Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky

bakalářská práce

2018 Lukáš Veverka

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Název bakalářské práce:

Ekonometrická analýza efektu komunikačních kanálů na webové vyhledávání

Autor: **Lukáš Veverka**

Katedra: **Katedra ekonometrie**

Obor: **Matematické metody v ekonomii**

Vedoucí práce: **Ing. Tomáš Formánek, Ph.D.**

Prohlášení

Prohlašuji, že jsem bakalářskou práci na téma *Ekonometrická analýza efektu komunikačních kanálů na webové vyhledávání* zpracoval samostatně. Veškerou literaturu a další podkladové materiály uvádím v seznamu použité literatury na stranách 47–49.

V Praze dne 3. května 2018 .....

Lukáš Veverka

Poděkování

Tímto bych chtěl poděkovat Ing. Tomášovi Formánkovi, Ph.D. za vedení mé bakalářské práce, za obrovskou trpělivost a skvělé rady, které práci obohatily. Ještě bych rád poděkoval doc. Ing. Janu Zouharovi, Ph.D. za ochotu, jelikož mi několikrát poradil, přestože nevedl mou práci a dále prof. RNDr. Ing. Michalovi Černému, Ph.D., který celému našemu ročníku ukázal zcela jiný pohled na psaní odborných prací. Velké díky patří i společnosti Nielsen Admosphere, a. s. za poskytnutá data.

Abstrakt

Název práce:	Ekonometrická analýza efektu komunikačních kanálů na webové vyhledávání
Autor:	Lukáš Veverka
Katedra:	Katedra ekonometrie
Vedoucí práce:	Ing. Tomáš Formánek, Ph.D.

V dnešní době, kdy data existují téměř o všem, jsou pro mnohé firmy velkou výzvou metody zpracování dat. Přestože data sbírají, neumí z nich většinou vyvozovat závěry. V marketingových odděleních je to například obtížné a nepřesné odhadování ROI nebo také problém nevyzpytatelné poptávky. K objasnění obou zmíněných problémů je vhodné použít ekonometrickou analýzu. Empirický výzkum bude zaměřen na analýzu poptávky po léku *ACC*, kterou ovlivňuje zajisté např. sezónnost, teplotní výchylky a mediální komunikace. Pomocí ekonometrie je možné odlišit vliv jednotlivých faktorů na poptávku a díky tomu poté vypočíst ROI, což je velmi cenný ukazatel pro marketingové ředitele.
Velmi výhodné je také využít možnosti *reproducible analysis*, která umožní daný postup kdykoli znovu zopakovat a navíc do něj nahlédnout. Jednou z možností takové analýzy je skript ve statistickém programovacím jazyce R.

Klíčová slova: data-driven marketing, ROI, komunikační mix, ekonometrická analýza.

Abstract

Title:	Econometric analysis of the effect of media types on searching the web
Author:	Lukáš Veverka
Department:	Department of Econometrics
Supervisor:	Ing. Tomáš Formánek, Ph.D.

In these days, when there are data almost about everything, it is a big challenge for companies to process all the data. Even though they collect them, they usually cannot infer decisions from them. For example in marketing departments it is a complicated and inaccurate estimation of ROI or unpredictable demand. It is appropriate to use econometric analysis in these cases. Empirical research will analyze the demand for medicament *ACC* which is affected by e.g. seasonality, temperature deviation and media communication. With econometrics it is possible to decompose the effect of each factor on demand and therefore calculate ROI which is a very valuable indicator for marketing directors.
Very convenient is to use *reproducible analysis* which allows not only to repeat the procedure anytime but also to look into it. One of the possibilities is to use script in the statistical programming language R.

Keywords: data-driven marketing, ROI, promotional mix, econometric analysis.

Obsah

Úvod **1**

1 O léku ACC **2**

2 Data **4**
2.1 Vyhledávání na Google . . . 5
2.1.1 Google Universal Search . . . 5
2.1.2 Zájem podle Google Trends . . . 6
2.2 Data o mediální komunikaci . . . 7
2.2.1 Dělení médií . . . 7
2.2.2 Měření médií . . . 8
2.2.3 Specifika televizní komunikace . . . 9
2.2.4 Data Ad Intel . . . 11
2.2.5 Analytický modul Adwind Kite . . . 12
2.3 Teplota . . . 12
2.4 Načtení dat do Rka . . . 12

3 Teorie **14**
3.1 Stacionarita a ADF test . . . 14
3.1.1 Stacionarita . . . 14
3.1.2 Augmented Dickey–Fuller test . . . 16
3.2 Fourierova transformace . . . 17
3.3 Frisch-Waugh-Lovell věta . . . 18
3.4 Carry-over efekt . . . 19

4 Empirická analýza **21**
4.1 Dekompozice a očištění dat . . . 21
4.1.1 Počasí a jeho vliv . . . 28
4.2 Započítání carry-over efektu . . . 32
4.3 Výběr nejlepšího modelu . . . 33
4.4 Testy a úprava modelu . . . 36
4.5 Interpretace výsledků . . . 39
4.6 ROI a optimální strategie . . . 42

5 Závěr **46**
5.1 Prostor pro zlepšení . . . 46

Literatura **49**

Příloha **50**
R kód pro teoretickou část 50
Tabulky 53

Úvod

Marketingové prostředí nabízí celou řadu možností, jak aplikovat ekonometrickou analýzu, a proto jsem se rozhodl, že svou závěrečnou práci budu směřovat právě k marketingu. Jednou z možností bylo získat data o mediální komunikaci a zjistit její vliv. Vliv na co? Data o prodejích jsou velmi citlivá, a proto by je firmy jen velmi málo pravděpodobně poskytly k akademickým účelům. Samozřejmě menší či rodinné firmy by s tím nemusely mít problém, na druhou stranu ale neinvestují do reklamy (a hlavně různých mediatypů) tolik, jako větší firmy. Při výběru tématu bakalářské práce zaměřené na ekonometrickou analýzu jsem se tedy ze začátku soustředil především na dostupnost dat.

Při práci v online marketingu jsem narazil na velmi zajímavý volný datový zdroj – *Google Trends for marketers*. Napadlo mě, že by mohl poskytnout data pro mou vysvětlovanou proměnnou, kterou by se stala vyhledávanost pojmu ACC na Googlu. Po přečtení několika odborných článků jsem zjistil, že poptávka (tudíž i prodej) je poměrně dost provázána s vyhledávaností na Googlu. Více bude zmíněno v kapitole Data.

Po získání dat pro vysvětlovanou proměnnou jsem začal hledat data, která by ji pomohla vysvětlit – především data o mediální komunikaci. Zde jsem měl štěstí, že mi společnost Nielsen Admosphere, a. s. umožnila použít poměrně velké množství dat o mediální komunikaci v kategorii *Léky proti kašli, chřipce a nachlazení* rozdělené podle mediatypů. Jedinou výjimku tvoří data o online reklamě, která jsou ve vlastnictví Sdružení pro internetový rozvoj v České repubice, z. s. p. o. (SPIR).

V celé práci bych chtěl určit faktory, které ovlivňují zájem o lék ACC. Těmito faktory může být mediální komunikace značky ACC, počasí, ale i komunikace konkurence. Vzhledem k provázanosti prodejních dat s daty Google Trends následně provedu modelový odhad ROI a doporučím postup pro příští strategii založenou na tzv. data-driven marketingu. Proč jen modelový odhad ROI? K přesnému výpočtu ROI by bylo nutné znát čistý zisk z jednoho výrobku a to je také velmi citlivá informace. Nicméně i odhadnutý zisk poslouží jako vzor k výpočtu.

Zpracování práce bude provedeno ve statistickém programovacím jazyce R [1], který umožňuje zapsat skript a následně provést výpočet znovu. Důvodem opětovné analýzy může být například získání většího objemu dat a zpřesnění výsledků ekonometrie.

1 O léku ACC

ACC LONG je lék na rozpouštění a vykašlávání hlenů. Má za úkol ulevit nemocnému s hleny v dýchacích cestách. Odborně se tento druh léku nazývá mukolytikum. Léčivá látka acetylcystein rozpouští všechny složky, které způsobují vazkost (větší viskozitu) hlenu. Tím se vazkost hlenu sníží a usnadní se vykašlávání.

Využití přípravku probíhá především při léčbě akutních a chronických onemocnění dýchacích cest, provázených intenzivní tvorbou hustého hlenu a vlhkým kašlem jako jsou například chřipka, akutní i chronické bronchitidy (zánět průdušek), záněty průdušnice, průduškové astma, bronchiektázi (rozšíření průdušek), bronchiolitidy (zánět průdušinek), mukoviscidóza (cystická fibróza, vrozená porucha zkapalňování hlenů).

Dále se přípravek může užívat při léčbě akutní i chronické sinusitidy (zánět vedlejších nosních dutin), zánětu hrtanu a zánětu středního ucha s výpotkem. Využití najde i u předoperačních a pooperačních stavů, u kterých by docházelo k větší tvorbě vazkého hlenu v dýchacích cestách.

U akutních onemocnění dýchacích cest, jako jsou chřipka nebo akutní zánět průdušek či průdušnice, lze užívat přípravek bez porady s lékařem. Naopak u chronických onemocnění dýchacích cest (např. u mukoviscidózy, průduškového astmatu a u bronchiektázií) smí být přípravek užíván pouze na doporučení lékaře. Přípravek je možné užívat od 14 let. [2]

Účinná látka N-acetylcystein

N-acetylcystein (NAC) je ve svých indikacích terapeuticky využíván již více než čtyřicet let. Jde především o jeho využití jako prekurzoru pro výrazně antioxidačně působící endogenně přítomný glutathion (GSH) v případě intoxikace paracatemolem, halogenovanými uhlovodíky nebo při otravě houbami. V posledních letech však je ceněn především pro své mukolytické účinky a setkáváme se s ním rovněž v léčbě některých virových onemocnění nebo u neurologických onemocnění (Parkinsonova choroba aj.) či v psychiatrii (Alzheimerova choroba aj.). V rámci testovaných indikací v pediatrii jde mj. o akutní selhání jater nezpůsobené paracetamolem, o léčbu Prader-Willi syndromu, lipofuscinózu aj. Po chemické stránce jde o velmi jednoduchou molekulu sestávající ze zbytku kyseliny octové a z esenciální aminokyseliny cysteinu. [3]

Obrázek 1.1: obal ACC LONG, šumivé tablety

zdroj: www.benu.cz

Využití NAC v pediatrii

Dnes jsou přípravky s obsahem NAC využívány především k léčbě akutních či chronických onemocnění dolních dýchacích cest, která jsou charakteristická tvorbou viskózního hlenu se zvýšeným vykašláváním – využíván je tedy zejména u zánětlivých onemocnění nebo u cystické fibrózy. [3]

Akutní či chronické infekce dolních dýchacích cest

N-acetylcystein je v těchto případech užíván zejména za účelem odstranit vlhký kašel. Paradoxem je, že využití léčivých látek s podobným mechanismem je běžné prakticky ve všech zemích Evropy, zatímco ve Spojených státech amerických je využití zanedbatelné. U dětí s infekcí dolních cest dýchacích proběhlo hodnocení mukolytického účinku NAC v rámci několika studií, kde se ukázalo především na zkrácení epizody kašle, nebo alespoň snížení jeho intenzity. Bohužel se často vůbec neporovnávala reálná účinnost na větším souboru nemocných a placebo efekt. Z pohledu účinnosti NAC v případech zánětu dolních cest dýchacích byla provedena významná meta-analýza pod záštitou Cochranovy knihovny. Analyzován byl vzorek 497 osob léčených NAC nebo karbocysteinem v rámci šesti studií zaměřených na účinnost obou těchto látek. Bylo zde prokázáno mírné potlačení kašle. Co se týče bezpečnosti, tak NAC je obecně příznivá látka s výjimkou dětí do 2 let, kde díky nedostatku informací neexistuje jednotné vyjádření, případně se uvádí vyšší riziko výskytu nežádoucích účinků. [3]

Cystická fibróza

V případě cystické fibrózy je účinnost NAC hodnocena již od poloviny 60. let minulého století. Za tímto účelem je podáván nejen v zemích Evropské Unie, ale schválil ho i americký Úřad pro potraviny a léčiva (FDA). [3]

2 Data

V této kapitole budou popsána data a způsob jejich získávání. Protože v empirické analýze bude zkoumáno, co vše má vliv na vyhledávání pojmu ACC na Googlu, budou zde popsáná data popisující zájem o určitý produkt (z Google Trends), dále data o mediální komunikaci a o teplotě.

Data jsou údaje popisující nějaký jev. Mohou mít formu čísel, slov, zvuků nebo obrazů. Nejčastěji jsou to však čísla nebo text. Mohou být uložena na záznamových médiích v papírové nebo elektronické podobě. V případě záznamů o mediální komunikaci jsou data například v podobě databáze, jakožto systému s pevně danou strukturou záznamů. [4] Pokud proběhne v televizi reklamní spot, tak společnost zabývající se monitoringem reklamy zaznamená vše okolo této události – krásný příklad pořízení dat. Data mohou mít ale i povahu obrázku, zvukové nahrávky či videa (např. konkrétní ukázka reklamního spotu). Z nich lze poté získat informace obvykle kategoriální povahy, jako například strategie brandingu určité značky.

Pro řešení problému bude zapotřebí použít několik druhů dat:

- **Strukturovaná data** – jsou ta, která jsou přesně definována, dají se snadno ukládat a analyzovat. Většina databází pracuje právě se strukturovanými daty.
- **Semistrukturovaná data** – jedná se o typ strukturovaných dat bez přesného modelu. Příkladem mohou být HTML stránky, které se skládají z nestrukturovaných dat (textového obsahu a obrázků) a strukturovaného obsahu (hlavičky s přesně definovanou strukturou – titulek, popis, klíčová slova atd.).
- **Nestrukturovaná data** – nemají jakkoliv definovanou strukturu (např. obrázky, videa nebo obsah proběhlé komunikace). Naprostá většina generovaných dat jsou tvořena právě tímto druhem. Známý pojem *big data* se vztahuje především k nim.

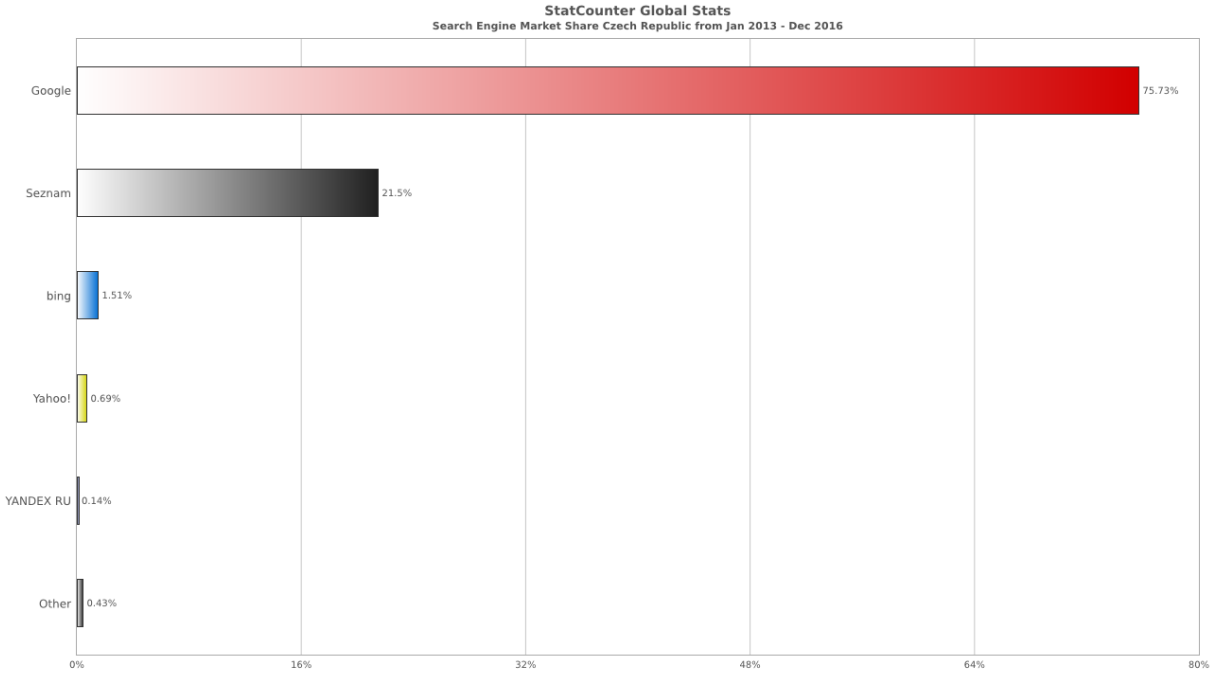
Způsob zpracování dat se také liší podle typu proměnné. Proměnná představuje vždy jeden údaj v databázi (např. konkrétní rok nebo značku produktu). Zde je také důležité vymezit jejich dělení:

- **Kvantitativní** – Hodnoty jsou měřitelné, tzn. lze u nich jasně určit o kolik je dané kritérium vyšší či nižší. Jedná se například o věk, plat nebo investice do mediální komunikace.
- **Nominální** – U těchto hodnot můžeme pouze určit, zda jsou různé. Nemůžeme stanovit jejich pořadí. Příkladem nominální proměnné může být pohlaví, značka produktu nebo název společnosti.

- **Ordinální** – V této kategorii lze stanovit pořadí, avšak není možné určit, o kolik je jedna hodnota větší či menší než druhá. Jako příklad poslouží stupeň dosaženého vzdělání nebo velikost oblečení.

2.1 Vyhledávání na Google

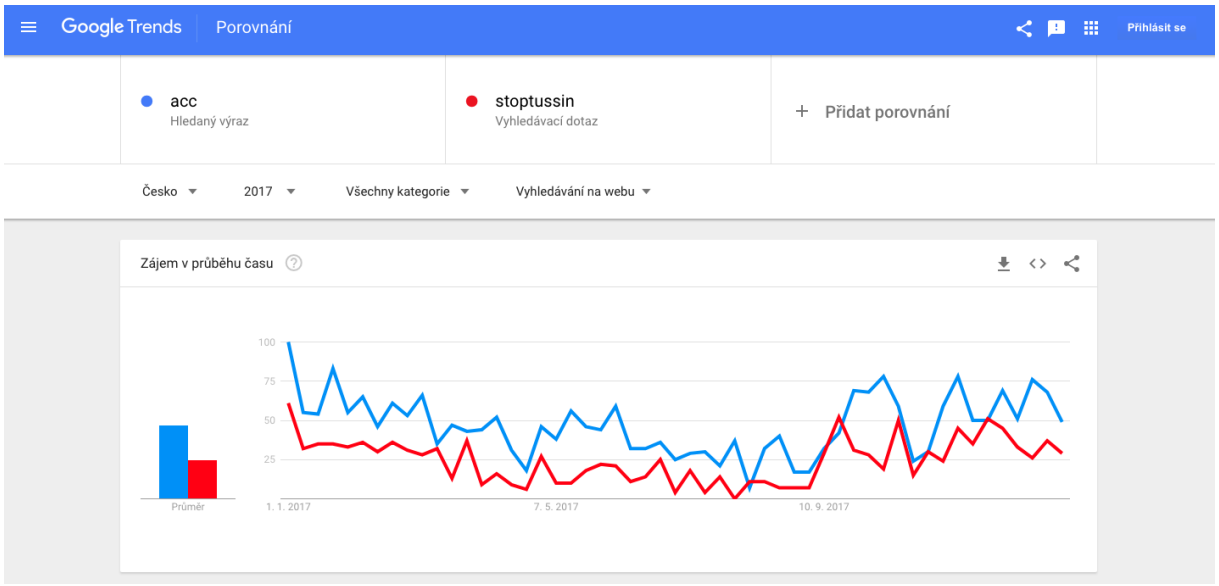
Existuje mnoho druhů vyhledávačů, ale Google dosáhl jako vyhledávač celosvětově dominantního postavení. V České republice i přes velkou popularitu Seznamu jasně vede Google (viz obrázek 2.1). V dnešní době je jen málo zemí, kde tomu tak není – např. Rusko (Yandex 54 %) nebo Čína (Baidu 58,5 %). [5] Ve většině zemí má však takovou převahu, že ostatní vyhledávače jsou prakticky zanedbatelné. O vyhledávači Google je také známo, že preferuje jednoduchost, a proto si v roce 2009 nechal patentovat svou domovskou stránku na následujících 14 let. [6]



Obrázek 2.1: Tržní podíl vyhledávačů v České republice
zdroj: gs.statcounter.com

2.1.1 Google Universal Search

Univerzální hledání je vyhledávání nejen v běžných webových stránkách, ale také v obrázcích, zprávách, videu nebo na sociálních sítích. Toto vyhledávání je nyní nastaveno jako standard, a proto pokud uživatel chce hledat jen např. v obrázcích, musí si takové hledání nastavit. Přínosy pro uživatele jsou ovšem rozporuplné. Na jednu stranu může snadno a rychle získat relevantní informace včetně obrázků a map, avšak na druhou stranu může být požadovaná informace jen v jednom zdroji (nejčastěji pouhý text) a dostat se k ní bude mnohem složitější. [7]



Obrázek 2.2: Ukázka Google Trends

zdroj: trends.google.com

2.1.2 Zájem podle Google Trends

Data poskytovaná veřejně a volně od Googlu jsou krásnou ukázkou semistrukturovaných dat – uživatel zadá do vyhledávače text, Google ho následně analyzuje a poskytne podle jeho uvážení nejlepší a nejrelevantnější možné výsledky. Znamená to tedy zařazení nestrukturovaného vyhledávacího dotazu do určité kategorie. Data o těchto zařazeních jsou poté uveřejněna na stránce Google Trends. Využívají toho především lidé v marketingu, kteří tak mají velmi jednoduchou možnost zkontrolovat aktuální témata (potažmo skandály) kdekoliv na světě na základě Google vyhledávání. Data jsou poskytována velmi rychle (možnost sledovat poslední hodinu po minutách). Na obrázku 2.2 je vidět porovnání zájmu v průběhu roku 2017 o lék ACC a Stoptussin v České republice. Je patrné, že větší popularitu měl lék ACC.

Již mnoho vědců v minulosti potvrdilo provázanost vyhledávanosti na Googlu s mnoha ukazateli v různých odvětvích. [8] Proto v této práci jako vysvětlovaná proměnná poslouží právě Google index, o kterém se dá předpokládat, že je silně zkorelovaný s reálnými prodeji. Konkrétně se tedy bude jednat o vyhledávání léku ACC v letech 2013–2016 rozdělené po týdnech.

Trendy Google udávají, kolikrát bylo klíčové slovo vyhledáno jako zlomek celkového počtu vyhledávacích dotazů ve stejném období. Týdenní ukazatel GI (Google index) se pak vypočítá tak, že se nejprve shromáždí týdenní trendové hodnoty všech klíčových slov, které byly vyhledány v konkrétní (předem definované) oblasti, a poté se vydělí každá agregovaná týdenní hodnota součtem všech týdenních hodnot za dané období. Získaný údaj lze pak vynásobit 100 pro lepší interpretaci. Pokud by nová data měla přesáhnout hranici 100 %, tak by Google Trends časovou řadu opět předimenzoval tak, aby se maximum rovnalo 100 % a předešlé hodnoty by adekvátně snížil. Vzorec pro získání indexu Google $GI_{t,i}$

v týdnu t v zeměpisné oblasti i, je zde:

$$GI_{t,i} = \left[\frac{\frac{S_{t,i}}{R_{t,i}}}{\max_t(\frac{S_{t,i}}{R_{t,i}})} \right] \times 100, \quad t \in [1, N_w] \tag{2.1}$$

kde $S_{t,i}$ představuje počet vyhledávacích dotazů a $R_{t,i}$ označuje celkový počet vyhledávacích dotazů v geografické oblasti i v každém týdnu t v celkových N_W týdnech. [9]

2.2 Data o mediální komunikaci

Jedna z důležitých věcí hned na začátku je definovat pojem média, aby nedošlo k nedorozuměním. Rozumějme tím sdělovací prostředky masové komunikace, jejichž cílem je šíření informací a zábavy mezi široké vrstvy obyvatelstva. Jednou z významných funkcí médií je i funkce nosiče reklamního (propagačního) sdělení. [4]

2.2.1 Dělení médií

Existuje mnoho kritérií, jak rozdělovat média. Jednou z možností je například podle jejich vlastnictví:

- **Veřejnoprávní média** je označení pro sdělovací prostředky, které mají příjem z veřejných peněz a právně definovaný rozsah a povinnosti. Cílem jejich provozování je zabránit zneužití médií politickými stranami a podpořit média, jako čtvrtý pilíř demokracie. [10] V případě České republiky se jedná o Český rozhlas a Českou televizi.
- **Komerční média** jsou vlastněna soukromou osobou či subjektem a jejich cílem je generování finančního zisku. Nejčastěji je hlavním zdrojem zisku příjem z reklam nebo přímých prodejů. V České republice se jedná například o TV Nova s. r. o. (dříve CET 21 spol. s r. o.) či MAFRA, a. s.

Dalším kritériem, podle kterého lze média dělit, je způsob a podoba jejich šíření. Existují tyto základní mediatypy:

- **Televize:** Jednotlivé televizní stanice. Například Nova, Nova 2, ČT1 atd.
- **Rádio:** Rozhlasové stanice. Například Evropa 2, Radio Blaník, Rádio Impuls atd.
- **Tisk:** Časopisy (tiskovina opakovaně vycházející v určitém intervalu) a deníky (denně vycházející tiskovina). Například Hospodářské noviny, Blesk, Forbes, ABC atd.
- **Internet:** Jednotlivé webové stránky (provozovatele) nabízející mediální obsah včetně reklamního prostoru. Rozdělují se ještě na klasická média, která vytváří svůj vlastní obsah (Novinky.cz, Stream.cz, Penize.cz atd.) a média odkazující na další stránky (například Google nebo Seznam).
- **OOH:** Reklamní nosiče, které je možné zahlédnout mimo domov (odtud zkratka Out Of Home). Dělí se zejména z hlediska různých typů nosičů (billboardy, zastávky městské hromadné dopravy, plakáty ve vlacích, sloupy pouličního osvětlení atd.)

- **OOH-TV:** Specifický případ předešlého mediatypu OOH vyznačující se de facto televizí ve veřejném prostoru. Příkladem mouhou být televizory v obchodech (např. v lékárnách promítající léky na aktuální nejčastější nemoc) nebo velká promítací plátna ve městech.

2.2.2 Měření médií

Pro rozvoj médií (stejně jako každého jiného subjektu na trhu) je zapotřebí mít spolehlivé informace, aby bylo možné učinit správná strategická rozhodnutí. Cílem médií je získat co nejvíce diváků, posluchačů či čtenářů. Právě proto je naprostá většina výzkumů o médiích zaměřena na hodnocení atraktivity obsahu pro konzumenty. Samozřejmě existují různé metody výzkumů a měření, které mohou mít odlišné (někdy i velmi odlišné) výsledky. Lidé zodpovědní za správná rozhodnutí zajisté narazí na problém, které z měření je to „správné“. Vyspělé trhy proto mají zájem o jeden trhem uznávaný výzkum na základě odsouhlasené metodiky. Trhem uznávané mediální výzkumy a měření jsou někdy označována jako *one currency*, protože slouží k získání objektivní představy o výkonu jednotlivých médií. Výsledky těchto výzkumů jsou důležitým podkladem k tvorbě ceníků reklamního prostoru v médiích. Zajímají se o ně i mediální agentury, které je využívají k mediálnímu plánovaní a tvorbě tzv. mediamixu (kombinaci médií pro maximální zásah dané cílové skupiny). Existují vesměs dva druhy výzkumu sledovanosti: [4]

Meření médií je zaznamenávání mediálního chování pomocí softwaru nebo měřících zařízení po určitou dobu nebo kontinuálně. Typickým příkladem je v současné době měření sledovanosti televize nebo rádia pomocí speciálních zařízení – tzv. peoplemetrů pro TV (obrázek 2.3) a audiometrů pro rádia. V současné době jsou k dospizici dva různé způsoby měření s využitím elektronického zařízení pro identifikaci vysílání:

- Audiomatching – porovnává nahrané zvukové stopy s tzv. referenčními zvuky (uloženými záznamy původního vysílání)
- Watermarking – do televizního či rozhlasového vysílání se vloží neslyšitelná značka, která je následně identifikována

Příkladem takovéhoto výzkumu je *PEM (2013–2017)* a *PCEM (2018–2022)* – měření televizní sledovanosti, kde je využita technologie audiomatchingu a zadavatelem je Asociace televizních organizací (ATO).

Mediální výzkum je založený na pravidelném nebo jednorázovém dotazování respondentů. I zde jsou dvě nejrozšířenější metody:

- Deníčkový výzkum – při tomto výzkumu reprezentativní vzorek zaznamenává svou mediální konzumaci do tzv. deníčku. Většinou se nyní využívá online verze, kde odpadá nutnost posílání fyzického předmětu.
- Recall – respondenti jsou dotazováni na svou mediální konzumaci v uplynulém časově omezeném období. Nejčastější přístup je *day after recall*, kdy respondenti uvádějí, na co se včera dívali, co četli nebo poslouchali.

Jako příklad lze uvést *Radioprojekt* – výzkum poslechovosti rádií, jejímž zadavatelem je Asociace provozovatelů soukromého vysílání (APSV) a Rozhlasová sekce Sdružení komunikačních a mediálních agentur (SKMO) a využívá se metody recall.

Obrázek 2.3: Ukázka peoplemetru (TV-metru)
zdroj: www.peoplemetry.cz

2.2.3 Specifika televizní komunikace

Televizi jakožto nejvlivnějšímu mediatypu je rozumné věnovat více pozornosti. Zpravidla se do ní totiž investuje nejvíce. Specifická pro ni je poměrně komplikovaná tvorba ceny. Vzhledem k mnoha slevám či cenovým přirážkám je často matoucí výsledná ceníková cena uváděná v monitoringu mediální komunikace. Před vysvětlením samotné tvorby ceny je vhodné vymezit a vysvětlit několik pojmů.

Rating je počet lidí (v populaci nebo cílové skupině), kteří v daném časovém intervalu sledovali program nebo reklamní sdělení. Nejčastější vyjádření je v procentech.

$$Rating(\%) = \frac{S}{P} \times 100, \tag{2.2}$$

kde S je počet lidí sledujících daný program nebo spot a P je velikost populace nebo cílové skupiny.
Příklad: Velikost populace je 10 000 000, sledovanost pořadu je 4 000 000. Rating pořadu je tedy $\frac{4000000}{10000000} \times 100 = 40$ %.

GRP *(Gross rating point)* je 1 procento z cílové skupiny, které bylo zasaženo daným médiem (velikost zásahu reklamní kampaně). Obecně se používá pro cílovou skupinu 15+, která je považována za ekvivalent koupěschopné populace. Tato veličina vzniká jako součet všech *ratingů* (sledovaností), proto může být výsledná hodnota vyšší než 100 (viz tabulka 2.1). Výpočet by se dal zapsat jako: $\sum \frac{S}{P} \times 100$, kde značení je stejné jako ve vzorci 2.2.

TRP *(Target rating point)* je veličina, která se používá pro popis zásahu kampaně v konkrétní cílové skupině – ta je definována zadavatelem reklamy (např. u pánského sprchového gelu to může být cílová skupina: muži 25–39 let). TRP vzniká stejně jako GRP jako součet *ratingů* (sledovaností), avšak s rozdílem že jak velikost populace, tak počet sledujících je omezen na konkrétní cílovou skupinu. Oproti GRP by výpočet vypadal takto: $\sum \frac{S_{tg}}{P_{tg}} \times 100$, kde P_{tg} je celková velikost cílové skupiny

tg (např. M 25–39: 1 000 000 lidí) a S_{tg} je počet lidí z dané cílové skupiny, kteří sledovali daný program nebo spot. Důležité je zmínit, že TRP může být větší než GRP (v tabulce 2.1 je názorný příklad).

Afinita je index popisující vhodnost konkrétního reklamního nosiče pro cílovou skupinu. Charakterizuje, jak konkrétní cílová skupina sleduje dané médium ve srovnání s obecnou populací. Z toho lze usoudit, že čím vyšší afinita, tím je médium vhodnější pro oslovení konkrétní cílové skupiny. Výpočet afinity není složitý:

$$Afinita = \frac{TRP}{GRP} \tag{2.3}$$

Příklad: Je zapotřebí určit, zda je TV kanál Nova Action vhodný pro reklamní kampaň zaměřenou na cílovou skupinu mužů 25–39 let. K dispozici máme pozorování z tabulky 2.1. V případě celé populace měly reklamní spoty *rating* (sledovanost) 40 % a 70 % (což v součtu dává 110 GRPs), zatímco u cílové skupiny byla sledovanost vyšší – 70 % a 80 % (v součtu 150 TRPs). Za použití vzorce 2.3 je snadné dosadit a zjistit, že afinita $= \frac{150}{110} = 1,36$. Což udává, že muži ve věku 25–39 let sledují Nova Action relativně více než populace a médium je tedy vhodné pro použití v kampani.

Tabulka 2.1: GRP vs. TRP

GRP				TRP			
Nova Action	1. spot	2.spot		Nova Action	1.spot	2.spot	
zásah	4×10^6	7×10^6	$\sum =$	zásah	7×10^5	8×10^5	$\sum =$
rating	40 %	70 %	110 GRPs	rating	70 %	80 %	150 TRPs
populace (A 15+) = 10^7				populace (M 25–39) = 10^6			

Nyní je možné vysvětlit podrobněji tvorbu ceny reklamy v televizi. Nákupní jednotkou reklamy v TV je právě výše zmíněný GRP. Cena za jeden GRP se nazývá CPP (cost per point), která udává, kolik klient zaplatí za oslovení 1 % osob z populace. Základním faktorem výše CPP je objem investice. Pokud se například investice bude pohybovat v rozmezí 30–40 milionů Kč, bude průměrné CPP 19 832,- Kč. Samozřejmě výše základního CPP je ovlivněna i jinými faktory především vyjednávacího charakteru (např. včasný podpis, předplatba, nový klient, apod.). Následně je CPP upravováno pomocí koeficientů, které berou v potaz období (např. duben nebo květen mají koeficienty 1,25 díky jedné z hlavních sezón reklamní komunikace). Dále je pak zohledněna část dne – např. off prime time (doba mimo hlavní vysílací čas) má keoficient 0,8. I délka spotu (tzv. stopáž) má vliv na výsledné CPP (např. jen 15 sekundový spot má CPP snížené na 0,75 původní hodnoty). Aby toho nebylo málo, tak si zadavatel může připlatit za změnu podílu stanic v rámci jedné skupiny stanic (např. u skupiny CET 21 může zažádat o větší podíl komunikace na stanici Nova Sport na úkor Nova Cinema) nebo za umístění reklamního sdělení na vyžádané pozici (tzn. první, druhá, předposlední nebo poslední pozice v rámci reklamního bloku). Pro větší přehlednost jsou všechny základní koeficienty zobrazené v obrázku 2.4. Uvedené příklady jsou konkrétní čísla z veřejně dostupných cenových podmínek televize Nova. [11]

OBJEMOVÝ INTERVAL (CZK)	PRŮMĚRNÉ REF. CPP (CZK) – NOVA GROUP
0 MIL. – 1 999 999	22354
2 MIL. – 4 999 999	22010
5 MIL. – 9 999 999	21437
10 MIL. – 19 999 999	20921
20 MIL. – 29 999 999	20348
30 MIL. – 39 999 999	19832
40 MIL. – 49 999 999	19431
50 MIL. – 59 999 999	19030
60 MIL. – 69 999 999	18571
70 MIL. – 79 999 999	18112
80 MIL. – 89 999 999	17654
90 MIL. – 109 999 999	17023

Období	Měsíční koeficient
Leden	0,7
Únor	0,8
Březen	1,05
Duben	1,25
Květen	1,25
Červen	1,15
Červenec	0,8
Srpen	0,85
Září	1,2
Říjen	1,35
Listopad	1,35
Prosinec (1.–22.)	1,15
Prosinec (23.–31.)	0,8

Část dne	Koeficient
Prime Time (18:00–19:30 a 21:30–23:00)	1
Off Prime Time (23:00–18:00)	0,8
Super Prime Time (19:30–21:30)	1,2

Délka reklamního spotu (v sekundách)	Koeficient
10	0,50
15	0,75
20	0,90
25	0,95
30	1,00
35	1,20
40	1,35
45	1,50
50	1,65
55	1,85
60	2,00

Obrázek 2.4: Faktory ovlivňující cenu spotu
zdroj: TV Nova s. r. o.

Z předešlého odstavce je patrné, že porovnání reálných výdajů určitého zadavatele s konkurencí, je pomocí ceníkové ceny uvedené v databázi spíše orientační. Daleko lepším, přesnějším a více vypovídajícím ukazatelem je počet nakoupených GRPs.

2.2.4 Data Ad Intel

Společnost Nielsen Admosphere, a. s. monitoruje komerční mediální komunikaci v České republice. Výsledkem jsou strukturovaná data v rozsáhlé databázi, která obsahuje popis všech komerčních sdělení.

Popisem reklamního sdělení je myšleno kdo, kde, kdy, co, jakým způsobem a za kolik komunikoval (tj. ceníková cena reklamního sdělení). Komunikace mohla proběhnout například skrze televizní spot či sponzorský vzkaz, tištěný inzerát nebo billboard u silnice.

V této práci bude zapotřebí jen následující popis a rozčlenění dat:

- **Rok a číslo týdne** (kvantitativní proměnná) – Data o časovém průběhu komunikace. V případě této práce data z let 2013–2016 rozdělená po týdnech.
- **Firemní značka** (nominální proměnná) – Firemní značka je značka produktové řady, pod kterou výrobce vystupuje a sjednocuje skupiny a podskupiny výrobků (např. ACC, Hasbro nebo L'oréal PARIS)
- **Kategorie produktu** (nominální proměnná) – Základem víceúrovňového stromu produktových segmentů je přizpůsobená evropská kategorizace CZ-NACE Rev.2 tak, aby vyhovovala potřebám reklamního trhu. V této práci konkrétně kategorie *Léky proti kašli, chřipce a nachlazení*
- **Mediatyp** (nominální proměnná) – Komerční sdělení jsou popsána v těchto mediatypech: *televize* (reklamní spoty, product placement, sponzoring atd.), *tisk* (plošná inzerce, PR článek atd.), *rozhlas* (rozsah poskytnutý smluvními partnery), *internet* (tato data nebudou použita, protože nejsou ve vlastnictví Nielsen Admosphere, a. s.,

ale vlastní je Sdružení pro internetový rozvoj v České republice, z. s. p. o. – SPIR) a *OOH sektor* (outdoor, OOH-TV, kinoreklama atd.)

- **Ceníková cena reklamy** (kvantitativní proměnná) – Základní cena před odečtením slev a přičtením poplatků zprostředkovávající agentuře. Tato cena se používá při konkurenčních analýzách, protože je srovnatelná. Často se také nazývá Ratecard price nebo Gross cost.

2.2.5 Analytický modul Adwind Kite

Ve světě existuje několik softwarových nástrojů zabývajících se zpracováním dat. Jednolitvé programy se liší rozsahem analýz, rychlostí a způsobem práce s daty. Jedním z takových je i modul společnosti Nielsen Admosphere, který umožňuje počítat základní analýzy ve formě kontingenčních tabulek. Analýza umožní dávat proměnné do řádků, sloupců a listů a to i ve více úrovních. Ukázka prostředí tohoto programu je na obrázku 2.5, kde je nastavená analýza konkrétně pro tuto práci.

Software Adwind Kite zpracovává: [4]

- demografická data o domácnostech a respondentech,
- živou sledovanost i odloženou sledovanost na televizorech i na počítačích,
- popisná data o pořadech, upoutávkách a spotech,
- census data o sledovanosti na internetu.

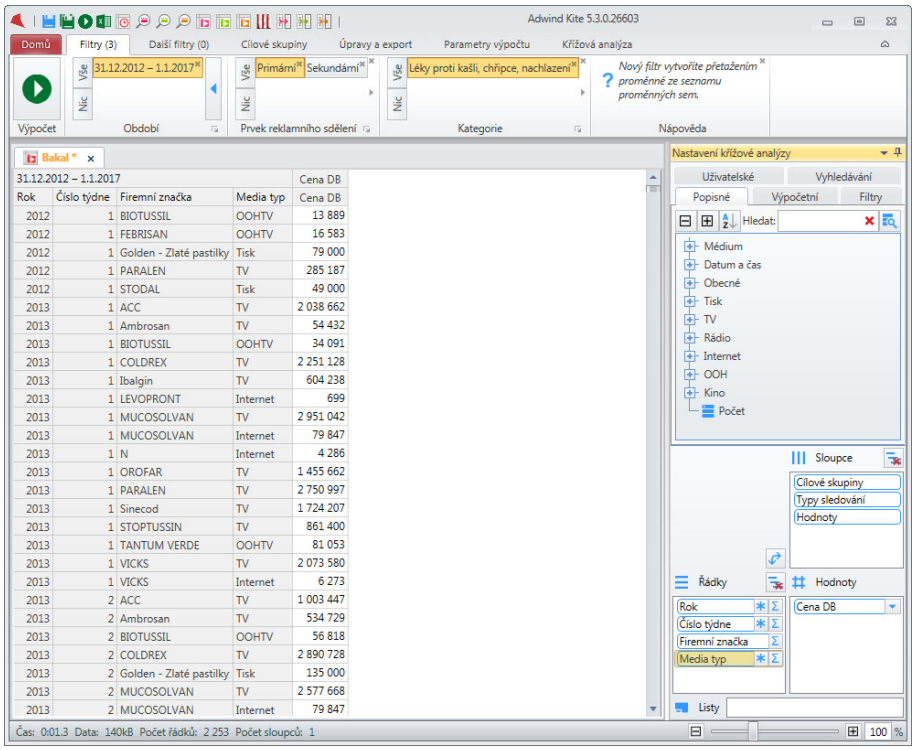
2.3 Teplota

Data o teplotě jsou k dispozici na webu Českého hydrometeorologického ústavu (ČHMÚ). Jedná se o data denní průměrné, maximální a minimální teploty vzduchu, relativní vlhkosti vzduchu, tlaku vzduchu, průměrné rychlosti větru, denního úhrnu srážek, doby trvání slunečního svitu a výšky celkové sněhové pokrývky od roku 1961. Tato data jsou zveřejněna na základě datové politiky Světové meteorologické organizace. Aktualizace datových souborů probíhá jednou ročně. [12]

2.4 Načtení dat do Rka

Ze začátku budou v práci použité 2 mírně odlišné časové řady. Důvod bude objasněn v sekci 4.1 (Dekompozice a očištění dat). Pro rychlejší načtení .xlsx souborů do Rka je vhodné využít balíčku *openxlsx*. [13]

```
install.packages("openxlsx")
library(openxlsx) # zrychli nacteni dat
setwd("~/bakalarska_prace")
data<-read.xlsx('Bakalarska_prace_data.xlsx',sheet = 1)
data52<-read.xlsx('Bakalarska_prace_data.xlsx',sheet = 2)
```



Obrázek 2.5: Nastavení analýzy pro export dat pro práci
zdroj: Adwind Kite software

```
head(data) # moznost prohlednuti dat
names(data) # nazvy jednotlivych polozek

# Zlogaritmovani vysvetlovane promenne
data$log_Zajem_ACC = log(data$Zajem_ACC)
data52$log_Zajem_ACC = log(data52$Zajem_ACC)
```

3 Teorie

V této části bakalářské práce se budu věnovat nejdůležitějším teoretickým předpokladům, které jsou využity v následující empirické analýze. Vzhledem k tomu, že zkoumám časovou řadu, bude zapotřebí objasnit pojem *stacionarita* a s ní spojený *Augmented Dickey–Fuller test*. V časové řadě se nachází také silná sezónnost, kterou budu aproximovat pomocí *Fourierovy transformace*. Vhodné je také zmínit *Frisch-Waugh-Lovell větu*, protože její princip bude použit při sezónním očišťování pozorovaných řad. Na závěr bude popsán *carry-over efekt*, který se vyskytuje v mediálním prostředí a při výzkumu efektivity komunikace je důležité s ním počítat.

Všechny grafy a pomocné obrázky byly vytvořeny pomocí jazyka R a kód je možné naleznout v příloze na straně 50. Pro zobrazení grafů byl využit balíček *ggplot2* [14], který tvoří graficky atraktivnější výstupy.

3.1 Stacionarita a ADF test

Vzhledem k tomu, že tato práce analyzuje časovou řadu, je nutné objasnit pojem stacionarní a nestacionární časová řada. Stacionarita je totiž důležitou sadou vlastností, které je nutné dosáhnout před samotnou analýzou.

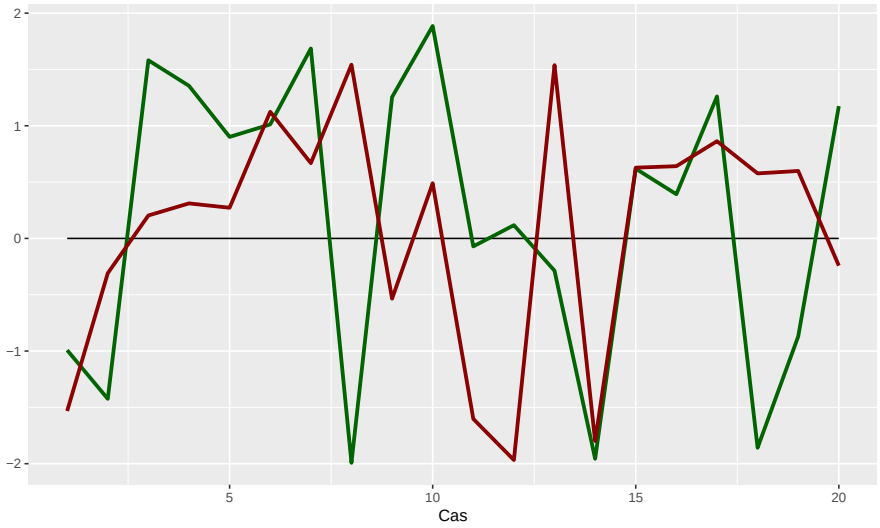
3.1.1 Stacionarita

Je vlastnost stochastického procesu vyžadující, aby její pravděpodobnostní rozdělení bylo stabilní v čase. Matematicky lze vyjádřit jako [15]:

- Střední hodnota je konstantní v čase: $\mathrm{E}(y_t) = \mathrm{E}(y_{t-h}) = \mu, \quad h = \ldots, -1, 0, 1, \ldots$
- Rozptyl se v čase nemění: $\mathrm{E}[(y_t - \mu)]^2 = \mathrm{E}[(y_{t-h} - \mu)]^2 = \sigma_y^2, \quad h = \ldots, -1, 0, 1, \ldots$
- Kovariance mezi y_t a y_{t-h} závisí pouze na h a ne na t:
  $\mathrm{E}[(y_t - \mu)(y_{t-h} - \mu)] = \gamma_h, \quad h = \ldots, -1, 0, 1, \ldots$

Slabě a silně závislé procesy:

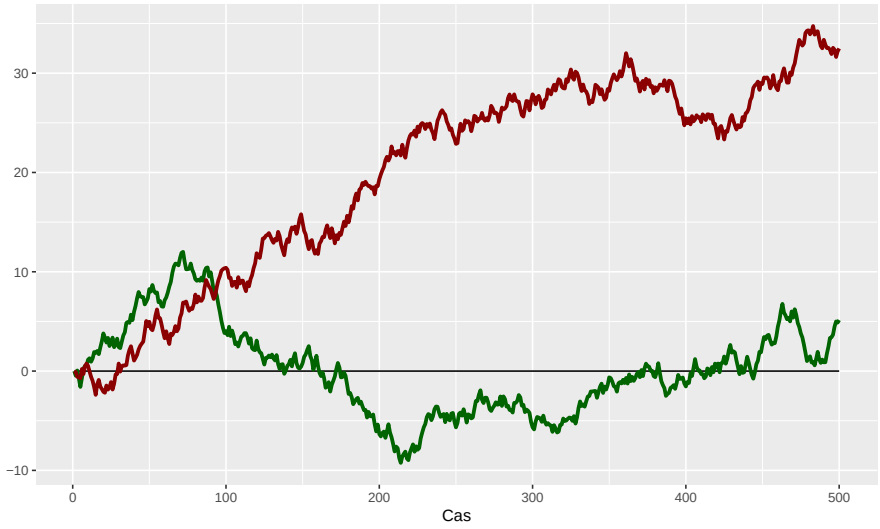
Ve slabě závislém procesu závislost mezi y_t a y_{t+h} zmizí, pokud se h blíží k nekonečnu. Slabá nebo žádná závislost je zapotřebí, aby bylo možné použít zákon velkých čísel a centrální limitní větu. V opačném případě (tzn. při silně závislém procesu) by nebylo možné použít např. t-testy (kvůli směrodatným chybám), tím pádem ani testování hypotéz a p-hodnoty. Navíc by hrozil výskyt zdánlivé regrese. Uvedu zde 2 ukázkové procesy, na kterých jsou jasně vidět jejich vlastnosti.



Obrázek 3.1: Proces bílého šumu

zdroj: autor

Bílý šum (obrázek 3.1) je příkladem nejjednoduššího stochastického procesu, který je řadou nekorelovaných náhodných veličin, jednoho pravděpodobnostního rozdělení, s nulovou střední hodnotou a konstantním rozptylem. Jeho simulaci lze provést jako: $y_t = R$, kde R je náhodné číslo z normálního rozdělení s nulovou střední hodnotou – šum.



Obrázek 3.2: Proces náhodné procházky

zdroj: autor

Náhodná procházka (obrázek 3.2) je někdy také nazývána „opilcova chůze“, protože ji velmi dobře připomíná. Představme si opilce, který jde a jeho každý krok je učiněn náhodným směrem. Záznam jeho cesty by se dal přirovnat k autoregresní funkci řádu 1 – značeno AR(1). Tento proces je silně závislý, jelikož každá jeho další hodnota závisí na té minulé. Formálně by se tento vztah zapsal: $y_t = y_{t-1} + R$, kde R představuje šum. Jedná se

o *nestacionární* proces, jelikož zde již není patrný žádný návrat k původní hodnotě. Jak již bylo zmíněno výše, u nestacionárních procesů může dojít k zdánlivé regresi. Příkladem je červená časová řada na obrázku 3.2, kde by se mohlo zdát, že má lineární trend.

Abychom dostali správné výsledky, o které je možné opřít další tvrzení, musíme tedy pracovat se stacionární časovou řadou. Můžeme ji ověřit vizuálním zhodnocením grafu (konkrétně korelogramu ACF a PACF) nebo otestovat pomocí testu jednotkových kořenů. Příkladem takového testu je ADF test.

3.1.2 Augmented Dickey–Fuller test

Tento test vychází z Dickey-Fuller testu, ale navíc započítává i další zpoždění. Samotný Dickey-Fuller test existuje ve 3 verzích, kde vychází z nestacionárního AR(1) modelu:

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \ldots \tag{3.1}$$

Předpokládejme, že řada e_t má průměr roven nule, tedy:

$$E(e_t | y_{t-1}, y_{t-2}, \ldots, y_0) = 0, \tag{3.2}$$

kde y_0 je počáteční hodnota pozorování. Pokud je splněna podmínka (3.2), tak má časová řada jednotkový kořen pouze pokud $\rho = 1$. Pokud navíc $\alpha = 0$, tak se jedná o výše zmíněný proces náhodné procházky.

Vynecháme-li α, můžeme upravit rovnici (3.1) na tvar:

$$\begin{aligned} y_t &= \rho y_{t-1} + e_t \qquad / - y_{t-1} \\ \Delta y_t &= \rho y_{t-1} - y_{t-1} + e_t \\ \Delta y_t &= (\rho - 1) y_{t-1} + e_t \qquad / \theta = \rho - 1 \\ \Delta y_t &= \theta y_{t-1} + e_t \end{aligned} \tag{3.3}$$

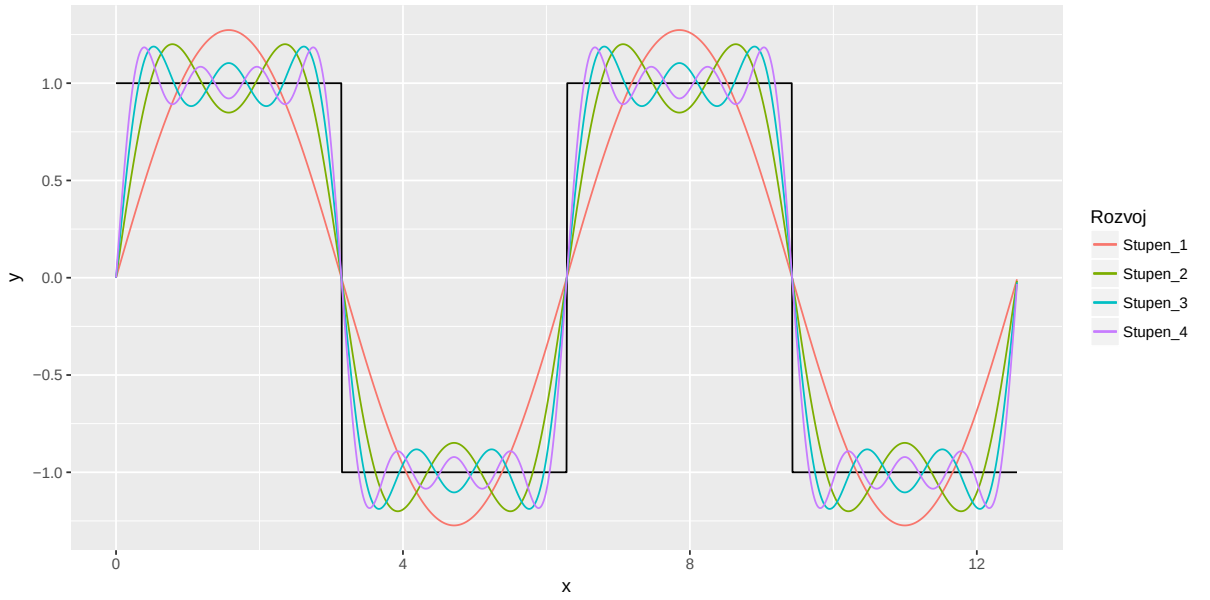
Poslední řádek rovnice (3.3) představuje první verzi Dickey-Fuller testu, který testuje jen samotný jednotkový kořen (proces náhodné procházky).

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t \tag{3.4}$$

Další verze testuje jednotkový kořen s driftem (posunem). Taková časová řada obsahuje lineární trend a právě tato verze DF testu dokáže rozpoznat nestacionární chování náhodných pozorování kolem trendu. Pokud bychom tedy řadu očistili o trend a následně použili první verzi DF testu, dostali bychom stejné výsledky. Lineární trend je v rovnici (3.4) zobrazen jako $\Delta y_t = \alpha$.
Poslední verze umožňuje pokročilejší práci s trendem a lze matematicky zapsat jako:

$$\Delta y_t = \alpha + \delta t + \theta y_{t-1} + e_t \tag{3.5}$$

Nezbytně důležité je také popsat hypotézy testu. Jak lze vidět již rovnice (3.3), tak časová řada obsahuje jednotkový kořen, pokud $\theta = 0$ (ekvivalent k původnímu AR(1) modelu by bylo $\rho = 1$). Proto se testuje nulová hypotéza H_0: $\theta = 0$ oproti alternativní H_1: $\theta < 0$. Přestože nemůžeme použít klasické kritické hodnoty t-rozdělení, můžeme použít t-test pro $\hat{\theta}$ v (3.3) (3.4) a (3.5). Kritické hodnoty byly vypočítány několika autory, ale



Obrázek 3.3: Trigonometrický rozvoj (obdélníkový průběh funkce)
zdroj: autor

původní práce náleží právě pánům Dickey a Fuller. Nulovou hypotézu H_0 zamítáme pokud $t_{\hat{\theta}} < c$, kde c je speciální kritická hodnota právě pro DF test, která se liší podle verze DF testu – tj. pro (3.3), (3.4) a (3.5).

Augmented Dickey-Fuller test je rozšířením DF testu, které počítá i se zpožděnou autoregresí (viz rovnice 3.6).

$$\Delta y_t = \alpha + \delta t + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + \ldots + \gamma_{p-1} \Delta y_{t-p+1} + e_t \tag{3.6}$$

Kritické hodnoty a pravidla o hypotézách zůstavají stejná jako u DF testu, ale musíme vzít v potaz fakt, že čím více zpoždění započítáme, tím méně počátečních pozorování budeme mít. Pokud započítáme příliš mnoho zpoždění, tak malý vzorek negativně ovlivní sílu testu. [16]

3.2 Fourierova transformace

Přestože byla Fourierova transformace vynalezena za účelem řešení rovnice stacionárního rozložení teploty, našla uplatnění v mnoha dalších oborech. Především v oblasti zpracování signálů, elektrotechnice, ale i ekonometrii. V této práci bude využita k aproximaci sezónnosti. Funguje podobně jako Taylorův polynom, který dokáže aproximovat složitou funkci kolem daného bodu pomocí mocninné řady. Fourierova řada využívá trigonometrických funkcí, které při vhodné kombinaci zadaných parametrů dokáží konvergovat k jakékoliv periodické funkci. Extrémním příkladem je obdélníkový či pilový průběh funkce, které jsou de facto hranaté. Právě trigonometrický rozvoj dokáže konvergovat i k nim. Na obrázku 3.3 představuje černá řada zmiňovaný obdélníkový průběh funkce a je vidno, že s roustoucím počtem řádů se k němu Fourierova řada blíží. [17]

Odvození Fourierovy transformace vychází z výše zmíněné Fourierovy řady periodické funkce. Základní myšlenka spočívá ve skládání anizochronních (tzn. majících rozdílnou

frekvenci) harmonických kmitů téhož směru s takovými frekvencemi, aby výsledná funkce mohla být periodická, tedy $T_1 = nT_n$, kde n je celé číslo. Předpis takovéto funkce bude:

$$f(t) = B_0 + \sum_{n=1}^{\infty} A_n \sin(\omega_1 nt) + \sum_{n=1}^{\infty} B_n \cos(\omega_1 nt), \tag{3.7}$$

kde A_n, B_n jsou koeficienty tvořící tzv. spektrum funkce. [18] Právě tyto koeficienty jsou poté odhadovány pomocí OLS (metody nejmenších čtverců), aby se přiblížily co nejvíce k sezónnosti v dané časové řadě. Tento rozvoj byl aplikován i k tvorbě obrázku 3.3, do jehož kódu je možné nahlédnout v příloze na straně 50.

3.3 Frisch-Waugh-Lovell věta

Základní zápis ekonometrického modelu by se dal maticově zapsat takto:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \tag{3.8}$$

kde $\mathbf{y}$ je vektor vysvětlované proměnné (resp. matice $n \times 1$, kde n je počet pozorování), $\mathbf{X}$ matice $n \times (k+1)$ vysvětlujících proměnných (k je počet regresorů), $\boldsymbol{\beta}$ je vektor parametrů a $\mathbf{u}$ představuje náhodnou složku.

Představme si nyní matici regresorů jako:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X_1} & \boldsymbol{X_2} \end{bmatrix}$$

$$\boldsymbol{X} = \left[\begin{array}{ccccc|ccc} 1 & x_{11} & x_{12} & \cdots & x_{1s} & x_{1(s+1)} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2s} & x_{2(s+1)} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{ns} & x_{n(s+1)} & \cdots & x_{nk} \end{array}\right] \tag{3.9}$$

Kde matice $\boldsymbol{X_1}$ představuje matici $n \times (s+1)$ s úrovňovou konstantou a s dummy proměnnými udávající sezónnost (např. $s = 4$, pokud bychom započítali dummy proměnné pro každé čtvrtletí). Matice $\boldsymbol{X_2}$ o rozměrech $n \times (k-s)$ obsahuje všechny zbylé regresory. Nyní již není problém provést regresní rozklad úvodní rovnice 3.8 a získat:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \; = \begin{bmatrix} \boldsymbol{X_1} & \boldsymbol{X_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{bmatrix} + \boldsymbol{u} \; = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \boldsymbol{u} \tag{3.10}$$

Aby bylo možné vysvětlit F-W-L větu, bude ještě nutné zavést 2 typy matic. První maticí bude matice $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ a druhou bude $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$. Matice $\boldsymbol{P}$ promítá vektor vysvětlovaných proměnných $\boldsymbol{y}$ do podprostoru span($\boldsymbol{X}$) jako vyrovnané hodnoty $\hat{\boldsymbol{y}}$. Podobně matice $\boldsymbol{M}$ promítá do do ortogonálního doplňku tohoto podprostoru rezidua $\hat{\boldsymbol{u}}$. Zaveďme ještě matice $\boldsymbol{P_1}$ a $\boldsymbol{M_1}$ s podprostorem span($\boldsymbol{X_1}$) a $\boldsymbol{P_2}$ a $\boldsymbol{M_2}$ s podprostorem span($\boldsymbol{X_2}$).

Nyní je již možné uvažovat následující 2 modely:

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \boldsymbol{u} \tag{3.11}$$

$$\boldsymbol{M_1 y} = \boldsymbol{M_1 X_2 \beta_2} + \boldsymbol{u} \tag{3.12}$$

V modelu 3.11 provádíme regresi $\boldsymbol{y}$ na sezónních i ostatních vysvětlujících proměnných. Model 3.12 provádí regresi už jen na reziduích z modelu $\boldsymbol{y} = \boldsymbol{X_1}\boldsymbol{\beta_1} + \boldsymbol{u}$, které lze zkráceně zapsat jako $\boldsymbol{M_1}\boldsymbol{y}$. Matice $\boldsymbol{M_1}$ totiž promítne rezidua, která jsou očištěna od sezónnosti (matice $\boldsymbol{X_1}$). Stejně tak je očištěna od sezónnosti i matice zbylých regresorů $\boldsymbol{X_2}$, jejíž rezidua se dají zapsat jako $\boldsymbol{M_1}\boldsymbol{X_2}$. U tohoto modelu (3.12) je vynechaná úrovňová konstanta, protože na obou stranách rovnice jsou rezidua modelů, která mají střední hodnotu rovnou nule.

Za zmínku ještě stojí napsat, že matice $\boldsymbol{P}$ i $\boldsymbol{M}$ jsou idempotentní. Tato vlastnost lze snadno zapsat matematicky jako $\boldsymbol{P^2} = \boldsymbol{P}$, stejně tak $\boldsymbol{M^2} = \boldsymbol{M}$. Je nutné si uvědomit zásadní rozdíl mezi idempotencí a impotencí! Idempotence je totiž mnohem horší vlastnost. Narozdíl od impotence, kde když se sejdou dva, tak zůstanou stále dva, tak idempotence je, pokud se sejdou dva a zůstane jen jeden. [19] Této vlastnosti bude v empirické analýze využito, když bude vysvětlovaná proměnná očištěna od sezónnosti a znovu vložena do modelu i s regresory určujícími sezónnost. Nicméně právě díky tomu, že se jedná o idempotentí operaci, tak se s očištěným regresorem nic nestane. V podstatě bude „očištěna“ již očištěná časová řada, což znamená, že se z ní již více nevysvětlí.

Frisch-Waugh-Lovell věta tedy říká, že v obou modelech 3.11 a 3.12 bude odhad vektoru koeficentů $\boldsymbol{\beta_2}$ stejný. Navíc rezidua ve zmíněných modelech budou také shodná. [20]

3.4 Carry-over efekt

Během analýzy časových řad o úspěšnosti mediální komunikace se musí vzít v potaz mnoho faktorů a je i možné, že její závěry nebudou spolehlivé právě díky vývoji situace na trhu (např. prodeje nemusí být ovlivněny jen investicemi do reklamy, ale i aktivitou konkurence). Jedním z faktorů, který nemůže být opominut, je projev investic do komunikace s časovým zpožděním (tzv. *carry-over efekt*). [21]

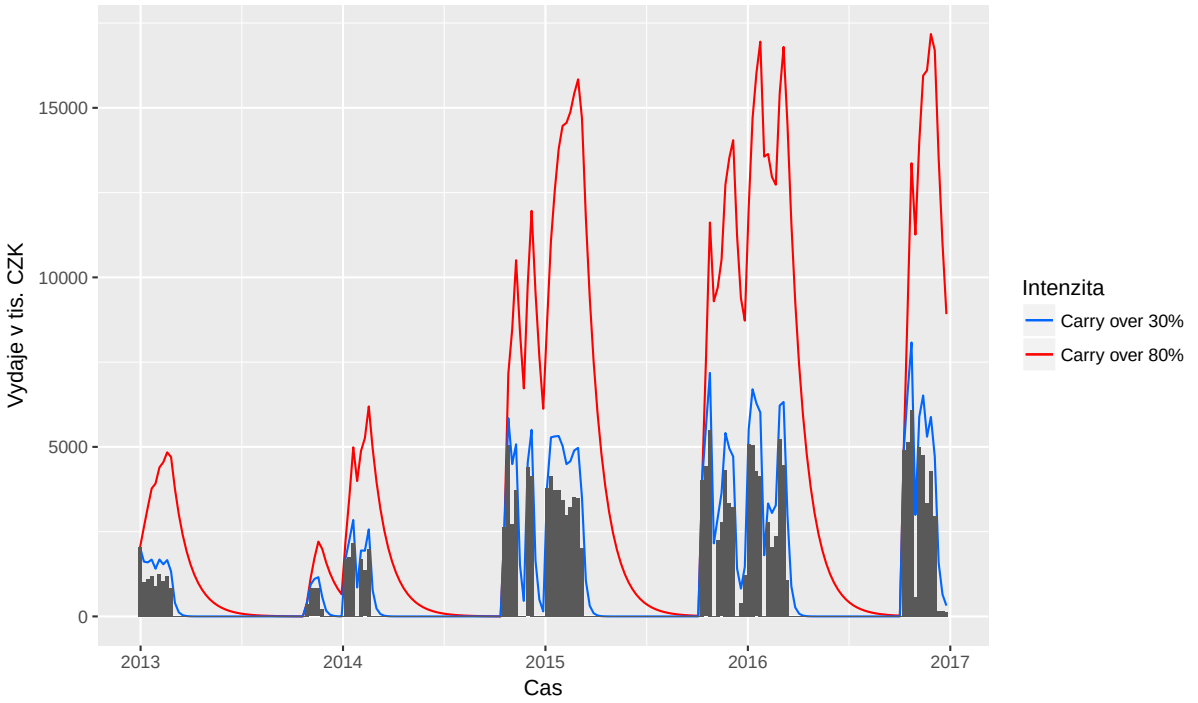
Mohou se vyskytnout dva typy tohoto jevu. První typ nastane, pokud vznikne interval mezi okamžikem, kdy jsou vynaloženy náklady na reklamu a časem, kdy jsou zaznamenány výsledné prodeje. Druhým typem je setrvačnost nových zákazníků, kteří byli získáni na základě současných výdajů na reklamu a opakovaně budou nakupovat v budoucích obdobích. [22]

V této práci bude carry-over efekt počítán kumulativně. Pokaždé, když se imprese zopakuje, tak se stane silnější dvěma způsoby. Zaprvé existuje tendence, aby povědomí o reklamě mizelo pomaleji než samotné invetsice. Zadruhé se zkombinuje vliv nových impresí se zbytky té předešlé. [23] Matematicky tento efekt lze zapsat jako:

$$E_t = I_t + cE_{t-1}, \tag{3.13}$$

kde E je intenzita efektu způsobeného reklamou, I je velikost investic v daném období a c je síla carry-over efektu. Hodnota c musí nabývat hodnot v intervalu $\langle 0; 1\rangle$ (např. $c = 0,84$ by znamenalo, že 84 % intenzity z předešlého období by se přeneslo i na aktuální období). Počáteční intenzita E_0 bude rovna jen velikosti počáteční investice I_0. Zde se nabízí prostor pro chybný výsledek, protože pokud by proběhla před sledovaným obdobím masivní kampaň, její dozvuk by v modelu nebyl započítán.

Ze vzorce 3.13 je patrné, že intenzita reklamy po skončení investic geometricky klesá a konverguje k nule, protože $c \in \langle 0; 1\rangle$. Tento jev je patrný i z obrázku 3.4, kde je možné



Obrázek 3.4: Carry-over efekt

zdroj: autor

porovnat důsledky síly carry-over efektu. Pokud je silný (např. 80 % – na obrázku 3.4 červená čára), tak intenzita kampaně bude výrazně silnější než při slabším CO efektu (např. 30 % – na obrázku 3.4 modrá čára).

Síla carry-over efektu také závisí na frekvenci dat (resp. délce jednoho období v časové řadě). Pokud bychom uvažovali měsíční frekvenci, síla carry-over efektu by byla menší než při týdenní frekvenci. Důvodem je logická úvaha, že zpožděný vliv reklamy bude například zhruba 3 týdny a pokud se bude pracovat s týdenními daty, tento efekt se projeví velmi silně. Ovšem u dat s měsíční frekvencí bude tento vliv pohlcen delším obdobím. Druhým faktorem, který ovlivňuje sílu carry-over efektu je i použitý mediatyp. Nelze například počítat s tím, že síla carry-over efektu bude u OOH (out of home) reklamy stejně velká jako u TV (ta je zpravidla nejsilnější).

Problémy způsobené carry-over efektem

- **Inzerční problémy** – Toto se týká jen málo dynamických mediatypů (např. tisk), protože pokud se zjistí problém s inzerátem, tak je již pozdě ho řešit. Narozdíl od internetové reklamy, kde je možné ihned změnit kreativu.
- **Problémy s trhem** – Změní-li se situace na trhu a je nutné na tuto změnu reagovat svou marketingovou strategií (např. optimálním komunikačním mixem), pak je nevyhnutelná ztráta způsobená carry-over efektem reklamy.
- **Analytické problémy** – Komplikované určení začátku a délky sledovaného období.

4 Empirická analýza

V této části bude analyzována časová řada zájmu podle Google Trends o hledaný výraz ACC, který zahrnuje nejen samotný pojem ACC, ale i pojmy s ním související (např. ACC Long, ACC pro děti atd.). Vysvětlujícími proměnnými (tzv. regresory) budou údaje o teplotě a mediální komunikaci ACC a jeho relevantní konkurence. Ta byla vybrána mezi léky proti kašli s největšími investicemi a srovnatelnými hodnotami zájmu v Google Trends. Jedná se tedy o značky: Mucosolvan, Robitussin a Stoptussin. Následně budou interpretovány výsledky a vypočten ukazatel ROI.

4.1 Dekompozice a očištění dat

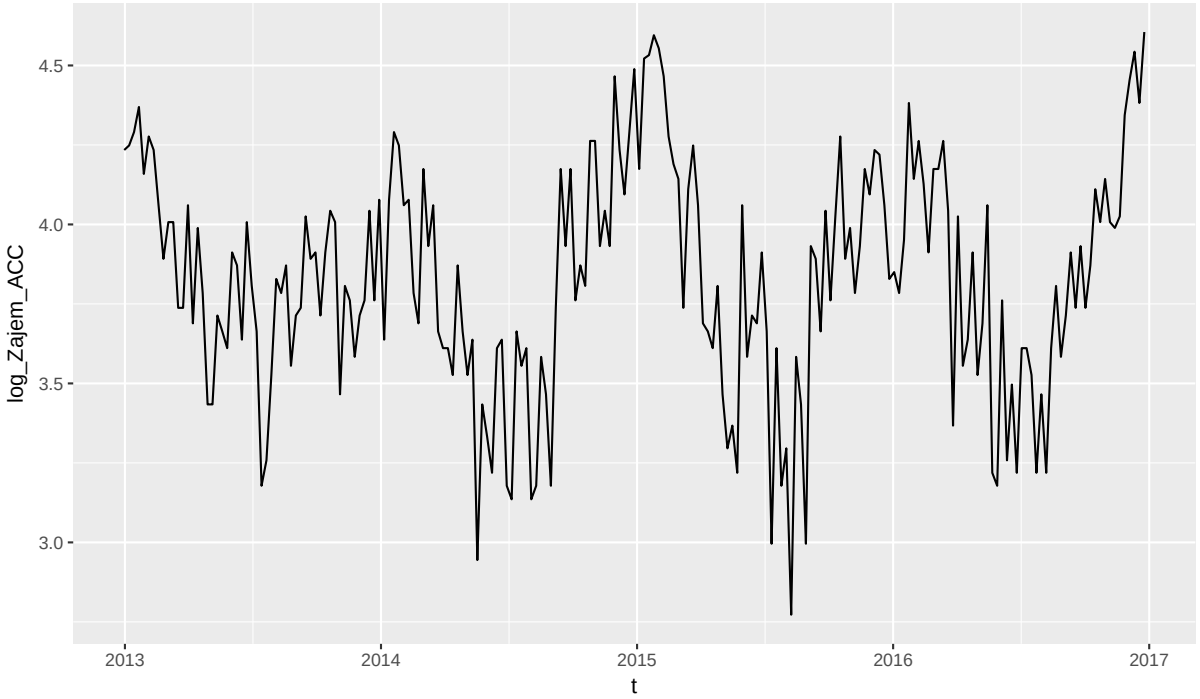
Prvním důležitým krokem bude dekompozice časové řady, identifikace případného trendu nebo sezónnosti a jejich očištění. Abychom toho mohli v Rku dosáhnout, je nutné upravit data na časovou řadu. Již zde se naskytl první problém – rok 2015 totiž měl 53 týdnů a zbylé roky jen 52. Řešení, které není z mého hlediska úplně korektní, je vymazat jeden týden v roce 2015, kde nikdo nekomunikoval, teploty jsou podobné s týdny před a po a zájem je také podobný. Takový případ v roce 2015 byl naštěstí 25. týden. Následně použít běžnou transformaci dat na časovou řadu s frekvencí 52 týdnů.

```
data52TS=data52
for (i in 1:(ncol(data))) {
  data52TS[,i]=ts(data52TS[,i], freq=52, start=2013)
}
```

Daleko elegantnější řešení je použít balíček *lubridate*. [24] Výhoda tohoto balíčku je fakt, že počítá frekvenci časové řady i s necelými čísly. Vzhledem k tomu, že použitá data jsou přesně po týdnech (tzn. začínají i končí celým týdnem), byla tato možnost jasnou volbou. Aby byl započítán i 53. týden, zvolila se frekvence 52,214 týdne za rok.

```
library(lubridate)
dataTS=data
roky = (decimal_date(ymd("2017-1-1"))-
          decimal_date(ymd("2012-12-31")))
for (i in 1:(ncol(data))) {
  dataTS[,i]=ts(dataTS[,i], freq=nrow(data)/roky,
                start=decimal_date(ymd("2012-12-31")))
}
```

Pro potřeby dalších výpočtů ještě bylo nutné zaznamenat začátek i konec sledovaného období. Zároveň tyto hodnoty posloužily jako kontrola správnosti manipulace s daty.



Obrázek 4.1: Průběh zájmu o ACC v letech 2013–2016

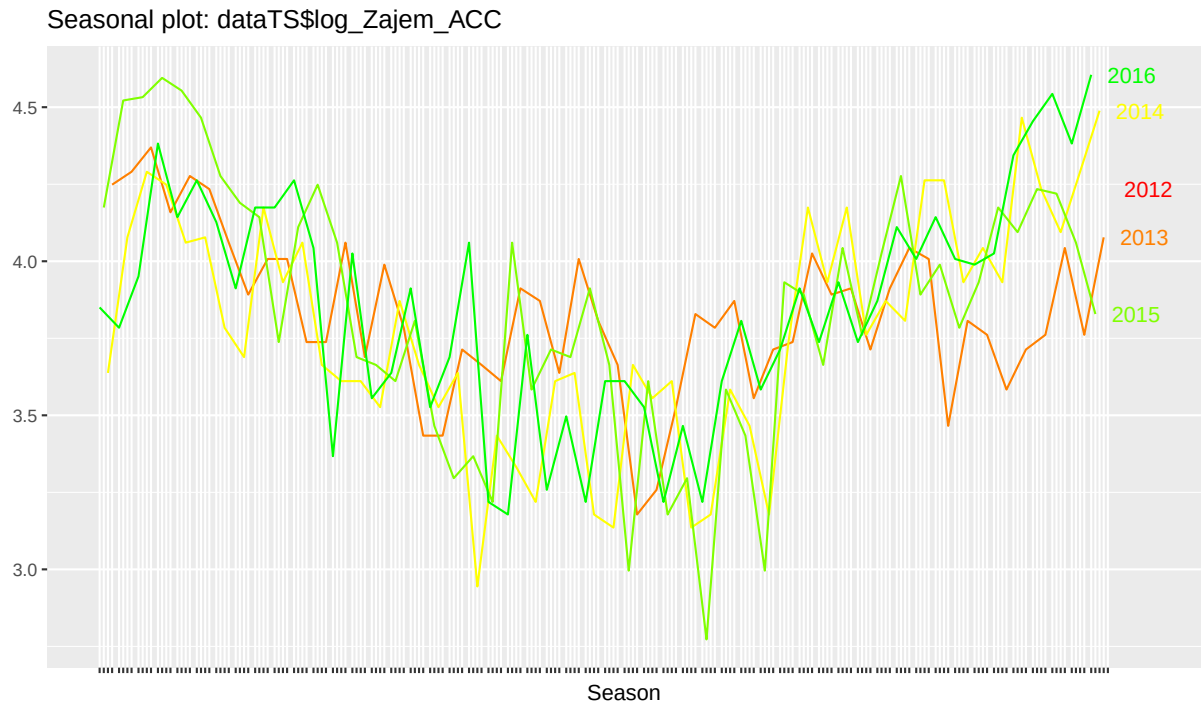
zdroj: autor

```
t=time(dataTS$Zajem_ACC)
start=start(t)
end=end(t)
date_decimal(start)
date_decimal(end+1/frequency(t))  # Ke koncovemu
# datu musim pricist "jedno" pozorovani, protoze
# je to od end(t) cely nasledujici tyden
```

Nyní se již můžeme podívat na průběh časové řady na obrázku 4.1. Hned na první pohled je zde patrná silná sezónnost, kterou potvrzuje i graf 4.2. Funkce na tvorbu tohoto grafu je součástí balíčku *forecast*. [25]

```
library(ggplot2)
graf1 <- ggplot(data = dataTS, aes(x = t,y = log_Zajem_ACC)) +
  geom_line()
pdf("graf1.pdf",width = 8.5, height = 5) #export
graf1
dev.off() #export

library(forecast)
pdf("grafS.pdf",width = 8.5, height = 5) #export
ggseasonplot(dataTS$log_Zajem_ACC, col=rainbow(12),
             year.labels=TRUE) # Dukaz existence sezonnosti
dev.off() #export
```

Pokud bychom využili jednu z nejjednodušších možností dekompozice časové řady v Rku, dostali bychom podle mě špatné výsledky. Lépe řečeno špatné nejsou, ale ne-



Obrázek 4.2: Průběh zájmu o ACC v letech 2013–2016

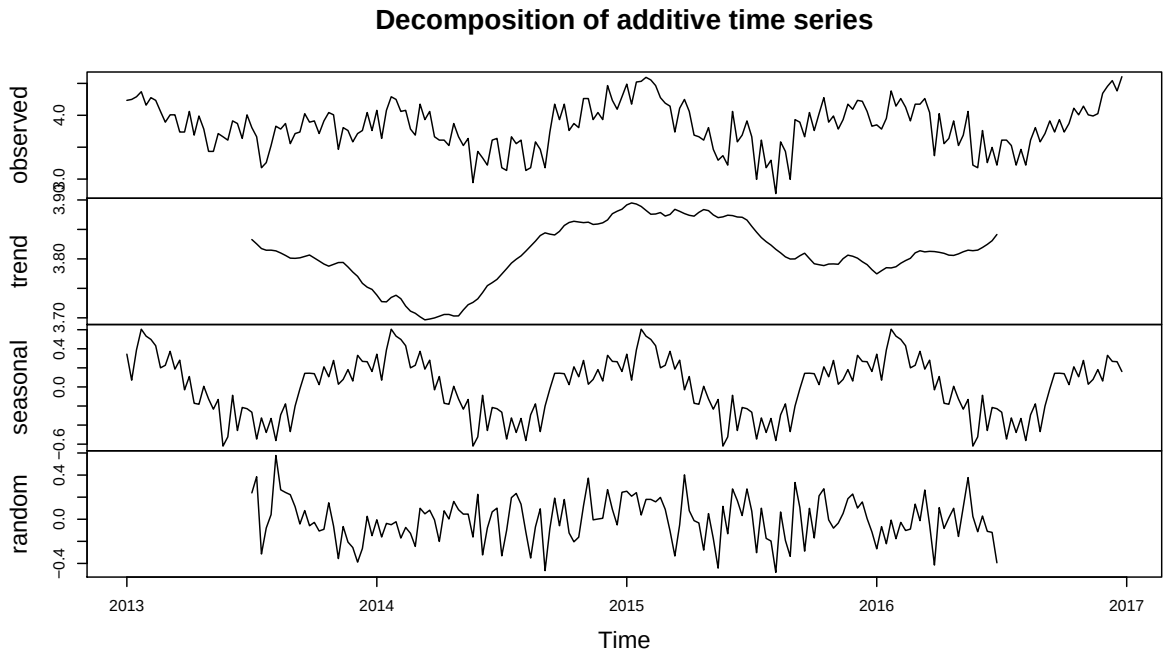
zdroj: autor

odpovídají realitě a ekonomická interpretace by nebyla možná. Jedná se o dekompozici využívající MA (klouzavý průměr) pro určení trendu a následně aditivní metodu výpočtu sezónních koeficientů. Rozklad zlogaritmované časové řady je vidět na obrázku 4.3. Ekonomicky zdůvodnit například propad mezi 14. a 17. týdnem je podle mého názoru nemožné. Jaký asi může mít vliv zrovna 15. a 16. týden na zájem o ACC? V případě delší časové řady by tyto výkyvy byly zřejmě mnohem menší.

```
pdf("grafDE.pdf",width = 8.5, height = 5) #export
plot(decompose(data52TS$log_Zajem_ACC))
dev.off() #export
names(decompose(data52TS$log_Zajem_ACC))
```

Lépe interpretovatelná je „vyhlazená“ sezónnost. Té je možné dosáhnout pomocí Fourierovy transformace, která byla popsána v teoretické části v sekci 3.2. Budeme uvažovat 4 modely, kde bude použit pokaždý vyšší stupeň Fourierovy transformace. V tabulce 5.1 v příloze na straně 53 je možné vidět výstupy modelů zahrnujících trigonometrický rozvoj. K tvorbě tabulek bylo využito opět dalšího balíčku v Rku *stargazer*. [26]

```
m = lm(dataTS$log_Zajem_ACC ~ 1 + cos(2*pi*(t-start)) +
       sin(2*pi*(t-start)) ) # prvni stupen

m2 = lm(dataTS$log_Zajem_ACC ~ 1 + cos(2*pi*(t-start)) +
        sin(2*pi*(t-start)) + cos(4*pi*(t-start)) +
        sin(4*pi*(t-start))) # druhy stupen

m3 = lm(dataTS$log_Zajem_ACC ~ 1 + cos(2*pi*(t-start)) +
        sin(2*pi*(t-start)) + cos(4*pi*(t-start)) +
```



Obrázek 4.3: Dekompozice pomocí funkce *decompose* v Rku

zdroj: autor

```
          sin(4*pi*(t-start)) + cos(6*pi*(t-start)) +
          sin(6*pi*(t-start))) # treti stupen

m4 = lm(dataTS$log_Zajem_ACC ~ 1 + cos(2*pi*(t-start)) +
          sin(2*pi*(t-start)) + cos(4*pi*(t-start)) +
          sin(4*pi*(t-start)) + cos(6*pi*(t-start)) +
          sin(6*pi*(t-start)) + cos(8*pi*(t-start)) +
          sin(8*pi*(t-start))) # ctvrty stupen

library(stargazer)
stargazer(m,m2,m3,m4,omit.stat=c("LL","ser","f"))
```

Pro porovnání modelů je vhodné využít analýzu rozptylu (ANOVA) součtu čtverců reziduí z regrese. Nulová hypotéza H_0 říká, že oba modely jsou rovnocenné. Z tabulky 4.1 je patrné, že mezi modelem s jedním stupněm rozvoje a modelem s 2 stupni rozvoje není na 5% hladině významnosti rozdíl. Tento výsledek byl očekávatelný, protože při porovnání upraveného koeficientu determinance těchto modelů má první model vyšší hodnotu R^2_{adj}.

Tabulka 4.1: ANOVA modelů s 1 a 2 stupni rozvoje

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
m	206	12.570				
m2	204	12.528	2	0.042	0.342	0.711

Podobným způsobem je porovnán model s jedním stupněm trigonometrického rozvoje s modelem se 3 stupni. Výsledky je možné vidět v tabulce 4.2 a je z nich patrné, že 3 stupně rozvoje jsou lepším odhadem sezónnosti než model s jedním stupněm. Na 5% hladině významnosti totiž zamítám H_0 o shodě obou modelů.

Tabulka 4.2: ANOVA modelů s 1 a 3 stupni rozvoje

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
m	206	12.570				
m3	202	11.893	4	0.677	2.875	0.024**

Poznámky: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

A konečně v tabulce 4.3 porovnám poslední model se 4 stupni trigonometrického rozvoje s modelem se 3 stupni. Poměrně těsně nezamítám nulovou hypotézu o rovnocennosti modelů, a proto volím odhad sezónnosti se 3 stupni rozvoje, jelikož 4. stupeň nevycházel v regresní analýze signifikantní (model č. 4 ve výstupu z regresní analýzy v souhrné tabulce všech modelů 5.1 na straně 53).

Tabulka 4.3: ANOVA modelů s 3 a 4 stupni rozvoje

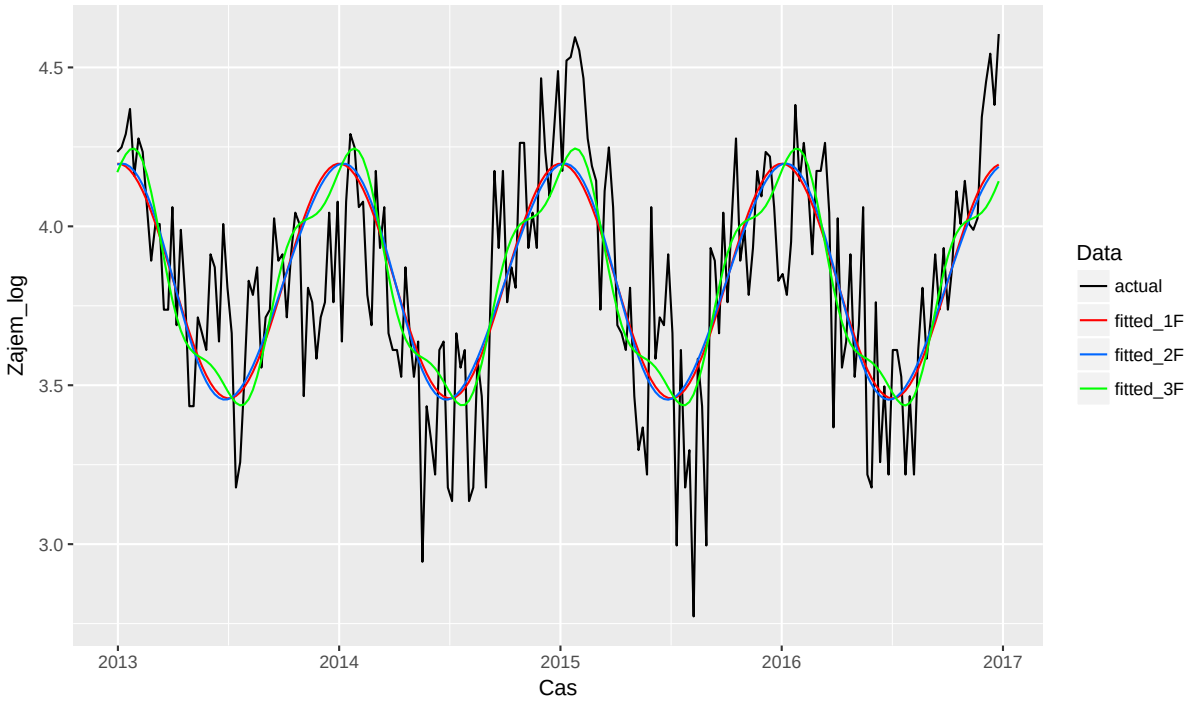
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
m3	202	11.893				
m4	200	11.599	2	0.294	2.535	0.082

V Rku toto rozhodování i s exportem do LaTeXu vypadalo následovně:

```
summary(m)
summary(m2)
stargazer(anova(m,m2), summary = FALSE) # 2. stupen zamitam

summary(m3)
stargazer(anova(m,m3), summary = FALSE) # 1. stupen zamitam

summary(m4)
stargazer(anova(m3,m4), summary = FALSE) # 4. stupen zamitam
```

Pro jasnější porozumění odhadu sezónnosti je lepší uvedené modely vizualizovat. Na obrázku 4.4 je zřetelný rozdíl mezi 3. stupněm trigonometrického rozvoje a zbylými 2 modely. Odpovídá to i nepatrné „špičce“ v grafu 4.3 hned v prvních týdnech každého roku. Ekonomicky by se dala interpretovat zřejmě jako hlavní sezóna všech nemocí (především kašle v tomto případě).

Po volbě odhadu sezónnosti již mohu získat očištěnou časovou řadu. Jsou to rezidua ze zvoleného modelu (m3), která si uložím do samostatné proměnné v Rku. Očištěný průběh časové řady je možné vidět na obrázku 4.5.

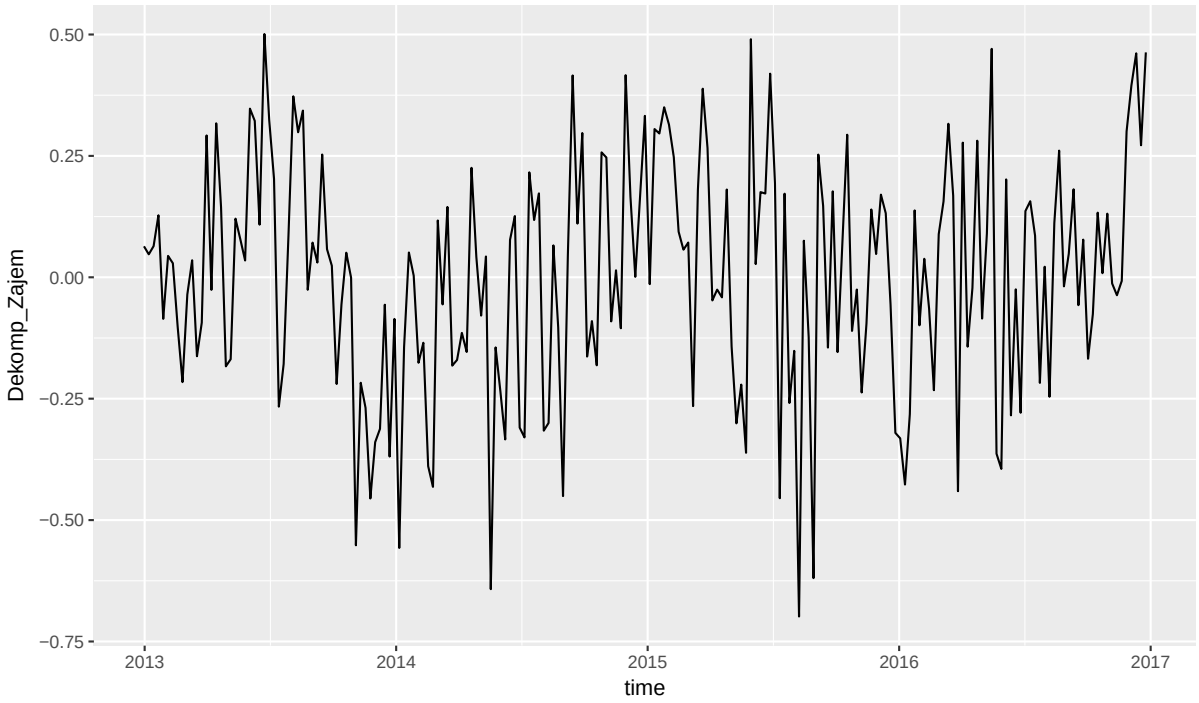


Obrázek 4.4: Dekompozice pomocí Fourierovy transformace

zdroj: autor

Další část kódu, která vytvořila obrázek 4.4, zapsala rezidua do nové proměnné a vizualizovala jejich průběh na obrázku 4.5:

```
l = length(dataTS$log_Zajem_ACC)
graf2 = data.frame(Cas = c(t,t,t,t),
                   Zajem_log = c(dataTS$log_Zajem_ACC,
                                 m$fitted.values,m2$fitted.values,
                                 m3$fitted.values),
                   Data = c(rep("actual",times = l),
                            rep("fitted_1F",times = l),
                            rep("fitted_2F",times = l),
                            rep("fitted_3F",times = l)))
graf2 <- ggplot(data = graf2, aes(x = Cas, y = Zajem_log ,
                                  group = Data)) +
  geom_line(aes(color=Data))
pdf("graf2.pdf",width = 8.5, height = 5) #export
graf2+scale_color_manual(values=c("#000000", "#ff0000",
                                  "#0066ff","#00ff00","#ff00ff"))
dev.off() #export

DEKOMP_DATA = m3$residuals
DEKOMP_DATA=ts(DEKOMP_DATA, freq=nrow(data)/roky,
               start=decimal_date(ymd("2012-12-31")))

graf3_data = data.frame(time = t,
                        Dekomp_Zajem = DEKOMP_DATA)
```



Obrázek 4.5: Průběh očištěné časové řady od sezónnosti

zdroj: autor

```
graf3 <- ggplot(data = graf3_data, aes(x = time, y = Dekomp_Zajem)) +
  geom_line()
pdf("graf3.pdf",width = 8.5, height = 5) #export
graf3      # data ocistena o~sezonnost pomoci Fourierovy transformace
dev.off() #export
```

Jedním z klíčových momentů při práci s časovou řadou je zjištění stacionarity. Ta již byla popsána v sekci 3.1. Pokud odhadneme model se zpožděním řádu 1 vysvětlované proměnné (v kódu $DEKOMP_DATA$), dostaneme ρ z rovnice 3.1. Víme, že pokud se ρ blíží k 1, mohlo by se jednat o nestacionární řadu. Z tabulky 4.4 víme, že $\rho = 0.263$, což znamená slabou závislost.

Tabulka 4.4: Regresní model AR(1)

	Závislá proměnná:
	DEKOMP_DATA
$DEKOMP_DATA_{t-1}$	0.263^{***}
	(0.068)
Počet pozorování	208
R^2	0.068
Poznámky:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Pro jasné rozhodnutí nám pomůže ADF test – taktéž popsaný v teoretické části v podsekci 3.1.2 na straně 16. Nulová hypotéza H_0 říká, že řada je nestacionární, naproti tomu alternativní hypotéza H_1 znamená stacionaritu časové řady. Výstup ADF testu zaznamenaný v tabulce 4.5 svědčí o stacionaritě časové řady (p-hodnota $< 0,05$), což znamená, že bez dalších větších komplikací je možné pokračovat v analýze.

Tabulka 4.5: Augmented Dickey-Fuller Test

Dickey-Fuller	Lag order	p-value
−3.9003	5	0.01514**
Poznámky:		*p<0.1; **p<0.05; ***p<0.01

Pro výpočet modelu se zpožděním řádu 1 byl použit balíček *dynlm* [27], který umí provést regresi časové řady i se zpožděnými hodnotami. Dále jsem využil balíčku *tseries*, který obsahuje funkci schopnou vypočítat ADF test. [28]

```
library(dynlm)
dfm = dynlm(DEKOMP_DATA ~ 0 + L(DEKOMP_DATA,1))
summary(dfm)  # Rho vychazi 0.26 - uplne v~poradku
stargazer(dfm,omit.stat=c("LL","ser","f"))

library(tseries)
adf.test(DEKOMP_DATA) # Zamitam H0 -> stacionarni
```

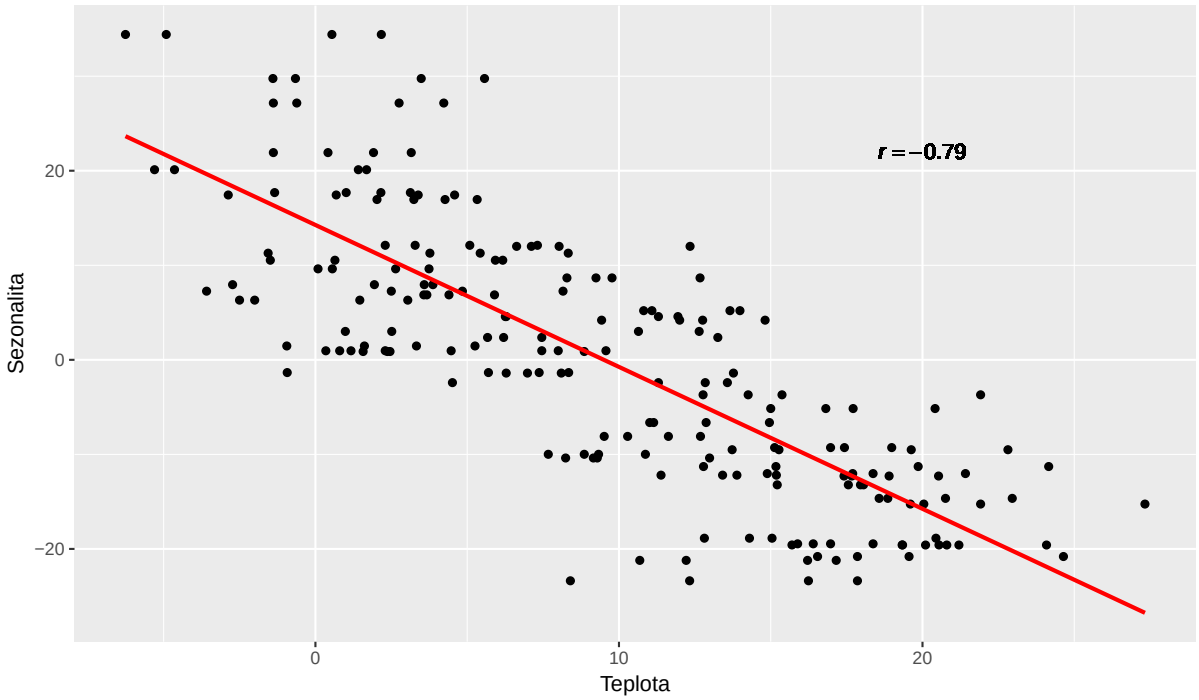
4.1.1 Počasí a jeho vliv

Vzhledem k tomu, že ACC je nepochybně sezónní produkt, který závisí na nemocnosti lidí, stojí za úvahu vzít v potaz počasí. Nízké teploty v zimě přinášejí sezóny viróz a různých prochladnutí, u kterých je jedním z příznaků kašel. To znanemená i větší zájem o ACC. Důkazem je i fakt, že sezónnost (pro názornost a jednoduchost bude brána pouze sezónnost určená funkcí *decompose* – v tabulce 4.6 $Zajem_season$) je silně negativně zkorelovaná s teplotou (viz rovnice 4.1 a obrázek 4.6). Na tomto jednoduchém příkladě je také možné ukázat vztah mezi korelačním koeficientem a koeficientem determinance (z tabulky 4.6) regrese s pouhou jednou vysvětlující proměnnou. Platí totiž vztah z rovnice 4.2, který je i numericky dokázán.

$$\mathrm{corr}(Zajem_season, Teplota) = -0.7921616 \tag{4.1}$$

$$\begin{aligned} corr^2 &= R^2 \\ (-0.7921616)^2 &= 0.628 = 0.628 \end{aligned} \tag{4.2}$$

Jednou z možností, jak by bylo možné předpovídat vliv počasí na zájem (a tím i poptávku) o ACC, by bylo započítat zpožděné hodnoty teploty očištěné o sezónnost. [29] Vyplývá to z logiky, že lidé nejsou nemocní hned, když jsou nízké teploty, ale až s nějakým zpožděním (např. 1–2 týdny by mohla být inkubační doba menší virové epidemie). Proto

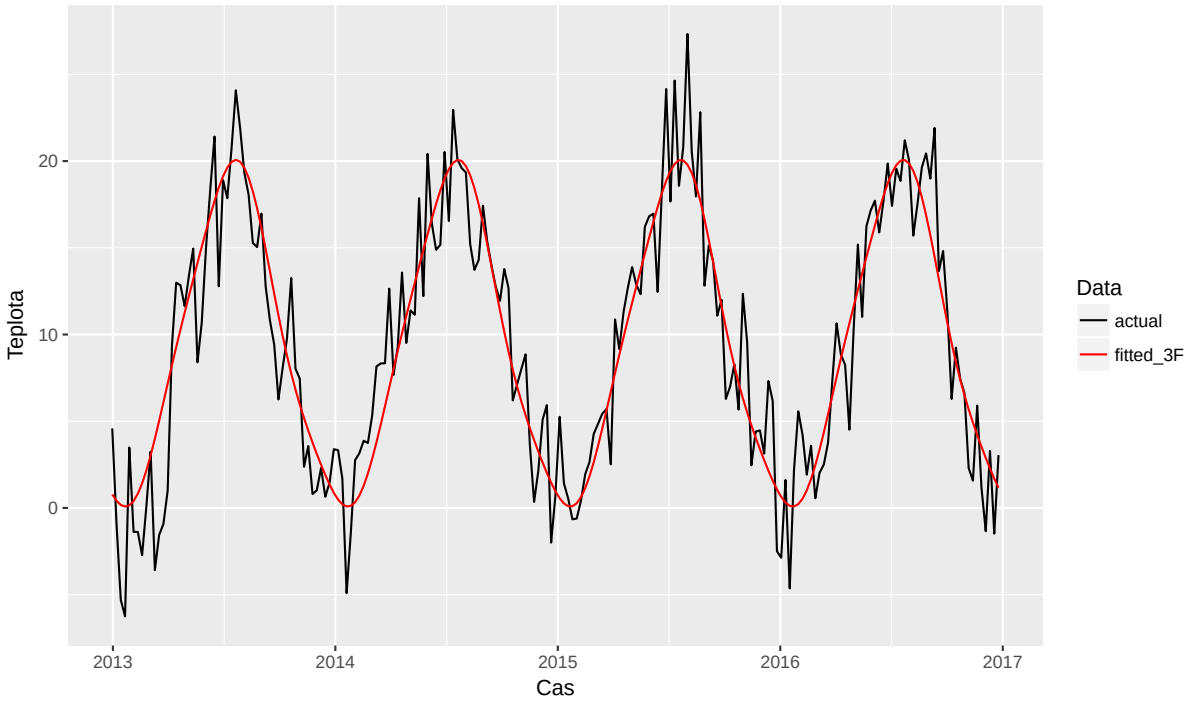


Obrázek 4.6: Korelace sezónní složky a teploty

zdroj: autor

Tabulka 4.6: Regresní analýza sezónní složky a teploty

	Závislá proměnná:
	Zajem_season
Teplota	-1.501^{***}
	(0.081)
Konstanta	14.262^{***}
	(0.971)
Počet pozorování	208
R^2	0.628
Upravené R^2	0.626
Poznámky:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$



Obrázek 4.7: Očištění teploty od sezónní složky

zdroj: autor

je nutné nejprve očistit řadu hodnot teploty o sezónnost a následně ji vložit do celkového regresního modelu se zpožděním. Jak asi všichni ví, teplota má silnou sezónní složku, a proto se modelem s trigonometrickým rozvojem podařilo vysvětlit značnou část její variability (výstupy z regresní analýzy jsou v tabulce 5.2 v příloze na straně 54). Tento závěr však lze konstatovat už jen při letmém pohledu na graf 4.7, kde je patrné, že Fourierova transformace je vhodnou aproximací sezónnosti.

Jak by bylo možné na teplotu reagovat?

Poptávka po sezónních produktech jako je např. vytápění, chladicí zařízení, cestovní balíčky nebo právě léky je citlivá na počasí. Právě to může negativně ovlivnit využití těchto položek a tím snížit poptávku. Tento fakt, spojený se skutečností, že se prodejci musí rozhodovat o úrovni zásob a o prodejní ceně ještě před pravidelnou sezónou, staví ziskovost prodejců do rizika.

Některé z těchto rizik však mohou prodejci přenést na zákazníky tím, že nabídnou propagační akce, které vyvolají počáteční prodej. Jednou z takovýchto akcí může být slevový program reagující na počasí, který poskytne slevy na předčasné nákupy, pokud je stav počasí nepříznivý. [30]

R kód k celé sekci Počasí a jeho vliv je zde:

```
# Souvislost s~teplotou
Zajem_season=decompose(data52TS$Zajem_ACC)$season
# Korelacni koeficient
x=cor(Zajem_season,data52TS$Teplota,use = "pairwise.complete.obs")
lm = lm(Zajem_season ~ data52TS$Teplota)
```

```
stargazer(lm,omit.stat=c("LL","ser","f"))
summary(lm)$r.squared
x^2 # = summary(lm)$r.squared

#Scatter plot
x = round(x, digits = 2)
x = paste("italic(r) == ", x)
graf4 = data.frame(Sezonnost = as.numeric(Zajem_season),
                   Teplota = as.numeric(data52TS$Teplota),
                   label = x)
graf4 = ggplot(data = graf4, aes(x = Teplota, y = Sezonnost)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  geom_text(data = graf4, aes(x = 20, y = 22, label = label),
                          parse = TRUE)
pdf("graf4.pdf",width = 8.5, height = 5) #export
graf4
dev.off() #export

# Vytvoreni odychlek pocasi od cos sezonnosti
m_pocasi_dekomp = lm(dataTS$Teplota ~ cos(2*pi*(t-start)) +
                     sin(2*pi*(t-start)) + cos(4*pi*(t-start)) +
                     sin(4*pi*(t-start)) + cos(6*pi*(t-start)) +
                     sin(6*pi*(t-start)) )
summary(m_pocasi_dekomp)
stargazer(m_pocasi_dekomp,omit.stat=c("LL","ser","f"))
odchylky_pocasi = m_pocasi_dekomp$residuals

l = length(dataTS$log_Zajem_ACC)
graf5 = data.frame(Cas = c(t,t),
                   Teplota = c(dataTS$Teplota,
                               m_pocasi_dekomp$fitted.values),
                   Data = c(rep("actual",times = l),
                            rep("fitted_3F",times = l)))
graf5 <- ggplot(data = graf5, aes(x = Cas, y = Teplota ,
                 group = Data)) +
  geom_line(aes(color=Data))
pdf("graf5.pdf",width = 8.5, height = 5) #export
graf5+scale_color_manual(values=c("#000000", "#ff0000"))
dev.off() #export
```

4.2 Započítání carry-over efektu

Abychom mohli vybrat model, který vysvětlí nejvíce variability zájmu o ACC, bude potřeba vypočítat jednotlivé carry-over efekty pro každý mediatyp všech zadavatelů. Následně odhadnout modely, do kterých budou postupně vstupovat všechny kombinace sil carry-over efektů.
Prvním krokem je vybrat z datového souboru pouze časové řady s mediální komunikací. Následně někam zapsat hodnoty s carry-over efektem. Za tímto účelem jsem se rozhodl využít několik datových struktur v Rku – vektor, data frame a list. Jednotlivé síly carry-over efektů jsem zapisoval do vektorů, které jsou uložené v data framu. Pro jeden mediatyp jednoho zadavatele tedy existuje jeden data frame obsahující 100 vektorů carry-over efektů (síla 0–99 %). Takto tedy vznikne dohromady 8 data framů (pro každou jedinečnou kombinaci zadavatele a mediatyp jeden). Pro přehlednost a možnost dalších algoritmických výpočtů v Rku je nutné tyto data framy zastřešit. K tomu slouží list, do kterého je možné nahrát různé druhy objektů včetně data framu. Výsledkem této části tedy bude jeden list obsahující napočítané všechny síly carry-over efektů pro všechny zadavatele a jejich jednotlivé využité mediatypy.

Ještě před samotným výpočtem je vhodné zkrátit vysvětlující proměnné o započítané zpoždění. Důvod manuální úpravy zpoždění bude vysvětlen v části Výběr nejlepšího modelu na straně 33.

```
names(data)
komunikace = dataTS[-c(1:4,13)] # Vyber jen komunikace
lag = 1
odchylky_pocasi_L1=odchylky_pocasi[-length(odchylky_pocasi)]
komunikace = komunikace[-lag,]
odchylky_pocasi=odchylky_pocasi[-lag]

komunikace_CO = list()
for(k in 1:ncol(komunikace)){
  komunikace_CO[[k]] = data.frame(komunikace[k]) # Z~k-teho dataframu

  for (i in seq(0.01,0.99,by = 0.01)) {
    komunikace_CO[[k]][,paste("CarryOver_",i,sep = "")] = 0
  }

  for(j in 1:(length(komunikace_CO[[k]])-1)) {    # j = pocet CO efektu
    komunikace_CO[[k]][1,j+1]=komunikace_CO[[k]][1,1]
     # do kazdeho dataframu zanesu prvni hodnotu z~vychozi komunikace
    for (i in 1:(nrow(komunikace_CO[[k]])-1)) {   # i = pocet pozorovani
      komunikace_CO[[k]][i+1,j+1]=(komunikace_CO[[k]][i+1,1]+
                                   komunikace_CO[[k]][i,j+1]*(j/100))
    }
  }
}

colnames(komunikace_CO[[1]][32]) # Zjistim Carryover
colnames(komunikace_CO[[7]][1]) # Zjistim nazev signifikantni promenne
```

4.3 Výběr nejlepšího modelu

Nejvhodnější model bude vybrán na základě koeficientu determinance R^2. Upravený koeficient determinance nebude potřeba, jelikož do modelu vstupuje stejný počet regresorů (liší se pouze sílou carry-over efektu). Výpočet je velmi rozsáhlý díky nutnosti odhadnout regresní model pro každý mediatyp všech zadavatelů. Pokud bychom spustili celý výpočet bez průběžných omezení, tak by vzniklo 256 triliónů modelů (256×10^{12}). K výsledkům se tedy musí člověk dopracovat ručním postupným omezováním jednotlivých regresorů, až dojde k řešení, které je optimální. Tento postup zajisté není nejvhodnější a naskýtá se zde obrovský prostor pro zjednodušení výpočetní náročnosti volbou vhodného algoritmu.

Aby bylo možné vytvářet modely v Rku pomocí for cyklu je nutné upravit k tomu i vysvětlující proměnné. Ke každé se musí přidat identifikátor (v mém případě písmeno), pomocí kterého se bude ve for cyklu měnit index data framu uloženého v listu. Dalším omezením je, že z neznámého důvodu nelze do for cyklu zakomponovat funkci *dynlm*, která dokáže sama vytvořit zpožděné proměnné. Řešením je vytvořit si tyto proměnné ručně a zbylé proměnné omezit, aby byl dodržen stejný počet pozorování. Když jsou všechny proměnné připravené, tak může být vytvořen vzorec, který bude vstupovat do funkce *lm* vnořené do for cyklu.

Výsledná rovnice regresního modelu bude vypadat takto:

$$\begin{aligned} y_{dekomp} =& \beta_0 + \boldsymbol{\beta_{komunikace}} \boldsymbol{x_{komunikace}} + \sum_{n=1}^{3} \beta_{sin_n} \sin(2\pi n(t - start)) + \\ & \sum_{n=1}^{3} \beta_{cos_n} \cos(2\pi n(t - start)) + \beta_t teplota_{dekomp} + \beta_{t-1} teplota_{t-1_{dekomp}} + u, \end{aligned} \tag{4.3}$$

kde β_0 je úrovňová konstanta, $\boldsymbol{\beta_{komunikace}}$ je vektor parametrů týkající se komunikace, $\boldsymbol{x_{komunikace}}$ je vektor (v kódu list *komunikace_CO*) jedinečných kombinací mediatypů a zadavatelů (např. ACC TV, Stoptussin Tisk, apod.), β_{sin_n} je n parametrů týkající se sinusu v n-tém rozvoji Fourierovy transformace, stejně tak β_{cos_n} pro cosinus v n-tém rozvoji, β_t spolu s β_{t-1} jsou parametry pro teplotu očištěnou od sezónnosti a její zpoždění úrovně 1 (v kódu pod proměnnou *odchylky_pocasi* a *odchylky_pocasi_L1*) a y_{dekomp} je vysvětlovaná proměnná očištěná od sezónnosti (v kódu jako *DEKOMP_DATA*).

```
DEKOMP_DATA = DEKOMP_DATA[-1]
t=t[-1]
promenne = paste("komunikace_CO[[",1:length(komunikace_CO),"]]")
promenne2=promenne
for(i in 1:length(komunikace_CO)){
  promenne2[i] = paste(promenne[i],"[[",LETTERS[i],"]]")
}

vzorec = as.formula(paste("DEKOMP_DATA~",
                          paste(promenne2, collapse = "+"),
                          "+ cos(2*pi*(t-start)) + sin(2*pi*(t-start))+
                          cos(4*pi*(t-start)) + sin(4*pi*(t-start))+
                          cos(6*pi*(t-start)) + sin(6*pi*(t-start))+
                          odchylky_pocasi + odchylky_pocasi_L1"))
```

Nyní již k tvorbě samotného for cyklu. Je nezbytné ručně zapsat jednotlivé for cykly, což shledávám další velkou nevýhodou (mimo výpočetní náročnost) použití for cyklu na tento problém. Pro každou jedinečnou kombinaci mediatypu a zadavatele tedy musí existovat jeden for cyklus. Díky tomu je ale získána možnost nastavení hranic carry-over efektů a tím i jejich postupné omezování. Jak bylo zmíněno v teoretické části Carry-over efekt na straně 19, u některých mediatypů je síla carry-over efektu slabší a u některých výrazně silnější. Vzhledem k tomu, že se jedná o týdenní data, bude u TV počítáno s carry-over efektem až 99 %, zatímco u ostatních mediatypů to bude maximálně 40 %. Tyto hodnoty jsou odhadnuté z praktické zkušenosti získané z podobných výzkumů.

Abychom ušetřili výkon a paměť počítače, tak se z výsledného modelu bude zapisovat jen samotné R^2, podle kterého se následně určí nejvhodnější model a indexy všech for cyklů, které určí kombinaci sil carry-over efektů pro daný model.

```
names(komunikace) # Zjistit posloupnost zadavatelu a mediatypu

modely=c()
komb=list()
i=1
for(A in 1:2){ #nastaveni CO pro ACC_OOHTV
  for(B in 38:41){ #nastaveni CO pro ACC_Outdoor
    for(C in 90:92){ #nastaveni CO pro ACC_TV
      for(D in 97:99){ #nastaveni CO pro MUCOSOLVAN_TV
        for(E in 38:41){ #nastaveni CO pro Robitussin_Radio
          for(F in 96:98){ #nastaveni CO pro Robitussin_TV
            for(G in 1:2){ #nastaveni CO pro Stoptussin_Tisk
              for(H in 94:96){ #nastaveni CO pro Stoptussin_TV
                modely[i] = summary(lm(vzorec))$r.squared
                komb[[i]]=c(A,B,C,D,E,F,G,H)
                print(c(A,B,C,D,E,F,G,H))
                i=i+1
              }
            }
          }
        }
      }
    }
  }
} # omezeno na temer finalni podobu, aby se to rychleji pocitalo
```

Nejvhodnější model nalezneme jako maximum ze všech R^2 zapsaných ve vektoru *modely*. Následně bude možné zjistit i kombinaci sil jednotlivých carry-over efektů, které jsou zapsány v listu *komb* pod stejným indexem jako maximální hodnota R^2. Jednotlivé síly carry-over efektu se ještě musí zmenšit o jednotku, protože pod indexem 1 se skrývá 0% carry-over efekt (tedy čisté ceníkové investice). Přehled sil carry-over efektu je možné najít v tabulce 4.7. Když už jsou zjištěné jednotlivé efekty, tak můžeme zapsat intenzity jednotlivých komunikačních kanálů (mediatypů) do proměnné, která následně bude vstupovat do výsledného (téměř finálního) modelu. Pro snazší interpretaci a možnost číst ve výsledcích ještě změníme jednotky u intenzity komunikace. Z tabulky 4.8 lze vidět, že se započítaným carry-over efektem se jednotky dostanou až do milionů Kč. Vydělením 10^6

všech hodnot matice intenzit přeškálujeme jednotky a získáme údaje v mil. Kč.

Tabulka 4.7: Nejleší zvolené carry-over efekty

Zadavatel a mediatyp	Carry-over efekt
ACC OOHTV	0
ACC Outdoor	40
ACC TV	90
MUCOSOLVAN TV	97
Robitussin Rádio	40
Robitussin TV	96
STOPTUSSIN Tisk	0
STOPTUSSIN TV	94

Tabulka 4.8: Intenzita investic způsobená carry-over efektem

Zadavatel a mediatyp	Průměr	St. Dev.	Min	Max
ACC OOHTV	4 914	20 034	0	118 125
ACC Outdoor	1 017	6 165	0	50 553
ACC TV	8 065 565	8 090 287	167 799	28 526 149
MUCOSOLVAN TV	9 310 643	5 541 229	2 362 671	32 342 707
Robitussin Rádio	3 206	30 531	0	406 720
Robitussin TV	6 897 899	5 171 994	0	19 578 562
STOPTUSSIN Tisk	543	5 641	0	68 000
STOPTUSSIN TV	12 334 839	6 633 637	1 555 299	29 657 356

Nyní nám již nic nebrání v sestavení modelu s kombinací carry-over efektů, které vyšly jako nejvhodnější možnost v předešlé části. Zapíšeme tedy znovu vzorec a odhadneme model.

```
best = which(modely==max(modely)) #Naleznu nejlepsi model podle R2

vysledek = c()
for(i in 1:length(komunikace_CO)){
  vysledek[i]=komb[[best]][i]
}
Vysl = data.frame(Zadavatel = names(komunikace),
                  CarryOver = vysledek - 1)
stargazer(Vysl,summary = FALSE)

komunikace2 = komunikace
for(i in 1:length(komunikace2)){
  komunikace2[i] = komunikace_CO[[i]][vysledek[i]]
} # Zapisu vsechny hodnoty regresoru v nejvhodnejsim CO efektu
```

```
komunikace2 = komunikace2/(10^6) # Rozumnejsi jednotky
stargazer(komunikace2,digits = 0,digit.separator = " ")

promenneF = paste("komunikace2[,",1:length(komunikace_CO),"]")
vzorecF = as.formula(paste("DEKOMP_DATA ~ ",
                        paste(promenneF, collapse = "+"),
                        "+ cos(2*pi*(t-start)) + sin(2*pi*(t-start)) +
                        cos(4*pi*(t-start)) + sin(4*pi*(t-start)) +
                        cos(6*pi*(t-start)) + sin(6*pi*(t-start)) +
                        odchylky_pocasi + odchylky_pocasi_L1"))
vzorecF # Dostanu vzorec s nejlepsim modelem

FINAL = lm(vzorecF)
summary(FINAL)
```

Vzhledem k tomu, že odhadnuté parametry β u teploty nevycházejí signifikantní (výsledky regresní analýzy v tabulce 5.3 model (1) v příloze na straně 55), tak odhadneme ještě jeden model bez odchylek počasí. Porovnání obou modelů je ve výše zmíněné tabulce a už při pohledu na upravený koeficient determinance (R^2_{adj}) je jasné, že model bez počasí je lepší. Potvrdí to i ANOVA obou modelů, jejíž výsledky jsou zaznamenány v tabulce 4.9. Podle p-hodnoty lze říct, že mezi modely není na 5% hladině významnosti rozdíl.

Tabulka 4.9: ANOVA modelů s a bez teplotních odchylek

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	191	9.800				
2	193	9.845	-2	-0.045	0.438	0.646

```
vzorecF_bez_pocasi = as.formula(paste("DEKOMP_DATA ~ ",
                        paste(promenneF, collapse = "+"),
                        "+ cos(2*pi*(t-start)) + sin(2*pi*(t-start))+
                        cos(4*pi*(t-start)) + sin(4*pi*(t-start)) +
                        cos(6*pi*(t-start)) + sin(6*pi*(t-start))"))
FINAL_bezPocasi = lm(vzorecF_bez_pocasi)
stargazer(FINAL, FINAL_bezPocasi ,keep = c(1:9,15:17))

anova(FINAL,FINAL_bezPocasi) # Odchylky teploty nehraji roli
stargazer(anova(FINAL,FINAL_bezPocasi), summary = FALSE)
```

4.4 Testy a úprava modelu

Odhadnutím modelu ještě práce nekončí. Důležité je otestovat zda nejsou rezidua autokorelovaná a zda se v modelu nevyskytuje multikolinearita. Pro test autokorelace reziduí využijeme Durbin-Watson test a Breusch-Godfrey test, které jsou součástí balíčku *lmtest*. [31]

Durbin-Watson test, který počítá s rezidui, má stejné předpoklady jako klasický lineární model, tedy i normalitu a homoskedasticitu reziduí. To, zda-li jsou rezidua rozdělena normálně, se dá zjistit pomocí Shapiro-Wilkova testu, který má nulovou hypotézu o normálně rozdělené populaci. [32] Homoskedasticita se zkontroluje vizuálně – pohledem na graf reziduí v závislosti na čase nebo White testem na heteroskedasticitu. [16] Samotná interpretace výsledků D-W testu je poměrně obtížná, jelikož tento test má navíc i oblast neprůkaznosti, jejíž kritické hodnoty se udávají v tabulkách. Další možností testu autokorelace reziduí je Breusch-Godfrey test, který má méně předpokladů a navíc nepočítá jen se zpožděním reziduí řádu 1. [16]

```
library(lmtest)
shapiro.test(FINAL$residuals) # Rezidua normalne rozdelena
plot(FINAL$residuals~t) # Homoskedasticita
dwtest(FINAL) # DW-test vychazi 1.7 - test je neprukazny
bgtest(FINAL,order = 4) # nezamitam H0: neni zadna autokorelace rezidui
```

V našem modelu vychází hodnota D-W testu 1,7476. Ta se nachází v zóně neprůkaznosti určených dolní a horní mezí, které se rovnají $D_L = 1,61293$ a $D_U = 1,93947$. [33] I proto je lepší použít Breusch-Godfrey test, jehož nulová hypotéza svědčí o nulové autokorelaci reziudí. P-hodnota vyšla 0,1446 a tudíž nezamítáme nulovou hypotézu.

Nepřítomnost multikolinearity se ověří pomocí VIF (variance inflation factor), jehož obrovská výhoda oproti párové multikolinearitě (tzn. pouze mezi dvěmi regresory) spočívá v tom, že započítá vliv všech zbylých regresorů (mimo testovanou proměnnou). Příklad interpretace VIF: hodnota $\text{VIF}_j = 4$ znamená, že rozptyl $\hat{\beta}_j$ bude $4\times$ vyšší, než kdyby proměnná x_j nebyla korelovaná. [16] Existuje pravidlo, že $\text{VIF} \leq 10$ značí ještě přijatelnou multikolinearitu. Pro náš model jsou hodnoty VIF v tabulce 4.10. V Rku se všechny hodnoty spočítají jednoduše pomocí funkce *vif* v balíčku *car*. [34]

```
library(car)
vif(FINAL)
names(komunikace2)
summary(FINAL)
```

Tabulka 4.10: Hodnoty VIF pro jednotlivé parametry

Zadavatel a mediatyp	VIF
ACC OOHTV	1.898
ACC Outdoor	1.352
ACC TV	14.790
Mucosolvan TV	2.248
Robitussin Rádio	1.487
Robitussin TV	3.197
Stoptussin Tisk	1.211
Stoptussin TV	27.062
Odchylky počasí	1.128
Odchylky počasí$_{t-1}$	1.167

Kritérii k očištění modelu od nesignifikantních proměnných budou p-hodnoty a VIF. Všechny proměnné, které mají p-hodnotu větší než 0,05 a zároveň VIF ≤ 10, z modelu vynechám. Na hodnotu VIF je důležité koukat, protože přímo ovliňuje t-test, který počítá s rozptylem $\hat{\beta}_j$. Za zmínku při rozhodování stojí upozornit na to, že komunikace ACC TV by vyšla nesignifikantní, pokud bychom počítali s oboustrannou alternativní hypotézou $H_1 \neq 0$. Od vlastní mediální komunikace však očekáváme jasný pozitivní vliv, proto $H_1 > 0$ a tudíž pro jednostranný test dělíme p-hodnotu dvoustranného testu 2. Poté již televizní komunikace ACC vychází signifikantní. Naopak od konkurence očekávám buď slabý pozitivní nebo spíše negativní vliv, proto zde alternativní hypotéza zůstává stejná a Mucosolvan TV nebude vycházet statisticky významný. Pozitivní vliv mediální komunikace konkurence může nastat v případě, kdy připomene určitý druh produktu zákazníkům, kteří jsou věrní jiné značce – ta pak čerpá pozitivní vliv konkurenční reklamy.

Výsledná rovnice modelu bude vypadat takto (značení stejné jako u rovnice 4.3):

$$\begin{aligned} y_{dekomp} =& \beta_0 + \beta_{ACC_{TV}} ACC_{TV} + \beta_{Robitussin_{TV}} Robitussin_{TV} + \beta_{Stoptussin_{TV}} Stoptussin_{TV} + \\ & \sum_{n=1}^{3} \beta_{sin_n} \sin(2\pi n(t - start)) + \sum_{n=1}^{3} \beta_{cos_n} \cos(2\pi n(t - start)) + u, \end{aligned} \tag{4.4}$$

Kód k této části v Rku:

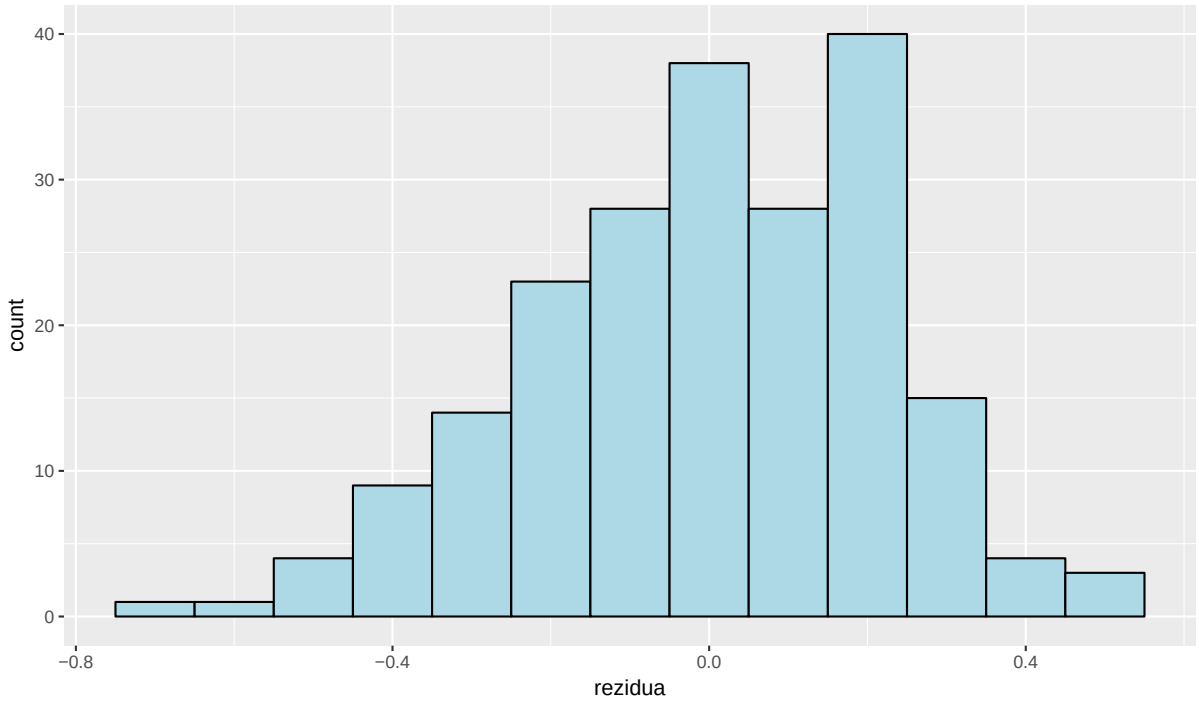
```
vzorecF2 = as.formula(paste("DEKOMP_DATA ~ komunikace2[,3] +
                            komunikace2[,6] + komunikace2[, 8] +
                            cos(2*pi*(t-start))+sin(2*pi*(t-start))+
                            cos(4*pi*(t-start))+sin(4*pi*(t-start))+
                            cos(6*pi*(t-start))+sin(6*pi*(t-start))"))

FINAL2 = lm(vzorecF2)
summary(FINAL2)
stargazer(anova(FINAL,FINAL2),summary = FALSE)
```

Z tabulky 4.11 je vidět, že v modelech není na 5% hladině významnosti rozdíl. Proto použijeme očištěný model a opět se provedou testy autokorelace reziduí a multikolinearity.

Tabulka 4.11: ANOVA očištěného modelu

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	191	9.800				
2	198	10.191	-7	-0.391	1.089	0.372

Rezidua při pohledu na histogram na obrázku 4.8 zřejmě nebudou normálně rozdělena (jsou negativně sešikmena). Tuto domněnku potvrzuje i Shapiro-Wilkův test, který zamítá nulovou hypotézu o normálně rozdělené populaci. Díky tomu není možné použít Durbin-Watsonův test. Jedinou možností je tedy použít Breusch-Godrey test, který



Obrázek 4.8: Histogram reziduí výsledného modelu

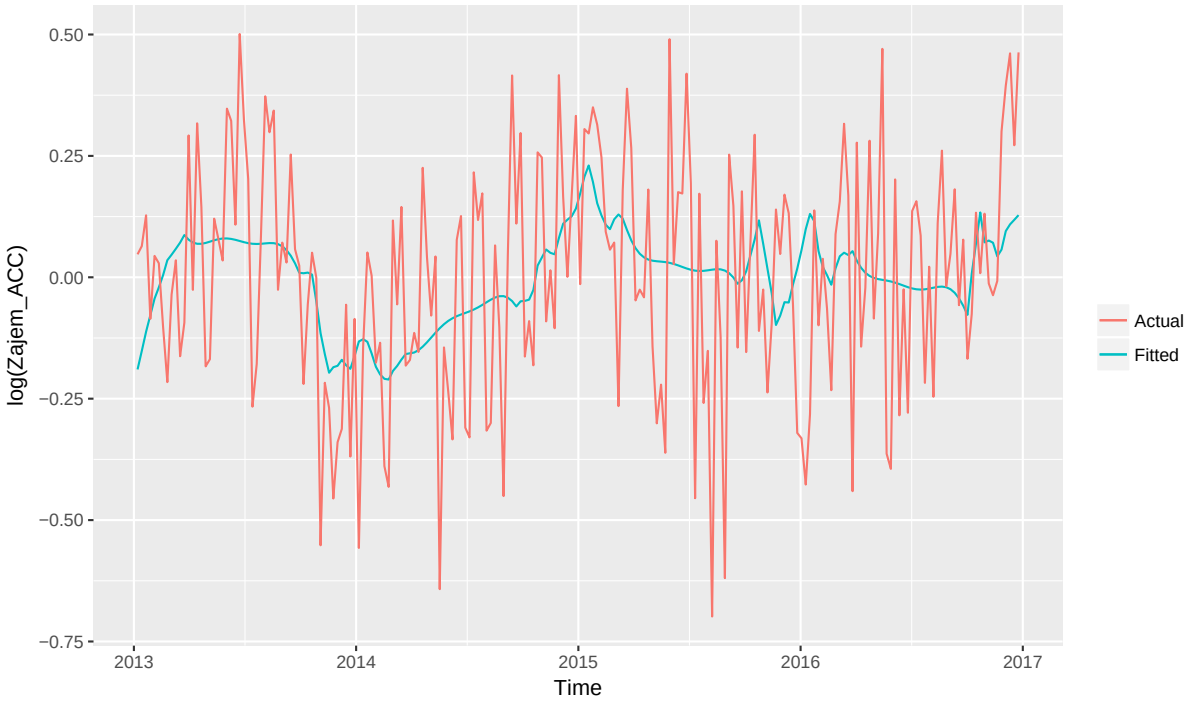
zdroj: autor

na 5% hladině významnosti nezamítá H_0 o neexistenci autokorelace. Multikolinearita se v modelu vyskytuje v únosné míře, avšak může ovlivnit t-testy.

```
pdf("hist.pdf",width = 8.5, height = 5) #export
qplot(FINAL2$residuals,
      geom = "histogram",
      binwidth = 0.1,
      xlab = "rezidua",
      fill = I("lightblue"),
      col = I("black"))
dev.off()

shapiro.test(FINAL2$residuals)
# dwtest(FINAL2) - nemuzu pouzit
bgtest(FINAL2,order = 4) # H0: neni zadna autokorelace rezidui

vif(FINAL2)
```

4.5 Interpretace výsledků

Před samotnou interpretací výsledků se ještě podíváme na obrázek 4.9, který zobrazuje, kolik variability se nám podařilo vysvětlit na sezónně očištěné časové řadě zájmu o ACC. Oproti obrázku 4.10, na kterém je započítána i odhadnutá sezónnost, jsou zde lépe viditelné detaily vlivu mediálních investic. Propad v roce 2014 byl způsoben zahá-



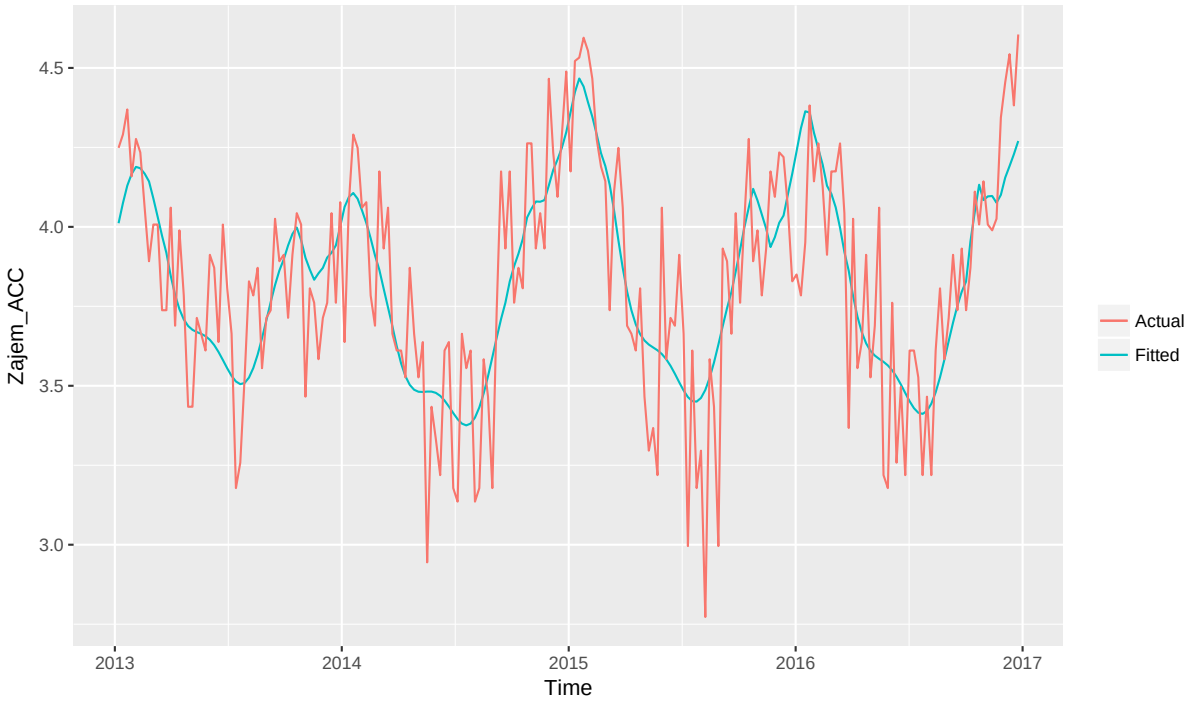
Obrázek 4.9: Vyrovnané vs. původní hodnoty očištěné řady

zdroj: autor

jením televizní kampaně Robitussinu. V roce 2015 došlo k většímu růstu, protože ACC významně podpořilo televizní komunikaci a Robitussin naopak své investice snížil. V roce 2016 došlo opět k mírnému poklesu i přesto, že ACC investovalo ve stejném rozsahu jako v roce 2015. Důvodem bylo silné navýšení investic Robitussinu.

Z popisu událostí ovlivňujících zájem o ACC v letech 2013–2016 bylo možné povšimnout si faktu, že největší konkurencí ACC je Robitussin. Z tabulky 4.12 je možné vidět, že efekt televizní reklamy Robitussinu je mnohem větší než samotného ACC. Naštěstí pro ACC je objem investic Robitussinu menší a intenzita kampaně je tak po započítání carry-over efektu slabší. Patrné to je z tabulky 4.8, kde maximální intenzita ACC TV je adekvátní 28 milionům Kč investovaných v jeden okamžik. I průměrná intenzita kampaní Robitussinu je více jak o jeden milion nižší. ACC by tedy mělo sledovat svého největšího konkurenta v podobě Robitussinu a jeho TV investic a v případě větších investic buď zvýšit i své investice nebo počítat s menším zájmem.

Hodnota 0,008 v tabulce 4.12 u ACC TV znamená, že pokud by intenzita kampaně po carry-over efektu byla o 1 milion Kč větší a vše ostatní zůstalo stejné, tak by zájem o ACC stoupnul o $(e^{0,008} - 1) \times 100 = 0,79$ %. Stejně tak pokud by intenzita televizní kampaně Robitussinu po carry-over efektu vzrostla o jeden milion Kč ceteris paribus, tak by zájem o ACC klesnul o $(e^{-0,023} - 1) \times 100 = -2,27$ %. Pokud bychom chtěli změnu vyjádřit v původních jednotkách modelu (a ne v relativních změnách), bavili bychom se o procentních bodech, protože původní jednotky modelu jsou procenta. To znamená, že pokud by zájem byl v jednom období například 60 % a intenzita ACC TV by byla o milion Kč vyšší ceteris paribus, tak by zájem vzrostl o $60 \times 0,79$ % $= 0,47$ p. b.



Obrázek 4.10: Vyrovnané vs. původní hodnoty zájmu o ACC

zdroj: autor

Tabulka 4.12: Výsledný model (směrodatné chyby jsou ovlivněny slabou multikolinearitou)

	Závislá proměnná:
	DEKOMP_DATA
ACC TV	0.008
	(0.006)
Robitussin TV	-0.023^{***}
	(0.005)
Stoptussin TV	0.018^{*}
	(0.010)
Počet pozorování	208
R^2	0.143
Upravené R^2	0.104
F Statistika	3.665^{***} (df = 9; 198)
Poznámky:	(i) Model obsahoval i trigonometrický rozvoj
	(ii)$^{*}p$<0.1; $^{**}p$<0.05; $^{***}p$<0.01

```
R = round(summary(FINAL2)$r.squared,digits = 2)
graf6 = data.frame(Fitted = FINAL2$fitted.values,
                   Actual = DEKOMP_DATA,
                   Time = t)
graf6 = ggplot(data = graf6, aes(Time)) +
  geom_line(aes(y = Fitted, colour = "Fitted")) +
  geom_line(aes(y = Actual, colour = "Actual")) +
  ggtitle(substitute(paste("Fitted values vs. Actual data (",
                           R^2," = ",W,")"),list(W = R))) +
  ylab("log(Zajem_ACC)") +
  labs(colour="")
# pdf("graf6.pdf",width = 8.5, height = 5) #export
graf6
# dev.off()

FIT = FINAL2$fitted.values
summary(m3)
FIT2 = m3$fitted.values[-1] + FIT

R = round(cor(dataTS$log_Zajem_ACC[-1],FIT2)^2,digits = 2)
graf7 = data.frame(Fitted = FIT2,
                   Actual = dataTS$log_Zajem_ACC[-1],
                   Time = t)
graf7 = ggplot(data = graf7, aes(Time)) +
  geom_line(aes(y = Fitted, colour = "Fitted")) +
  geom_line(aes(y = Actual, colour = "Actual")) +
  ggtitle(substitute(paste("Fitted values vs. Actual data (",
                           R^2," = ",W,")"),list(W = R))) +
  ylab("Zajem_ACC") +
  labs(colour="")
# pdf("graf7.pdf",width = 8.5, height = 5) #export
graf7
# dev.off()
```

4.6 ROI a optimální strategie

Jediným statisticky významným mediatypem se ukázala být TV a to jak u ACC, tak u jeho konkurentů. Ostatní mediatypy využité ACC nebyly signifikantní (viz tabulka 5.3), a proto nevstupovaly do výsledného modelu. Optimální komunikační mix by tedy měl obsahovat jen televizní kamapaně, protože mají jako jediný mediatyp signifikantní vliv na zájem o ACC.

Pro výpočet ROI bude nutné předpokládat několik věcí. První je provázanost zájmu podle Google Trends s reálnými prodeji, která se dá velmi dobře předpokládat a již byla i dokázána. [8] Druhou bude, že 1 % zájmu na Google Trends znamená cca 4 000 prodaných balení ACC a čistý zisk z jednoho balení bude průměrně 100 Kč.

Ukazatel ROI (Return on investment) je jedním ze základních ukazatelů v marketingu a podnikové ekonomice. Udává výnos v procentech z utracené částky. Tento ukazatel lze

spočítat podle vzorce 4.5. [35]

$$\text{ROI}(\%) = \frac{\text{čistý zisk}}{\text{investice}} \times 100, \tag{4.5}$$

Vzhledem k tomu, že model má zlogaritmovanou vysvětlovanou proměnnou a my budeme potřebovat reálné hodnoty, tak je nutné exponovat vyrovnané hodnoty. Zde se ale objevuje problém v odhadu náhodné složky.

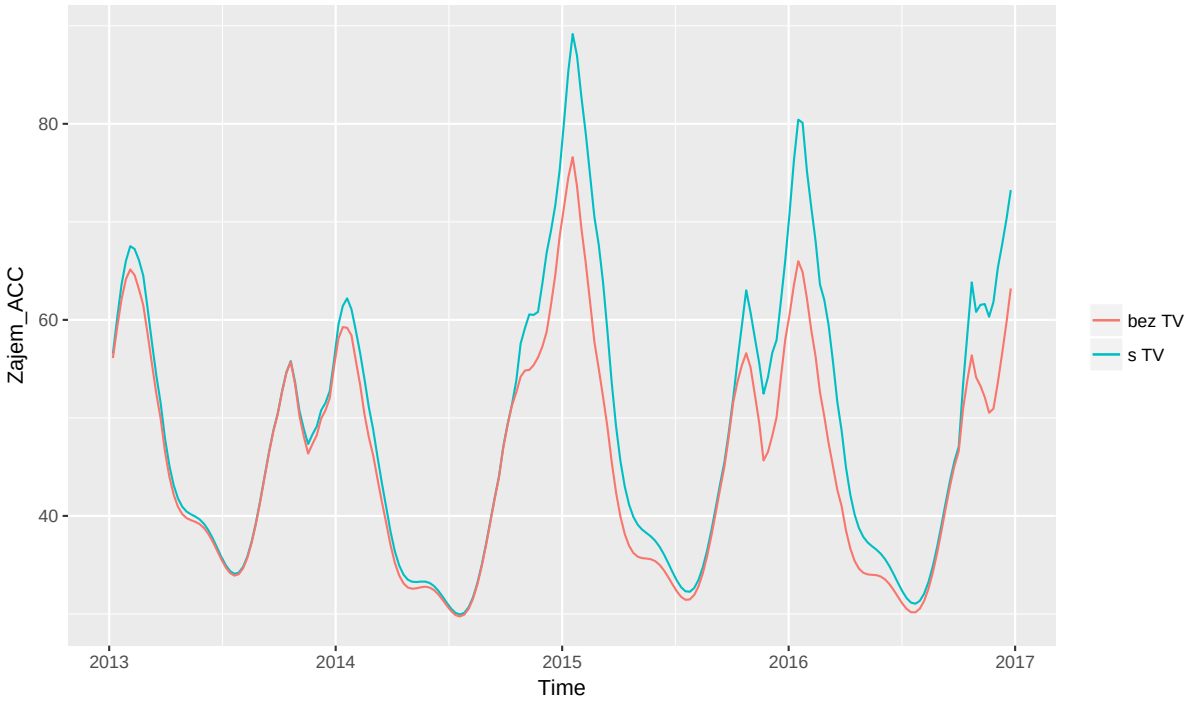
$$\begin{aligned} \mathrm{E}[y|\boldsymbol{x}] &= \mathrm{E}[e^{\boldsymbol{\beta x}+\boldsymbol{u}}|\boldsymbol{x}] \\ &= e^{\boldsymbol{\beta x}}\mathrm{E}[e^{u}|\boldsymbol{x}] \end{aligned} \tag{4.6}$$

Ze vzorce 4.6 není problém získat část $e^{\boldsymbol{\beta x}}$, ale střední hodnotu exponované náhodné složky nemáme jak odhadnout. Na pomoc však přichází Duanův estimátor, který odhaduje $\mathrm{E}[e^{u}|\boldsymbol{x}]$ jako průměr exponovaných reziduí z modelu se zlogaritmovanou vysvětlující proměnnou. [16]

$$\widehat{\mathrm{E}[e^{u}|\boldsymbol{x}]} = n^{-1}\sum_{i=1}^{n} e^{\hat{u}_i} \tag{4.7}$$

Abychom zjistili, jak velký přírůstek byl způsoben TV komunikací, vypočítáme vyrovnané hodnoty pro výsledný model, kde vynecháme celou proměnnou ACC TV. Součet rozdílů těchto vyrovnaných hodnot a hodnot z modelu se všemi proměnnými bude představovat inkrementální nárůst zájmu o ACC. Na obrázku 4.11 je vidět, kolik procentních bodů zájmu o ACC bylo způsobených televizní komunikací.

```
unlog = exp(FIT2) * mean(exp(FINAL2$residuals)) # Odlogaritmovano

FIT_bezTV = (FINAL2$coefficients[1] +
               FINAL2$coefficients[3]*komunikace2[,6] +
               FINAL2$coefficients[4]*komunikace2[,8] +
               FINAL2$coefficients[5]*cos(2 * pi * (t - start)) +
               FINAL2$coefficients[6]*sin(2 * pi * (t - start)) +
               FINAL2$coefficients[7]*cos(4 * pi * (t - start)) +
               FINAL2$coefficients[8]*sin(4 * pi * (t - start)) +
               FINAL2$coefficients[9]*cos(6 * pi * (t - start)) +
               FINAL2$coefficients[10]*sin(6 *pi * (t - start)))
FIT2_bezTV = FIT_bezTV + m3$fitted.values[-1] # pridam sezonnost
unlogTV = exp(FIT2_bezTV) * mean(exp(FINAL2$residuals))

increaseTV = sum(unlog-unlogTV)

graf8 = data.frame(sTV = unlog,
                   bezTV = unlogTV,
                   Time = t)
graf8 = ggplot(data = graf8, aes(Time)) +
  geom_line(aes(y = sTV, colour = "s TV")) +
  geom_line(aes(y = bezTV, colour = "bez TV")) +
  ggtitle("TV prinos") +
```



Obrázek 4.11: Inkrementální nárůst zájmu o ACC způsobený TV
zdroj: autor

```
  ylab("Zajem_ACC") +
  labs(colour="")
# pdf("graf8.pdf",width = 8.5, height = 5) #export
graf8
# dev.off()
```

Máme-li vypočítaný inkrementální nárůst zájmu o ACC způsobený televizní kampaní a předpokládáme-li, že 1 % zájmu na Google Trends představuje zhruba 4 000 prodaných balení ACC, kde jedno balení generuje čistý zisk průměrně 100 Kč, tak můžeme vypočítat návratnost investice do televizní komunikace.

```
TV_krabicky = increaseTV * 4000 # prinos TV
TV_prinos = TV_krabicky*100 # cena baleni = 100 Kc
invest = sum(dataTS$ACC_TV) # celkova suma investic

ROI = TV_prinos / invest
print(ROI)
```

Ukazatel ROI vychází 1,55 za výše zmíněných předpokladů. To znamená, že z každé investované peněžní jednotky získáme navíc 55 %. Vzhledem k tomu, že předpis funkce popisující vliv TV je lineární (i po započítání carry-over efektu), tak by bylo nejlepší možnou volbou investovat všechny disponibilní peníze do TV reklamy, abychom měli co největší zisk. Toto přesně odpovídá jedné z geniálních ekonomických myšlenek, které jsem se na VŠE naučil a totiž prvnímu axiomu racionálního chování investora – více peněz je lepší než méně peněz. [36] Z této myšlenky zřejmě později vzešlo i podobné lidové moudro a to, že je lepší být zdravý a bohatý, než nemocný a chudý. Ovšem ve funkci popisující vliv

TV na zájem o ACC není započítána maximální hranice nasycenosti trhu ani frekvence reklamy, kdy začíná její množství zákazníky odrazovat. Takže se dá předpokládat, že její vliv by neměl stoupat lineárně, ale konkávně s určitou limitní hodnotou. To by znamenalo nelineární regresní analýzu, která by mohla být předmětem dalšího výzkumu založeného na složitější teorii, kterou bych se rád v budoucnu zabýval.

Předchozí odstavec popisoval optimistickou variantu výsledku ukazatele ROI. Je však možné, že ROI < 1, takže by investice do reklamy byly větší než příjmy, které způsobila. To znamená, že by negenerovaly žádný zisk a naopak způsobovaly ztrátu. Zde je však vhodné se zamyslet nad tím, že reklama může přinést i jiné efekty než jen prodej daného produktu. Výdaje větší než příjmy se vyplatí, když například získáme kontaktní informace, které povedou k budoucím nákupům na základě cross-sellu nebo podrobně cílené reklamy. Ke zhodnocení návratnosti investice do reklamy je tedy zapotřebí započítat i čistou současnou hodnotu budoucího zisku či výnosu daného zákazníka (vyjadřuje za kolik spokojený zákazník v dané firmě nakoupí za celý život), která se může u každého jednotlivce lišit. [37] Bohužel toto lze také jen velmi složitě modelovat.

5 Závěr

V empirické analýze jsem určil faktory ovlivňující zájem o ACC na webu. Nejsilnějším faktorem je sezónnost časové řady, která souvisí s faktem, že ACC je převážně sezónní produkt. Odhad a očištění od sezónnosti jsem provedl pomocí Fourierovy transformace, která odhalila průměrný vrchol zájmu o ACC, který je těsně po začátku roku. Zřejmě se jedná o období s průměrně nejvyšším výskytem nemocí doprovázených kašlem. Překvapivě teplota po očištění od sezónnosti ani její zpožděná hodnota nebyla statisticky významná. Dalšími faktory ovlivňující vyhledávanost pojmu ACC na webu je televizní komunikace. Ostatní mediatypy nevyšly signifikantní ani u jednoho ze zadavatelů.

Optimální strategií komunikačního mixu pro následující období by tedy bylo využití pouze televizní reklamy. Vzhledem k silným carry-over efektům (kolem 90 %), které způsobují kumulování intenzity reklamy je vhodné investovat souvisle a hned ze začátku hlavní sezóny.

Největší konkurencí ACC podle výsledného modelu je Robitussin, který ale naštěstí pro ACC investuje do reklamy výrazně méně než ACC. Určitě by ale investice Robitussinu měly být sledované, protože v případě nárůstu by zájem o ACC byl znatelně menší.

Modelový odhad ukazatele ROI pro televizní komunikaci ACC byl vypočítán a vychází 1,55. Znamená to tedy, že čistý zisk způsobený TV se bude rovnat 55 % původní hodnoty investic do reklamy.

5.1 Prostor pro zlepšení

Co se týče samotného výpočtu, tak největší prostor se naskýtá v části, kde se nejlepší model odhaduje několika do sebe vnořenými for cykly. Tento postup je velmi náročný na výpočetní kapacitu. Ideální by bylo použít jiný algoritmus, který bohužel neznám. Další možností co zlepšit, je určitě numerické napočítání všech carry-over efektů. Zde dokonce i vím, že v Rku by měl existovat balíček, který umí vytvořit geometricky klesající zpožděné proměnné.

Zmínit také musím linearitu celého modelu. Závislost mezi mediálními investicemi a zájmem nebo prodeji generovanými reklamou velmi pravděpodobně nebudou lineární, ale konkávní. Přece jen existuje určitá konečná velikost trhu, takže zájem či prodej nemohou stoupat s vyššími investicemi do nekonečna, ale dojdou na maximální hodnotu, přes kterou není možné se dostat. U vysvětlujících proměnných existuje také prostor pro zlepšení. Místo ceníkových cen investic do televizní reklamy by bylo vhodnější počítat s počtem nakoupených GRPů. Ceníkové ceny jsou totiž ovlivněny například částí dne nebo pozicí reklamního spotu. S tím souvisí i výpočet ROI, kde bylo také počítáno s ceníkovými cenami. Ty se však mohou lišit od čistých investic díky množstevním slevám.

Literatura

[1] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[2] "ACC LONG, šumivé tablety." *SÚKL* [online], citováno 12.2.2017, dostupné z: www.sukl.eu.

[3] J. Slíva, "Význam N-acetylcysteinu jako léčivé látky v pediatrii," *Praktické lékárenství*, vol. 10, no. 5, pp. 171 – 173, 2014.

[4] R. T. a kolektiv, *Marketingový výzkum*. Grada, 2017. ISBN: 978-80-271-0206-8.

[5] "Search Engine Market Share." *StatCounter* [online], citováno 3.3.2017, dostupné z gs.statcounter.com.

[6] "The Google Homepage Design Patent." *Design Patent* [online], citováno 3.3.2017, dostupné z: www.patentadesign.com.

[7] V. Janouch, *Internetový marketing*. Computer Press, 2011. ISBN: 978-80-251-2795-7.

[8] N. Askitas and K. F. Zimmermann, "Google econometrics and unemployment forecasting," *Applied Economics Quarterly*, vol. 55, pp. 107–120, Apr 2009.

[9] C. I. Papanagnou and O. Matthews-Amune, "Coping with demand volatility in retail pharmacies with the aid of big data exploration," *Computers & Operations Research*, 2017.

[10] K. Hvížďala, *Moc a nemoc médií*. Praha: Máj, 2003. ISBN: 80-86643-07-7.

[11] "Cenové podmínky propagace na televizních kanálech Nova." *Ministerstvo vnitra ČR - registr smluv* [online], citováno 2.3.2018, dostupné z: www.smlouvy.gov.cz.

[12] "Historická data - meteorologie a klimatologie." *Český hydrometeorologický ústav* [online], citováno 29.11.2017, dostupné z: portal.chmi.cz.

[13] A. Walker, *openxlsx: Read, Write and Edit XLSX Files*, 2017. R package version 4.0.17.

[14] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. ISBN: 978-0-387-98140-6.

[15] J. Arlt, M. Arltová, E. Rublíková, *Analýza ekonomických časových řad s příklady*. Oeconomica, 2004. ISBN: 80-245-0777-3.

[16] J. M. Wooldridge, *Introductory Econometrics: a Modern Approach.* Cengage Learning, 2013. ISBN: 978-1-111-53104-1.

[17] Nerlove, Marc; Grether, David M.; Carvalho, Jose L. , *Analysis of Economic Time Series. Economic Theory, Econometrics, and Mathematical Economics.* Elsevier, 1995. ISBN: 0-12-515751-7.

[18] A. Kufner, J. Kadlec, *Fourierovy řady.* Academia, 1969.

[19] J. Zouhar. Osobní sdělení přednášejícího (Vysoká škola ekonomická v Praze, nám. W. Churchilla 1938/4, Praha 3) dne 22. února 2018.

[20] R. Davidson, J. G. MacKinnon, *Econometric Theory and Methods.* Oxford University Press, 2005. ISBN: 978-0195123722.

[21] S. Miloslav, a kol., *Manažerská ekonomika.* Grada, 2011. ISBN: 978-80-247-3494-1.

[22] L. J. Parsons, "A rachet model of advertising carryover effects," *Journal of Marketing Research*, vol. 13, pp. 76–79, Feb 1976.

[23] D. S. Tull, "The carry-over effect of advertising," *Journal of Marketing*, vol. 29, pp. 46–53, Apr 1965.

[24] G. Grolemund and H. Wickham, "Dates and times made easy with lubridate," *Journal of Statistical Software*, vol. 40, no. 3, pp. 1–25, 2011.

[25] H. RJ, "Automatic time series forecasting: the forecast package for R.," *Journal of Statistical Software*, vol. 26, no. 3, pp. 1–22, 2018.

[26] M. Hlavac, *stargazer: Well-Formatted Regression and Summary Statistics Tables.* Central European Labour Studies Institute (CELSI), Bratislava, Slovakia, 2018. R package version 5.2.1.

[27] A. Zeileis, *dynlm: Dynamic Linear Regression*, 2016. R package version 0.3-5.

[28] A. Trapletti and K. Hornik, *tseries: Time Series Analysis and Computational Finance*, 2017. R package version 0.10-41.

[29] J.-L. Bertrand, X. Brusset, and M. Fortin, "Assessing and hedging the cost of unseasonal weather: Case of the apparel sector," *European Journal of Operational Research*, vol. 244, no. 1, pp. 261–276, 2015.

[30] O. C. Demirag, "Performance of weather-conditional rebates under different risk preferences," *Omega*, vol. 41, no. 6, pp. 1053–1067, 2013.

[31] A. Zeileis and T. Hothorn, "Diagnostic checking in regression relationships," *R News*, vol. 2, no. 3, pp. 7–10, 2002.

[32] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[33] "Critical Values for the Durbin-Watson Test." *Stanford University* [online], citováno 12.4.2018, dostupné z: https://web.stanford.edu/~clint/bench/dwcrit.htm.

[34] J. Fox and S. Weisberg, *An R Companion to Applied Regression.* Thousand Oaks CA: Sage, second ed., 2011.

[35] P. Phillips and J. Phillips, *Return on Investment (ROI) Basics.* ASTD Press, 2006. ISBN: 978-1562864064.

[36] Marek, Petr a kolektiv, *Studijní průvodce financemi podniku.* Ekopress, 2006. ISBN: 80-86119-37-8.

[37] K. P. Onn and A. Mercer, "The direct marketing of insurance," *European Journal of Operational Research*, vol. 109, no. 3, pp. 541 – 549, 1998.

Příloha

R kód pro teoretickou část

```
#-------------------------------------------------------------
# Fourierova transformace
y1 = c()
y2 = c()
y3 = c()
y4 = c()
l = seq(from = 0, to = 4*pi, by = 0.01)
i=1
for(x in seq(from = 0, to = 4*pi, by = 0.01)){
  y1[i]=4*sin(x)/(pi)
  y2[i]=4*sin(x)/(pi)+4*sin(3*x)/(3*pi)
  y3[i]=4*sin(x)/(pi)+4*sin(3*x)/(3*pi)+4*sin(5*x)/(5*pi)
  y4[i]=4*sin(x)/(pi)+4*sin(3*x)/(3*pi)+4*sin(5*x)/(5*pi
        +4*sin(7*x)/(7*pi)
  i=i+1
}

y5 = c(1,rep(1,times = pi/0.01),rep(-1,times = pi/0.01),
       rep(1,times = pi/0.01),rep(-1,times = pi/0.01))

library(ggplot2)
FourierG = data.frame(x = l,
                      Square_wave = y5,
                      Stupen_1 = y1,
                      Stupen_2 = y2,
                      Stupen_3 = y3,
                      Stupen_4 = y4)
FourierG <- ggplot(data = FourierG, aes(x)) +
  geom_line(aes(y = Square_wave),colour = "black") +
  geom_line(aes(y = Stupen_1, colour = "Stupen_1")) +
  geom_line(aes(y = Stupen_2, colour = "Stupen_2")) +
  geom_line(aes(y = Stupen_3, colour = "Stupen_3")) +
  geom_line(aes(y = Stupen_4, colour = "Stupen_4")) +
  # ggtitle("Fourierova rada") +
  ylab("y") +
  xlab("x") +
```

```
  labs(colour="Rozvoj")
pdf("Fourier.pdf",width = 10, height = 5) #export
FourierG
dev.off() #export

#-------------------------------------------------------------
# Random walk & White noise
y=runif(20,min=-2,max=2) # Generovani 20 nahodnych cisel
y2=runif(20,min=-2,max=2)
x=1:20
w=c(0)
w2=c(0)
x2=1:500
for(i in 2:500){
  w[i]=w[i-1]+runif(1,-1,1)
}
for(i in 2:500){
  w2[i]=w2[i-1]+runif(1,-1,1)
}

library(ggplot2)
WhiteNoise = data.frame(y1 = y,
                        y2 = y2,
                        0~= rep(0, times = 20),
                        Cas = x)
WhiteNoise = ggplot(data = WhiteNoise, aes(Cas)) +
  geom_line(aes(y=0),color = "black") +
  geom_line(aes(y=y1),color = "darkgreen",size = 1.2) +
  geom_line(aes(y=y2),color = "darkred",size = 1.2) +
  ylab("")
pdf("White_noise.pdf",width = 8.5, height = 5) #export
WhiteNoise
dev.off() #export

RandomWalk = data.frame(y1 = w,
                        y2 = w2,
                        0~= rep(0, times = 500),
                        Cas = x2)
RandomWalk = ggplot(data = RandomWalk, aes(Cas)) +
  geom_line(aes(y=0),color = "black") +
  geom_line(aes(y=y1),color = "darkgreen",size = 1.2) +
  geom_line(aes(y=y2),color = "darkred",size = 1.2) +
  ylab("")
pdf("Random_walk.pdf",width = 8.5, height = 5) #export
RandomWalk
dev.off() #export
```

```
#-----------------------------------------------------------
#Carry Over efekt
sloupec=dataTS$ACC_TV
NOVYsloupec=sloupec
NOVYsloupec2=sloupec
CO=0.8
CO2=0.3

for (i in 2:length(sloupec)) {
  NOVYsloupec[i]=NOVYsloupec[i]+NOVYsloupec[i-1]*CO
}
for (i in 2:length(sloupec)) {
  NOVYsloupec2[i]=NOVYsloupec2[i]+NOVYsloupec2[i-1]*CO2
}

graf_CO = data.frame(Cas = t,
                     Vydaje = sloupec/1000,
                     Carry_Over_efekt_80 = NOVYsloupec/1000,
                     Carry_Over_efekt_30 = NOVYsloupec2/1000)
grafCO <- ggplot(data = graf_CO,aes(Cas)) +
  geom_line(aes(y = Carry_Over_efekt_80,colour = "Carry_over_80%")) +
  geom_line(aes(y = Carry_Over_efekt_30,colour = "Carry_over_30%")) +
  geom_bar(aes(x = Cas, y = Vydaje),stat="identity") +
  # ggtitle("Porovnani carry over efektu") +
  ylab("Vydaje_v_tis._CZK") +
  labs(colour="Intenzita")
pdf("Carry_over.pdf",width = 8.5, height = 5) #export
grafCO + scale_color_manual(values=c("#0066ff", "#ff0000"))
dev.off() #export
```

Tabulky

Tabulka 5.1: Fourierova transformace využitá k odhadu sezónnosti závislé proměnné

	Závislá proměnná:			
	log_Zajem_ACC			
	(1)	(2)	(3)	(4)
cos(2 ∗pi ∗(t - start))	0.369***	0.369***	0.369***	0.369***
	(0.024)	(0.024)	(0.024)	(0.024)
sin(2 ∗pi ∗(t - start))	0.009	0.009	0.009	0.009
	(0.024)	(0.024)	(0.024)	(0.024)
cos(4 ∗pi ∗(t - start))		−0.003	−0.003	−0.003
		(0.024)	(0.024)	(0.024)
sin(4 ∗pi ∗(t - start))		0.020	0.020	0.020
		(0.024)	(0.024)	(0.024)
cos(6 ∗pi ∗(t - start))			−0.024	−0.024
			(0.024)	(0.024)
sin(6 ∗pi ∗(t - start))			0.074***	0.074***
			(0.024)	(0.024)
cos(8 ∗pi ∗(t - start))				0.046*
				(0.024)
sin(8 ∗pi ∗(t - start))				−0.026
				(0.024)
Konstanta	3.828***	3.828***	3.828***	3.828***
	(0.017)	(0.017)	(0.017)	(0.017)
Počet pozorování	209	209	209	209
R^2	0.532	0.533	0.557	0.568
Upravené R^2	0.527	0.524	0.544	0.551
Poznámky:				$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Tabulka 5.2: Očištění teploty od sezónnosti pomocí Fourierovy transformace

	Závislá proměnná:
	Teplota
cos(2 ∗pi ∗(t - start))	-9.287^{***}
	(0.287)
sin(2 ∗pi ∗(t - start))	-2.406^{***}
	(0.287)
cos(4 ∗pi ∗(t - start))	0.459
	(0.287)
sin(4 ∗pi ∗(t - start))	0.316
	(0.287)
cos(6 ∗pi ∗(t - start))	0.083
	(0.287)
sin(6 ∗pi ∗(t - start))	-0.564^{*}
	(0.287)
Konstanta	9.522^{***}
	(0.203)
Počet pozorování	209
R^2	0.848
Upravené R^2	0.843
Poznámky:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Tabulka 5.3: Porovnání modelů s a bez vlivu teplotních odchylek

	Závislá proměnná:	
	DEKOMP_DATA	
	(1)	(2)
ACC OOHTV	1.262	1.342
	(1.083)	(1.074)
ACC Outdoor	1.009	1.333
	(2.933)	(2.902)
ACC TV	0.013*	0.014*
	(0.007)	(0.007)
Mucosolvan TV	0.008*	0.008**
	(0.004)	(0.004)
Robitussin Rádio	−0.293	−0.309
	(0.626)	(0.615)
Robitussin TV	−0.024***	−0.025***
	(0.005)	(0.005)
Stoptussin Tisk	−3.087	−2.541
	(3.071)	(2.995)
Stoptussin TV	0.013	0.012
	(0.012)	(0.012)
Odchylky počasí	−0.004	
	(0.006)	
Odchylky počasí$_{t-1}$	−0.003	
	(0.006)	
Konstanta	−0.172**	−0.172**
	(0.081)	(0.080)
Počet pozorování	208	208
R^2	0.176	0.172
Upravené R^2	0.107	0.112
F statistika	2.545*** (df = 16; 191)	2.862*** (df = 14; 193)
Poznámky:	(i) Oba modely obsahovaly i trigonometrický rozvoj (ii)*p<0.1; **p<0.05; ***p<0.01	