku.

Význam tohoto odvětví byl od počátku obrovský. Prvotní spoje realizované telegrafy a telefony přinesly nevídané možnosti komunikace na velké vzdálenosti s minimálním zpožděním. Také rozvoj hromadných komunikačních prostředků jako rozhlasu a televize byl nezanedbatelným milníkem v historii. V dnešní situaci masivního rozšíření počítačů získává pojem telekomunikace další rozměr díky existenci počítačových sítí a internetu, které jsou snadno dostupné médium pro přenos informací k jedinci i masám.

Samotný telekomunikační sektor, respektive telekomunikační činnost, je dnes podle klasifikace Organisation for Economic Co-operation and Development (OECD) zařazen v tzv. Information and Communication Technologies (ICT) sektoru. Tento sektor je definován jako kombinace činností produkujících výrobky a poskytujících služby, jež jsou primárně určeny ke zpracování, komunikaci a distribuci informací elektronickou cestou, včetně jejich zachycení, ukládání, přenosu a zobrazení [9]. Samotná telekomunikační činnost pak v rámci tohoto

zařazení pokrývá poskytování telekomunikačních služeb, mezi které patří přenos hlasu, dat, textu, obrazu a zvuku. Současně sem patří také poskytování přístupu k internetu realizované provozovateli telekomunikačních sítí.

Téma této diplomové práce nepokrývá samozřejmě celou historii vývoje telekomunikací. Na následujících stránkách se budu zabývat rozvojem a analýzou telekomunikačního odvětví v horizontu posledních 15 let, a to především na území České republiky. I v tomto časovém horizontu prošly telekomunikace obrovskými změnami. Především v tom, že se stále více a více rozpadá tradiční monopolní struktura trhu telekomunikací a na trh vstupují stále noví a noví poskytovatelé telekomunikačních služeb. Telekomunikační trh tak patří v současnosti mezi nejrychleji se rozvíjející odvětví českého i světového hospodářství.

Tento nutný rozvoj, který vede především k rozšiřování a zkvalitňování služeb, samozřejmě úzce souvisí s technickou stránkou věci. Pro dosažení kvalitního zázemí, které uspokojuje poptávku v telekomunikačním odvětví, tak stále roste význam investic vkládaných právě do telekomunikačních prostředků. A právě analýzou telekomunikačního odvětví v České republice z pohledu investic se budu zabývat v této diplomové práci.

Kapitola 2

Telekomunikační sektor v České republice

2.1 Charakteristika odvětví telekomunikací

Vedle internetové infrastruktury je v dnešní době telekomunikační infrastruktura základním prvkem převážné části informačních technologií a vůbec moderní společnosti s potřebou komunikovat. Na následujících stránkách jsou vysvětleny některé základní pojmy související s telekomunikační infrastrukturou. Dále uvádím stručné zhodnocení a grafickou reprezentaci dat charakterizujících telekomunikační sektor v České republice. Podkladová data jsou získána především z databází, které provozuje Český statistický úřad (ČSÚ) a Český telekomunikační úřad (ČTÚ).

2.1.1 Účastník veřejné telefonní sítě

Účastníka veřejné telefonní sítě můžeme definovat jako fyzickou nebo právnickou osobu, která je připojena k veřejné telefonní síti na základě smlouvy s poskytovatelem přístupu k této síti [14]. Soubor účastníků veřejné telefonní sítě tedy zahrnuje všechny hlavní telefonní stanice pro veřejnou pevnou telefonní síť a veškeré aktivní Subscriber Identity Module (SIM) karty pro mobilní telefonní síť. V dnešní době musíme do tohoto odvětví zahrnout i účastnické stanice Voice over Internet Protocol (VoIP) (stanice pro IP telefonii), protože se jedná o službu, která umožňuje přenos hlasu po datových sítích prostřednictvím protokolu Internet Protocol (IP). Jedná se tak o alternativu k hlasovým službám provozovaným prostřednictvím klasické telefonní sítě.

V případě hlavních telefonních stanic se jedná o takové koncové body sítě, které mají jednoznačně určené ústřednové zakončení. Tedy například více telefonů v domácnosti nebo podnikové telefony za podnikovou telefonní ústřednou s jedním veřejným telefonním číslem jsou považovány za jednoho účastníka veřejné telefonní sítě. Dále je jisté, že jeden člověk může používat více linek s veřejným telefonním číslem, případně více aktivních SIM karet, proto počet účastníků veřejné telefonní sítě může převyšovat počet osob na zkoumaném území.

2.1.2 Typy pevných telefonních linek

Celkový soubor telefonních linek se skládá ze čtyř následujících typů linek.

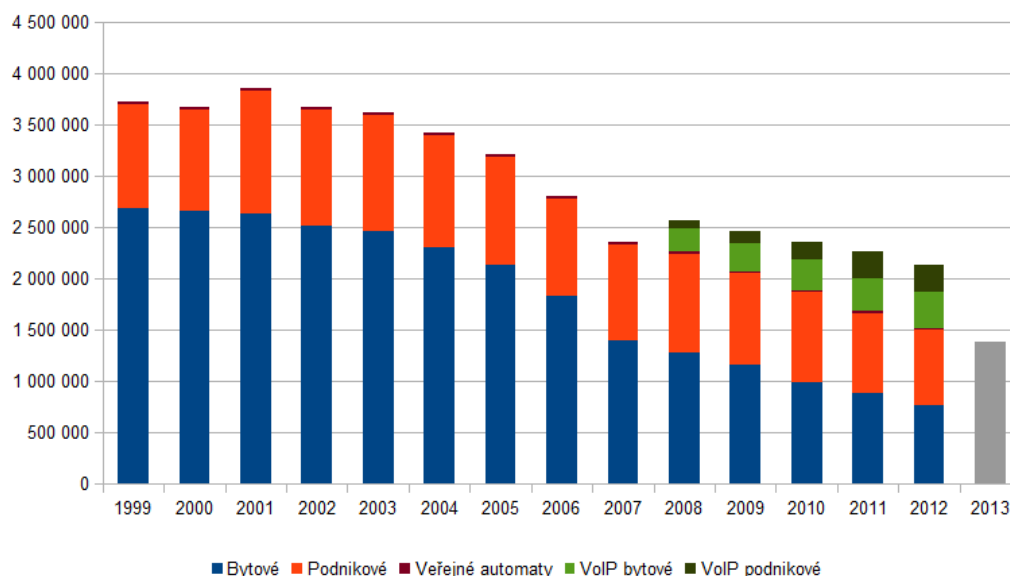
- Bytové linky zahrnují telefonní stanice nebo přípojné body veřejné telekomunikační sítě, které jsou určeny pro soukromé potřeby fyzických osob, nikoliv pro podnikání. Patří sem, jak již název napovídá, převážně linky v domácnostech a nepodnikových prostorech.
- Podnikové linky, které zahrnují telefonní stanice nebo přípojné body určené pro podnikatelské účely fyzických nebo právnických osob. Jsou určeny jednak pro podnikatelskou a jinou výdělečnou činnost a také pro potřeby státních institucí, spolků, sdružení a organizací.
- Veřejné telefonní automaty, hovorny a veřejné stanice, například na poštách.
- Nově zařazujeme mezi pevné telefonní linky také linky VoIP pro IP telefonii. Jedná se o linky přenášené prostřednictvím datových sítí s možností přenosu nejen hlasu, ale také dat a videa.

2.1.3 Vývoj počtu pevných telefonních linek a VoIP

Graf 2.1 zobrazuje vývoj počtu pevných linek od roku 1999, od roku 2008 pak i VoIP stanic. Údaje jsou získány z databáze ČSÚ [21], pro rok 2013 nebyla data dosud zahrnuta do statistik, celkový počet pevných linek bez ohledu na typ a bez zahrnutí IP telefonie vychází z finančních výsledků Telefónica Czech Republic, a.s. za rok 2013. Jak je z grafu patrné, počet podnikových pevných linek klesá zatím minimálně, naproti tomu počet pevných linek v domácnostech vykazuje v posledních letech značný meziroční pokles, a to zhruba o 6 – 7%. Vzhledem k neustálému rozvoji mobilních sítí a narůstajícímu počtu aktivních SIM karet, jak bude uvedeno dále, je logickým vyústěním pokles potřeby veřejných telefonních automatů. Jejich počet klesl od roku 2000 na téměř polovinu a v současné době je jich v provozu asi 15 000. Podle údajů ČSÚ je zhruba jedna třetina automatů nevýdělečná a je dotována z Fondu univerzální služby [13].

2.1.4 Aktivní SIM karty pro mobilní telefonní síť

Počet aktivních SIM karet samozřejmě nekoresponduje s celkovým počtem uživatelů mobilního telefonu. Počet uživatelů je zjišťován prostřednictvím výběrového šetření mezi jednotlivci od 16 let, kteří mají přístup k mobilnímu telefonu na rozdíl od údaje o počtu SIM karet, který vychází z dat jednotlivých operátorů. SIM karty pro mobilní telefonní síť dělíme na karty předplacené a smluvní. U karet předplacených je nutné, aby zákazník nejprve zaplatil určitou částku (kredit), kterou postupně čerpá ať už voláním, psáním SMS nebo využíváním mobilních dat. Všechny služby tak má zákazník zaplacený předem. Naproti tomu u SIM karet smluvních musí mít zákazník s operátorem mobilní sítě uzavřenou smlouvu a na jejím základě platí paušální poplatek, obvykle měsíční. Všechny využití služby mobilní sítě jsou mu tak vyúčtovány zpětně [14].

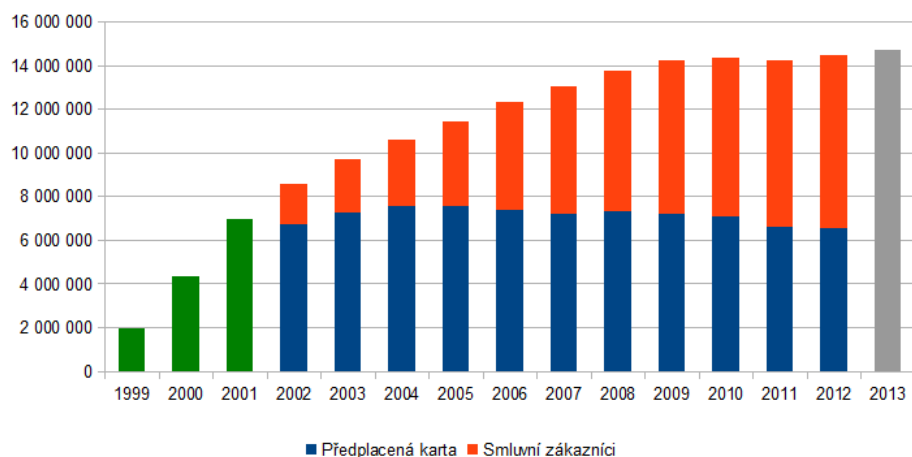


Obrázek 2.1: Vývoj počtu pevných linek a VoIP [21]

2.1.5 Vývoj počtu aktivních SIM karet

Ve sledovaném období došlo k rapidnímu nárůstu počtu SIM karet. V roce 1999 bylo v České republice aktivních 1,94 milionu SIM karet, v roce 2012 pak více než 14,4 milionu, což dělá po přepočtení na počet obyvatel 138 SIM karet na 100 obyvatel. Největší nárůst je možné pozorovat v letech 1999 – 2002, kdy se zvýšil počet během čtyř let přibližně čtyřnásobně. Tento trend je dán především tím, že v počátcích mobilní komunikace připadal většinou jeden mobilní telefon a tedy i SIM karta na jednoho uživatele. S postupným rozšiřováním mobilních služeb ale dochází k tomu, že jeden uživatel používá i několik SIM karet najednou. Důvodem může být výhodnost různých služeb od různých operátorů, další SIM karta na datové přenosy pro mobilní zařízení, služební telefony paralelně se soukromými a v poslední době též různé vestavné zařízení jako řídicí systémy domácností, alarmy a sledovací zařízení vozidel, které všechny využívají přístupu do sítě Groupe Spécial Mobile (GSM).

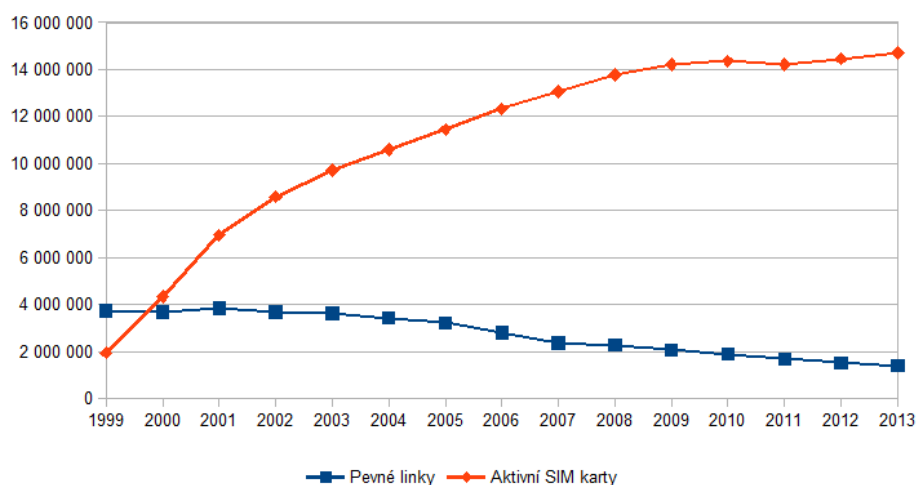
Graf 2.2 zachycuje vývoj počtu aktivních SIM karet i podle jejich typu. V letech 1999 – 2001 nejsou tyto údaje dostupné v rozložené formě, v grafu je tak zachycen pouze celkový počet SIM karet. Hodnota roku 2013 nebyla dosud Českým statistickým úřadem ani Českým telekomunikačním úřadem zveřejněna, vychází tedy pouze z očekávaného meziročního růstu. Hodnoty v grafu jsou kvůli konzistenci zobrazení získána z výpočtů podle klasické metodologie ČTÚ. Od roku 2009 byla zavedena nová metodologie počítání aktivních SIM karet na základě doporučení Evropské unie, pro předešlé roky však tento údaj nelze získat.



Obrázek 2.2: Vývoj počtu aktivních SIM karet [22]

2.1.6 Srovnání vývoje počtu pevných telefonních linek a aktivních SIM karet

Zajímavým srovnáním, které odráží moderní trend vývoje komunikační společnosti, může být také porovnání vývoje počtu SIM karet a pevných linek. Zlomovým okamžikem pro mobilní technologie se zdá být rok 2000, kdy počet aktivních SIM karet překročil počet pevných linek. Do roku 2009 pak můžeme charakterizovat vývoj jako prudce rostoucí, v následujících letech již není nárůst tak výrazný, což je podle mého názoru dáno nasycením trhu. V tomto odstavci popsaný trend je zřejmý z grafu 2.3.



Obrázek 2.3: Srovnání počtu pevných linek a SIM karet

2.2 Ekonomická struktura telekomunikačního sektoru

Abychom mohli alespoň stručně charakterizovat ekonomickou strukturu telekomunikačního sektoru v České republice, musíme definovat některé základní pojmy. Při definici těchto pojmů vycházím především z terminologie zavedené Českým statistickým úřadem, především proto, že právě odtud čerpám data pro navazující analýzu.

2.2.1 Základní pojmy

Informační a komunikační technologie, ICT

ICT definujeme jako veškeré technologie a produkty, jejichž primární funkcí je umožnění komunikace nebo přenosu informací elektronickou cestou.

Výdaje v ICT

Výdaje jsou finanční prostředky vložené do urychlení rozšiřování ICT a rovněž přispívají k růstu HDP.

Investice

Investicí rozumíme vklad vložený do dlouhodobých statků, který umožní budoucí zvýšení výnosů, ale nemusí přinášet okamžitý prospěch. Též lze říci, že napomáhají k investičnímu prohlubování kapitálu [2]. V kontextu užití Českého statistického úřadu budeme pod pojmem investice rozumět tvorbu hrubého fixního kapitálu, právě takto jsou investice kalkulovány v databázi ČSÚ.

Tvorba hrubého fixního kapitálu se skládá podle metodiky ČSÚ z *pořízení mínus úbytků fixních aktiv u rezidentských výrobců v průběhu daného období plus určité zvýšení k hodnotě nevyráběných aktiv, které vzniká v důsledku produkční činnosti výrobců nebo institucionálních jednotek* [16]. Fixními aktivy potom budeme rozumět taková hmotná nebo nehmotná aktiva, která jsou výstupem výrobního procesu a používají se opakovaně po dobu delší než jeden rok.

ICT produkty a služby

ICT produkty a služby jsou zahrnuty ve statistice ročních národních účtů, které poskytují data o výdajích a investicích v oblasti ICT. Vymezení tohoto sektoru je založeno na Klasifikaci produkce podle činností (CZ-CPA) a obsahuje následující položky [17].

- ICT vybavení
 - Počítače a periferní zařízení (CZ-CPA 262)
 - Komunikační zařízení (CZ-CPA 263)
- Telekomunikační služby
- IT služby

- Programování a poradenství (CZ-CPA 62)
- Software (CZ-CPA 582, CZ-CPA 62)

Pro účely této práce budou zajímavé především položky Komunikační zařízení a Telekomunikační služby. Ne vždy jsou ale statistická data dostupná a úplná a například pro investice do telekomunikačních služeb neuvádí Český statistický úřad relevantní data. Pro potřeby analýzy v této práci by bylo nejvhodnější hodnotit pouze telekomunikační sektor, ale díky výše uvedené použité metodice bude zapotřebí vzít v úvahu hodnocení celého ICT sektoru.

2.2.2 Struktura sektoru z pohledu subjektů

Celá informační společnost je složena z několika skupin, které tvoří strukturu ICT, posléze telekomunikačního sektoru. Tato struktura je zobrazena na obrázku 2.4 a zachycuje skupiny podílející se na existenci a tvorbě telekomunikačního trhu, návazně investic do něj. Uživatelé telekomunikační infrastruktury jsou samozřejmě stěžejním prvkem celého telekomunikačního ekosystému. Právě ti vytvářejí poptávku po službách a motivují tak ostatní skupiny k činnosti. Investoři pak zajišťují financování projektů s účelem zisku a jejich blízká vazba na operátory a výrobce je tímto zřejmá. Zejména operátoři mají přímý vliv na zisky investorů. Telekomunikační sektor byl v minulosti výhradně monopolním trhem, aby byla hospodářská soutěž i v této oblasti spravedlivá, je celý sektor regulován státem [18].

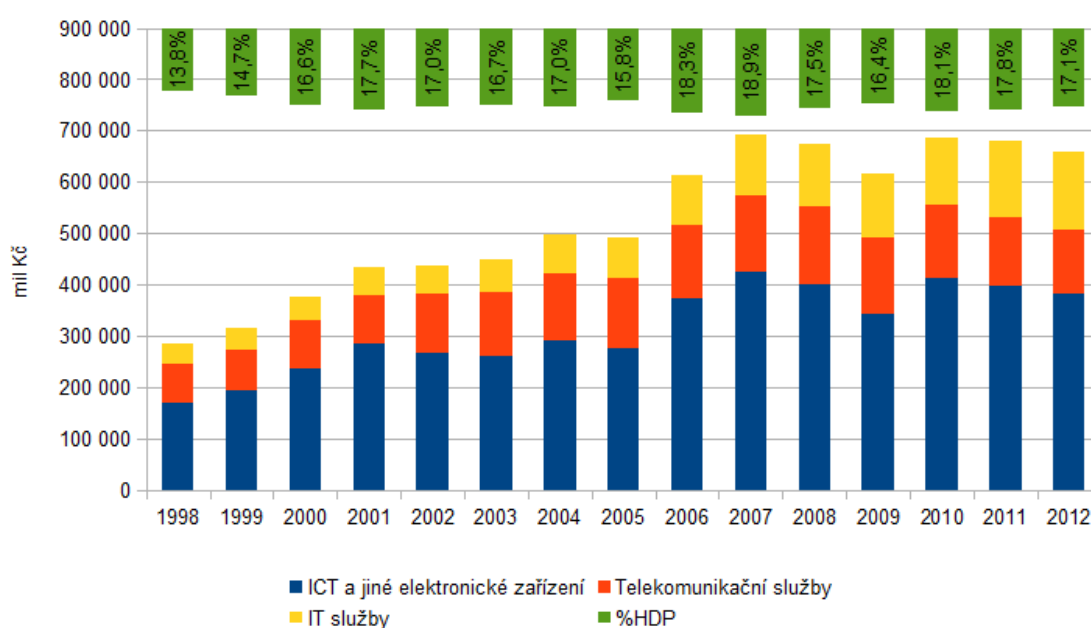


Obrázek 2.4: Struktura telekomunikačního sektoru

2.2.3 Investice a výdaje ICT v posledních 15 letech

2.2.3.1 Výdaje za ICT

Jak je patrné z grafu 2.5, v posledních patnácti letech mají hodnoty výdajů za ICT rostoucí trend, od roku 2007 pak přibližně konstantní. Graf zachycuje celou strukturu ICT sektoru tak, jak je uvedeno podle CZ-CPA výše. Na první pohled je vcelku patrné, že investice do IT služeb mají rostoucí tendenci. Telekomunikační služby mají naproti tomu z hlediska výdajů tendence konstantní až mírně klesající. Pro přibližnou představu graf obsahuje také procentuální podíl výdajů v ICT na tvorbu hrubého domácího produktu. Toto číslo není úplně malé a pohybuje se prakticky v celém sledovaném období nad 15%.



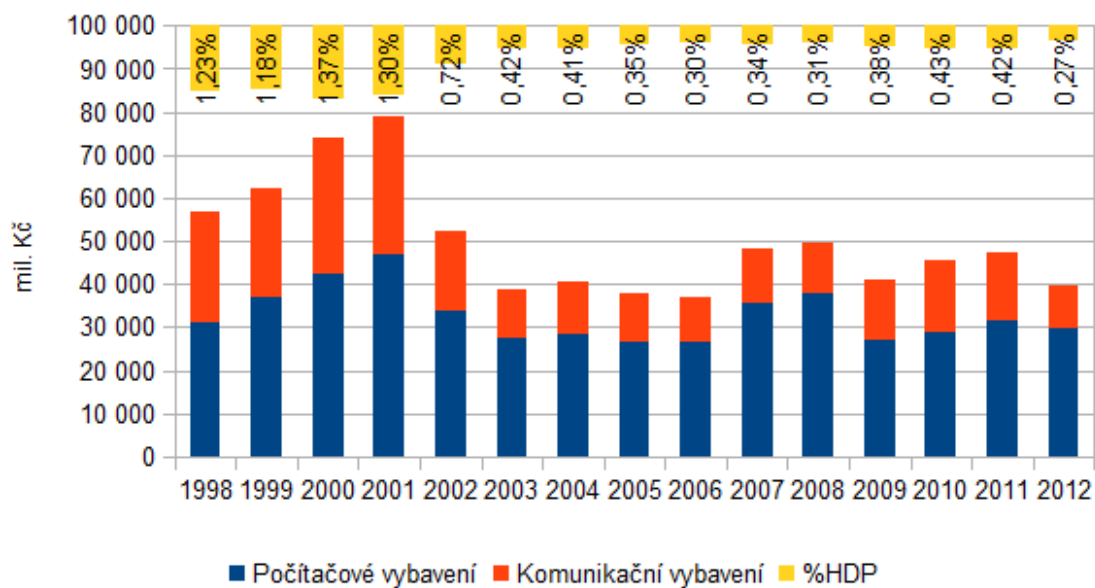
Obrázek 2.5: Výdaje za ICT [19]

Pro potřeby této další práce je podstatný zejména údaj o výdajích v telekomunikačních službách a ICT zařízeních. IT služby jsou nad rámec této práce a dále se jimi nebudeme zabývat. Pro přesnější představu ještě rozebereme strukturu výdajů právě v oddíle ICT a jiné elektronické zařízení, kde se dostaneme k výši investic do komunikačního vybavení.

2.2.3.2 Investice v ICT

Na základě grafu 2.6 lze konstatovat, že markantní meziroční snížení investic o více než 40% do komunikačního vybavení nastalo v roce 2002 a od této doby má výše investic přibližně konstantní trend. Domnívám se, že tento zvrát vysvětluje vyvrcholení budování infrastruktury pro mobilní komunikační sítě v rámci konkurenčního boje, protože právě v roce 2000 došlo ke vstupu nového operátora Vodafone na český trh. V horní části grafu je

opět pro představu uveden podíl investic, tentokrát již pouze investic do komunikačního vybavení, na tvorbě hrubého domácího produktu [19].



Obrázek 2.6: Výše investic v ICT sektoru [19]

Kapitola 3

Hrubý domácí produkt

3.1 Charakteristika HDP

Hrubý domácí produkt (HDP) je jedním ze základních makroekonomických ukazatelů, pomocí kterého jsme schopni výkonnost celé ekonomiky zhodnotit a porovnávat. Jedná se pravděpodobně o nejsledovanější ukazatel při hodnocení ekonomického růstu země nebo při porovnávání s ostatními zeměmi. Jeho velká obliba především v médiích vychází především z toho, že lze jedním číslem snadno ukázat ekonomický stav země. V praxi je ale důležité sledovat kromě samotné hodnoty HDP i její chování a vývoj, jako meziroční růst nebo pokles. Pro mezinárodní srovnání se též používá hodnota HDP na obyvatele, která kompenzuje různý počet obyvatel v porovnávaných zemích.

Definovat hrubý domácí produkt můžeme jako *peněžní vyjádření celkové hodnoty nově vytvořených statků a služeb v daném období na určitém území*[15]. Dále jsou uvedeny faktory, které HDP charakterizují nebo mají na vznik jeho hodnoty přímý vliv.

3.1.1 Tržní ceny

Hodnota služeb a statků započtených do HDP musí být stanovena v tržních cenách. To proto, aby bylo možné stanovit společnou hodnotu statků různého typu. V praxi jde především o započtení nepřímých daní různého typu podle charakteru statku nebo služby.

3.1.2 Stínová ekonomika

Dále je důležité zahrnout do HDP veškeré statky a služby vzniklé v daném období. V praxi je tento požadavek značným problémem ekonomických statistik všech států. Existují určité skupiny ekonomických vztahů, které se oficiálně do statistiky nedostanou a nelze je tak jednoduše do HDP zahrnout. Jedná se o statky a služby vzniklé produkcí domácností, šedou a černou ekonomikou. Společně se tyto ekonomické vztahy označují jako stínová ekonomika.

Produkce domácností

U statků a služeb, které se běžně neprodávají na trzích, je jejich hodnota určena empiricky na základě statistických odhadů, nebo se do HDP vůbec nepromítne. Jako příklad

lze uvést práce na vlastní zahradě, za kterou není nutné platit nájem. Je to bezesporu nově vytvořená služba, ale protože ji člověk poskytuje sám sobě, je obtížné ji do HDP zahrnout. Prostřednictvím čerpání takovýchto služeb, které se nedostanou na trh a nejsou pak ekonomicky zachytitelné, snižuje výsledné HDP země.

Šedá ekonomika

Negativně se tak na tvorbě HDP podílí především tzv. šedá ekonomika. Patří sem zejména činnosti, které jsou ve své podstatě legální, ale nezdaněné a tedy postižitelné. Jako příklad lze uvést nezdaněnou platbu v hotovosti bez vystaveného dokladu, práce načerno, ze které není odvedena daň z příjmu nebo prodej alkoholu a tabákových výrobků bez kolků. Za rok 2013 činí objem šedé ekonomiky cca 600 miliard Kč, což je zhruba polovina rozpočtu ČR a jde o 16 % HDP [11].

Černá ekonomika

Stejným způsobem se k HDP chová i tzv. černá ekonomika. Zde se jedná již o činnosti, které porušují zákony dané země, jako například prodej drog, prostituce, neoficiální hazard a další činnosti, které generují zisk nezákonným způsobem [12].

3.1.3 Období výpočtu HDP

Jako období výpočtu se obvykle stanovuje jeden rok, případně čtvrtletí. Do HDP za toto období jsou zahrnuty pouze služby a statky vyprodukované během tohoto období. Nelze zahrnout transakce vzniklé na trhu za již dříve vyrobené statky, došlo by tak k dvojímu započtení dané hodnoty a umělému navýšení HDP. Jako příklad může posloužit kniha vytištěná v minulém roce, která je již započtena do hodnoty vyprodukovaných statků daného roku. Její případný prodej v následujícím roce nemá na utváření hodnoty HDP roku dalšího žádný vliv.

3.1.4 Území daného státu

Do výpočtu HDP jsou započteny pouze statky a služby vyprodukované na území daného státu bez ohledu na to, kdo je vlastníkem výrobního nástroje. Například firma, jejíž produkty budou započteny do HDP, může mít sídlo a mít další produkci jak v dané zemi, tak mimo ni. Podstatné je, že do HDP budou započteny pouze a jen ty služby a statky, které vznikly na území daného státu.

3.1.5 Reálný a nominální HDP

Pokud chceme porovnávat hrubý domácí produkt v různých letech nebo obdobích, je potřeba vzít v úvahu časovou změnu hodnoty peněz, která je především v meziročním srovnání bezpochyby nezanedbatelná. Jedná se především o očištění hodnoty od inflace či deflace, protože právě ty mohou hodnotu HDP zkreslovat.

Nominální hrubý domácí produkt je ukazatel vyjádřený v běžných cenách. Běžné ceny jsou takové ceny, které jsou platné právě v tom období, kdy byla hodnota statků a služeb

do HDP započítána. Pro krátké časové období, jako hodnocení mezi kvartály jednoho roku, může být tato hodnota dostačující, pokud nedochází k extrémním změnám inflace či deflace.

Naproti tomu reálný hrubý domácí produkt je vyjádřen v cenách stálých. Stálé ceny mohou být stanoveny například k určitému základnímu roku, ostatní roky jsou pak vhodně přepočteny podle tohoto základu. Dosáhne se tak hodnoty HDP očištěné od vlivů inflace a deflace a změna HDP pak ukazuje skutečný nárůst objemu produkce statků a služeb v daném období. Přepočet na reálný HDP je tedy vhodný pro porovnávání v dlouhých časových intervalech, nebo v kratších obdobích, kdy dochází k velkým změnám hodnoty peněz [10].

3.2 Metodiky výpočtu HDP

Hrubý domácí produkt může být spočítán jednou ze tří metod. Jedná se o metodu produkční, výdajovou a důchodovou [15].

3.2.1 Produkční metoda

Produkční metoda vychází z faktu, že všechny statky a služby na trhu jsou produkovány výrobci. Každý výrobce na jednotlivém stupni výroby se podílí na výsledku nějakou přidanou hodnotou. Součtem všech těchto přidaných hodnot se pak dostaneme k hodnotě HDP. Dále bere výpočet v úvahu ještě mezipotřebu. Vzhledem k tomu, že produkce je oceňována v základních cenách a užití v kupních cenách, je celková produkce doplněna o daně z produktů snížené o dotace na výrobky. Výpočet HDP produkční metodou pak probíhá podle vzorce 3.1, kde PR je výše produkce, M mezipotřeba, DZP daně z produktů a DNP jsou dotace na produkty.

$$HDP = PR - M + DZP - DNP \quad (3.1)$$

3.2.2 Výdajová metoda

Výdajovou metodou můžeme vypočítat HDP tak, že sečteme všechny výdaje vynaložené na produkci všech statků a služeb. Celkové výdaje pak můžeme rozčlenit na výdaje domácností na spotřebu, výdaje na investice, výdaje vlády na nákupy výrobků a služeb a čistý export. Hodnota čistého exportu představuje celkový objem exportu po odečtení importu, který nezvyšuje HDP, protože byl vyroben v jiné zemi. Výpočet HDP výdajovou metodou probíhá podle vzorce 3.2, kde V je výše výdajů na konečnou spotřebu, THK tvorba hrubého kapitálu, E export výrobků a služeb a I je import výrobků a služeb.

$$HDP = V + THK + E - I \quad (3.2)$$

3.2.3 Důchodová metoda

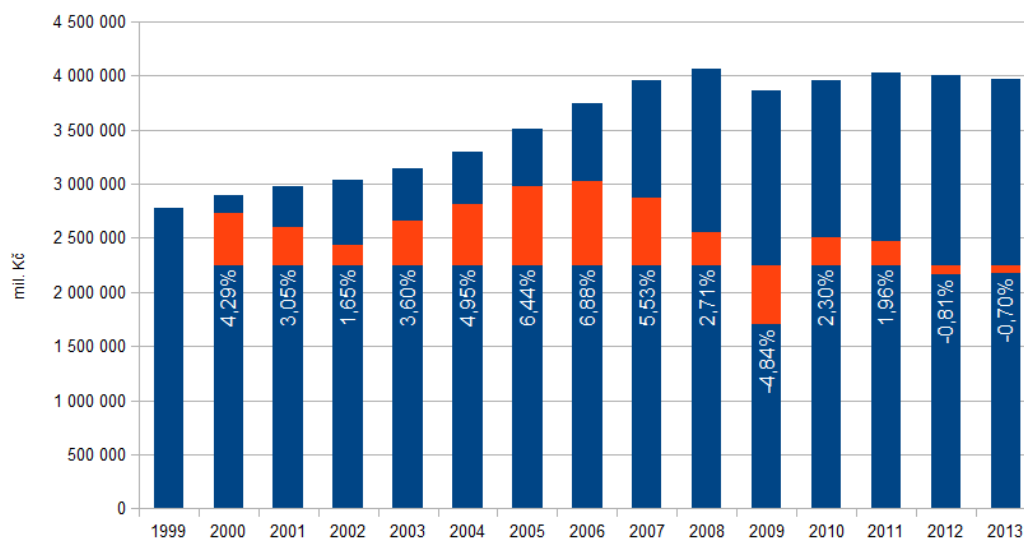
Výpočet důchodovou metodou vychází ze součtu prvotních důchodů v celém hospodářství. Vychází se z toho, že každý výdaj je zároveň něčím příjmem (důchodem). Jejich součtem tak získáme vyšší HDP. Jde o náhrady zaměstnancům, daně z výroby a z dovozu snížené o

dotace, čistý provozní přebytek, smíšený důchod a spotřebu fixního kapitálu. Výpočet HDP důchodovou metodou pak probíhá podle vzorce 3.3, kde N je výše náhrad zaměstnancům, DV daně z výroby a z dovozu, DO dotace, CPP je čistý provozní přebytek, CSD čistý smíšený důchod a S je spotřeba fixního kapitálu.

$$HDP = N + DV - DO + CPP + CSD + S \quad (3.3)$$

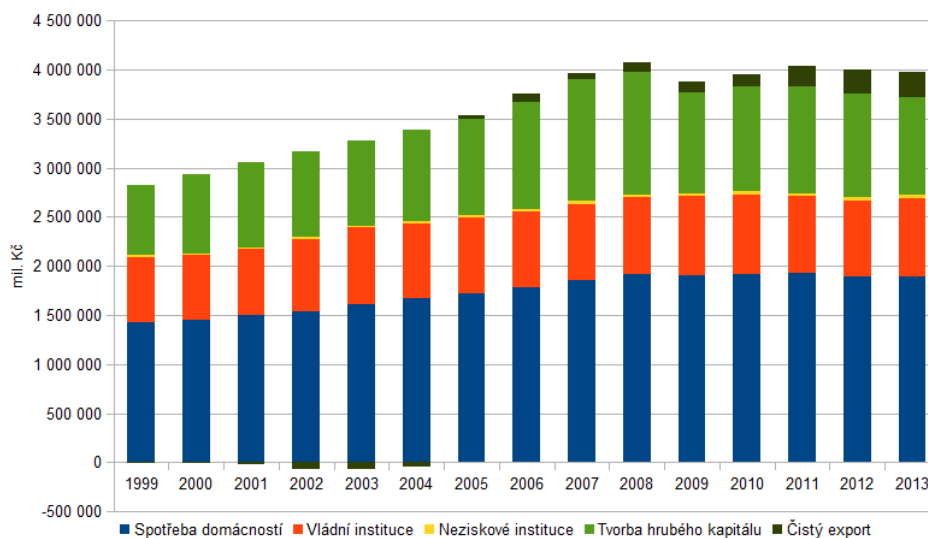
3.3 Vývoj HDP v České republice za posledních 15 let

Graf 3.1 ukazuje vývoj výše hrubého domácího produktu v České republice od roku 1999, tedy za posledních 15 let. Data jsou získána z databáze Českého statistického úřadu a jsou zachycena v cenách roku 2000. Procentuálně je zobrazen rovněž meziroční přírůstek HDP. Jak je z grafu vidět, od roku 2002 do 2007 vykazovala výše HDP meziroční růst o 5 % i více, v roce 2007 došlo k ustálení hodnoty a další výkyvy jsou minimální.



Obrázek 3.1: Vývoj HDP v ČR v cenách roku 2000 [20]

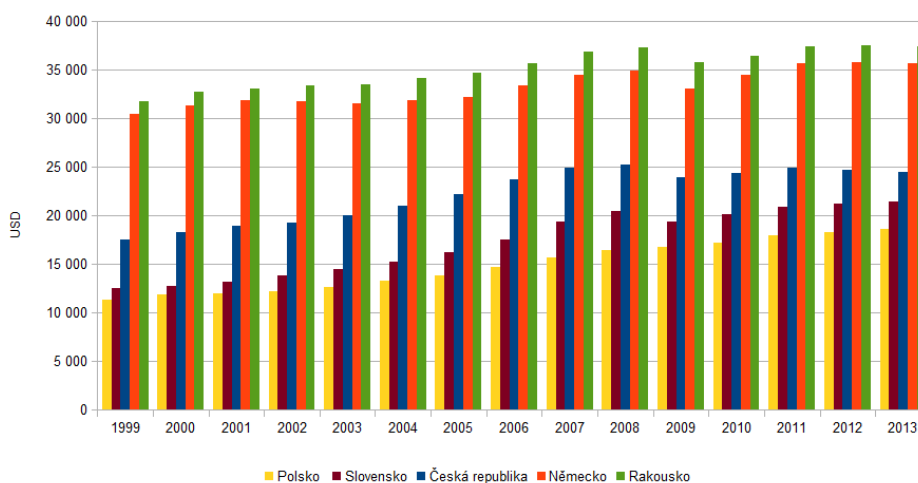
Na základě výpočtu HDP výdajovou metodou můžeme pro představu o skladbě hodnoty HDP sestavit graf 3.2. Hodnoty jsou získány opět z databáze Českého statistického úřadu a jsou uvedeny v cenách roku 2000. Položka čistý export zahrnuje rozdíl exportu a importu. Z grafu je tak patrné, že do roku 2005 docházelo k většímu objemu importu než exportu, naopak v posledních letech roste objem exportu. Dále je vidět, že majoritní složkou je spotřeba domácností, což je předpokladatelný stav. Pod složkou tvorba hrubého kapitálu se skrývá změna zásob, čisté pořízení cenností a tvorba hrubého fixního kapitálu, jako jsou nové investice, rekonstrukce, modernizace a pořízení nehmotných fixních aktiv. Vládní výdaje a výdaje na neziskové instituce jsou v daném období víceméně konstantní.



Obrázek 3.2: Struktura HDP v ČR v cenách roku 2000 [20]

3.3.1 HDP na obyvatele

Pro úplnost je určitě zajímavým ukazatelem také přepočtení HDP na obyvatele (na hlavu). Tento ukazatel nám umožní srovnat různé lidnaté státy spravedlivým způsobem. Data tentokrát pocházejí z databáze OECD.StatExtracts [8], jsou v cenách roku 2005 a pro snadné srovnání je výše HDP vypočtena v amerických dolarech (USD). Z grafu 3.3 je velmi dobře vidět podobný trend vývoje HDP ve všech hodnocených státech.



Obrázek 3.3: Porovnání HDP na obyvatele v ČR s okolními státy [8]

Kapitola 4

Analýza časových řad

4.1 Časové řady obecně

Analýza časových řad je v poslední době velmi diskutovanou statistickou disciplínou. Je to především proto, že u dat, která v mnoha oblastech lidské činnosti získáváme, potřebujeme zkoumat kromě aktuální hodnoty i jejich dynamiku. Analýza časových řad je pro takovýto úkol vhodným nástrojem. Získaná data, která tvoří časovou řadu, můžeme charakterizovat jako chronologicky uspořádaný soubor pozorování nějaké náhodné veličiny. Jedná se o časovou řadu statistickou, tedy takovou, jejíž chování je zatíženo nejistotou, na rozdíl od časové řady deterministické, kterou můžeme matematicky přesně popsat a tudíž jí i přesně předpovídat. Data ve formě časových řad nás obklopují doslova na každém kroku a v souvislosti s v podstatě libovolnou činností. Časovou řadu tak tvoří například vývoj cen na burze, záznam denních teplot, počet nehod na silnicích nebo výše produkce výrobního podniku.

Výsledkem analýzy časové řady je ve většině případů vytvoření modelu, který co nejméněji modeluje chování dané časové řady. Účelem modelu je především simulace mechanismu časové řady, na jejímž základě hledáme některé známky chování časové řady jako je například pravidelné opakování nebo určitý trend vývoje řady. Rovněž eliminujeme faktory, které jsou v zaznamenaných hodnotách zahrnuty, ale na vývoj časové řady nemají vliv. Dobře postavený model časové řady můžeme také využít k prognózování budoucího vývoje hodnot této řady. Vždy je ale nutné brát v úvahu, že data časové řady jsou zachycena pozorováním náhodné veličiny a tak ani sebelepší model časové řady nedovede spolehlivě předpovědět budoucí hodnoty, je schopen pouze bodového nebo intervalového odhadu [3].

4.1.1 Časová proměnná řady

Časovou řadu zaznamenáváme jako posloupnost hodnot ukazatelů získaných v daných časových intervalech, které jsou zpravidla rovnoměrné. Zapsat pak můžeme časovou řadu jako $y_t, t = 1, \dots, n$, kde y vyjadřuje hodnotu daného ukazatele a t je časová proměnná řady, n je pak celkový počet zaznamenaných pozorování. Podle charakteru časové proměnné můžeme rozlišit dva základní typy časových řad.

Prvním typem je časová řada spojitá (rovněž tok), kde známe hodnoty sledovaného ukazatele v každém okamžiku. Příkladem může být například sledování odběru vody nebo plynu z určitého odběrného místa. Časová proměnná pak nabývá hodnoty z nějakého intervalu.

Druhým typem je pak časová řada diskretní (stav). Jedná se o takovou časovou řadu, kde známe hodnoty sledovaného ukazatele jen v určitých časových okamžicích. Jako příklad lze uvést například sledování teploty každou hodinu, meziroční vývoj hodnoty HDP nebo počet nově narozených za každý měsíc. U tohoto typu řad nabývá časová proměnná diskretních hodnot, buď korespondujících s počtem měření, tedy $t = 1, 2, 3, \dots$ nebo například $t = 1999, 2000, 2001, \dots$ při analýze dat z jednotlivých let. Možné je i vícerozměrné číslování, které je vhodné při periodickém opakování zaznamenaného ukazatele s dalším menším rozlišením, například $t = 2002, 11$, pokud jde o hodnotu zaznamenanou v listopadu roku 2002[4].

4.2 Metody analýzy časových řad

Nejdůležitějším rozhodnutím před samotnou analýzou časové řady je volba vhodné metody analýzy. Toto rozhodnutí závisí na několika faktorech. Prvotním faktorem, který zásadně ovlivňuje výběr vhodné metodiky, je účel analýzy. Jednotlivé metody poskytují různé výsledky a výstupy, musíme tedy předem vědět, co od výsledků analýzy očekáváme a jak s nimi budeme dále pracovat. Neméně důležitým faktorem je samotný charakter časové řady, který je určen jednak počtem zpracovávaných hodnot, ale také očekávaným rozdělením zaznamenané náhodné veličiny. Dále je nutné zmínit zkušenost analytika, který časovou řadu zpracovává, jeho možnosti a technické vybavení, které rovněž určitou měrou ovlivňuje možnost zpracování dat.

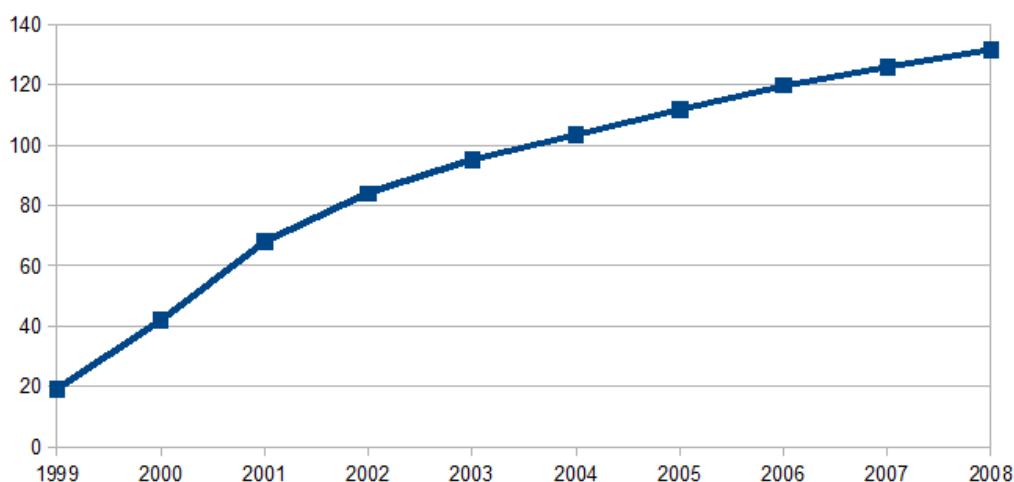
4.2.1 Grafické zobrazení časové řady

Při samotné analýze většinou nelze aplikovat univerzální jednoznačný postup na zpracování dané časové řady. Záleží tedy na přístupu analytika, který se pro daný postup rozhodne. Takovéto rozhodnutí nelze provést náhodným výběrem, ale je vhodné se seznámit alespoň se základní strukturou analyzované časové řady. Prvním krokem v analýze proto většinou bývá grafické znázornění zkoumaných dat ve spojnicovém grafu. Na ose x pak zachycujeme časovou posloupnost a na ose y zaznamenané ukazatele. Na základě takto zobrazených dat je v některých případech možné odhadnout základní parametry jako je trend, perioda či sezónní opakování. Chování těchto parametrů pak dále charakterizujeme postupy popsány níže.

Jako příklad zadání a zobrazení časové řady uvádím vývoj počtu aktivních SIM karet na 100 obyvatel. Jedná se o diskretní časovou řadu, protože hodnoty jsou sečítány jednou ročně. Data jsou zadána v tabulce 4.1, pro představu následuje graf 4.1, ze kterého je na první pohled vidět rostoucí trend časové řady.

t	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
y_t	19	42	68	84	95	104	112	120	124	132

Tabulka 4.1: Vývoj počtu aktivních SIM karet na 100 obyvatel [22]



Obrázek 4.1: Vývoj počtu aktivních SIM karet na 100 obyvatel [22]

4.2.2 Popisné charakteristiky časových řad

Na začátku samotné analýzy je kromě zobrazení časové řady vhodné také zjistit některé její statistické parametry, které nám usnadní další analýzu a rozhodování, například při volbě trendové funkce. Následující přehled obsahuje pouze vybrané charakteristiky vhodné pro rámec této práce [1]. Popisné charakteristiky můžeme dále rozdělit na několik skupin:

4.2.2.1 Charakteristiky polohy

Charakteristiky polohy vyjadřují polohu středu datového souboru, kolem něhož hodnoty kolísají. Výhodou zde uvedeného aritmetického průměru je především snadný výpočet a názorný význam, jako značná nevýhoda při analýze časových řad může být především vysoká citlivost na odlehlé hodnoty, takové je pak vhodné do průměru nezahrnovat [6].

- Prostý aritmetický průměr 4.1

$$\bar{y} = \frac{\sum_{t=1}^n y_t}{n} \quad (4.1)$$

- Vážený aritmetický průměr 4.2, v_t je váha hodnoty y_t v čase t

$$\bar{y} = \frac{\sum_{t=1}^n y_t v_t}{\sum_{t=1}^n v_t} \quad (4.2)$$

4.2.2.2 Charakteristiky variability

Charakteristiky variability vypovídají o tom, jak jsou hodnoty časové řady vzájemně blízké či vzdálené, hodnotí rozptýlenost hodnot časové řady kolem nějaké střední hodnoty.

- Rozptyl [4.3](#)

$$s_y^2 = \frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n} \quad (4.3)$$

- Směrodatná odchylka [4.4](#)

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n}} \quad (4.4)$$

4.2.2.3 Charakteristiky dynamiky

Tyto charakteristiky časových řad vypovídají o základních rysech chování časové řady.

- Absolutní přírůstek (první diference) [4.5](#) a průměrný absolutní přírůstek [4.6](#)

$$\Delta y_t = y_t - y_{t-1} \quad (4.5)$$

$$\bar{\Delta} = \frac{\sum_{t=2}^n \Delta y_t}{n-1} = \frac{y_n - y_1}{n-1} \quad (4.6)$$

- Druhá diference [4.7](#), vhodná pro určení trendové funkce

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \quad (4.7)$$

- Tempo (koeficient) růstu [4.8](#) a průměrný koeficient růstu [4.9](#)

$$k_t = \frac{y_t}{y_{t-1}}, t = 2, 3, \dots, n \quad (4.8)$$

$$\bar{k} = \sqrt[n-1]{k_2 k_3 \dots k_n} = \sqrt[n-1]{\frac{y_n}{y_1}} \quad (4.9)$$

- Meziroční koeficient růstu [4.10](#), např. pro měsíční hodnoty

$$k_{(12),t} = \frac{y_t}{y_{t-12}}, t = 13, 14, \dots, n \quad (4.10)$$

- Relativní přírůstek [4.11](#) a průměrný relativní přírůstek [4.12](#)

$$\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 \quad (4.11)$$

$$\bar{\delta} = \bar{k} - 1 \quad (4.12)$$

4.2.2.4 Korelace

Korelace 4.13 vyjadřuje míru závislosti ve vývoji dvou časových řad. Tzv. Pearsonův korelační koeficient k se pohybuje v rozmezí $-1 \leq k \leq 1$ a je koeficientem ve vztahu $y_t = kx_t$. Pokud je hodnota tohoto koeficientu -1 , je mezi hodnotami dvou časových řad nepřímá závislost (antikorelace). Hodnota korelačního koeficientu $+1$ značí naopak přímou závislost. Nulová hodnota znamená tzv. nekorelovanost, kdy mezi hodnotami časových řad neexistuje zjištěitelná lineární závislost [4].

$$\rho_{xy} = \frac{\sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y})}{\sigma_x \sigma_y} \quad (4.13)$$

4.2.3 Dekompozice časových řad

Základní myšlenka dekompozice časové řady spočívá v tom, že analyzovaná časová řada obsahuje některé ze čtyř složek – trend, sezónní složka, cyklická složka a reziduální složka. Předpokládáme potom, že zaznamenané hodnoty časové řady jsou tvořeny těmito složkami, ne však nutně všemi. Cílem dekompozice je pak rozložit časovou řadu na následující složky a určit jejich parametry. Motivace k tomuto postupu je založena na tom, že pomocí chování některé ze složek budeme moci snadněji analyzovat chování původní časové řady před dekompozicí.

4.2.3.1 Trend časové řady

Trendová složka T_t časové řady charakterizuje dlouhodobé změny ve vývoji časové řady. Trend může být rostoucí, klesající, případně konstantní. Můžeme říci, že trend vzniká dlouhodobým působením sil, které mají stejný směr. Například v případě sledování vývoje cen na trhu bude jistě trendovou složkou rostoucí míra inflace. Trend můžeme obvykle popsat matematickou funkcí v celém rozsahu časové řady. Typickým způsobem získání funkce trendu na daných datech je regresní analýza.

4.2.3.2 Sezónní složka

Sezónní složkou S_t časové řady rozumíme v hodnotách obsažené periodické změny, které nastávají v průběhu nějakého časového období. Typicky to bývá rok, měsíc, týden nebo den. Změny pak souvisejí například s průběhem pracovního týdne, ročním vývojem teploty, denní dobou atd. Sezónní složku je problematické analyzovat především u řad, kde jsou zachycená data řídká. Sezónní vývoj z nich nemusí být patrný a jeho odečtením může docházet ke zkreslení.

4.2.3.3 Cyklická složka

Cyklická složka C_t vyjadřuje kolísání hodnot časové řady kolem jejího trendu. Zachycuje pravidelné fáze poklesu a růstu, jejichž perioda může být různě dlouhá a na rozdíl od sezónní složky nelze jednoznačně stanovit závislost na nějakém časovém období. Vzhledem k tomu, že na rozdíl od sezónní složky není známa přímá příčina vzniku cyklické složky, je v některých případech její nalezení a charakterizace nesnadným úkolem.

4.2.3.4 Náhodná složka

Reziduální, též náhodná složka ε_t představuje náhodné změny hodnot časové řady, které nemají žádný systematický charakter. Zůstává z časové řady po odstranění tzv. systematických složek - trendu, sezónní a cyklické složky. Její charakter vychází z faktu, že časová řada je složena z pozorování náhodné veličiny. Rovněž může zahrnovat chybu měření získaných hodnot, případně zaokrouhlování pro účely zpracování dat. V případě, že je střední hodnota náhodné složky nulová, její rozptyl konstantní a její hodnoty nejsou korelované, můžeme náhodnou složku nazvat bílým šumem. V případě, že hodnoty ε_t mají normální rozdělení, jde pak o normální bílý šum [4].

4.2.3.5 Typy dekompozice časové řady

Možností dekompozice do výše popsaných složek existuje několik způsobů, nám však postačí dvě základní metody – dekompozice aditivní a dekompozice multiplikativní.

Tvar aditivní dekompozice časové řady je zachyceno vzorcem 4.14. Časovou řadu si tak lze představit jako trend, ke kterému jsou přidány ostatní složky. Všechny složky jsou v aditivním rozkladu zahrnuty ve své skutečné absolutní hodnotě a v jednotkách, ve kterých je dána hodnota časové řady.

$$y_t = T_t + S_t + C_t + \varepsilon_t \quad (4.14)$$

Jako příklad můžeme uvést opět časovou řadu zachycující počet aktivních SIM karet. Jednotkou této časové řady je počet SIM karet a tato jednotka je shodná pro všechny složky časové řady.

V případě multiplikativní dekompozice časové řady je pak hodnota pozorování vyjádřena vzorcem 4.15. Řadu si pak můžeme představit jako trend ovlivňovaný dalšími faktory, které jsou reprezentovány ostatními složkami řady. V takovémto rozkladu je ve své skutečné absolutní hodnotě a v jednotkách většinou uveden právě pouze trend. Ostatní složky jsou v hodnotách relativních a jsou to bezrozměrné veličiny.

$$y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t \quad (4.15)$$

4.3 Analýza složek časové řady

4.3.1 Analýza trendové složky

Cílem analýzy trendové složky T_t pro nás bude nalezení a následné odebrání trendové složky z časové řady. Tomuto postupu říkáme vyrovnání, nebo též vyhlazení časové řady. Pro nalezení trendu existují dva základní přístupy. Prvním z nich je globální metoda. Tato metoda využívá k nalezení trendové funkce především regresní analýzu a předpokládá, že pro celou zkoumanou časovou řadu existuje jedna jediná funkce, která realizuje funkci trendu. Lokální metoda naproti tomu pracuje s předpokladem, že sice existuje funkce trendu, ale je platná pouze lokálně, tedy nikoliv na celém rozsahu hodnot časové řady. Pro popis trendu na

celém rozsahu hodnot se tato funkce nebo její parametry postupně mění. K jejímu nalezení používáme převážně metody klouzavých průměrů a exponenciálního vyrovnání.

Samotných metod nalezení trendu časové řady je několik. Pro úplnost pouze vyjmenuji například metodu průměrování horních a dolních bodů zvratu, která patří mezi tzv. subjektivní nebo též expertní metody, protože není pevně dáno, co považovat za body zvratu. Záleží tak značnou měrou na výběru analytika. Další možností je využití diferencování, které se využívá u předpokládaného polynomiálního trendu řady a využívá výpočtu i několika řádů difference funkce trendu. Pro naše účely bude postačující využití metod klasického popisu trendu matematickými křivkami a pomocí klouzavých průměrů, které nadále rozvedu.

4.3.1.1 Modelování trendu analytickými funkcemi

Metody modelování trendu pomocí matematických křivek využívají popisu trendu časové řady analyticky pomocí některé jednoduché křivky jako je přímka nebo exponenciála. V případě nalezení odpovídající křivky je pak snadné spočítat její budoucí hodnoty a odhadnout tak budoucí vývoj trendu, pokud se jeho charakter v čase nemění. Při modelování matematickými křivkami předpokládáme, že časová řada je tvořena trendovou složkou a reziduální složkou, případně že byla na tento tvar upravena, tj. očištěna od ostatních složek.

Pro nalezení obvyklých tvarů modelujících křivek, tedy parametrů β v rovnicích uvedených dále, používáme lineární regresní analýzu. Abychom byli schopni určit vhodnou popisnou křivku trendu, je vhodné začít předběžným prozkoumáním časové řady, jako je její grafické zobrazení, které nám může hodně napovědět o vývoji trendu. Z grafu nemusí ale být trend řady vždy patrný, můžeme proto využít další nástroje, jako teorii vývoje sledované veličiny, která nějakým způsobem trend popisuje, nebo některé typy testů trendu. Volba funkce trendu není vždy jednoznačná a tak nelze přímo stanovit, v kterém případě jakou konkrétní funkci použít. Pro rozhodování ale může pomoci například test znamének diferencí, test bodů zvratu nebo mediánový test. Dále následuje přehled nejčastějších funkcí pro popis trendu [3] [4]. Obrázek 4.2 ilustruje průběhy jmenovaných trendových funkcí.

Konstantní trend V případě konstantní funkce trendu je odhad parametru této funkce jednoduchý. Jedná se o prostý průměr hodnot časové řady tak, jak zachycuje vzorec 4.16. Tento trend volíme, pokud se hodnota prvních diferencí pohybuje kolem nuly.

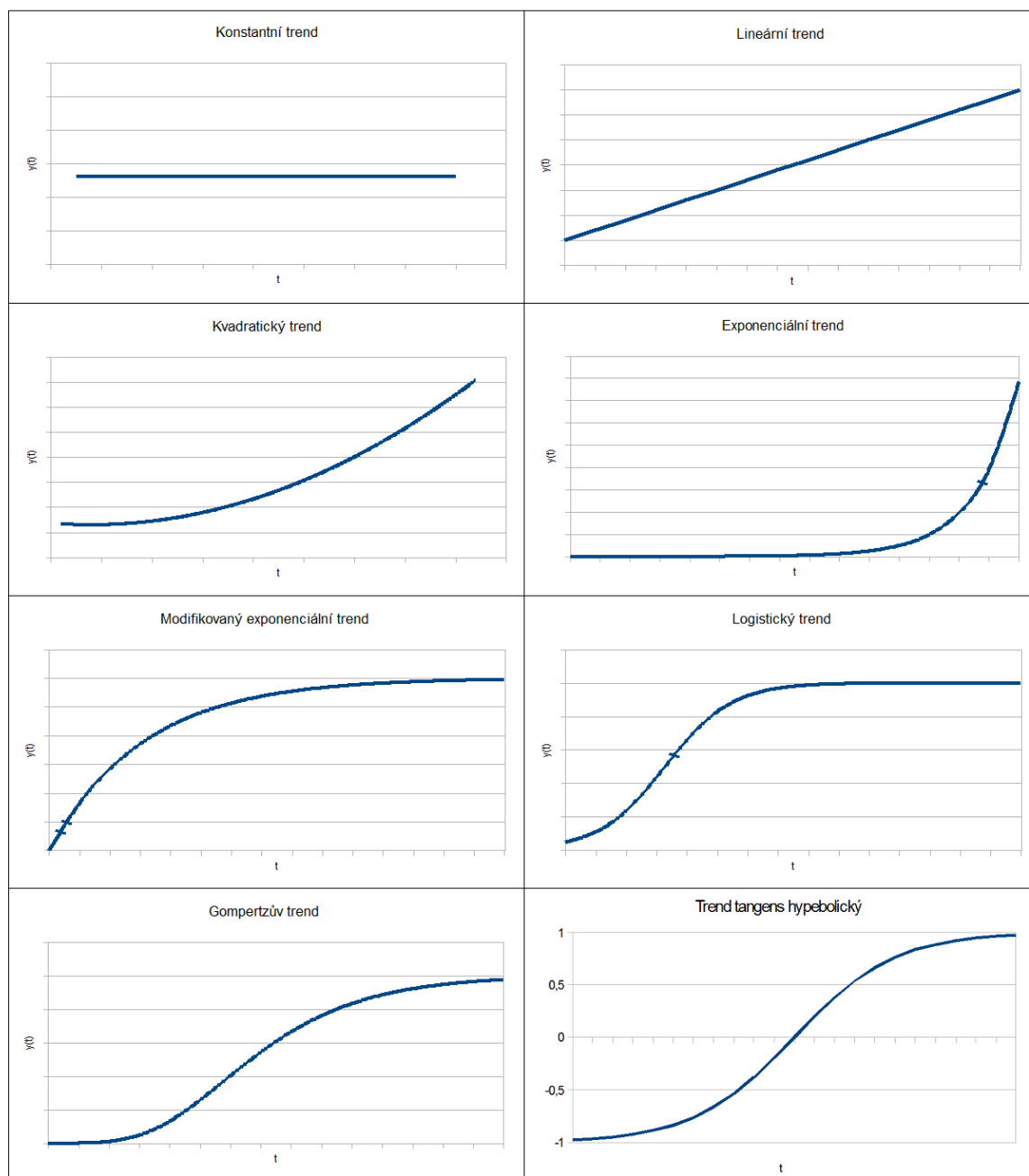
$$T_t = \beta_0 = \bar{y}_t = \frac{\sum_{t=1}^n y_t}{n} \quad (4.16)$$

Lineární trend Lineární trend reprezentuje trend ve tvaru přímky, která má rovnici 4.17. Lineární trend volíme, pokud se hodnota prvních diferencí pohybuje kolem nenulové konstanty.

$$T_t = \beta_0 + \beta_1 t \quad (4.17)$$

Kvadratický trend Tento trend je principiálně shodný s lineárním trendem, pouze obsahuje navíc kvadratický člen, tedy rovnice popisující parabolu má tvar 4.18. Je vhodné využít tuto trendovou funkci, pokud řady prvních diferencí vykazuje lineární trend a řada druhých diferencí trend konstantní.

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad (4.18)$$



Obrázek 4.2: Průběhy trendových funkcí

Exponenciální trend Jedná se o dvouparametrický trend s tvarem 4.19. Parametry odhadujeme tak, že logaritmováním převedeme trend na lineární tvar a postupujeme jako u lineárního trendu s odlogaritmováním na konci výpočtu. Teno trend volíme v případě, že se

koeficienty růstu pohybují kolem nenulové konstanty.

$$T_t = \beta_0 \beta_1^t, \quad \beta_0, \beta_1 > 0 \quad (4.19)$$

Trend hyperbolický tangens Tato funkce je symetrická ohraničená rostoucí křivka daná rovnicí 4.20.

$$T_t = \beta_0 \frac{e^{2t} - 1}{e^{2t} + 1} \quad (4.20)$$

Modifikovaný exponenciální trend Jde o zobecnění exponenciálního trendu a je vhodný pro modelování trendu, jehož sousední diference mají konstantní podíl. Tvar funkce modifikované exponenciály zachycuje vzorec 4.21. Funkci nelze žádnou transformací linearizovat, proto se její parametry odhadují iterativní metodou. Tuto trendovou funkci využijeme, pokud se podíly sousedních diferencí soustředí kolem nenulové konstanty.

$$T_t = \gamma + \beta_0 \beta_1^t, \quad \beta_0 < 0, 0 < \beta_1 < 1, \gamma > 0 \quad (4.21)$$

Logistický trend Graficky ho charakterizuje trendová křivka, která má inflexní bod. Navíc jde o tzv. S-křivku, která je podle bodu inflexe symetrická. Rovnice této křivky má tvar 4.22. Inverzní funkce má tvar modifikovaného exponenciálního trendu, čehož se využívá při odhadu parametrů. Logistický trend je vhodný v případě, že řada $\ln(y_t)$ vykazuje přibližně hyperbolický průběh.

$$T_t = \frac{1}{\gamma + \beta_0 \beta_1^t}, \quad \beta_1 > 0, \gamma > 0 \quad (4.22)$$

Gompertzův trend Křivka má podobný průběh jako u logistického trendu, pouze není symetrická podle inflexního bodu. Charakteristickou rovnicí pak zachycuje vztah 4.23. Po transformaci se lze dostat na tvar shodný s modifikovaným exponenciálním trendem a stejným způsobem odhadnout parametry. Gompertzův trend volíme konečně v případě, pokud se podíly sousedních diferencí logaritmovaných hodnot soustředí kolem nenulové konstanty.

$$T_t = \gamma \beta_0^{\beta_1^t}, \quad \beta_1 > 0 \quad (4.23)$$

4.3.1.2 Metoda klouzavých průměrů

Jak již vyplývá z předchozího textu, metoda klouzavých průměrů patří mezi tzv. lokální metody analýzy trendové funkce. Tyto metody také označujeme jako metody adaptivní, což znamená, že pomocí nich můžeme zpracovat trendovou funkci, která v průběhu mění svůj charakter, není tedy možné využít k jejímu popisu matematické křivky se statickými parametry.

Z nadpisu je patrné, že se budeme zabývat výpočtem klouzavých průměrů. Pod pojmem klouzavý průměr si můžeme představit lineární kombinaci prvků původní časové řady, jde vlastně o vyrovnání řady polynomem určitého řádu. Klouzavé průměry můžeme rozdělit na jednoduché a vážené.

Jednoduchý klouzavý průměr použijeme v případě lineárního trendu a vypočítáme ho jako obyčejný aritmetický průměr u původních hodnot časové řady. V případě, že chceme

získat klouzavý průměr hodnoty zaznamenané v čase t , vypočítáme aritmetický průměr z hodnoty y_t a určitého počtu jejích okolních hodnot, poté se přesuneme na hodnotu y_{t+1} a pokračujeme stejným způsobem. Teoreticky můžeme aplikovat na časovou řadu výpočet klouzavých průměrů různého řádu, v praxi je ale právě tato volba kritickým parametrem výpočtu. Při špatné volbě řádu může dojít k nechtěnému zkreslení analyzovaných dat. Obvykle proto preferujeme průměry co nejnižšího řádu s ohledem na požadavek vyhlazení řady hodnot. Délka klouzavého průměru totiž přímo souvisí s jemností vyhlazení. Důležitou zásadou volby řádu klouzavých průměrů je fakt, že délka by měla odpovídat periodě sezónních a cyklických složek, které chceme vyhladit.

Tento jednoduchý postup je snadno aplikovatelný, pokud počítáme klouzavé průměry liché délky. Pak vypočtený klouzavý průměr odpovídá dané hodnotě y_t , která leží uprostřed intervalu, ze kterého je vypočítáván průměr. V případě, že potřebujeme počítat klouzavé průměry sudého řádu, nastane nepříjemný jev, a to sice, že vyrovnaná hodnota, tedy vypočtená hodnota průměru, neodpovídá původnímu času zachyceného měření. Tedy například při analýze dat sebraných během let, kdy jedna hodnota odpovídá jednomu kvartálu, se nabízí zvolit délku klouzavého průměru 4. Tím ale získáme aritmetickým průměrem I. až IV. kvartálu hodnotu, která náleží doprostřed intervalu mezi II. a III. kvartál. To je krajně nevhodné, protože pak není možné určit, kterému kvartálu hodnota náleží. Řešením tohoto problému je pak zprůměrovat takovéto dvě hodnoty vedle sebe, čímž jednoznačně určíme hodnotu pro II. i III. kvartál. Takto vypočtené hodnoty nazýváme centrovaným klouzavým průměrem. Ten je často využíván především v ekonomických časových řadách, které se často vztahují právě k obdobím se sudým počtem zaznamenaných hodnot časové řady jako je rok s dvanácti měsíci a kvartály.

V případě, že trend není lineární, je též možné použít metodu klouzavých průměrů k očištění časové řady, ale pro různé trendy bude nutné použít různé váhy klouzavých průměrů [3].

Jednoduché klouzavé průměry můžeme rovněž využít pro odhad budoucích hodnot časové řady. Předpokladem ovšem je přibližně konstantní trend $T_t = \beta_0$ dané časové řady. Odhad parametru β_0 pak určíme jako prostý aritmetický průměr hodnot y_t , jak zachycuje vzorec 4.24 [1].

$$\hat{\beta}_0 = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad (4.24)$$

4.3.1.3 Exponenciální vyrovnání

Další velmi často používanou adaptivní metodou vyrovnání je vyrovnání exponenciální. Na rozdíl od vyrovnání klouzavým průměrem je zde odbourán problém s volbou délky klouzavého průměru, která je často velice subjektivní. Metoda exponenciálního vyrovnání tento problém nemá, protože každá vyrovnaná hodnota vychází ze všech dostupných předchozích pozorování zaznamenaných v časové řadě.

Základem je metoda nejmenších čtverců, která je upravena tak, že se váhy hodnot z minulosti pro výpočet dané vyrovnané hodnoty exponenciálně snižují, odtud také název metody. Pokud si symbolem \hat{y}_t označíme vyrovnanou hodnotu řady v čase t , tak musíme pro získání těchto hodnot provádět minimalizaci výrazu 4.25. Důležitým parametrem výpočtu

vyrovnaných hodnot je tzv. vyrovnávací konstanta α , která nabývá hodnot $0 < \alpha < 1$. Její volba stanovuje, jak je zřejmé z vzorce výše, rychlost klesání vah minulých pozorování [3].

$$(y_t - \hat{y}_t) + (y_{t-1} - \hat{y}_{t-1})^2\alpha + (y_{t-2} - \hat{y}_{t-2})^2\alpha^2 + \dots \quad (4.25)$$

V literatuře je klasicky popisováno exponenciální vyrovnání jednoduché, dvojitě a trojitě. Můžeme si pod těmito názvy představit vyrovnání podle typu trendové složky a to konstantní, lineární a kvadratické. V následujícím textu rozeberu pouze první dvě metody, protože trojité vyrovnání v praxi téměř nepoužíváme a nevyužiji ho ani v této práci.

Jednoduché exponenciální vyrovnání Jednoduché exponenciální vyrovnání používáme na takové řady, kde předpokládáme v kratších úsecích řady konstantní trend. Proto můžeme takovou časovou řadu zapsat vztahem 4.26, kde v případě konstantního trendu nahradíme tuto složku konstantou $y_t = \beta_0$. Při vyrovnávání tedy hledáme hodnotu konstanty β_0 , která je v případě jednoduchého vyrovnání rovna vyrovnané hodnotě \hat{y} . Hodnoty vyrovnané řady pak určujeme podle vztahu 4.27. Tento vztah můžeme rovněž zapsat jako 4.28, kde je ze zápisu patrné, že vyrovnanou hodnotu vytváříme opravením předchozí vyrovnané hodnoty o chybu, kterou určuje rozdíl mezi zaznamenanou hodnotou časové řady a předchozí vyrovnanou hodnotou.

$$y_t = T_t + \varepsilon_t \quad (4.26)$$

$$\hat{y}_t = (1 - \alpha)y_t + \alpha\hat{y}_{t-1} \quad (4.27)$$

$$\hat{y}_t = \hat{y}_{t-1} + (1 - \alpha)(y_t - \hat{y}_{t-1}) \quad (4.28)$$

Čím více se parametr α blíží hodnotě 1, tím roste vliv minulých hodnot na vypočtenou současnou vyrovnanou hodnotu. Empiricky bylo zjištěno, že nejvhodnější volba parametru α leží v rozsahu $0.7 \leq \alpha < 1$. Pro větší přesnost můžeme hodnotu určit tak, že provedeme vyrovnání řady s různými parametry z daného intervalu (typicky hodnoty z celého intervalu se zvoleným krokem) a poté vypočteme součet čtvercových chyb (SSM), který určí přesnost vyrovnání. Tuto hodnotu se snažíme minimalizovat vhodným výběrem parametru α .

Společným problémem rekurentních výpočtů je vždy stanovení první hodnoty, tedy v našem případě \hat{y}_0 . Nejjednodušší možností je stanovit tuto hodnotu jako první hodnotu skutečných dat z časové řady. Tato možnost se také velmi často používá, protože z principu metody exponenciálního vyrovnání dojde postupem výpočtu k opravení této hodnoty. Druhou metodou je opět na zkušenostech založený odhad počáteční hodnoty ve dvou fázích. První fáze nastává na samotném počátku vyrovnávání, kdy zjišťujeme výše popsanou metodou s hodnocením SSM vhodnost koeficientu α . V tom případě je vhodné použít jako počáteční hodnotu průměr prvních 6 až $\frac{n}{2}$ prvků řady, kde n je celkový počet prvků časové řady. V druhé fázi výpočtu, kdy již máme stanoven vhodný koeficient α , se poté osvědčuje volit jako počáteční hodnotu průměr všech n prvků časové řady [4].

Dvojité exponenciální vyrovnání Dvojité exponenciální vyrovnání používáme na takové řady, kde předpokládáme v kratších úsecích řady lineární trend, jehož tvar znázorňuje vztah 4.29. V praxi se předpoklad linearit v krátkém časovém období často používá. Naším cílem je tedy určení parametrů β_0 a β_1 . Pro jejich určení vycházíme z minimalizace výrazu 4.30, kde parametr α má stejný význam a funkci jako u vyrovnání jednoduchého.

$$T_t = \beta_0 + \beta_1 t \quad (4.29)$$

$$\sum_{j=0}^{\infty} [y_{t-j} - (\beta_0 + \beta_1(-j))]^2 \alpha^j \quad (4.30)$$

Pro zjednodušení dalších výpočtů a značení obvykle zavádíme dvě veličiny. První z nich – jednoduchá vyrovnávací statistika $S1$ má tvar shodný jako hodnota vyrovnané řady metodou jednoduchého exponenciálního vyrovnání 4.31. Druhá definovaná veličina je pak dvojitá vyrovnávací statistika $S2$, která je stanovena podobně, ale místo vyrovnané hodnoty bere v úvahu první vyrovnávací statistiku 4.32.

$$S1_t = (1 - \alpha)y_t + \alpha S1_{t-1} \quad (4.31)$$

$$S2_t = (1 - \alpha)S1_t + \alpha S2_{t-1} \quad (4.32)$$

Odhady hledaných parametrů beta pak mají po zjednodušení tvar 4.33 a 4.34. Na jejich základě pak můžeme stanovit vyrovnanou hodnotu i předpovědi další hodnoty podle tvaru 4.35. Hodnoty $S1$ a $S2$ v čase $t = 1$ lze stanovit vícero způsoby, je však možné je aproximovat hodnotou y_1 , protože jejich chybný odhad je vyrovnáním brzy opraven [3] [4].

$$\beta_0(t) = 2S1_t - S2_t \quad (4.33)$$

$$\beta_1(t) = \frac{1 - \alpha}{\alpha}(S1_t - S2_t) \quad (4.34)$$

$$\begin{aligned} \hat{y}_{t+\tau}(t) &= \beta_0(t) + \beta_1(t)\tau = (2S1_t - S2_t) + \frac{1 - \alpha}{\alpha}(S1_t - S2_t)\tau = \\ &= \left(2 + \frac{(1 - \alpha)\tau}{\alpha}\right)S1_t - \left(1 + \frac{(1 - \alpha)\tau}{\alpha}\right)S2_t \end{aligned} \quad (4.35)$$

4.3.2 Analýza sezónní složky

Analýza sezónní složky S_t časové řady má v zásadě dva významy. Prvním důvodem analýzy je nalezení samotné sezónní složky pro rozšíření znalostí o sezónním vývoji časové řady. Sezónní závislost je u ekonomických časových řad poměrně častým jevem a vychází z toho, že velká část sledovaných ukazatelů se váže ke kalendářním časovým obdobím jako týden, měsíc, čtvrtletí nebo rok. Identifikací sezónního faktoru si tak můžeme pomoci k lepšímu pochopení vývoje časové řady a rovněž ke konstrukci dokonalejší předpovědi budoucího vývoje. Druhým

důvodem analýzy sezónní složky je očištění dané časové řady právě o sezónní složku. Zde nám jde především o maximální možnou eliminaci, ideálně úplné odstranění sezónního vlivu na data zaznamenaná v časové řadě. Na takto očištěné časové řadě pak můžeme mnohem snáze analyzovat dlouhodobé vlivy a trendy vývoje.

Na úvod je třeba říci, že v předchozí kapitole popsaná metoda klouzavých průměrů má na výstupu rovněž určitým způsobem očištěnou časovou řadu. Touto metodou jsme se ale snažili potlačit především náhodné změny hodnot v časové řadě, nikoliv změny periodické. Z principu samotné metody však dochází k určitému samovolnému očištění časové řady od sezónních ale i cyklických složek, pokud řada takové složky obsahuje. Tento jev nastává především pro periodické jevy v časové řadě, jejichž perioda není delší než délka klouzavého průměru. Například tedy centrovaný klouzavý průměr délky 5 bude celkem efektivně očišťovat časovou řadu získanou kvartálním měřením od sezónní složky. Tato metoda ale není vždy vhodná, protože není možné snadno očistit počáteční a koncové hodnoty a rovněž nevhodnou délkou klouzavého průměru může dojít k nechtěnému zkreslení dat. Byly proto vyvinuty speciální metody, nebo též filtry, které jsou pro očištění od sezónní složky vhodnější. Jako příklad lze uvést například metodu Census X-11, která využívá k očištění speciální klouzavé průměry různých délek, a na časovou řadu jsou aplikovány v několika krocích za sebou. Zdokonalení tohoto postupu vede například k metodám X-12 nebo X-11 ARIMA, kde je časová řada vhodně extrapolována vhodným ARIMA modelem. Tyto metody spolehlivě očišťují i počáteční a koncové hodnoty časové řady.

Pro účely této práce bude dále uveden pouze základní klasický přístup k eliminaci sezónní složky. Příkladem komplexnějšího postupu zde pouze zmiňuji Wintersovu metodu jako zástupce adaptivní metody sezónního očištění časové řady, které je vhodné převážně tam, kde se sezónní charakter složky může časem změnit [3].

4.3.2.1 Očištění od sezónní složky metodou klouzavých průměrů

Prvním způsobem očištění časové řady od sezónní složky využívá právě výše zmíněných klouzavých průměrů. Prvním krokem takovéto analýzy je tedy právě výpočet klouzavých průměrů délky odpovídající délce kalendářní periody dat zachycených v časové řadě. V případě sudé délky periody (měsíční, kvartální) je nutné provést výpočet centrovaných klouzavých průměrů, jak bylo popsáno dříve. Tím provedeme v podstatě sezónní očištění časové řady, které bylo popsáno na začátku této kapitoly. Výchozím předpokladem je tedy fakt, že nyní máme časovou řadu očištěnou od sezónní složky a rovněž náhodné složky, což je důsledek právě užití klouzavých průměrů.

Předpokládáme nyní tvar časové řady při aditivním přístupu 4.36. Na základě tohoto předpokladu snadno získáme sezónní a náhodnou složku časové řady.

$$y_t = T_t + C_t \quad (4.36)$$

Odečtením těchto centrovaných klouzavých průměrů od původních hodnot zaznamenané časové řady se dostáváme k tvaru 4.37.

$$y_t = S_t + \varepsilon_t \quad (4.37)$$

Dalším krokem je zprůměrování odpovídajících hodnot těchto složek v odpovídajících časových obdobích. Tedy například v případě kvartální časové řady zjistíme hodnotu sezónního faktoru pro první kvartál jako aritmetický průměr hodnot ze všech lednových zaznamenaných hodnot. Tímto zprůměrováním odstraníme dostatečně vliv reziduální složky a získáme tak samotnou sezónní složku. Pro správnost rozkladu je obvykle nutné sezónní složku časové řady znormalizovat, aby se sezónní vliv v rámci jednoho roku vykompenzoval. V našem případě aditivní časové řady je tedy nutné, aby součet sezónních faktorů v jednom roce byl roven nule. Pro kvartální časovou řadu je zachycena normalizační podmínka vzorcem 4.38. Prostým odečtením takto získané sezónní složky od původní časové řady získáme řadu sezónně očištěnou.

$$S_1 + S_2 + S_3 + S_4 = 0 \quad (4.38)$$

V případě multiplikativní dekompozice budeme vycházet z tvaru 4.39 časové řady, který jsme získali stejně jako v předchozím případě výpočtem centrovaných klouzavých průměrů.

$$y_t = T_t \cdot C_t \quad (4.39)$$

Další postup je podobný jako u aditivní časové řady, pouze se liší matematickými operacemi. Prvním krokem je nyní vydělení původních hodnot vypočtenými centrovanými klouzavými průměry, čímž se dostáváme analogicky k tvaru 4.40.

$$y_t = S_t \cdot \varepsilon_t \quad (4.40)$$

Následující krok zprůměrování odpovídajících hodnot je shodný s minulým případem, opět tak dosáhneme eliminace reziduální složky. Normování vychází nyní z předpokladu, že součet sezónních faktorů je roven počtu časových období v jedné periodě. V případě kvartálních faktorů má normalizační podmínka tvar 4.41. Normalizací opět dosáhneme kompenzace sezónního vlivu v rámci jednoho roku. Pokud nyní vydělíme původní hodnoty časové řady tímto sezónním faktorem, získáme opět sezónně očištěnou časovou řadu, tentokrát v multiplikativním tvaru [3].

$$S_1 + S_2 + S_3 + S_4 = 4 \quad (4.41)$$

4.3.3 Analýza cyklické složky

Poslední složkou, která nám zbývá k analýze, je cyklická složka C_t . Na základě výsledků z předchozí analýzy sezónní složky se poměrně snadno dostaneme i k odhadu cyklické složky a budeme tak schopni vytvořit předpovědi časové řady. Jako základní data využijeme sezónně očištěnou časovou řadu z předchozích odstavců. Pokud provedeme nyní vyrovnání některou matematickou funkcí, jak jsem rovněž popsal dříve, získáme poměrně přesný odhad trendové složky časové řady. Jen pro připomenutí, v této fázi máme již vypočtené hodnoty sezónní složky S_t a nyní po odhadu trendové funkce i hodnoty trendu T_t . Nyní můžeme přejít k poslednímu kroku, tedy zjištění samotné cyklické složky. Opět postup rozdělíme podle dekompozičního tvaru časové řady [6].

V případě aditivní časové řady ve tvaru 4.42 získáme odečtením známých složek T_t a S_t tvar 4.43.

$$y_t = T_t + S_t + C_t + \varepsilon_t \quad (4.42)$$

$$y_t = C_t + \varepsilon_t \quad (4.43)$$

V případě multiplikativní dekompozice pak vyjdeme z tvaru 4.44 a postupným vydělením složkami T_t a S_t získáme nový tvar 4.45.

$$y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t \quad (4.44)$$

$$y_t = C_t \cdot \varepsilon_t \quad (4.45)$$

Eliminaci reziduální složky v obou případech provedeme opět využitím klouzavých průměrů. Délku klouzavého průměru volíme kratší než je perioda zkoumaného období, ale zároveň dostatečně dlouhou na to, aby bylo zaručeno, že spolehlivě vyloučíme možnost, že reziduální složka je v této délce systematicky vychýlena stejným směrem. Například při roční časové řadě není vhodné volit délku 4, případně 6, protože kvartálně nebo pololetně může mít reziduální složka podobné chování. Například délka klouzavých průměrů 7 či 8 je v pořádku a měla by dostatečně potlačit vliv reziduální složky. Po provedení tohoto kroku jsme analyzovali všechny složky časové řady a na základě jejich významnosti můžeme založit konstrukci modelu analyzovaná časové řady [3].

4.4 Kvalita vyrovnání a vhodnost modelu

Cílem výše popsaných postupů analýzy složek časových řad je tvorba modelu, který popíše chování co nejvěrněji chování časové řady a rovněž nám umožní zkonstruovat předpovědi budoucího vývoje. Jak vyplývá z předchozího přehledu, který obsahuje pouze vybrané metody, způsobů tvorby modelu je velké množství. Navíc lze jednotlivé metody mezi sebou kombinovat podle charakteru analyzované časové řady. Výsledkem různých druhů modelování chování časové řady mohou být také odlišné výsledky a odlišné hodnoty předpovědí. Otázkou tedy zůstává, jak vybrat vhodný model a především, jak vhodnost modelu objektivně posoudit.

Samotná volba nejvhodnějšího modelu nemá jednoznačný výsledek. Důvodem je především to, že každý model časové řady může být konstruován s jiným záměrem. V některých případech je potřeba sestavit model tak, aby například co nejlépe vyrovnával již naměřené hodnoty časové řady. Takovýto model může sloužit k pochopení mechanismu vzniku naměřených hodnot, aby bylo možné na základě toho dále časovou řadu řídit. V jiných případech je záměrem vytvořit model takový, aby jeho výsledkem byla co nejpřesnější predikce budoucích hodnot. V tomto případě jsme schopni míru vhodnosti modelu vzhledem k predikovaným hodnotám posoudit až zpětně, když získáme skutečná data, což už může být pozdě. Vzniklá chyba v případě predikce je obvykle dána reziduální složkou, kterou nejsme schopni, na rozdíl od ostatních složek časové řady, předpovídat. Naším cílem tedy bude najít model, ve kterém

bude minimální rozptyl reziduální složky. Dále se tedy budu zabývat kritérii, kterými jsme schopni zhodnotit kvalitu modelu na základě analýzy reziduální složky.

Následující parametry hodnocení vhodnosti modelu nazýváme obvykle interpolační kritéria. Jejich funkcí je zhodnotit rozdíl mezi skutečnými hodnotami a hodnotami vyrovnané časové řady [4]. Nejčastěji používané interpolační kritéria jsou:

- Součet čtvercových chyb SSE (Sum of Squared Error), 4.46

$$SSE = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (4.46)$$

- Střední čtvercová chyba MSE (Mean Squared Error), 4.47

$$MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} = \frac{SSE}{n} \quad (4.47)$$

- Střední absolutní chyba MAE (Mean Absolute Error), 4.48

$$MAE = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \quad (4.48)$$

- Střední absolutní procentní chyba MAPE (Mean Absolute Percentage Error), 4.49

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (4.49)$$

4.4.1 Kvalita prognózy

Jak již vyplývá z výše uvedeného, předchozí metody se hodí pro zhodnocení modelu vzhledem k známým hodnotám, což je logické, protože predikované hodnoty není s čím porovnávat. Aby byla zajištěna dostatečná kvalita modelu i v předpovědní rovině, používá se proto tradičně způsob rozdělení časové řady na dvě části. Na první části dat konstruujeme model dříve popsanými způsoby a snažíme se dosáhnout s ohledem na požadované užití maximální možné věrnosti. V případě, že jsme spokojeni s kvalitou vyrovnaní daným modelem, využijeme druhou část dat původní časové řady.

Na základě zkonstruovaného modelu budeme predikovat data v takovém horizontu, který je pokryt ještě původními daty z časové řady. Právě k těmto původním hodnotám budeme přistupovat jako ke kontrolnímu souboru a porovnáme s ním hodnoty naší predikce. Tím jsme schopni s určitou předností analyzovat rovněž kvalitu predikování námi vytvořeného modelu. Opět lze samozřejmě použít předchozí hodnotící kritéria, nicméně pro hodnocení kvality predikce jsou vhodné i další ukazatele.

- Střední vychýlení AFE (Average Forecasting Error), 4.50 hodnotí vychýlení našich predikcí. Pokud se tento ukazatel pohybuje v kladných hodnotách, námi vytvořený model prognózy systematicky nadhodnocuje, naopak záporné hodnoty značí systematické podhodnocování.

$$AFE = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)}{n} \quad (4.50)$$

- Theilův koeficient nesouladu TIE (Theil's Inequality Coefficient), 4.51 rovněž slouží k hodnocení kvality prognóz. Vychází z hodnocení středních kvadratických odchylek. Po odmocnění ukazuje TIE procentní odchylku prognózy od skutečné hodnoty.

$$TIE^2 = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{\sum_{t=1}^n y_t^2} \quad (4.51)$$

Kapitola 5

Ekonometrické modely

5.1 Ekonometrický model obecně

Na úvod této kapitoly je vhodné zmínit, čím se vědní obor ekonometrie zabývá. Ekonometrie je ekonomická vědní disciplína, která je založena na využití metod ekonomie, matematiky a statistiky při zkoumání a analýze vztahů mezi ekonomickými veličinami. Jedná se v podstatě o aplikaci matematických a statistických metod nad daty, která získáme na základě měření nebo pozorování reálných ekonomických veličin. Kromě samotné analýzy těchto veličin je ekonometrie také silným nástrojem pro ověření a hlubší pochopení ekonomické teorie. V této práci využijí ekonometrické metody pro analýzu vývoje HDP a výše investic do telekomunikačního odvětví České republiky. V těchto, ale i všech ostatních případech užití ekonometrických metod, je základem sestavit výchozí ekonomickou hypotézu, tzv. ekonomický model.

Ekonomický model je aparát, který nám umožňuje využít matematické a statistické nástroje pro formální zavedení teoretických ekonomických vztahů. Při samotné tvorbě ekonomického modelu je naším cílem nalézt co nejvěrnější vyjádření počáteční hypotézy avšak s co nevyšší možnou mírou zjednodušení. Výsledkem této specifikace počátečního ekonomického modelu je ekonometrický model, který bude hlavním nástrojem dalšího zkoumání. Model je tvořen algebraickými vztahy, například rovnicemi stochastického typu, které popisují základní ekonomickou hypotézu [7].

5.1.1 Formulace a specifikace modelu

Formulace a specifikace ekonometrického modelu vyžaduje, jak vyplývá už z předchozích odstavců, znalost ekonomické teorie i matematického aparátu pro konstrukci požadovaného modelu. Stejně jako v případě analýzy časových řad není v této problematice možné uvést konkrétní postup specifikace modelu. Záleží opět na typu zkoumaných dat, zkušenostech analytika a také na požadavcích na výsledný model. Obvyklým postupem může být také specifikace modelu několika způsoby, kdy se nejvhodnější varianta vybere na základě simulace modelu. Abychom byli schopni specifikovat ekonometrický model, musíme projít postupně tři fáze.

Určení typu proměnných

Prvním a zásadním krokem při specifikaci ekonometrického modelu je určení typu proměnných. Proměnné využití v modelu můžeme rozdělit několika způsoby. Pro účely této práce popíšeme pouze dva způsoby dělení proměnných, které budou pro tvorbu modelu klíčové.

Prvním typem dělení je rozdělení na endogenní a exogenní proměnné. Endogenní proměnnou rozumíme takovou proměnnou, která je generována samotným modelem, jedná se tedy v podstatě o vnitřní proměnné modelu, které mohou v průběhu výpočtu měnit své hodnoty. Naproti tomu exogenní proměnné vstupují do modelu zvenčí, mají vliv na samotnou funkci modelu, ale jejich hodnota není modelem měněna. Z ekonomické teorie přímo nevyplývá, jak určit tento typ proměnných, proto velmi záleží na cílové funkčnosti konkrétního modelu, podle které je typ proměnných určen.

Druhým typem dělení je rozdělení proměnných na závislé proměnné a nezávislé proměnné. Na základě předchozího odstavce můžeme říci, že exogenní proměnné jsou vždy nezávislé, nebo jak je v literatuře uváděno, vysvětlující. Proměnné endogenní mohou být jak nezávislé, tak závislé, vysvětlované.

Abychom sestavovaný model a jeho rovnice zbytečně nezatěžovali méně podstatnými vlivy a proměnnými, zahrnujeme do něj obvykle pouze nejdůležitější nezávislé proměnné. Vliv ostatních méně podstatných činitelů pak shrnujeme spolu s vlivem náhodných faktorů do tzv. stochastických složek, tedy proměnných, jejichž hodnota vzniká náhodně a nelze je získat měřením nebo pozorováním. Součástí modelu pak ale musí být hypotézy o charakteru rozdělení těchto náhodných složek, které musíme otestovat [5].

Stanovení předpokládaných hodnot

Druhým krokem specifikace modelu je stanovení předpokládaných znamének a hodnot parametrů modelu. Tyto předpoklady vycházejí většinou z teoretického základu, případně z jiných modelových situací, ze kterých je možné stanovit, s jakým znaménkem bude parametr v modelu vystupovat. U hodnot můžeme předpokládat, opět na teoretickém základě, jaké hodnoty bude parametr nabývat, případně určit intervalové meze těchto hodnot. Rovněž můžeme v některých případech dopředu stanovit očekávané hodnoty součtu, rozdílu případně podílu některých parametrů. Všechny tyto předem určené parametry mohou pomoci při následné ekonomické verifikaci modelu.

Volba formy modelu

Třetím a posledním krokem specifikace modelu je volba analytické formy modelu a jeho jednotlivých rovnic. Ani zde nemůžeme pokaždé vycházet z ekonomické teorie, protože z ní přímo nevyplývá jakou matematickou formu rovnice, případně rovnic, volit. V ekonometrii rozlišujeme tyto základní typy analytických modelů.

- Jednorovnicový model, který má obvykle tvar stochastického regresního modelu, tedy zobrazuje závislost jedné závislé endogenní proměnné na jedné nebo více exogenních proměnných a na stochastické (náhodné) složce.

- Vícerovnicový model je realizovaný soustavou rovnic, kde můžeme zkoumat buď jednotlivé rovnice jako v případě jednorovnicového modelu nebo se zabývat soustavou jako celkem.
- Simultánní model se skládá rovněž ze soustavy rovnic, které jsou navíc simultánně závislé. To znamená, že endogenní proměnné se v modelu objevují jako závislé i jako nezávislé proměnné a zároveň jsou určeny soustavou rovnic daného modelu. Tyto soustavy rovnic navíc dělíme na dvě skupiny. První z nich jsou soustavy interdependentní, kde jsou mezi endogenními simultánně závislými proměnnými přímé a nepřímé zpětné vazby. Druhým typem jsou soustavy rekurentní, kde jsou mezi endogenními proměnnými pouze jednosměrné vazby a stochastické složky v jednotlivých rovnicích jsou vzájemně nezávislé.

Důležitou charakteristikou rovnic ekonometrického modelu je také jejich tvar. V běžných případech se snažíme volit, pokud možno, takový tvar rovnic, který bude obsahovat pouze lineární parametry, případně bude možné jednoduchou transformací přejít k lineární podobě parametrů. Tento požadavek vychází hlavně z toho, že lineární parametry lze snadno interpretovat a odhadovat a tato forma je pro člověka nejpřirozenější. Vzhledem k možnostem moderních výpočetních nástrojů ale není tento požadavek vždy nutné dodržovat [5].

5.1.2 Kvantifikace modelu

Kvantifikací ekonometrického modelu rozumíme krok analýzy, při kterém získáváme číselné hodnoty parametrů modelu a to včetně hodnot stochastických (náhodných). Základem tohoto kroku je vhodná volba ekonometrického odhadového postupu. V rozsahu termínu kvantifikace modelu rozumíme i proces získávání statistických dat ve vhodném množství a kvalitě, které následně do modelu vstupují.

Následujícím krokem je vlastní odhad parametrů modelu. Zde je stěžejní především výběr vhodného odhadového postupu, který bude akceptovat všechna zaznamenaná vstupní data. Jak již bylo zmíněno dříve, u modelů jednorovnicových je situace prostá, u vícerovnicových modelů lze k rovnicím v soustavě přistupovat individuálně nebo ke všem současně. Odhadové metody individuálního přístupu k jednotlivým rovnicím nazýváme obvykle metody s omezenou informací a metody pro simultánní přístup k celému modelu pak označujeme jako metody s úplnou informací. Klíčovými parametry výběru odhadové metody jsou například kvalita a kvantita vstupních dat, účel použití odhadovaného modelu nebo například výpočetní náročnost daného typu modelu.

Na základě těchto odhadnutých parametrů, mezi které patří jak všechny regresní koeficienty, tak odhadnuté parametry rozdělení stochastických složek, a vstupních dat jsme schopni vyčíslit teoretické hodnoty závislých endogenních proměnných v modelu. V této fázi jsme zkonstruovali požadovaný ekonometrický model.

5.1.3 Verifikace modelu

Ve fázi verifikace modelu vycházíme z odhadnutého modelu, který jsme získali v předchozím kroku, tedy kvantifikací modelu. Naším cílem je nyní ověřit, zda odhadnuté parametry jsou správné a rovněž statisticky významné. Výchozím zdrojem pro ověření správnosti je teoretická základna jednak ekonomická, dále statistická a ekonometrická. Podle těchto oborů také verifikaci modelu členíme.

Ekonomická verifikace

V rámci ekonomické verifikace ověřujeme správnost modelu vzhledem k ekonomické teorii popisující modelovaný jev. Ověřujeme především shodu vypočtených parametrů s předpokládanými znaménky jednotlivých parametrů a ověřujeme, zda stanovené numerické hodnoty odpovídají stanoveným hodnotám či intervalům z fáze specifikace modelu. Ověříme tak správnost vypočtených parametrů a zároveň interpretovatelnost v návaznosti na ekonomickou teorii, ze které model vychází. V případě že tento krok verifikace proběhne s kladnými výsledky, je model odpovídajícím zobrazením zkoumaného ekonomického jevu. V opačném případě je model nevyhovující a je nutné provést specifikaci jiným způsobem.

Statistická verifikace

Statistická verifikace, též testy prvního řádu, umožňuje statisticky posoudit kvalitu samotného ekonometrického modelu i jeho jednotlivých parametrů. Základním nástrojem tohoto kroku verifikace je využití metodik statistických testovacích kritérií. Nejběžnější používaná kritéria tohoto druhu jsou koeficienty korelace, výpočet chyby odhadu parametrů a autokorelační koeficienty reziduí. Cílem je tedy určit věrnost chování modelu na základě statistické analýzy. Statistická verifikace by měla následovat vždy až po ekonomické verifikaci, a to zejména proto, že jednotlivé kvalitativní statistické testy mohou pro daný model vycházet pozitivně i při neodpovídajícímu chování vzhledem k teoretické rovině modelovaného ekonomického jevu.

Ekonometrická verifikace

Ekonometrické verifikace, také testy druhého řádu, nám umožní ověřit předpoklady nutné k použití ekonometrických metod. Jak již napovídá název, testy druhého řádu, navazují na testy prvního řádu, tedy statistické testy. Cílem ekonometrického testování je především ověření statistických testů na základě ekonometrických kritérií, mezi které patří například podmínky identifikovatelnosti simultánních rovnic, kolineární testování nezávislých proměnných a kritéria pro posouzení síly statistických testů. Ty jsou na místě zejména v případě, kdy máme k dispozici pouze omezený rozsah vstupních dat a statistické testy by na těchto datech mohly vykazovat zkreslené výsledky.

5.1.4 Aplikace modelu

V předchozích odstavcích popsaná snaha o tvorbu ekonometrického modelu s maximální možnou věrohodností by měla vést především k jeho praktickému uplatnění. Jak již bylo zmíněno na začátku kapitoly, jednotlivé způsoby uplatnění modelu mohou být různé. Model tak můžeme využít především ve třech oblastech.

- Analýza vývoje v minulosti slouží k rozboru chování řady vstupních dat, která máme k dispozici. Můžeme tak například ověřit věrnost popisu modelu a provádět různé citlivostní analýzy na základě změny jeho parametrů.
- Predikce budoucích hodnot nám umožní odhadnout numerické hodnoty endogenních proměnných v budoucím vývoji. Předpokladem pro tento odhad je konstantní chování modelu v čase a též aplikovatelnost nezávislých proměnných i v období pro které chceme s modelem hodnoty predikovat. Stejně jako v případě analýzy časových řad je nutné otestovat prognostické schopnosti simulací na známých datech a posoudit vhodnost a přesnost této predikce.
- Ekonomické rozhodování můžeme pomocí ekonomického modelu analyzovat a optimalizovat například tak, že zkoumáme vliv změny hodnot nezávislých řídicích proměnných na hodnoty cílových endogenních proměnných. Model nám tak může napomoci ve volbě vhodné varianty plánu či opatření, abychom dosáhli požadovaných ekonomických výsledků

5.1.5 Data vstupující do ekonometrického modelu

Vstupní data pro ekonometrickou analýzu mohou být různého druhu. Podstatné je především rozdělení těchto dat na data kvalitativní a kvantitativní povahy. Převážná většina dat v ekonomii má charakter kvantitativní, protože se jedná zpravidla o výši některého pozorovaného ukazatele. Tyto data získáváme především pozorováním a měřením. Abychom mohli do modelu zavádět i data kvalitativního typu, musíme je nejprve kvantifikovat pomocí tzv. umělých proměnných.

Typickým příkladem hodnot, které je nutné kvantifikovat, je například pohlaví, národnost či kvalifikace. Tyto umělé proměnné nabývají obvykle dvou možných hodnot, ale nemusí to být pravidlem. Obecně lze použít umělou proměnnou pro všechny kvalitativní i kvantitativní proměnné, pro které jsme schopni všechny jejich hodnoty roztřídit do vzájemně vylučujících se skupin [7].

5.2 Konstrukce jednorovnicového modelu

V následujících odstavcích stručně teoreticky uvedu konstrukci základního lineárního jednorovnicového modelu. Takovýto model přímo zachycuje lineární závislost jedné endogenní proměnné na více exogenních nezávislých proměnných. Navíc je v něm zahrnuta aditivní složka, která představuje náhodnou složku. Rovnice 5.1 zachycuje obecný tvar takového modelu, který může být jak statického, tak dynamického typu.

$$Y = B_0 + x_1 B_1 + x_2 B_2 + \dots + x_k B_k + u \quad (5.1)$$

Význam proměnných v obecném modelu 5.1 je následující:

- Y představuje závislou endogenní proměnnou
- x_i je i -tá nezávislá exogenní proměnná
- B_i je i -tý regresní koeficient
- u je náhodná složka

Rovnici můžeme vyjádřit rovněž v maticovém tvaru, jak zachycuje vzorec 5.2.

$$\mathbf{Y} = \mathbf{xB} + \mathbf{u} \quad (5.2)$$

Jak je patrné ze rovnice 5.1, jednotlivé odhadnuté parametry B_i reálně vyjadřují vliv změny odpovídající nezávislé exogenní proměnné x_i na hodnotu závislé endogenní proměnné Y . Předpokladem tohoto tvrzení je ale fakt, že jiné exogenní proměnné než ta s právě analyzovaným vlivem, jsou konstantní. Pokud chceme odhadnout samotný vliv b_i jedné konkrétní i -té nezávislé proměnné, použijeme parciální derivaci podle dané proměnné x_i , jak ukazuje rovnice 5.3.

$$b_i = \frac{\partial y}{\partial x_i}, i = 1, 2, \dots, k \quad (5.3)$$

Na základě dříve uvedené teorie ekonometrických modelů je zřejmé, že odhadovat parametry modelu jsme schopni pouze na základě znalosti statistických dat. Proto je nutné tvar modelu uzpůsobit tak, aby akceptoval všechna tato vstupní data. Forma reprezentace vstupních dat je maticová a vychází z toho, že každý řádek matice je množina hodnot všech nezávislých proměnných v jednom okamžiku pozorování a každý sloupec matice je množina všech pozorovaných hodnot jedné proměnné. Tvar takového modelu je zachycen rovnicí 5.4 a 5.5.

Počet pozorování tedy značíme jako n a počet parametrů zachycuje proměnná k . Rozdíl mezi těmito hodnotami $n - k$, kde $n > k$ nazýváme stupněm volnosti modelu.

$$\mathbf{Y} = \mathbf{XB} + \mathbf{u} \quad (5.4)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{bmatrix} \cdot \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad (5.5)$$

Význam proměnných v modelu 5.5 je pak následující:

- \mathbf{Y} představuje vektor hodnot závislé endogenní proměnné
- \mathbf{X} je matice zaznamenaných hodnot nezávislých exogenních proměnných
- \mathbf{B} je vektor regresních koeficientů
- \mathbf{u} je vektor náhodné složky z každého pozorování

Pro odhad parametrů B_i tohoto modelu 5.5 využijeme metody nejmenších čtverců. Aby byly tyto odhady relevantní, je potřeba splnit následující podmínky.

1. V modelu vystupující stochastická složka má nezávislé rozdělení, nulový průměr 5.6 a konečný rozptyl σ^2 5.7. Střední hodnota vektoru \mathbf{u} , respektive jeho kovarianční matice splňuje podmínky:

$$E(\mathbf{u}) = \mathbf{0} \quad (5.6)$$

$$E(\mathbf{u} \cdot \mathbf{u}^T) = \sigma^2 \mathbf{I}_n \quad (5.7)$$

ve kterých \mathbf{u}^T je transponovaný vektor náhodné složky a \mathbf{I}_n je jednotková matice řádu n .

2. Všechny zaznamenaná hodnoty nezávislých exogenních proměnných, tedy matice \mathbf{X} , jsou nenáhodná čísla. Nejsou tedy závislá na stochastické složce \mathbf{u} , proto platí vztah 5.8.

$$E(\mathbf{X}^T \cdot \mathbf{u}) = \mathbf{0} \quad (5.8)$$

3. Všechny sloupce matice \mathbf{X} , a tedy i všechny nezávislé exogenní proměnné, jsou lineárně nezávislé. Matice \mathbf{X} tedy musí být regulární, 5.9 a matice 5.10 existuje.

$$h(\mathbf{X}) = h(\mathbf{X}^T \cdot \mathbf{X}) = k \quad (5.9)$$

$$(\mathbf{X}^T \cdot \mathbf{X})^{-1} \quad (5.10)$$

Pokud následující podmínky platí, na základě znalosti n hodnot endogenní proměnné a také k nezávislých exogenních proměnných jsme schopni odhadnout parametry lineárního modelu pomocí metody nejmenších čtverců [7].

5.2.1 Metoda nejmenších čtverců

V souladu s výše definovaným pravidlem, které říká, že střední hodnota vektoru náhodných složek je nulová, můžeme říci, že střední hodnota závislé endogenní proměnné se rovná pouze systematické složce modelu 5.11.

$$E\mathbf{Y} = \mathbf{X}\mathbf{B} + E(\mathbf{u}) = \mathbf{X}\mathbf{B} \quad (5.11)$$

Pokud nyní označíme odhadovou funkci vektoru koeficientů \mathbf{B} získanou metodou nejmenších čtverců jako \mathbf{b} , dostaneme se k tvaru 5.12, kde vektor reziduí \mathbf{e} odpovídá odhadu vektoru náhodné složky \mathbf{u} a vektor \mathbf{y} reprezentuje odhadové hodnoty endogenní proměnné, tedy platí tvar 5.13.

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (5.12)$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} \quad (5.13)$$

Odhad vektoru regresních parametrů \mathbf{B} získáme pak na základě definice odhadové funkce minimalizací čtverců reziduálních složek, které můžeme maticově zapsat jako 5.14. Parciální derivací výrazu podle \mathbf{b}^t pak získáme vztah vhodný pro určení odhadového vektoru \mathbf{b} . Jedná se o tzv. normální rovnice nejmenších čtverců, které zachycuje vztah 5.15.

$$\mathbf{e}^T \mathbf{e} = (\mathbf{Y} - \mathbf{Y}\mathbf{b})^T (\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}^T \mathbf{Y} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{X}\mathbf{b} \quad (5.14)$$

$$\mathbf{X}^T \mathbf{X}\mathbf{b} = \mathbf{X}^T \mathbf{y} \quad (5.15)$$

Po vyjádření odhadové funkce \mathbf{b} dospějeme k tvaru 5.16, kde matice $(\mathbf{X}^T \mathbf{X})^{-1}$ je maticí konstant, tedy můžeme konstatovat, že \mathbf{b} je sice lineární odhadovou funkcí, ale vzhledem k závislosti na náhodné složce je stochastického typu. To je jasně patrné po dosazení 5.4 do vztahu 5.16, tedy 5.17.

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5.16)$$

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\mathbf{b} + \mathbf{u}) = \mathbf{b} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \quad (5.17)$$

Konečně pro zhodnocení rozptylu a kovariance odhadové funkce \mathbf{b} musíme stanovit její kovarianční matici. Na základě předchozího pravidla o střední hodnotě se tak dostaneme ke vztahu 5.18.

$$\mathbf{V}(\mathbf{b}) = E(\mathbf{b} - \mathbf{B})(\mathbf{b} - \mathbf{B})^T = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{u}\mathbf{u}^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (5.18)$$

Abychom byli schopni tuto matici určit, potřebujeme znát odhad rozptylu náhodné složky. Skutečnou hodnotu nejsme schopni numericky vyčíslit, protože hodnoty náhodné složky nejsme schopni zaznamenat pozorováním. Proto si označíme odhad rozptylu σ^2 jako

s^2 . Na základě vztahu 5.17 vyjádříme vektor reziduí vztahem 5.19. Po úpravách se pak dostaneme k vyjádření střední hodnoty reziduálních složek ve tvaru 5.20. Odhad rozptylu pak je ve tvaru 5.21 a vychází z toho, že platí $E(s^2) = \sigma^2$.

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = (\mathbf{X}\mathbf{b} + \mathbf{u}) - \mathbf{X}[\mathbf{b} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}] \quad (5.19)$$

$$\sigma^2 = \frac{E(\mathbf{e}^T\mathbf{e})}{n - k} \quad (5.20)$$

$$s^2 = \frac{\mathbf{e}^T\mathbf{e}}{n - k} \quad (5.21)$$

Nyní, po naplnění všech podmínek, lze jednak vyčíslit ale i ověřit kovarianční matici a tím aplikovat i odhadovou funkci \mathbf{b} pro odhad paramterů lineárního regresního jednorovnicového modelu [7] [5].

Kapitola 6

Analýza časových řad prakticky

Prvním krokem praktické části mé práce bude analýza časových řad HDP a investic do telekomunikačního odvětví. Tyto analýzy jsou potřebné především pro získání modelů časových řad, na jejichž základě budu dále vytvářet předpovědi budoucího vývoje a rovněž následný model zohledňující vztah výše HDP a investic.

V následujícím textu bude vždy popsán postup ruční analýzy pomocí Excelu, ukázka volby vhodného modelu na základě zhodnocení kvality vyrovnání a následně pak postup analýzy ve statistickém softwaru R ¹. R jsem zvolil pro jednoduchost použití a efektivní a robustní práci s daty. V textu uvedené výpočty slouží pro ilustraci a zachycují především hlavní charakteristiky dat. Podrobné výsledky výpočtů budou uvedeny na CD v příloze k diplomové práci.

6.1 Analýza časové řady HDP

6.1.1 Získání vstupních dat

Data pro analýzu časové řady HDP jsem získal z databáze České národní banky, a to v kvartálním rozlišení. Zdroj jsem zvolil na základě požadavku získat data neočištěná, zdroje uvedené v teoretické části práce obsahovaly pouze časové řady HDP sezónně očištěné. Ve výpočtech vystupují hodnoty od roku 1995 do roku 2014. Oproti patnáctiletému horizontu danému v zadání jsem zvolil pro pracovní výpočty časový rámec delší, abych mohl přesněji ověřit odchylku předpovědi a chování modelu na větším množství dat.

Hodnoty HDP v milionech Kč jsou uvedeny v příloze B.1. Jedná se o data zaznamenaná kvartálně, nejsou nijak očištěna.

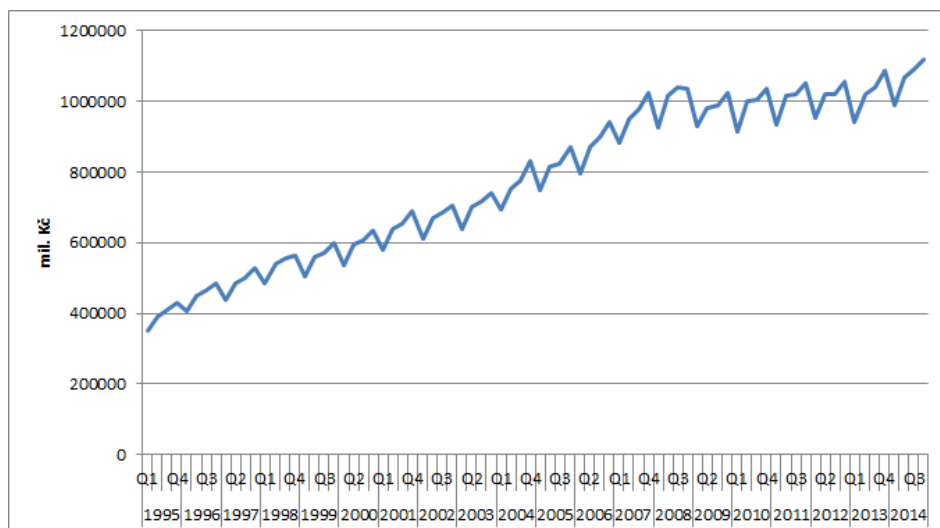
6.1.2 Základní parametry časové řady

Pro prvotní představu o chování dané časové řady je důležité a výhodné zobrazit ji před samotnou analýzou graficky a vypočítat některé základní popisné charakteristiky. Časovou řadu HDP z let 1995 až 2014 zobrazuje graf 6.1 a následující tabulka 6.1 pak shrnuje základní popisné charakteristiky. Jak je z grafu patrné, pro zadanou časovou řadu bude vhodné zvolit aditivní dekompozici.

¹<http://www.r-project.org/>

6.1.3 Dekompozice časové řady HDP

Při dekompozici časové řady HDP vyjdeme z postupu, který jsem již dříve popsal v teoretické části této práce. Dále postup stručně shrnu a uvedu parametry jednotlivých výpočtů. Podrobné tabulky s výpočty jsou uvedeny v přílohách.



Obrázek 6.1: Časová řada kvartálního vývoje HDP [?]

6.1.3.1 Trendová složka

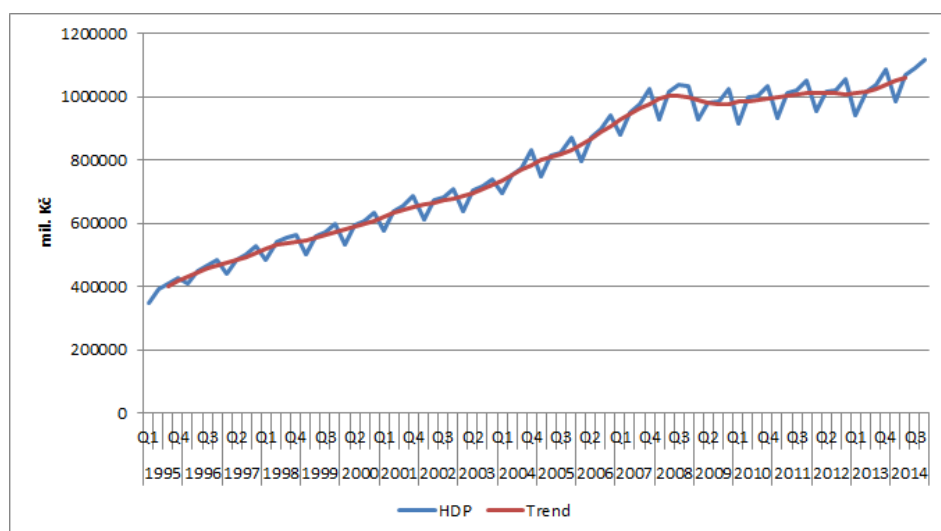
Prvním krokem očištění časové řady za účelem zisku sezónní složky, jejíž odstranění je pro další postup nejdůležitější, je analýza trendové složky. Trend vypočteme pomocí klouzavých průměrů. Vzhledem k tomu, že data jsou uvedena kvartálně, vhodná délka klouzavých průměrů je 4. Protože se tímto postupem dostaneme do situace, kdy vypočtená hodnota leží mezi intervaly, je nutné provést ještě centrování klouzavých průměrů. Takto získané hodnoty již představují trendovou složku časové řady. Výsledek výpočtu zobrazuje graf 6.2. Jak je vidět, trendová složka poměrně věrně modeluje chování časové řady HDP, můžeme, ji proto využít pro další výpočty.

6.1.3.2 Sezónní složka

Dalším krokem je výpočet sezónní složky časové řady. Po tom, co jsme získali hodnoty trendové složky, se jejím odečtením od původní časové řady a následným výpočtem průměrů hodnot odpovídajících kvartálů dostáváme právě k sezónní složce. Při klasické dekompozici je pro ni typické, že její perioda je shodná s periodou původní časové řady. V našem případě je perioda sezónní složky 4. Sezónní složku zobrazuje graf 6.3, řadu očištěnou o sezónní složku pak graf 6.4. Sezónní složka nedosahuje vzhledem k velikosti hodnot časové řady extrémních výkyvů, ale rozhodně není zanedbatelná a výsledný model jistě ovlivní.

HDP - parametry řady	
Střední hodnota	776 311,70
Chyba střední hodnoty	24 857,72
Medián	787 096,50
Modus	N/A
Směrodatná odchylka	222 334,20
Rozptyl výběru	4,94E+10
Špičatost	-1,36
Šikmost	-0,19
Rozdíl max - min	768 408
Minimum	349 683
Maximum	1 118 091
Součet	62 104 933
Počet	80

Tabulka 6.1: Popisná statistika časové řady HDP

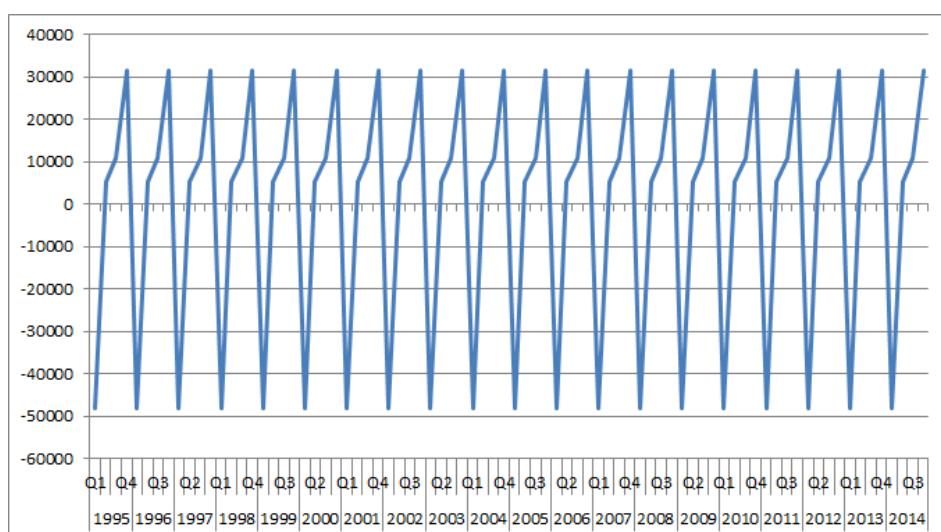


Obrázek 6.2: Trendová složka časové řady HDP

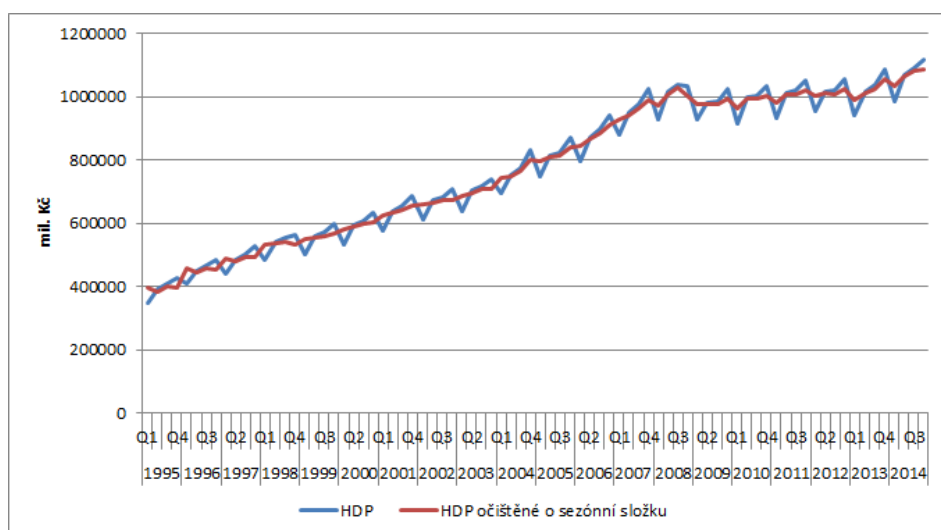
6.1.3.3 Náhodná a cyklická složka

Posledním krokem dekompozice je výpočet náhodné a cyklické složky. Pokud odečteme od původní časové řady výše získanou trendovou a sezónní složku, zbytkem budou právě náhodná a cyklická složka.

Pro izolaci cyklické složky využijeme opět klouzavých průměrů, které budou, jak je uvedeno již v teoretické části, délky kratší než perioda časové řady, ale zároveň tak dlouhé, aby byl vyloučen vliv dalšího sezónního vychýlení. Pro náš případ čtvrtletně zaznamenaných hodnot je vhodné zvolit délku klouzavých průměrů 3, protože eliminuje možné pololetní



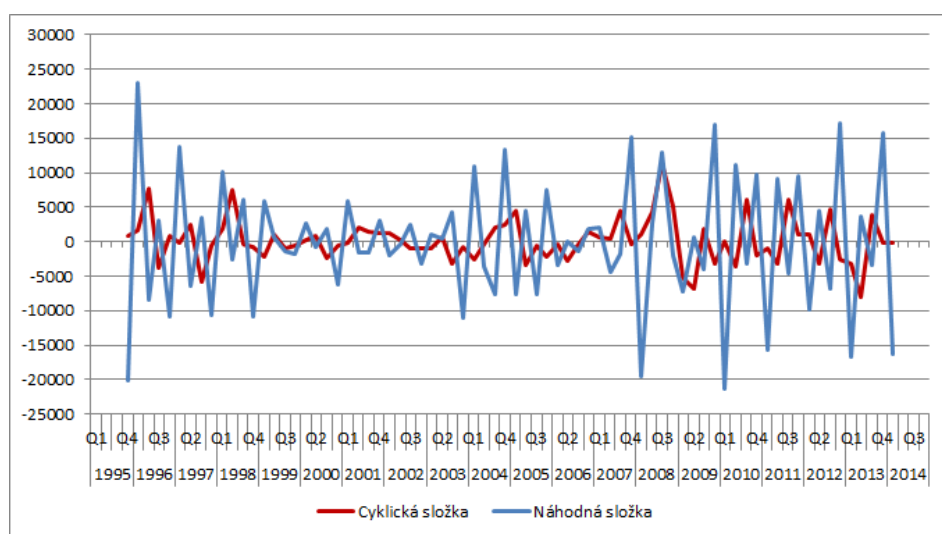
Obrázek 6.3: Sezónní složka časové řady HDP



Obrázek 6.4: Časová řada HDP očištěná o sezónní složku

vychýlení složek. Takto získaná data představují cyklickou složku časové řady.

Náhodnou složku poté získáme prostým odečtem cyklické složky od dat použitých pro její výpočet. Následující graf 6.5 shrnuje obě tyto složky časové řady. Vzhledem k charakteristice cyklické složky, kterou jsem již popsal v teoretické části práce, nejsme schopni přesně určit její závislost na časovém období a tak stanovit její periodu. Pro další výpočty ji proto nebudeme samostatně uvažovat. Zde je pro ilustraci uvedena úplná dekompozice časové řady.



Obrázek 6.5: Cyklická a náhodná složka časové řady HDP

6.1.4 Tvorba modelu

Dekompozice časové řady je prvním krokem, kde oddělíme jednotlivé složky, ze kterých se časová řada skládá. Jejich znalost je potřebná pro další práci s časovou řadou jako konstrukce modelu a následné předpovídání. Dalším úkolem je vytvoření modelu trendu časové řady.

Určitý druh modelu jsme vytvořili již při dekompozici a identifikaci trendové složky pomocí klouzavých průměrů. Tento druh modelování je ale nevhodný například pro předpovídání budoucích hodnot proto, že nejsme schopni identifikovat funkci, pomocí které další hodnoty vypočteme. Jak jsem popsal již v teoretické části, při odhadu budoucích hodnot pomocí klouzavých průměrů se dostáváme k přibližně konstantnímu trendu, což není pro většinu časových řad vhodné. Dalším omezením je také ztráta $n/2$ počátečních a koncových hodnot časové řady, což je dáno povahou výpočtu klouzavého průměru délky n .

V následujících odstavcích se budu zabývat postupem tvorby vhodného modelu a zhodnocením jeho chování vzhledem k původní časové řadě. Pro ukázkou tvorby modelu využiji tři základní postupy. Prvním bude modelování trendu pomocí přímky, druhým pak tvorba modelu zvoleného na základě popisných charakteristik časové řady a třetím přístupem bude adaptivní model reprezentovaný dvojitým exponenciálním vyrovnáním.

U všech těchto modelů budeme vycházet z časové řady očištěné od sezónní složky, ke které jsme se dostali v předchozích odstavcích při její dekompozici.

6.1.4.1 Modelování trendu přímkou

Pro vytvoření modelu pomocí regresní přímky využijeme v Excelu dostupný nástroj Analýza dat / Regrese. Vstupem pro tento výpočet budou hodnoty $1 - t_n$ označující pořadí hodnoty sezónně očištěné časové řady jako konstanta a samotné hodnoty sezónně očištěné časové řady jako závislá proměnná, jak ilustruje tabulka 6.2.

Výsledkem regrese pak získáme vypočtené konstanty β_0 a β_1 .

t_i	rok	čtvrtletí	sez. očištěná hodnota HDP v mil. Kč
1	1995	Q1	397 678,60
2		Q2	385 015,20
3		Q3	399 754,80
4		Q4	398 064,70
5	1996	Q1	456 368,60
6		Q2	445 298,20
7		Q3	456 137,80
8		Q4	455 215,70
⋮	⋮	⋮	⋮

Tabulka 6.2: Vstupní data pro regresní analýzu - přímka

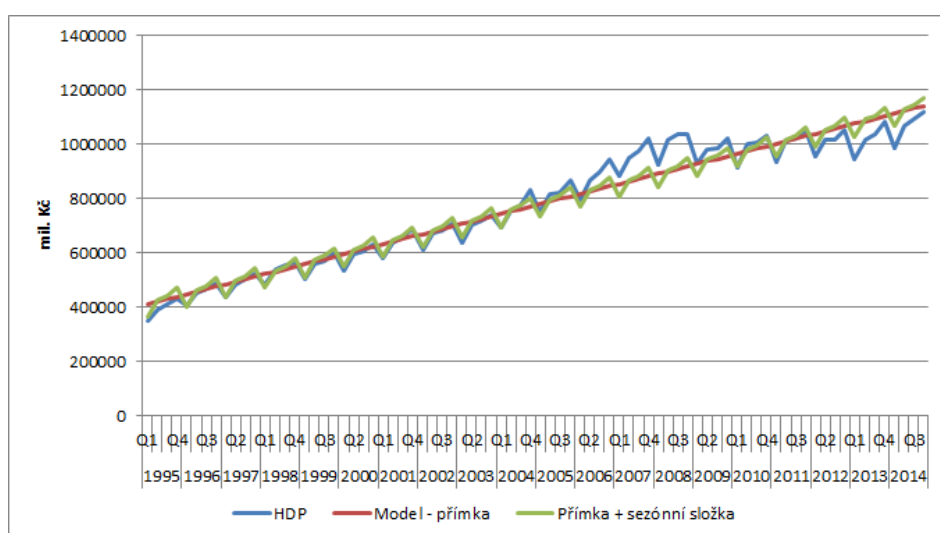
$$\beta_0 = 402\,108,200 \quad (6.1)$$

$$\beta_1 = 9\,242,049 \quad (6.2)$$

Z těch pak již snadno sestavíme rovnici regresní přímky 6.3, která bude mít po vyčíslení tvar 6.4.

$$HDP_{model} = \beta_0 + \beta_1 t_i \quad (6.3)$$

$$HDP_{model} = 402\,108,200 + 9\,242,049 t_i \quad (6.4)$$



Obrázek 6.6: Model HDP vytvořený regresní přímkou

Díky znalosti této rovnice dále snadno dopočítáme jednotlivé hodnoty regresní přímky modelující trend. Dalším krokem je přičtení sezónní složky k takto získaným hodnotám a tím i získání konečného modelu pomocí regresní přímky. Samotné vyrovnání přímkou a započtení vlivu sezónní složky zachycuje graf 6.6.

Abychom byli schopni dále zhodnotit kvalitu modelu a jednotlivé modely porovnat, na závěr vypočteme pro daný model součet čtvercových chyb (SSE) 6.5.

$$SSE_{\text{přímka}} = 143\,568\,549\,380,51 \quad (6.5)$$

6.1.4.2 Modelování trendu exponenciálou

Pro výběr druhé křivky reprezentující model trendu je využito výběru pomocí popisných charakteristik časové řady. Po zhodnocení první a druhé difference, jejich podílu a koeficientu růstu řady jsem jako druhou matematickou funkci vybral exponenciálu. Toto rozhodnutí jsem učinil na základě toho, že koeficient růstu sezónně očištěné časové řady kolísá kolem konstantní hodnoty 1.

Pro vytvoření modelu pomocí regresní exponenciály využijeme opět nástroj Excelu Analýza dat / Regrese. Vzhledem k tomu, že funkce exponenciálního trendu má rovnici 6.6,

$$HDP_{\text{model}} = \beta_0 \cdot \beta_1^{t_i} \quad (6.6)$$

je nutné provést na počátku převod rovnice na lineární tvar logaritmováním a po výpočtu koeficientů β opět provést odlogaritmování. Logaritmováním převedeme problém na lineární rovnici a budeme jej řešit podobně jako v minulém případě. Do regresní analýzy tedy budou vstupovat jako závislá proměnná logaritmus hodnoty časové řady a dále opět hodnoty $1 - t_n$ označující pořadí hodnoty sezónně očištěné časové řady jako konstanta, jak je vyznačeno v tabulce 6.3.

t_i	rok	čtvrtletí	sez. očištěná hodnota HDP v mil. Kč	log očištěné hodnoty
1	1995	Q1	397 678,60	5,599 532
2		Q2	385 015,20	5,585 478
3		Q3	399 754,80	5,601 794
4		Q4	398 064,70	5,599 954
5	1996	Q1	456 368,60	5,659 316
6		Q2	445 298,20	5,648 651
7		Q3	456 137,80	5,659 096
8		Q4	455 215,70	5,658 217
⋮	⋮	⋮	⋮	⋮

Tabulka 6.3: Vstupní data pro regresní analýzu - exponenciála

Výsledkem regrese získáme vypočtené logaritmy konstanty β_0 a β_1 .

$$\log_{10}\beta_0 = 5,645356 \quad (6.7)$$

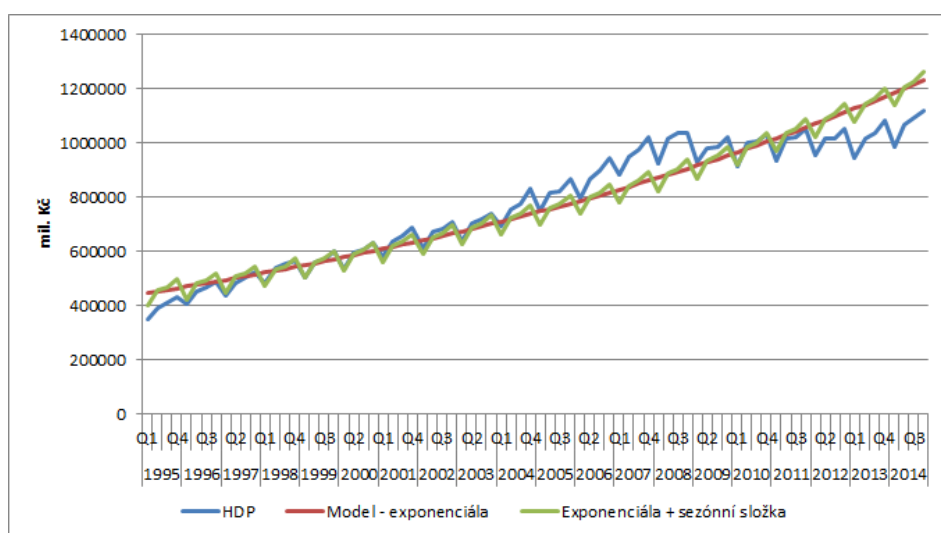
$$\log_{10}\beta_1 = 0,005571 \quad (6.8)$$

Po odlogaritmování získáme hodnoty

$$\beta_0 = 441\,932,300 \quad (6.9)$$

$$\beta_1 = 1,013 \quad (6.10)$$

Stejně jako v případě přímky se znalostí této rovnice dále dopočítáme jednotlivé hodnoty regresní exponenciály modelující trend a opět přičteme sezónní složku. Tím získáme požadovaný model. Samotné vyrovnání exponenciálou a započtení vlivu sezónní složky zachycuje graf 6.7.



Obrázek 6.7: Model HDP vytvořený regresní exponenciálou

Pro porovnání modelů a zjištění kvality vyrovnání opět vypočteme pro model součet čtvercových chyb (SSE) 6.11.

$$SSE_{\text{exponenciála}} = 355\,922\,601\,740,08 \quad (6.11)$$

6.1.4.3 Modelování trendu dvojitým exponenciálním vyrovnáním

Jako zástupce adaptivních metod vyrovnání jsem vybral dvojitě exponenciální vyrovnání, které je přesnější než vyrovnání jednoduché a zároveň je pro ruční výpočet dostatečně jednoduché.

Prvním krokem je volba vhodného koeficientu α . Pro prvotní odhad jsem zvolil koeficient $\alpha = 0,7$. Dále provedeme výpočet vyrovnávacích statistik $S1$ 6.12 a $S2$ 6.13.

$$S1_t = (1 - \alpha)y_t + \alpha S1_{t-1} \quad (6.12)$$

$$S2_t = (1 - \alpha)S1_t + \alpha S2_{t-1} \quad (6.13)$$

Pro jednoduchost jsem zvolil jako první hodnotu $S1$ a $S2$ odpovídající první hodnotu sezónně očištěné časové řady HDP. Na jejich základě jsou pak vypočteny hodnoty koeficientů β_0 a β_1 .

$$\beta_0(t) = 2S1_t - S2_t \quad (6.14)$$

$$\beta_1(t) = \frac{1 - \alpha}{\alpha}(S1_t - S2_t) \quad (6.15)$$

Tedy pro tento konkrétní model 6.16 a 6.17.

$$\beta_0 = 1\,087\,834 \quad (6.16)$$

$$\beta_1 = 11\,731,69 \quad (6.17)$$

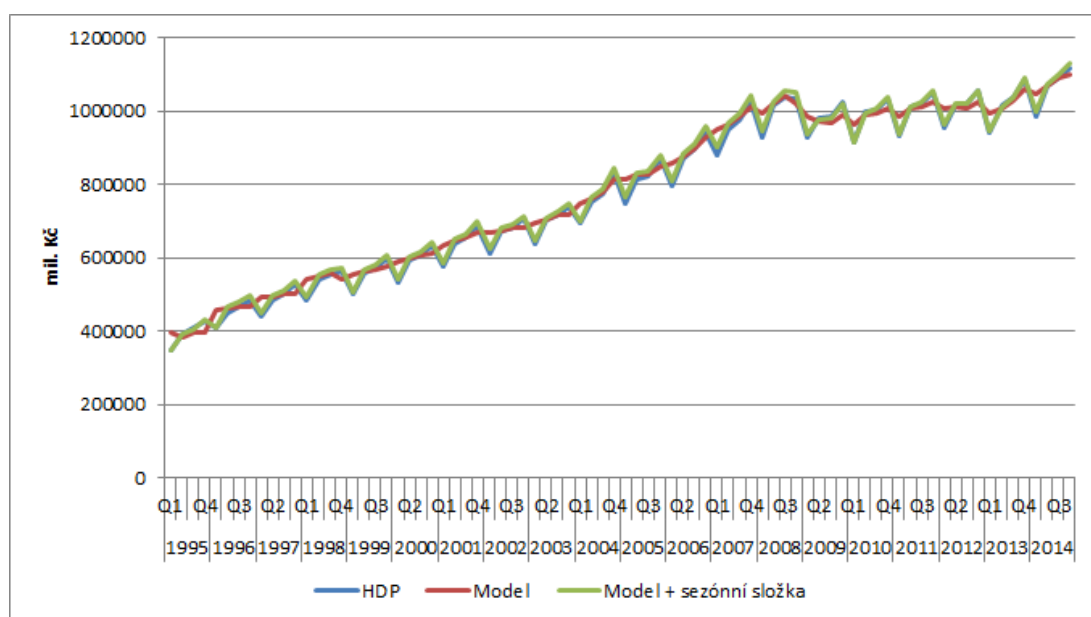
Pro takto vytvořený model vypočítáme hodnotu SSE a pomocí nástroje Řešitel v Excelu provedeme minimalizaci této hodnoty SSE na základě změny koeficientu α . Tím zajistíme volbu optimálního koeficientu. Pro danou časovou řadu HDP jsem tímto postupem získal hodnotu koeficientu 6.18:

$$\alpha = 0,497324 \quad (6.18)$$

S takto získanou hodnotou koeficientu přepočítáme hodnoty modelu a získáme tak optimální trendovou funkci. Dále následuje již přičtení sezónní složky a výsledný model je zobrazen v grafu 6.8.

Již dříve vypočtená hodnota SSE je pro tento model 6.19.

$$SSE_{dvojiteexp} = 8\,589\,454\,314 \quad (6.19)$$



Obrázek 6.8: Model HDP vytvořený dvojitým exponenciálním vyrovnáním

6.1.4.4 Porovnání kvality modelů

Jednotlivé výše získané modely nyní srovnáme prostřednictvím vypočtených hodnot SSE. Na základě hodnot v tabulce 6.4 je vidět, že chyba modelu vytvořeného pomocí dvojitého exponenciálního vyrovnání je o dva řády menší než u ostatních modelů, a proto se tento model jeví jako zdaleka nejlepší z těchto tří. Tento závěr je možné udělat již na základě grafu 6.8, kde je vidět značná blízkost hodnot původní časové řady a modelu s připočtenou sezónní složkou. Pro další výpočty a prognózování proto využijí právě tento model.

Model	SSE
Přímka	143 568 549 380,51
Exponenciála	355 922 601 740,08
Dvojité exponenciální vyrovnání	8 589 454 314,00

Tabulka 6.4: Porovnání kvality modelů HDP

6.1.5 Dekompozice v softwaru R

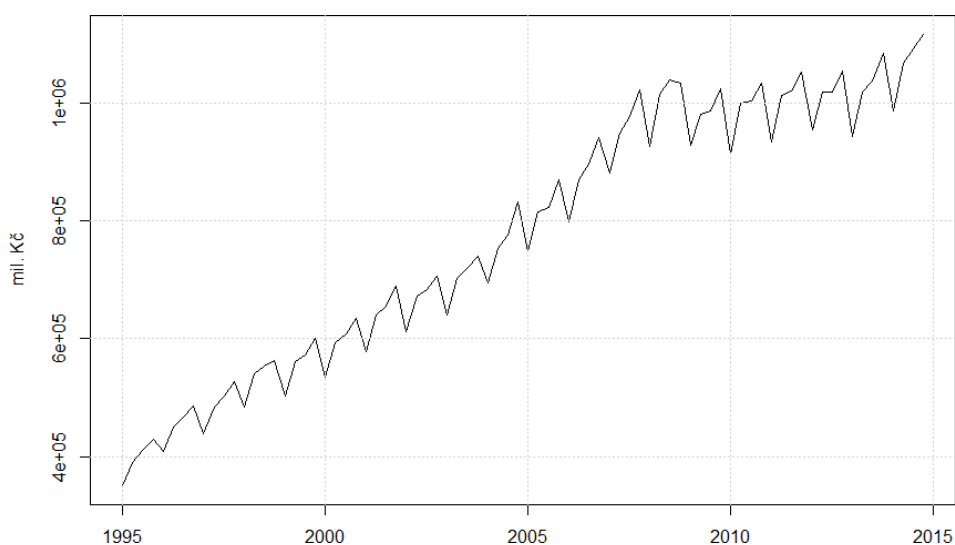
V následujících odstavcích provedu dekompozici a tvorbu modelu časové řady HDP ve statistickém softwaru R. V průběhu textu zmíním jednotlivé příkazy v konzoli R tak, aby bylo možné snadno postup zopakovat.

Prvním krokem bude načtení vstupních dat a jejich převedení do tvaru časové řady s periodou 4 a počátkem v roce 1995.

```
HDP <- scan("hdp-src.txt")
HDPts <- ts(HDP, frequency=4, start=c(1995,1))
```

Časovou řadu nyní můžeme snadno zobrazit v grafu 6.9.

```
plot.ts(HDPts, ylab="mil. Kč", xlab="", main="HDP ČR")
```



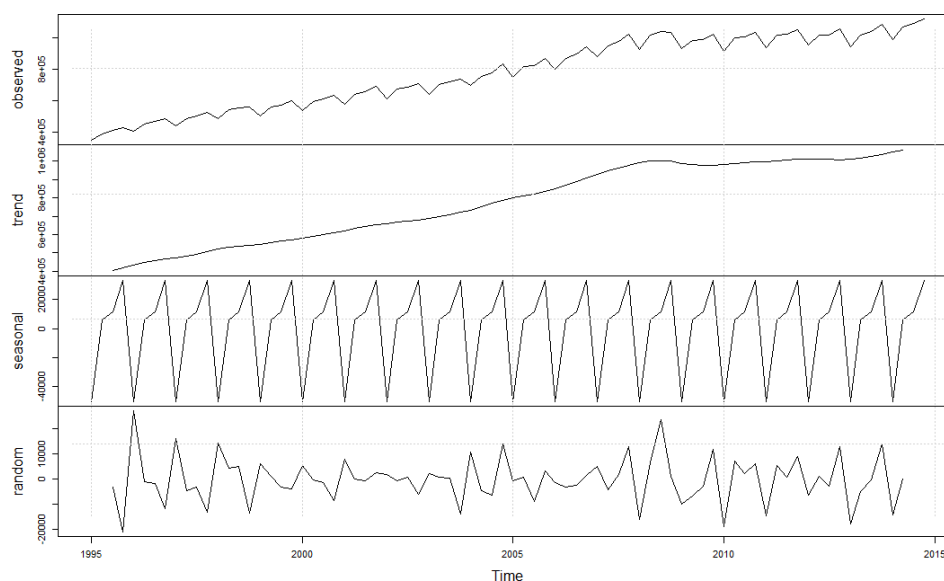
Obrázek 6.9: Časová řada HDP v R

Druhým krokem bude dekompozice časové řady na jednotlivé složky. V případě, že časová řada obsahuje sezónní složku, využijeme pro dekompozici časové řady funkci `decompose()`, která využije k nalezení trendu klouzavé průměry. Grafické zobrazení dekompozice je vidět v grafu 6.10.

```
HDPtsComp <- decompose(HDPts, type=c("additive"), filter = NULL)
```

6.1.6 Tvorba modelu v R

Díky analýze provedené v Excelu jsme došli k závěru, že vhodným modelem pro časovou řadu HDP je model vytvořený dvojitým exponenciálním vyrovnáním. Pomocí software R se proto nebudeme zabývat tvorbou modelu pomocí matematických funkcí, ale využijeme výpočetních možností software pro vytvoření modelu pomocí dvojitého a trojitého exponenciálního vyrovnání.



Obrázek 6.10: Dekompozice časové řady HDP v R

6.1.6.1 Modelování trendu dvojitým exponenciálním vyrovnáním

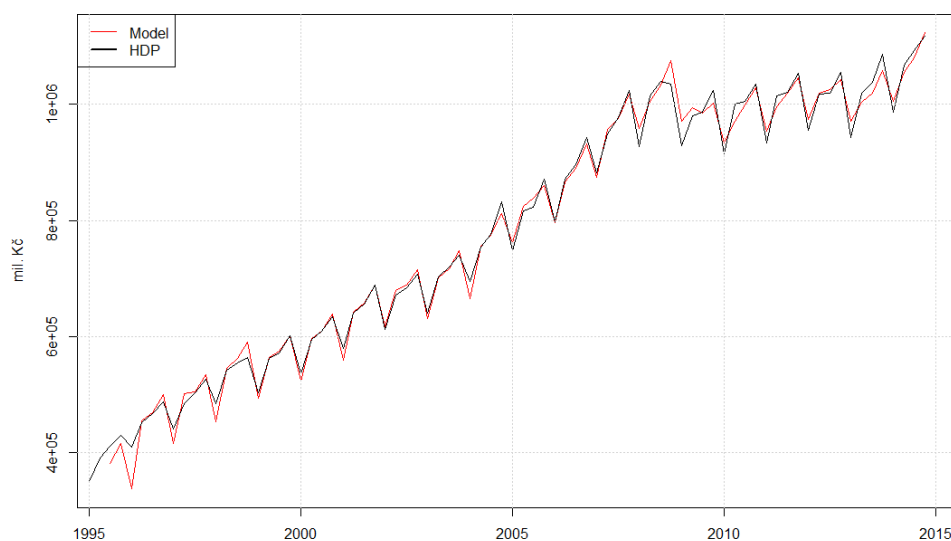
Postup dvojitého exponenciálního vyrovnání je stejný jako při ručním výpočtu. Nejprve odečteme od původní časové řady dekompozici zjištěnou sezónní složku, poté provedeme samotné vyrovnání a nakonec opět sezónní složku přičteme. Tento postup je v R realizovaný sekvencí následujících příkazů.

```
bezSez <- HDPts - HDPtsComp\$seasonal
model <- HoltWinters(bezSez, gamma=FALSE)
modelSez <- model\$fitted[,1] + HDPtsComp\$seasonal
```

Pro vykreslení grafu pak použijeme následující příkazy a výstupem bude graf 6.11.

```
plot(modelSez, ylab="mil. Kč", xlab="", main="Model - dvojitě exponenciální
vyrovnání", col="red")
lines(HDPts, col="black")
legend("topleft", c("Model", "HDP"), lwd=c(1,2), col=c("red", "black"))
```

Parametry samotného modelu jsou uvedeny ve výpisu z R dále. Jak je vidět, výpočetní funkce v R zvolila jinou hodnotu optimálního koeficientu α zřejmě proto, že sezónní očištění je provedeno pravděpodobně váženými klouzavými průměry a též vyrovnávací statistiky pro první člen jsou vypočteny sofistikovanější metodou než pouhý odhad pomocí první položky samotné časové řady.



Obrázek 6.11: Model HDP - dvojitě exponenciální vyrovnání v R

```
Holt-Winters exponential smoothing with trend and without seasonal comp.
```

```
Smoothing parameters:
```

```
alpha: 0.6967192
```

```
beta : 0.3445873
```

```
gamma: FALSE
```

```
Coefficients:
```

```
a      1086439.53
```

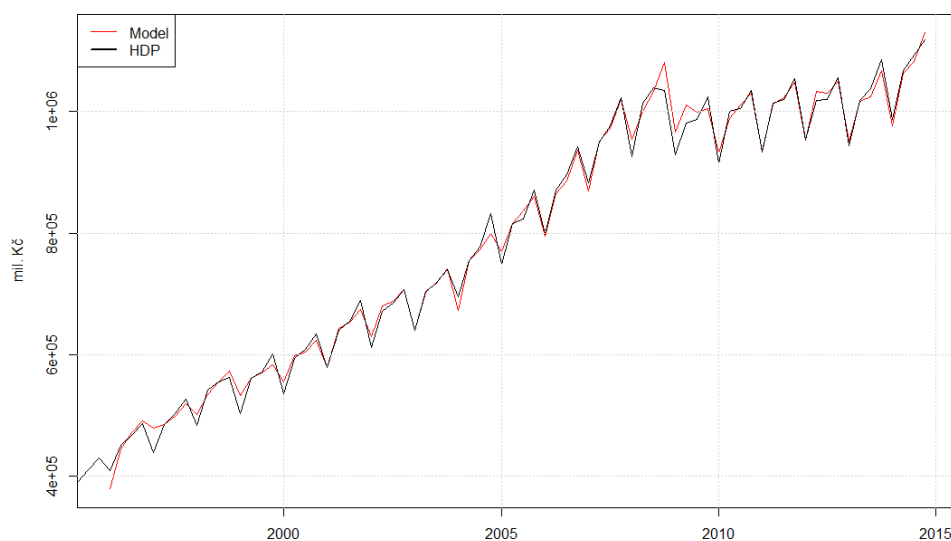
```
b      11548.46
```

```
SSE:
```

```
22 923 970 533
```

6.1.6.2 Modelování trendu trojitým exponenciálním vyrovnáním

Výpočetní možnosti softwaru nám snadno dovolí vytvořit i model pomocí trojitého exponenciálního vyrovnání. Postup je prakticky totožný jako u dvojitého exponenciálního vyrovnání, pouze při tvorbě modelu nenastavíme parametr *gamma* na FALSE jako v předešlém případě. Kvůli totožnému postupu nebudu uvádět sekvenci příkazů, dále uvedu pouze výsledky 6.12 tvorby modelu pomocí trojitého exponenciálního vyrovnání.



Obrázek 6.12: Model HDP - trojité exponenciální vyrovnání v R

```
Holt-Winters exponential smoothing with trend and additive seasonal comp.
```

```
Smoothing parameters:
```

```
alpha: 0.6982307
```

```
beta : 0.1629777
```

```
gamma: 1
```

```
Coefficients:
```

```
a      1081213.7382
```

```
b      6603.0458
```

```
SSE:
```

```
15 959 037 179
```

6.1.6.3 Porovnání kvality modelů

Oba výše získané modely nyní srovnáme prostřednictvím vypočtených hodnot SSE. Na základě hodnot v tabulce 6.5 je vidět, že chyba modelu vytvořeného pomocí trojitého exponenciálního vyrovnání je menší, a proto je model vhodnější pro další zpracování.

Model	SSE
Dvojité exponenciální vyrovnání	22 923 970 533
Trojité exponenciální vyrovnání	15 959 037 179

Tabulka 6.5: Porovnání kvality modelů HDP - R

6.2 Analýza časové řady investic

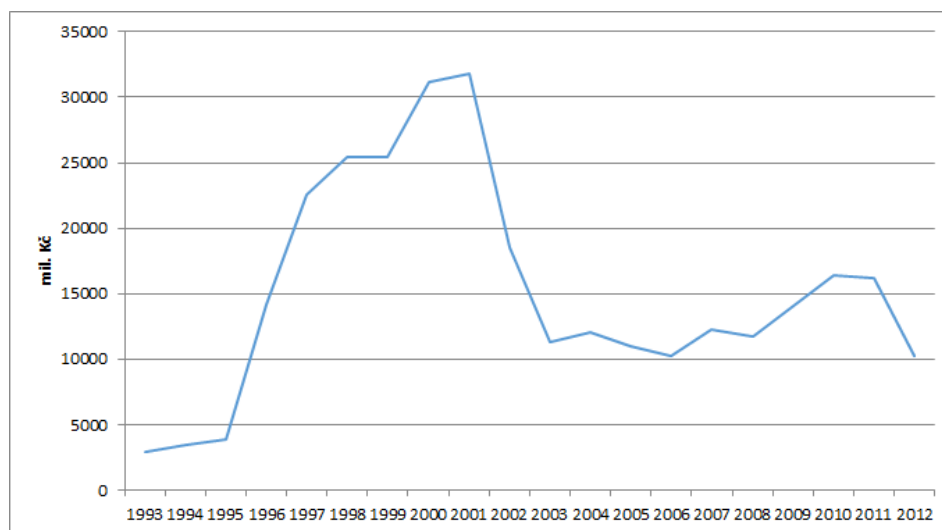
6.2.1 Získání vstupních dat

Druhým souborem dat, kterým se budeme dále zabývat, je údaj o výši investic do telekomunikačního sektoru. Tato data jsem získal z databáze Českého statistického úřadu, konkrétně ze souboru *Výdaje a investice do ICT v České republice*. Tabulka číslo 7 v této databázi ukazuje výši investic do ICT vybavení v České republice a je možné z ní získat data o výši investic do telekomunikačního vybavení dle klasifikace CZ CPA 263. Tato data jsem použil již v teoretické části při popisu ekonomického stavu telekomunikačního sektoru České republiky.

Samotná data jsou v databázi zpracována od roku 1993 do roku 2012. Výše investic jsou uvedeny v milionech Kč a jsou uvedeny v tabulce C.1. Jedná se o data zaznamenaná s roční periodou, nejsou nijak očištěna.

6.2.2 Základní parametry časové řady

Opět si nejprve pro představu o chování časové řady zobrazíme její hodnoty v grafu a uvedeme její základní popisné charakteristiky. Časovou řadu výše investic do telekomunikačního zařízení z let 1995 až 2012 zobrazuje graf 6.13 a následující tabulka 6.6 pak shrnuje základní popisné charakteristiky. Protože má ročně zaznamenaná časová řada periodu 1, nemá smysl hledat v datech sezónní složku, která vzniká periodickým opakováním sledovaných období. Ani dekompozice časové řady tedy v tomto případě nedává smysl.



Obrázek 6.13: Časová řada výše investic do telekomunikačního vybavení[?]

HDP - parametry řady	
Střední hodnota	16 577,11
Chyba střední hodnoty	1 822,67
Medián	14 119
Modus	N/A
Směrodatná odchylka	7 732,96
Rozptyl výběru	59 798 385,99
Špičatost	-0,21
Šikmost	0,72
Rozdíl max - min	27 956
Minimum	3 882
Maximum	31 838
Součet	298 388
Počet	18

Tabulka 6.6: Popisná statistika časové řady investic

6.2.3 Tvorba modelu

Protože výše zmíněná časová řada má periodu 1 rok a není možné hledat sezónní závislost, nebudeme provádět dekompozici časové řady a rovnou přikročíme k tvorbě modelu časové řady výše investic do telekomunikačního zařízení. Ten se opět bude hodit pro modelování chování řady a dále pro předpovídání budoucích hodnot.

Dále se opět budu zabývat postupem tvorby vhodného modelu a zhodnocením jeho chování vzhledem k původní časové řadě. Pro ukázkou tvorby modelu využiji opět tři základní postupy. Prvním bude modelování trendu pomocí přímky, druhým pak tvorba modelu zvoleného na základě popisných charakteristik časové řady a třetím přístupem bude adaptivní model reprezentovaný dvojitým exponenciálním vyrovnáním.

6.2.3.1 Modelování trendu přímkou

Pro vytvoření modelu pomocí regresní přímky budeme postupovat jako v případě časové řady HDP a využijeme v Excelu dostupný nástroj Analýza dat / Regrese. Vstupem pro tento výpočet budou hodnoty $1 - t_n$ označující pořadí hodnoty časové řady jako konstanta a samotné hodnoty časové řady jako závislá proměnná, jak ilustruje následující tabulka 6.7.

Výsledkem regrese pak získáme vypočtené konstanty β_0 a β_1 .

$$\beta_0 = 20\,990,860 \quad (6.20)$$

$$\beta_1 = -464,605 \quad (6.21)$$

t_i	rok	výše investic v mil. Kč
1	1995	3 882
2	1996	14 134
3	1997	22 554
4	1998	25 426
5	1999	25 417
6	2000	31 156
7	2001	31 838
8	2002	18 484
⋮	⋮	⋮

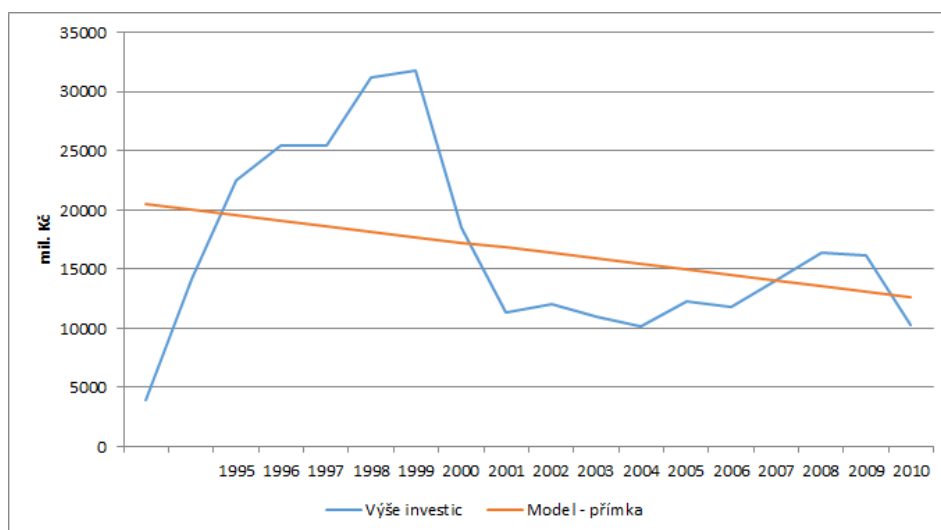
Tabulka 6.7: Vstupní data pro regresní analýzu - přímka

Z těch pak již sestavíme rovnici regresní přímky 6.22, která bude mít po vyčíslení tvar 6.23.

$$investice_{model} = \beta_0 + \beta_1 t_i \quad (6.22)$$

$$investice_{model} = 20\,990,860 - 464,605 t_i \quad (6.23)$$

Díky znalosti této rovnice dále dopočítáme jednotlivé hodnoty regresní přímky modelující trend. Samotné vyrovnání přímkou zachycuje graf 6.14.



Obrázek 6.14: Model investic vytvořený regresní přímkou

Abychom byli schopni dále zhodnotit kvalitu modelu a jednotlivé modely porovnat, na závěr vypočteme pro daný model součet čtvercových chyb (SSE) 6.24.

$$SSE_{primka} = 911\,994\,668,6 \quad (6.24)$$

6.2.3.2 Modelování trendu exponenciálou

Pro výběr druhé křivky reprezentující model trendu je využito výběru pomocí popisných charakteristik časové řady. Po zhodnocení první a druhé diference, jejich podílu a koeficientu růstu řady jsem jako druhou matematickou funkci vybral exponenciálu. Toto rozhodnutí jsem učinil na základě toho, že koeficient růstu časové řady kolísá kolem konstantní hodnoty 1.

Pro vytvoření modelu pomocí regresní exponenciály využijeme opět nástroj Excelu Analýza dat / Regrese. Vzhledem k tomu, že funkce exponenciálního trendu má rovnici 6.25,

$$investice_{model} = \beta_0 \cdot \beta_1^{t_i} \quad (6.25)$$

na počátku je nutné zlogaritmovat rovnici a po výpočtu koeficientů β opět provést odlogaritmování. Logaritmováním převedeme problém na lineární rovnici a budeme jej řešit podobně jako v minulém případě. Do regresní analýzy tedy budou vstupovat jako závislá proměnná logaritmus hodnoty časové řady a dále opět hodnoty $1 - t_n$ označující pořadí hodnoty sezónně očištěné časové řady jako konstanta, jak je vyznačeno v tabulce 6.8.

t_i	rok	sez. očištěná hodnota investic v mil. Kč	log očištěné hodnoty
1	1995	3 882	3,589 056
2	1996	14 134	4,150 265
3	1997	22 554	4,353 224
4	1998	25 426	4,405 278
5	1992	25 417	4,405 124
6	2000	31 156	4,493 542
7	2001	31 838	4,502 946
8	2002	18 484	4,266 796
⋮	⋮	⋮	⋮

Tabulka 6.8: Vstupní data pro regresní analýzu - exponenciála

Výsledkem regrese získáme vypočtené logaritmy konstanty β_0 a β_1 .

$$\log_{10}\beta_0 = 4,225876 \quad (6.26)$$

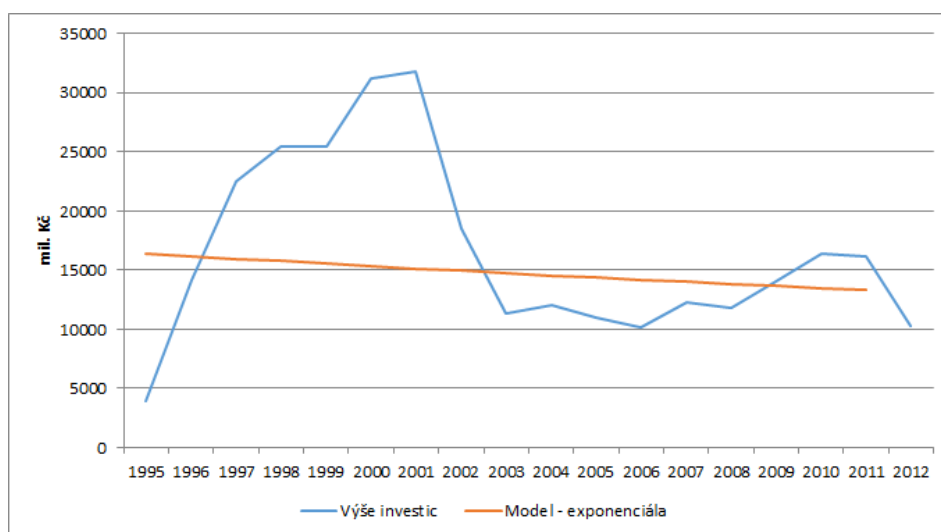
$$\log_{10}\beta_1 = -0,00569 \quad (6.27)$$

Po odlogaritmování získáme hodnoty

$$\beta_0 = 16\,821,927 \quad (6.28)$$

$$\beta_1 = 0,987 \quad (6.29)$$

Stejně jako v případě přímky se znalostí této rovnice dále dopočítáme jednotlivé hodnoty regresní exponenciály modelující trend. Tím získáme požadovaný model. Samotné vyrovnání exponenciálou a započtení vlivu sezónní složky zachycuje graf 6.15.



Obrázek 6.15: Model investic vytvořený regresní exponenciálou

Pro porovnání modelů a zjištění kvality vyrovnání opět vypočteme pro model součet čtvercových chyb (SSE) 6.30.

$$SSE_{exponenciala} = 1\,000\,883\,780 \quad (6.30)$$

6.2.3.3 Modelování trendu dvojitým exponenciálním vyrovnáním

Ze samotné povahy dat, která je vidět již v grafu 6.13, je zřejmé, že modelování chování řady s takovým průběhem nebude jednoduše realizovatelné matematickou funkcí s malým množstvím parametrů. Jistě vhodnějším modelem bude adaptivně vytvořený model pomocí dvojitého exponenciálního vyrovnání, které je přesnější než vyrovnání jednoduché a zároveň je pro ruční výpočet dostatečně jednoduché.

Prvním krokem je jako v případě časové řady HDP volba vhodného koeficientu α . Pro prvotní odhad jsem zvolil koeficient $\alpha = 0,7$. Dále provedeme výpočet vyrovnávacích statistik S1 6.31 a S2 6.32.

$$S1_t = (1 - \alpha)y_t + \alpha S1_{t-1} \quad (6.31)$$

$$S2_t = (1 - \alpha)S1_t + \alpha S2_{t-1} \quad (6.32)$$

Pro jednoduchost jsem zvolil jako první hodnotu $S1$ a $S2$ odpovídající první hodnotu sezónně očištěné časové řady HDP. Na jejich základě jsou pak jako v případě HDP vypočteny hodnoty koeficientů β_0 a β_1 .

Tedy pro tento konkrétní model 6.33 a 6.34.

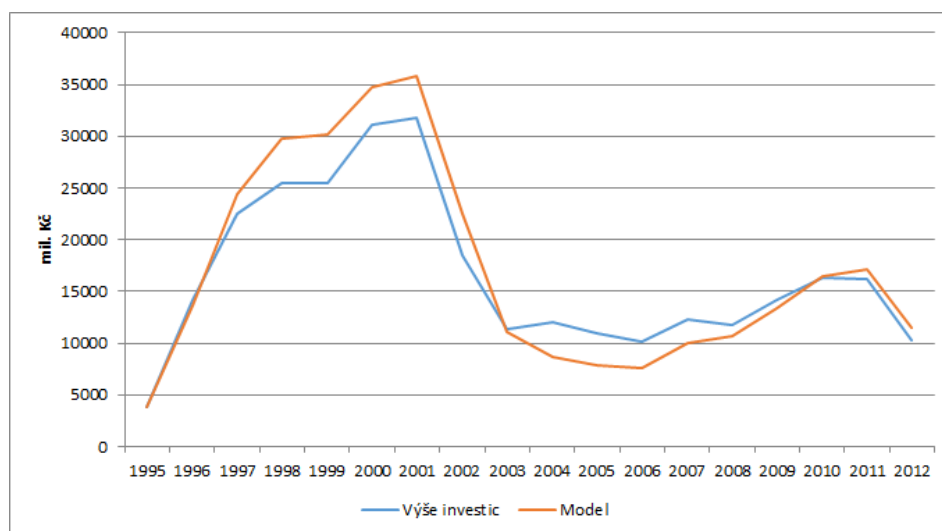
$$\beta_0 = 12\,154,370 \quad (6.33)$$

$$\beta_1 = -707,266 \quad (6.34)$$

Pro tyto hodnoty vypočítáme hodnotu SSE a pomocí nástroje Řešitel v Excelu provedeme minimalizaci této hodnoty SSE na základě změny koeficientu α . Tím zajistíme volbu optimálního koeficientu. Pro danou časovou řadu investic jsem tímto postupem získal hodnotu koeficientu 6.35:

$$\alpha = 0,525041 \quad (6.35)$$

S takto získanou hodnotou koeficientu přepočítáme hodnoty modelu a získáme tak optimální trendovou funkci. Výsledný model je zobrazen v grafu 6.16.



Obrázek 6.16: Model investic vytvořený dvojitým exponenciálním vyrovnáním

Již dříve vypočtená hodnota SSE je pro tento model 6.36.

$$SSE_{dvojexp} = 126\,270\,446,10 \quad (6.36)$$

6.2.3.4 Porovnání kvality modelů

Jednotlivé výše získané modely nyní opět srovnáme prostřednictvím vypočtených hodnot SSE. Na základě hodnot v následující tabulce 6.9 je vidět, že chyba modelu vytvořeného pomocí dvojitěho exponenciálního vyrovnání je o řád menší než model vytvořený exponenciálou, který se jeví jako nejméně vhodný. SSE modelu vytvořeného regresní přímkou je pak asi $8\times$ horší, a proto se model vytvořený dvojitým exponenciálním vyrovnáním jeví jako nejlepší z těchto tří. Pro další výpočty a prognózování proto využijí právě tento model.

Model	SSE
Přímka	911 994 668,60
Exponenciála	1 000 883 780,00
Dvojitě exponenciální vyrovnání	126 270 446,10

Tabulka 6.9: Porovnání kvality modelů investic

6.2.4 Tvorba modelu v R

Podobně jako v případě analýzy časové řady HDP jsem při analýze provedené v Excelu došel k závěru, že vhodným modelem pro časovou řadu investic je model vytvořený dvojitým exponenciálním vyrovnáním. Tento postup tedy zopakujeme opět v software R. Trojitě exponenciální vyrovnání nelze pomocí funkce `HoltWinters()` realizovat, protože řada neobsahuje sezónní složku.

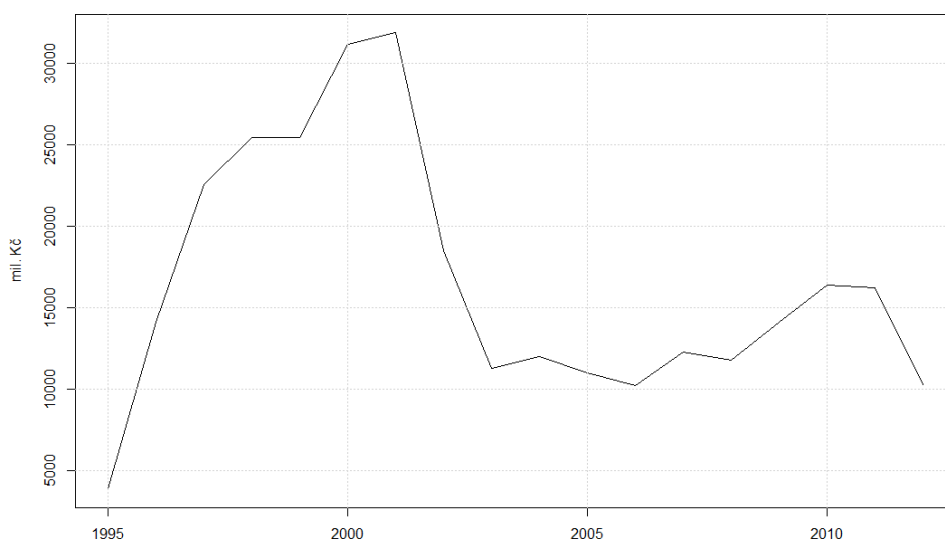
6.2.4.1 Modelování trendu dvojitým exponenciálním vyrovnáním

Protože v tomto případě neoperujeme se sezónní složkou časové řady, bude vytvoření modelu záležitostí jednoho příkazu. Před tím ještě načteme hodnoty časové řady do prostředí programu a necháme si vykreslit graf 6.17 časové řady.

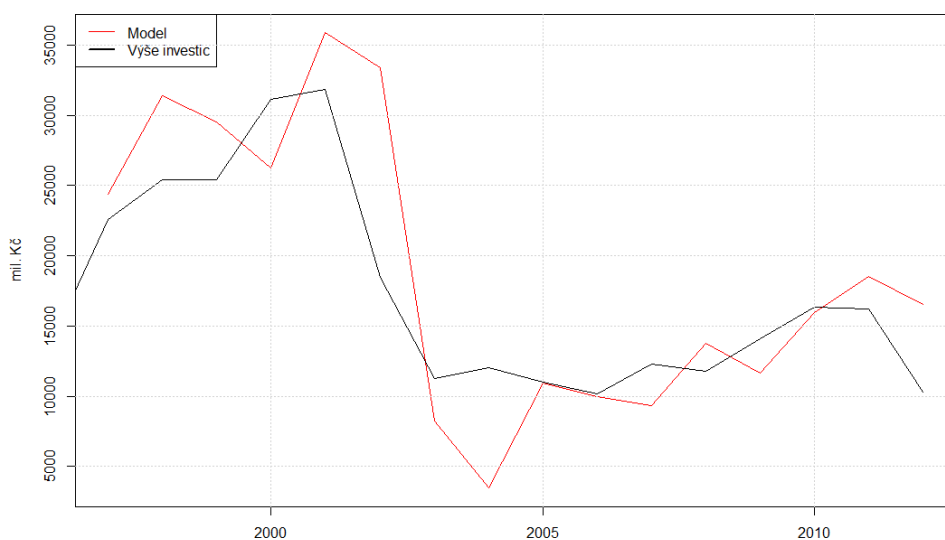
```
INV <- scan("investice-src.txt")
INVts <- ts(INV, frequency=1,start=c(1995,1))
plot.ts(HDPts, ylab="mil. Kč",xlab="",main="Výše investic do
        telekomunikačního vybavení")
```

Nyní již můžeme postoupit k samotnému vyrovnání a opět vykreslení grafu 6.18 výsledku.

```
modelINV <- HoltWinters(INVts, gamma=FALSE)
plot(modelINV$fit, ylab="mil. Kč", xlab="", main="Model - dvojitě
        exponenciální vyrovnání", col="red")
lines(INVts, col="black")
legend("topleft",c("Model", "Výše investic"),lwd=c(1,2),
        col=c("red","black"))
```



Obrázek 6.17: Výše investic do telekomunikačního vybavení



Obrázek 6.18: Model investic - dvojitě exponenciální vyrovnání

Parametry samotného modelu jsou uvedeny ve výstupu z R dále. Jak je vidět, výpočetní funkce v R zvolila jinou hodnotu optimálního koeficientu α zřejmě proto, že sezónní očištění je provedeno pravděpodobně váženými klouzavými průměry a též vyrovnávací statistiky pro první člen jsou vypočteny sofistikovanější metodou než pouhý odhad pomocí první položky samotné časové řady.

```
Holt-Winters exponential smoothing with trend and without seasonal comp.
```

```
Smoothing parameters:
```

```
alpha: 1
```

```
beta : 0.7895001
```

```
gamma: FALSE
```

```
Coefficients:
```

```
a      10276.00
```

```
b      -4605.343
```

```
SSE:
```

```
462 847 275
```

Kapitola 7

Krátkodobé prognózy budoucího vývoje

Za krátkodobé prognózování budeme považovat metodu odhadu budoucího vývoje s danou mírou spolehlivosti. Při tvorbě prognóz vývoje jednotlivých veličin vyjdeme z modelů vytvořených v předchozí kapitole práce.

Pro tvorbu prognózy využijeme v této části vždy nejlepší nalezený model. V případě HDP to bude model vytvořený trojitým exponenciálním vyrovnáním (v Excelu pouze dvojitým), v případě investic pak model vytvořený dvojitým exponenciálním vyrovnáním. Prognózané období jsem stanovil na 3 roky. Vzhledem k trendu výše investic do telekomunikačního vybavení bude i tento horizont dost dlouhý a spolehlivost prognózy bude podle mého názoru omezená.

7.1 Krátkodobé prognózy vývoje HDP

7.1.1 Prognóza na základě ručního výpočtu v Excelu

Pro úplnost řešení v této práci uvedu i prognózy vývoje HDP, ačkoliv zadání požaduje pouze prognózu vývoje investic. Při prognózování vyjdeme, jak již bylo řečeno, z modelu vytvořeného pomocí dvojitého exponenciálního vyrovnání.

Pro bodový odhad budoucího vývoje pomocí Excelu využijeme vypočtené parametry β . Pomocí těch vypočteme budoucí hodnoty modelu 7.1 tak, jak shrnuje následující tabulka 7.1.

$$HDP_{prognóza} = \beta_0 + \beta_1 \cdot (t_t - t_{t-1} + 1) \quad (7.1)$$

7.1.2 Kvalita prognózy

Abychom byli schopni zhodnotit kvalitu prognózy modelu, budeme postupovat tak, že vytvoříme model se shodnými parametry, ale část dat do jeho tvorby nezahrneme. Tyto data pak vypočteme pomocí prognózování a porovnáme vypočtená data se skutečnými hodnotami.

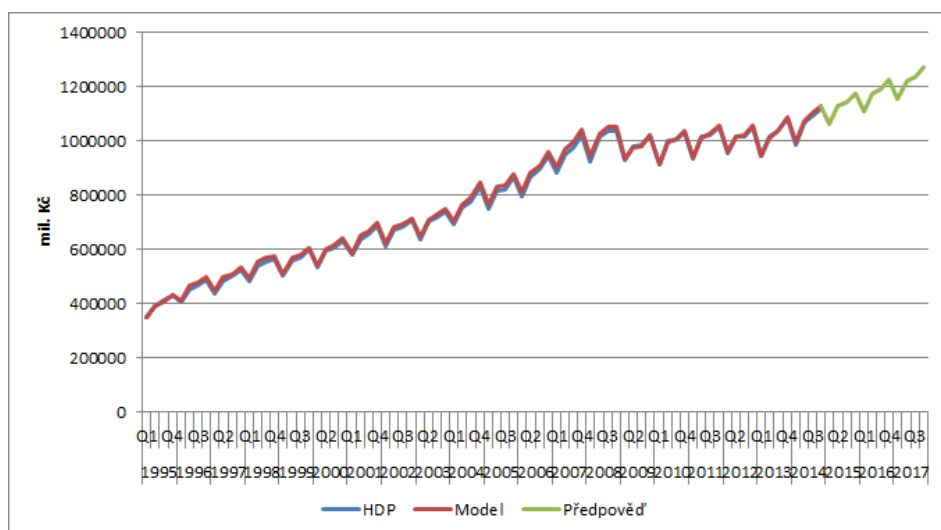
n	rok	čtvrtletí	predikovaná hodnota
81	2015	Q1	1 063 302
82		Q2	1 128 318
83		Q3	1 145 614
84		Q4	1 177 948
85	2016	Q1	1 110229
86		Q2	1 175 245
87		Q3	1 192 541
88		Q4	1 224 875
89	2017	Q1	1 157 155
90		Q2	1 222 171
91		Q3	1 239 467
92		Q4	1 271 801

Tabulka 7.1: Prognóza vývoje HDP

Pro číselné vyjádření kvality modelu využijeme Theilův koeficient nesouladu. Jako testovací období tedy zvolíme délku pěti let a vypočteme hodnoty prognózy, které porovnáme se skutečnými hodnotami 7.2. Pro tento konkrétní model vychází Theilův koeficient nesouladu 3,68%, což je dost nízká hodnota a prognóza modelu je dostatečně spolehlivá.

$$\begin{aligned} TIE^2 &= 0,001353 \\ TIE &= 0,036793 \end{aligned} \quad (7.2)$$

Výsledné predikované hodnoty ještě zobrazíme v grafu 7.1.



Obrázek 7.1: Prognóza budoucí hodnoty HDP

7.1.3 Prognóza v R

Pro vytvoření prognózy v R využijeme již zmíněný model vytvořený pomocí trojitého exponenciálního vyrovnání. Následující sekvencí příkazů vytvoříme předpověď pro 12 period (tedy 3 roky) na základě modelu vytvořeného dříve. V tabulce 7.4 i grafu 7.2 je zobrazena bodová předpověď a 80% a 95% interval spolehlivosti odhadu.

```
library(forecast)
forHDP <- forecast(modelSez, h=12)
plot(forHDP, ylab="mil. Kč", xlab="", main="Prognóza vývoje HDP na 3 roky")
```

rok	čtvrtletí	předpověď	80% dolní	80% horní	95% dolní	95% horní
2015	Q1	1 055 326	1 037 246	1 073 405	1 027 676	1 082 976
	Q2	1 099 970	1 072 798	1 127 141	1 058 414	1 141 525
	Q3	1 124 078	1 088 396	1 159 760	1 069 507	1 178 650
	Q4	1 162 278	1 116 006	1 208 549	1 091 512	1 233 044
2016	Q1	1 094 355	1 039 843	1 148 867	1 010 986	1 177 724
	Q2	1 138 999	1 075 958	1 202 039	1 042 587	1 235 411
	Q3	1 163 107	1 091 240	1 234 974	1 053 196	1 273 019
	Q4	1 201 307	1 118 647	1 283 967	1 074 890	1 327 724
2017	Q1	1 133 384	1 041 468	1 225 300	992 811	1 273 957
	Q2	1 178 028	1 076 533	1 279 522	1 022 805	1 333 250
	Q3	1 202 136	1 090 750	1 313 522	1 031 786	1 372 486
	Q4	1 240 336	1 117 318	1 363 354	1 052 196	1 428 476

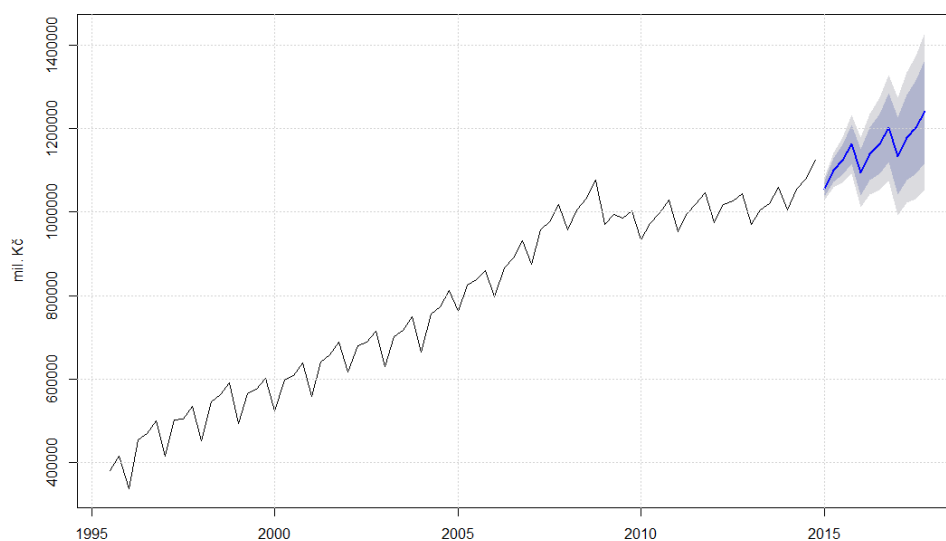
Tabulka 7.2: Prognóza budoucí hodnoty HDP v R

7.1.4 Kvalita prognózy

Kvalitu prognózy v R vypočítáme podobným způsobem jako v Excelu, tedy část původních dat srovnáme s jejich predikcí a vypočteme Theilův koeficient 7.3 pomocí funkce `accuracy`. Ten je nižší oproti hodnotě vypočtené ručně v Excelu, model tedy predikuje ještě lépe.

```
partHDP <- window(HDPTs, start=c(1995,1), end=c(2009,4))
restHDP <- window(modelSez, start=c(2010,1), end=c(2014,4))
forHDP2 <- forecast(partHDP, h=20)
accuracy(restHDP, forHDP2$mean)
```

$$TIE = 1,218\% \quad (7.3)$$



Obrázek 7.2: Prognóza budoucí hodnoty HDP v R

7.2 Krátkodobé prognózy vývoje investic

7.2.1 Prognóza na základě ručního výpočtu v Excelu

Pro bodový odhad budoucího vývoje výše investic pomocí Excelu využijeme opět vypočtené parametry β modelu získaného dvojitým exponenciálním vyrovnáním. Pomocí těch vypočteme budoucí hodnoty modelu 7.4 tak, jak shrnuje následující tabulka 7.3.

$$investice_{prognóza} = \beta_0 + \beta_1 \cdot (t_t - t_{t-1} + 1) \quad (7.4)$$

n	rok	predikovaná hodnota
19	2013	10 739,84
20	2014	10 032,57
21	2015	9 325,307

Tabulka 7.3: Prognóza vývoje investic

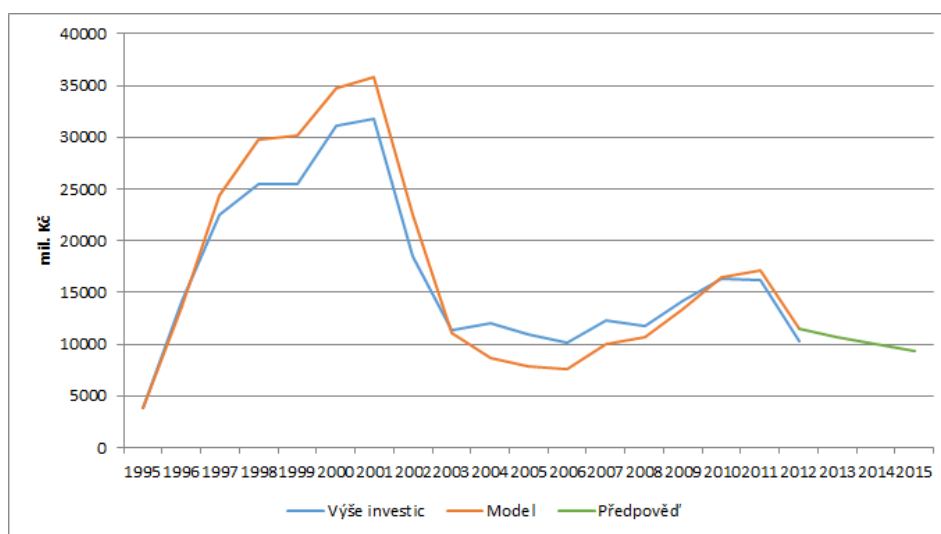
7.2.2 Kvalita prognózy

Kvalitu prognózy opět zhodnotíme výpočtem Theilova koeficientu 7.5 nad částí původních dat, která se pokusíme predikovat. Jako testovací období tedy zvolíme opět délku pěti let a vypočteme hodnoty prognózy, které porovnáme se skutečnými hodnotami. Pro tento konkrétní model vychází Theilův koeficient nesouladu 50,53%, což je hodnota velice vysoká a model bude prognózovat značně nespolehlivě. Tato hodnota je dána především tím, že v datech není žádný výrazný trend, jak je vidět již při pohledu na graf časové řady investic.

Zjednodušeně řečeno, chování této časové řady je na pozorovaném období a rozlišení dat značně nepředvídatelné a prognostický model správně extrapolující takoveto rozložení dat je prakticky nemožné sestrojít.

$$\begin{aligned} TIE^2 &= 0,255355 \\ TIE &= 0,505327 \end{aligned} \quad (7.5)$$

Výsledné predikované hodnoty ještě zobrazíme v grafu 7.3.



Obrázek 7.3: Predikce budoucí výše investic

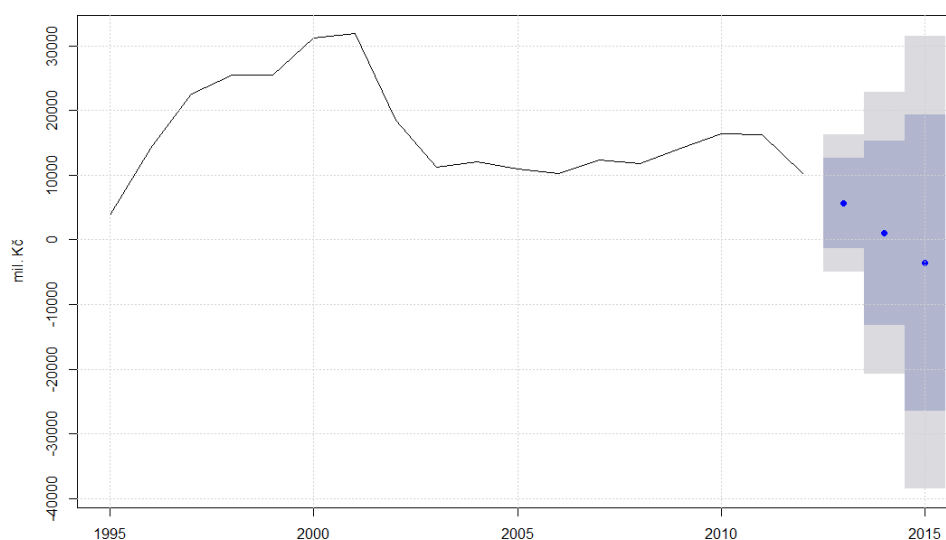
7.2.3 Prognóza v R

Pro vytvoření prognózy v R využijeme opět model vytvořený pomocí dvojitého exponenciálního vyrovnání, protože modeluje průběh nejspolehlivěji. Následující sekvencí příkazů vytvoříme předpověď pro 3 periody (tedy 3 roky) na základě modelu vytvořeného dříve. V tabulce 7.4 i grafu 7.4 je zobrazena bodová předpověď a 80% a 95% interval spolehlivosti odhadu. Hodnoty klesající pod 0 jsou samozřejmě nesmyslné a k takovému vývoji nemůže dojít.

```
library(forecast)
forINV<-forecast(modelINV, h=3)
plot(forINV, ylab="mil. Kč", xlab="", main="Prognóza vývoje výše investic
na 3 roky")
```

rok	předpověď	80% dolní	80% horní	95% dolní	95% horní
2013	5 670,66	-1 275,88	12 617,20	-4 953,16	16 294,47
2014	1 065,32	-13 174,80	15 305,40	-20 713,00	22 843,65
2015	-3 540,03	-26 425,20	19 345,16	-38 539,90	31 459,85

Tabulka 7.4: Prognóza budoucí hodnoty investic v R



Obrázek 7.4: Prognóza budoucí výše investic v R

7.2.4 Kvalita prognózy

Kvalitu prognózy v R vypočítáme opět jako v předchozím případě, tedy část původních dat srovnáme s jejich predikcí a vypočteme Theilův koeficient 7.6. Ten je ještě vyšší než hodnota vypočtená ručně v Excelu. O modelu opět říká, že jeho prognóza je velmi nespolehlivá a chování řady nelze spolehlivě popsat.

```
partINV<-window(INVts, start=c(1995), end=(2009))
restINV<-window(modelINV\\$fitted, start=c(2010), end=c(2012))
forINV2<-forecast(partINV, h=3)
accuracy(restINV, forINV2\\$mean)
```

$$TIE = 63,638\% \quad (7.6)$$

Kapitola 8

Ekonometrický model souvislostí HDP a investic

V poslední kapitole praktické části této práce se budu zabývat odhadem ekonomického modelu souvislostí HDP a investic do telekomunikačního sektoru. Pokusím se odhadnout co možná nejlepší jednorovnicový model pomocí lineární regrese, model verifikovat a využít ho pro intervalový odhad budoucího vývoje vysvětlované proměnné. Tvorbu modelu budu realizovat v software R.

8.1 Specifikace modelu

Jak již vyplývá z textu výše, zvolil jsem formu jednorovnicového modelu vytvořeného pomocí lineární regrese. Jak je patrné již z grafů uvedených v předchozích částech práce, pro modelové popsání vývoje investic do telekomunikací rozhodně nebude stačit samotná časová řada HDP, proto do modelu zavedeme další proměnné, tak jak zachycuje tabulka 8.1.

Vysvětlovaná proměnná	investice	Výše investic de telekomunikačního vybavení v mil. Kč
Vysvětlující proměnná	HDP	HDP v mil. Kč
	pevne	Počet pevných linek
	sim	Počet SIM karet
	broadband	Počet broadband připojení
	narodniHovory	Délka národních hovorů v min.

Tabulka 8.1: Proměnné modelu

Všechna data vychází z již v této práci dříve zpracovaných hodnot a jsou k nalezení v databázích Českého statistického úřadu a Eurostatu.

Nejvhodnějším modelem nazveme takový model, kde budou jednotlivé zahrnuté proměnné statisticky významné a zároveň se v modelu nebude vyskytovat heteroskedasticita (rozptyl reziduí je závislý na parametrické proměnné), autokorelace (vzniká regresní vztah

mezi rezidui) a multikolinearita (vzájemná lineární závislost vysvětlujících proměnných). Veškeré statistiky budeme testovat na hladině významnosti $\alpha = 0,05$.

8.2 Kvantifikace modelu

V tomto kroku se pokusíme nalézt takový model, aby jeho proměnné byly statisticky významné a bylo možné pokračovat verifikací modelu (ověřením jeho chování a vzájemného vztahu proměnných).

V software R nejprve načteme vstupní data, která jsou uvedena v příloze D.1 a poté vytvoříme lineární regresní model pomocí funkce `lm()`, poté vypíšeme parametry modelu.

```
data<-read.csv2("model-data.csv")
model<-lm(investice~HDP+pevne+sim+broadband+narodniHovory,data)
summary(model)
```

Dále zde uvedu pouze zkrácený výpis významnosti koeficientů modelu 8.1. Na jeho základě je vidět, že všechny vysvětlující proměnné až na národní hovory jsou málo významné, proto odebereme tu s nejvyšší hodnotou p-testu (nejnižší významností) a pokusíme se znovu o vytvoření modelu.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.990e+04  5.428e+04  -0.551 0.595079
HDP          4.799e-03  8.101e-03   0.592 0.568165
pevne        4.581e-03  1.198e-02   0.382 0.711064
sim          -2.368e-04  7.251e-04  -0.327 0.751426
broadband    3.713e-03  6.149e-03   0.604 0.560869
narodniHovory 3.186e-04  6.387e-05   4.988 0.000751 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obrázek 8.1: Model HDP a investic 1 - významnost proměnných

```
model2<-lm(investice~HDP+pevne+broadband+narodniHovory,data)
summary(model2)
```

Po odebrání proměnné vyjadřující počet SIM karet je výsledek takový, že významnost proměnných se oproti předchozímu modelu nezměnila 8.2.

Tento postup budeme opakovat, dokud v modelu nezůstanou jen statisticky významné proměnné. Po dalších dvou iteracích, kdy byly odebrány proměnné vyjadřující počet pevných linek a internetových přípojek, je výstup modelu 8.3 takový, že obě zbývající proměnné jsou statisticky významné a koeficient determinance 92,07% je poměrně vysoký.

Tento model tedy prohlásíme pro fázi kvantifikace modelu za odpovídající a budeme se dále zabývat jeho verifikací.

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.917e+04  4.123e+04  -0.465 0.651866
HDP          2.541e-03  4.028e-03   0.631 0.542347
pevne       2.705e-03  1.003e-02   0.270 0.792967
broadband   3.080e-03  5.569e-03   0.553 0.592365
narodniHovory 3.253e-04  5.775e-05   5.632 0.000218 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Obrázek 8.2: Model HDP a investic 2 - významnost proměnných

```

Residuals:
      Min       1Q   Median       3Q      Max
-4166.9 -1165.9   439.5  1356.5  3481.3

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.067e+04  8.406e+03  -2.459 0.03011 *
HDP          5.949e-03  1.920e-03   3.098 0.00922 **
narodniHovory 3.617e-04  4.596e-05   7.870 4.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2310 on 12 degrees of freedom
Multiple R-squared:  0.9207,    Adjusted R-squared:  0.9075
F-statistic: 69.7 on 2 and 12 DF,  p-value: 2.478e-07

```

Obrázek 8.3: Model HDP a investic 3 - významnost proměnných

8.3 Verifikace modelu

V rámci verifikace budeme, jak již bylo zmíněno výše, testovat nalezený model na následující jevy:

- **Heteroskedasticita** (rozptyl reziduí je závislý na parametrické proměnné)
- **Autokorelace** (vzniká regresní vztah mezi reziduí)
- **Multikolinearita** (vzájemná lineární závislost vysvětlujících proměnných)

8.3.1 Test heteroskedasticity

Pro provedení každého testu nejprve stanovíme hypotézy, které budeme ověřovat, zde tedy:

- H_0 : v nalezeném modelu není přítomna heteroskedasticita
- H_1 : v nalezeném modelu je přítomna heteroskedasticita

Pro samotný test využijeme funkci `bptest()` v R, která provede Breusch-Paganův test heteroskedasticity. Pro vyhodnocení testu porovnáme hodnotu testovací statistiky p s hladinou významnosti α , pokud bude p hodnota vyšší, přijmeme tak hypotézu H_0 .

```
bptest(model)

          studentized Breusch-Pagan test

data:  model
BP = 1.5055, df = 2, p-value = 0.4711
```

Na základě faktu, že $0,4711 > 0,05$ přijímáme hypotézu H_0 a v nalezeném modelu není přítomna heteroskedasticita.

8.3.2 Test autokorelace

Opět nejprve stanovíme hypotézy:

- H_0 : data v modelu nejsou autokorelovaná
- H_1 : data v modelu jsou autokorelovaná

Pro test využijeme v R funkci `dwtest()`, která realizuje Durbin-Watsonův test autokorelace. Hypotézu H_0 přijmeme, pokud bude hodnota DW statistiky vypočtená testem blízká hodnotě 2.

```
dwtest(model)

          Durbin-Watson test

data:  model
DW = 2.5991, p-value = 0.7676
```

Hodnota DW testu je sice mírně vyšší než 2, ale budeme ji akceptovat s tím, že v modelu je přítomná mírná autokorelace.

8.3.3 Test multikolinearity

Předposledním krokem verifikace bude test multikolinearity, opět nejprve stanovíme hypotézy:

- H_0 : data v modelu nejsou zatížena multikolinearitou
- H_1 : data v modelu jsou zatížena multikolinearitou

Test provedeme pomocí funkce v R `vif()` (Variance Inflation Factor), jejíž hodnota by neměla přesáhnout kritickou hodnotu 5 pro žádné proměnné modelu. Pokud se tak stane, přijmeme hypotézu H_0 .

```
vif(model)
      HDP      narodniHovory
4.992137 4.872123
```

Hodnota testu multikolinearity je sice těsně pod kritickou mezí, budeme ji však akceptovat a přijmeme hypotézu H_0 . V ekonomických časových řadách je multikolinearita častým jevem především proto, že značná část ekonomických veličin se vyvíjí jedním směrem, tedy roste, případně klesá a vytváří tak přibližně lineární závislosti.

8.3.4 Výsledek verifikace modelu

Na základě jednotlivých přijatých hypotéz můžeme konstatovat, že model je vhodným ekonometrickým modelem k vyjádření vztahu mezi výší investic do telekomunikačního vybavení, výše HDP a objemu národních hovorů realizovaných prostřednictvím telekomunikačního vybavení.

Na závěr provedeme pro kontrolu ještě obecný test kvality specifikace modelu, tzv. Ramsey RESET test. Protože tento test nevyžaduje alternativní hypotézu, budeme testovat pouze hypotézu:

- H_0 : model je specifikovaný správně

Test provedeme pomocí funkce `resettest()` a o přijetí či zamítnutí hypotézy H_0 rozhodneme na základě porovnání p hodnoty a hladiny významnosti α . Pokud bude p hodnota vyšší než α , přijmeme hypotézu H_0 .

```
resettest(model)

      RESET test

data:  model
RESET = 0.86638, df1 = 2, df2 = 10, p-value = 0.4498
```

Na základě porovnání $0,4498 > 0,05$ přijímáme hypotézu H_0 : model je specifikovaný správně.

8.4 Aplikace modelu

Posledním krokem, kvůli kterému se o samotný odhad modelu pokoušíme, je aplikace modelu. Jak jsem popsal již v teoretické části, základní aplikací ekonometrického modelu je

analýza dat získaných v minulosti, na základě kterých je možné zhodnotit chování modelu a následně jeho využití pro predikci dat v budoucnosti. Opět využijeme možnosti software R pro predikci nad lineárním modelem, tentokrát funkce `predict()` a následně vykreslíme graf 8.4, kde body značí původní hodnoty investic, červená čára pak modelem předpovězené hodnoty a modré přerušované čáry ohraničují interval 95% spolehlivosti předpovědi. Hodnoty vypočtené na základě modelu jsou uvedeny i v tabulce 8.2.

```
predpoved<-predict(model, interval="prediction")
plot(investice~rok,data, xlab="Rok", ylab="Výše investic", main="Aplikace
      modelu - předpověď")
lines(predpoved[,1]~data$rok,col="red")
lines(predpoved[,2]~data$rok,col="blue",lty=2)
lines(predpoved[,3]~data$rok,col="blue",lty=2)
```



Obrázek 8.4: Aplikace modelu - předpověď

Druhou možností využití modelu je predikce budoucích hodnot vysvětlované veličiny. Předpokladem ovšem je znalost, případně expertní odhad veličin vysvětlujících. Pro demonstraci využití modelu budeme tedy předpokládat, že jsme expertním odhadem zjistili výši objemu národních hovorů v roce 2013 ve výši 15 245 530 minut a na základě předpovědi HDP při analýze časové řady známe hodnotu HDP za rok 2013 ve výši 4 086 260 mil. Kč. Predikci budoucí výše investic pak provedeme následovně:

```
predict(model, newdata=data.frame(rok=2013,HDP=4086260,
      narodniHovory=15245530), interval="prediction")
```


	předpověď	95% dolní	95% horní
1	24 146,260	18 554,290	29 738,230
2	26 414,908	20 874,689	31 955,130
3	29 722,646	24 029,508	35 415,780
4	31 398,538	25 479,046	37 318,030
5	17 627,209	12 226,354	23 028,060
6	15 169,703	9 713,708	20 625,700
7	10 493,908	4 836,751	16 151,060
8	12 322,846	6 968,010	17 677,680
9	9 928,853	4 513,390	15 344,320
10	14 286,822	8 946,056	19 627,590
11	10 988,791	5 584,183	16 393,400
12	18 270,929	12 631,869	23 909,990
13	14 111,052	8 695,652	19 526,450
14	12 711,700	7 287,068	18 136,330
15	10 223,835	4 798,731	15 648,940

Tabulka 8.2: Hodnoty vypočtené modelem

Výstupem bude opět trojice hodnot uvedených v tabulce 8.3, kde první hodnota značí modelem odhadnutou výši investic a další hodnoty ohraničují pás spolehlivosti. Tímto způsobem je možné predikovat budoucí hodnoty a případně zpětně upravovat model na základě ověření jeho výsledků proti reálným získaným datům.

	předpověď	95% dolní	95% horní
16	19 156,825	3 698,349	14 615,300

Tabulka 8.3: Předpověď na základě modelu

Kapitola 9

Závěr

V první části práce jsem se zabýval teoretickým úvodem v oblasti telekomunikačního sektoru, jako je definice jednotlivých dále používaných pojmů, rozdělení typů telefonních linek a SIM karet a srovnání vývoje jejich počtu v posledních letech. Dále jsem popsal ekonomickou strukturu telekomunikačního sektoru a zaměřil jsem se na výše investic a výdajů v tomto sektoru, které jsou primárně tvořeny telekomunikačními a mobilními operátory v rámci budování infrastruktury sítí. Tuto skutečnost lze pozorovat i v grafech, které zobrazují vývoj výše investic. Kolem roku 2002 byla výše investic do sektoru právě nejvyšší, protože vstoupil na trh poslední mobilní operátor a budování telekomunikační infrastruktury dosáhlo svého vrcholu. Dále již výše těchto investic klesá.

Dále jsem charakterizoval hrubý domácí produkt České republiky a jednotlivé faktory, které ho definují a ovlivňují jako například stínová ekonomika. Stručně jsem zmínil rovněž metodiky výpočtu hrubého domácího produktu a na závěr této části jsem provedl analýzu vývoje HDP v České republice za posledních 15 let a srovnal růst a výši HDP s okolními zeměmi.

V následující části se zabývám teorií analýzy časových řad, která je v praktické části této práce dále využita. Stěžejní je především postup dekompozice časové řady na jednotlivé složky a analýza těchto jednotlivých složek. Tento postup je důležitý při tvorbě modelu popisujícího časovou řadu a dále k tvorbě modelu závislosti mezi jednotlivými řadami. Poslední kapitolou teoretické části mé práce je teoretický základ k tvorbě jednorovnicového ekonometrického lineárního modelu, protože právě tato metodika je dále použita v části praktické.

V praktické části práce jsem nejprve začal s analýzou časových řad HDP a výše investic do telekomunikací. U obou řad jsem se snažil nalézt co možná nejlepší model na základě porovnání několika variant nalezených modelů, a to jak ruční cestou pomocí výpočtů popsaných v teoretické části, tak analýzou ve statistickém software R. U obou řad jsem pak na základě zhodnocení variant prostřednictvím SSE vybral nejvhodnější model, který jsem dále využil v následujících částech práce.

Pomocí nalezených modelů jsem se dále zabýval krátkodobými prognózami jak vývoje HDP, tak investic do telekomunikací. Tyto prognózy jsem založil právě na nejlepších vybraných modelech z předchozí analýzy. Opět jsem při prognózování postupoval na základě ručních výpočtů realizovaných v MS Excel a dále jsem provedl predikci i s výpočtem intervalu spolehlivosti v software R. Pomocí Theilova koeficientu nesouladu jsem zhodnotil jednotlivé modely a jejich schopnost prognózy.

Závěrečným úkolem této práce bylo nalézt ekonometrický model souvislostí HDP a výše investic do telekomunikací. Vycházel jsem z modelu popsaného v teoretické části, tedy z lineárního jednorovnicového ekonometrického modelu, jehož parametry jsem se pokoušel odhadnout pomocí metody nejmenších čtverců. V průběhu práce jsem sestrojil několik modelů, do kterých jsem zahrnul kromě výše HDP i další vysvětlující proměnné tak, abych byl schopen vytvořit model s co nejvyšším koeficientem determinance a maximální významností jednotlivých proměnných. Protože samotná lineární závislost mezi výší HDP a výší investic byla zamítnuta již na základě vývoje pozorovatelného z grafů těchto časových řad, postupně jsem do modelu zkoušel přidávat proměnné jako počet SIM karet, počet pevných linek, množství broadband přípojek a také počet minut provolaných v národních sítích. Právě tato poslední proměnná spolu s výší hrubého domácího produktu umožnila vytvořit model s koeficientem determinance 92.07%, který jsem prohlásil za dostatečně kvalitní. Dále jsem model otestoval na přítomnost jevů jako je autokorelace a multikolinearita a na závěr jsem ověřil správnost jeho specifikace prostřednictvím Ramseyho RESET testu. Protože jsem všechny tyto testy vyhodnotil na dané hladině významnosti jako úspěšné, prohlásil jsem model za správně specifikovaný a věrohodný a demonstroval jsem jeho použití na předpovědi výše investic v následujícím období.

Možným zdokonalením daného modelu by bylo použití jiné metody odhadu parametrů modelu než je metoda nejmenších čtverců, případně specifikace modelu jiného typu než lineárního. V každém případě nebude výše HDP jako jediná vysvětlující proměnná dostačovat a model tak bude muset obsahovat další vstupní proměnné.

Literatura

- [1] ARTL, J. - ARTLOVÁ, M. - RUBLÍKOVÁ, E. *Analýza ekonomických časových řad s příklady*. Praha : VŠE, 2002.
- [2] BEDNAŘÍK, J. *Hodnocení efektivnosti investičního projektu a jeho financování*. Brno : ESF MU, 2012.
- [3] CIPRA, T. *Analýza časových řad s aplikacemi v ekonomii*. Praha : SNTL, 1986.
- [4] HANČLOVÁ, J. - TVRDÝ, L. *Úvod do analýzy časových řad*. Ostrava : VŠB-TU, 2003.
- [5] HUŠEK, R. *Ekonometrická analýza*. Praha : VOeconomica, 2007.
- [6] KAŇOK, M. *Statistické metody v managementu*. Praha : Vydavatelství ČVUT, 2005.
- [7] KLÍMEK, P. *Ekonometrie*. Zlín : UTB, 2010.
- [8] OECD. *OECD.StatExtracts* [online]. 2014. [cit. 2. 12. 2014]. Dostupné z: <<http://stats.oecd.org/>>.
- [9] OECD. *OECD Guide to Measuring the Information Society* [online]. 2007. [cit. 2. 12. 2014]. Dostupné z: <<http://www.oecd.org/sti/measuring-infoeconomy/guide>>.
- [10] PAVELKA, T. *Makroekonomie – základní kurz*. Praha : VŠEM, 2nd edition, 2007.
- [11] SCHNEIDER, F. *The Shadow Economy in Europe 2013*. Linz : A.T. Kearney, Inc., 2013. Dostupné z: <<http://www.atkearney.com/documents/10192/1743816/The+Shadow+Economy+in+Europe+2013.pdf>>.
- [12] TUČEK, J. *Šedá a černá ekonomika a její ekonomické implikace*. Brno : Masarykova univerzita, 2006.
- [13] Český statistický úřad. *Telekomunikační a internetová infrastruktura - Pevná telefonní síť, analýzy a komentáře* [online]. 2014. [cit. 2. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/pevna_telefonni_sit_telekomunikacni_a_internetova_infrastruktura/\protect\T1\textdollarFile/2013_pevna.pdf>.
- [14] Český statistický úřad. *Telekomunikační a internetová infrastruktura - Definice pojmů* [online]. 2013. [cit. 2. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/definice_pojmu/\protect\T1\textdollarFile/2013_definice_infra.pdf>.

- [15] Český statistický úřad. *Hrubý domácí produkt (HDP) - Metodika* [online]. 2014. [cit. 2. 12. 2014]. Dostupné z: <http://notes.czso.cz/csu/redakce.nsf/i/hruby_domaci_produk_t_%28hdp%29>.
- [16] Český statistický úřad. *Definice a vymezení výdajů a investic v ICT v ČR* [online]. 2013. [cit. 3. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/definice_a_vymezeni_vydaju_a_investic_v_ict/\protect\T1\textdollarFile/definice_a_vymezeni_vydaju_a_investic_v_ict.pdf>.
- [17] Český statistický úřad. *Informační ekonomika v číslech 2013* [online]. 2014. [cit. 3. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/2013edicniplan.nsf/publ/9707-13-r_2013>.
- [18] Český statistický úřad. *Informační ekonomika - Investice v ICT* [online]. 2014. [cit. 3. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/investice_v_ict>.
- [19] Český statistický úřad. *Informační ekonomika - Výdaje a investice v ICT v České republice (1993-2012)* [online]. 2014. [cit. 3. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/vydaje_a_investice_v_ict_v_ceske_republice/\protect\T1\textdollarFile/vydaje_a_investice_v_ict_1993_2012.xls>.
- [20] Český statistický úřad. *HDP, národní účty, Roční národní účty* [online]. 2014. [cit. 2. 12. 2014]. Dostupné z: <<http://apl.czso.cz/pll/rocenka/rocenka.indexnu>>.
- [21] Český statistický úřad. *Telekomunikační a internetová infrastruktura - Pevná telefonní síť, podrobná data* [online]. 2014. [cit. 2. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/pevna_telefonni_sit_xls/\protect\T1\textdollarFile/2014_pevna.xlsx>.
- [22] Český statistický úřad. *Telekomunikační a internetová infrastruktura - Mobilní telefonní síť, podrobná data* [online]. 2014. [cit. 2. 12. 2014]. Dostupné z: <http://www.czso.cz/csu/redakce.nsf/i/mobilni_telefonni_sit_xls/\protect\T1\textdollarFile/2014_mobil.xlsx>.

Příloha A

Seznam zkratek

AFE	Average Forecasting Error
ČSÚ	Český statistický úřad
ČTÚ	Český telekomunikační úřad
GSM	Groupe Spécial Mobile
HDP	Hrubý domácí produkt
ICT	Information and Communication Technologies
IP	Internet Protocol
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
OECD	Organisation for Economic Co-operation and Development
SIM	Subscriber Identity Module
SSE	Sum of Squared Error
TIE	Theil's Inequality Coefficient
VoIP	Voice over Internet Protocol

Příloha B

Vstupní data - HDP

rok	čtvrtletí	HDP v mil. Kč	rok	čtvrtletí	HDP v mil. Kč
1995	Q1	349683	2005	Q1	748169
	Q2	390304		Q2	816100
	Q3	410608		Q3	823330
	Q4	429520		Q4	870373
1996	Q1	408373	2006	Q1	797802
	Q2	450587		Q2	870410
	Q3	466991		Q3	896437
	Q4	486671		Q4	942482
1997	Q1	439588	2007	Q1	882421
	Q2	484563		Q2	949243
	Q3	502337		Q3	976901
	Q4	526823		Q4	1023254
1998	Q1	484005	2008	Q1	926672
	Q2	541239		Q2	1014549
	Q3	554510		Q3	1039324
	Q4	562833		Q4	1034801
1999	Q1	503329	2009	Q1	929368
	Q2	561412		Q2	980926
	Q3	571889		Q3	987262
	Q4	600670		Q4	1024271
2000	Q1	535093	2010	Q1	914538
	Q2	594227		Q2	1000113
	Q3	608866		Q3	1004700
	Q4	634444		Q4	1034300
2001	Q1	578875	2011	Q1	934091
	Q2	639594		Q2	1014428
	Q3	655533		Q3	1020472
	Q4	688677		Q4	1053419

2002	Q1	611813	2012	Q1	955024
	Q2	671613		Q2	1018386
	Q3	684430		Q3	1019127
	Q4	706778		Q4	1055138
2003	Q1	639690	2013	Q1	943436
	Q2	702666		Q2	1018633
	Q3	718964		Q3	1038376
	Q4	739843		Q4	1085815
2004	Q1	694424	2014	Q1	987177
	Q2	754045		Q2	1068175
	Q3	776391		Q3	1092698
	Q4	832800		Q4	1118091

Tabulka B.1: Neočistěné hodnoty HDP

Příloha C

Vstupní data - výše investic

rok	výše investic v mil. Kč
1993	2937
1994	3451
1995	3882
1996	14134
1997	22554
1998	25426
1999	25417
2000	31156
2001	31838
2002	18484
2003	11290
2004	12018
2005	10989
2006	10205
2007	12288
2008	11758
2009	14104
2010	16376
2011	16193
2012	10276

Tabulka C.1: Výše investic do telekomunikačního vybavení

Příloha D

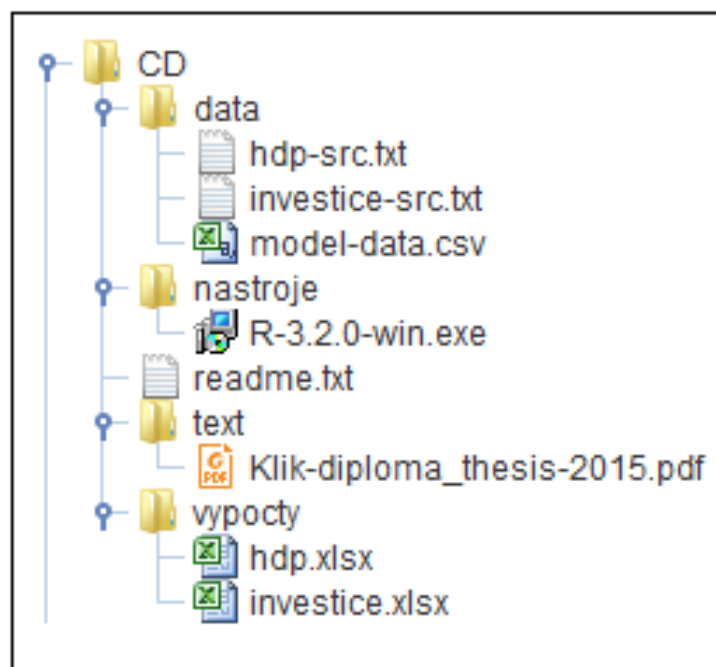
Vstupní data - model

rok	investice	HDP	pevné	SIM	broadband	hovory
1998	25426	2142587	3744776	960500	0	88653280
1999	25417	2237300	3724427	1944500	0	93367490
2000	31156	2372630	3676426	4338100	0	100286360
2001	31838	2562679	3853922	6946800	0	101793830
2002	18484	2674634	3669743	8581400	0	61879980
2003	11290	2801163	3619273	9708900	0	53004880
2004	12018	3057660	3423781	10584000	0	35859460
2005	10989	3257972	3210400	11450000	625921	37621250
2006	10205	3507131	2804437	12326000	1229962	26904840
2007	12288	3831819	2354137	13055000	1720891	33612830
2008	11758	4015346	2264351	13780165	2036256	21476550
2009	14104	3921827	2069756	14217162	2375624	43147020
2010	16376	3953651	1889000	14359000	2783000	27123120
2011	16193	4022410	1683000	14215000	3086000	26123560
2012	10276	4047675	1514000	14456000	3399000	18830020

Tabulka D.1: Vstupní data ekonometrického modelu

Příloha E

Obsah přiloženého CD



Obrázek E.1: Obsah přiloženého CD