

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITY KOMENSKÉHO V BRATISLAVE

DIPLOMOVÁ PRÁCA

BRATISLAVA 2003

JURAJ KUČERA

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITY KOMENSKÉHO V BRATISLAVE

Ekonomická a finančná matematika



**ANALÝZA DOPYTU A TVORBA PROGNÓZ
S APLIKÁCIAMI NA HYPERMARKETY**

Diplomant: Juraj Kučera

Vedúci diplomovej práce: Ing. Rastislav Rázus

Čestne prehlasujem, že som túto diplomovú prácu vypracoval samostatne, len s použitím uvedenej literatúry.

.....
Juraj Kučera

Rád by som poďakoval diplomovému vedúcemu Ing. Rastislavovi Rázusovi za pomoc pri práci a konštruktívne návrhy, Prof. RNDr. Pavlovi Brunovskému DrSc. za ochotu a vecné pripomienky a zamestnancom Jednoty za poskytnuté dáta.

OBSAH

Úvod.....	1
1. Dekompozícia.....	3
1.2 Trend.....	3
1.3 Inflácia.....	11
1.4 Sezónnosť.....	13
2. Regresný prístup.....	19
3. Box-Jenkinsova metodológia.....	32
3.1 Základné pojmy.....	32
3.2 Procesy MA, AR, ARMA.....	36
3.3 Identifikácia modelu.....	38
3.4 Nestacionárne procesy.....	39
3.5 Sezónne procesy.....	41
3.6 Prognózy.....	43
3.7 Aplikácie.....	45
4. Mikroekonomický prístup.....	50
Záver.....	54
Zoznam použitej literatúry.....	55

ÚVOD

Cieľom tejto diplomovej práce je nájsť a aplikovať vhodné metódy na odhad budúceho dopytu. Zameriame sa na krátke predpovede s cieľom urobiť prognózu použiteľnú na optimálne zásobenie skladu, ktoré je veľmi dôležité. Prebytok na sklade znamená neproduktívne uložené peniaze, primálo tovaru vedie k ušlému zisku a strate zákazníkov. Dobrá analýza dopytu prináša aj ďalšie výhody, napríklad možnosť pripraviť cenovú a letákovú akciu¹.

V práci sa zameriame na predajne hypermarketového typu. Výhoda analyzovania dát z veľkých predajní spočíva v ich množstve, dopyt nepodlieha náhodným fluktuáciám v takej miere ako napríklad pri malých lokálnych potravinách. Dáta použité v tejto práci pochádzajú z dvoch bratislavských hypermarketov, Terno 1 v Petržalke a Terno 2 v Auparku. Dáta pochádzajú z časového obdobia september 2001 - október 2002.

Prevádzky hypermarketového typu sú pomerne špecifické. Veľkú úlohu pri predaji totiž zohráva určitá psychológia, akási snaha presvedčiť zákazníka o potrebe kúpiť ten ktorý produkt. Dôležitá je aj otázka optimálneho rozmiestnenia tovaru a dostatočného zaplnenia regálov, obzvlášť pri cenových akciách. Psychológiu a optimálne rozmiestnenie tovarov nebudeme v tejto práci uvažovať, analýze dopytu a tvorbe prognóz sa budeme venovať len zo stránky matematickej a ekonomickej.

Predtým ako sa začneme venovať problému, krátko spomenieme riešenie optimálneho zásobenia skladu v praxi. Zdá sa, že realita nepotrebuje matematický pohľad. V praxi sa kladie dôraz na skúsenosti a prax vedúcich zodpovedných za určité úseky. Nemôžeme tvrdiť, že takto zásobený sklad by trpel nedostatočnou zásobou tovarov a to ani v prípade sviatkov, keď je dopyt zvýšený. Ťažko však zodpovedať otázku, či je tento stav optimálny a zásoba tovarov na sklade nie je príliš veľká a tým pádom neefektívna.

Práca je rozdelená do niekoľkých častí. V prvej časti sa budeme venovať časovým radom a dekompozičnému prístupu, skúsime odhadnúť trend a vplyv niektorých sezónnych

¹ Pod pojmom cenová akcia budeme rozumieť zníženie ceny výrobku spojené s upozornením na toto zníženie. Pod pojmom letáková akcia sa rozumie uvedenie výrobku v reklamnom letáku, väčšinou spojené so znížením ceny.

faktorov. V druhej časti budú prezentované regresné modely. Tretia časť je vyhradená modelom ARIMA a sezónnym modelom ARIMA (SARIMA). V poslednej štvrtej časti aplikujeme mikroekonomický pohľad a urobíme konštrukciu prognóz dopytu tovarov, ktoré sa predávajú v akciách.

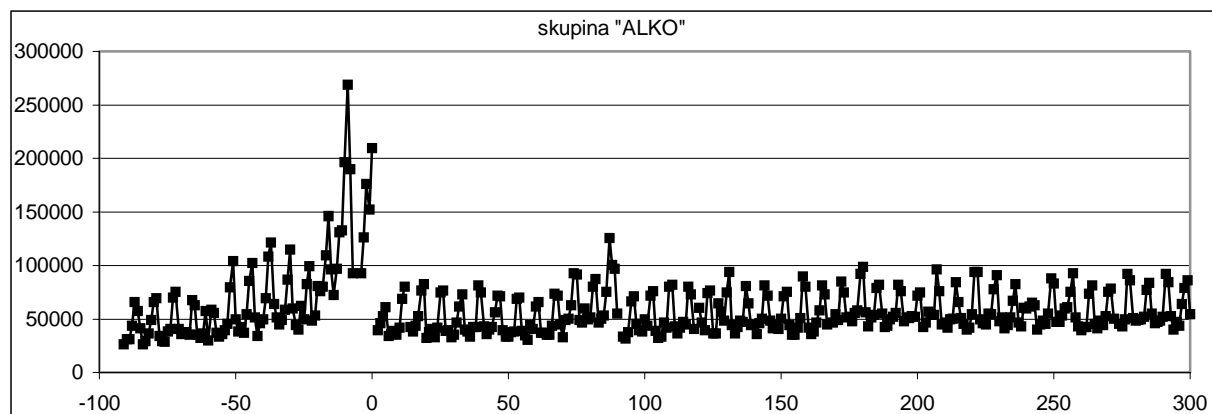
V prvých troch častiach budeme skúmať vývoj celých skupín tovarov, napríklad drogéria alebo alkoholické nápoje. Vývoj predaja konkrétnych tovarov budeme diskutovať v poslednej časti. Kým dopyt po jednotlivých tovaroch je citlivý na zmenu ceny, skupina tovarov pozostávajúca z niekoľkých desiatok až stoviek tovarov je na zmenu ceny jednotlivých tovarov necitlivá. Je to dané veľkým množstvom tovarov v skupine a možnou substitúciou medzi tovarmi. Za prognózy vývoja celých skupín tovarov hovoria aj iné argumenty. Prvým je pochopenie ďalšieho vývoja, postrehnutie dlhodobého trendu. Ďalším argumentom je spomínaná možnosť substitúcie tovarov v rámci jednej skupiny. Uvedme krátky príklad. Odhadnime dopyt po skupine nealkoholických nápojov. Pre jednoduchosť, nech je táto skupina tvorená len 2 výrobkami, napríklad Pepsi a Coca Cola. Nech predpoveď na týždeň dopredu hovorí o 100 predaných fľašiach. Objednajme 50 fliaš každého nápoja. Ak však bude skutočný dopyt po Pepsi 60 fliaš a po Coca Cole len 40 fliaš, nebude to mať až také negatívne následky. Nakoľko oba výrobky sú si navzájom dobrou náhradou, väčšinu z 10 prebytočných fliašiek Coca Coly predáme. Oveľa väčšou chybou by zrejme bolo mať na sklade spolu len 80 fliaš. Nakoľko v skupine nemusí existovať vhodné spoločné merné množstvo, dopyt po celej skupine vyjadrujeme v peňažných jednotkách.

1. KAPITOLA - dekompozícia

Uvažujme časový rad predaja skupín výrobkov. Predpokladáme, že časový rad je tvorený trendovou zložkou Tr , sezónnou zložkou Sz a náhodnou chybou E . Vzhľadom na malý počet pozorovaní nebudeme cyklickú zložku uvažovať. Kapitola sa delí na tri časti, v prvej sa budeme venovať trendovej zložke, v druhej časti krátko spomenieme infláciu, tretia časť je vyhradená pre sezónnu zložku.

1.1. Trend

Prvou zložkou, ktorou sa budeme pri dekompozícii časového radu zaoberať je trendová zložka. Práve táto nám udáva hlavný smer ďalšieho vývoja v čase. Z charakteru dát (viď obrázok 1.1) môžeme konštatovať, že v budúcnosti sa stačí zaoberať len konštantným a lineárnym trendom. Iný druh trendu (iná rýchlosť rastu / poklesu predaja) nie je podľa nás primeraný. Množstvo predaného tovaru v rámci jednej prevádzky je závislé hlavne od dvoch veličín. Je to počet zákazníkov a niečo, čo by sa dalo nazvať priemerným nákupom alebo priemerným košíkom. Nakoľko sa zaoberáme už zabehnutými prevádzkami, predpoklad, že majú svojich stálych zákazníkov sa zdá byť racionálny. Ak upustíme od vplyvu inflácie, rovnako nepredpokladáme, že by sa hodnota priemerného nákupu s časom výrazne menila. Preto upustíme od použitia exponenciálneho trendu, resp. polynomických trendov, ktoré majú rád vyšší ako 1. Isté opodstatnenie by mali tzv. S-krivky, ktoré by popisovali dopyt po nových výrobkoch. Takýmito sa však nebudeme zaoberať.



Obr. 1. 1

Na x-ovej osi grafu 1.1 je znázornený čas v dňoch (hodnota 1 zodpovedá 1.1.2002), na y-ovej osi je hodnota tržieb za celú skupinu ALKO²

Predpokladajme teraz konštantný trend a urobme jeho odhad. Minimalizujeme výraz:

$$\sum_{t=0}^n (y_t - Tr)^2 \rightarrow \min \quad (1.1)$$

kde trendová zložka je konštantná v čase, teda: $Tr_t = \beta_0$ pre $t = 1, 2, 3, \dots, n$

Metódou najmenších štvorcov dostávame odhad b_0 parametra β_0 v tvare:

$$b_0 = \bar{y} = \sum_{t=1}^n y_t / n \quad (1.2)$$

$$\text{Predikcia budúcich hodnôt je potom nemenná a nadobúda tvar: } \hat{y}_T = b_0 \quad (1.3)$$

Tvar (1-p).100% intervalu spoľahlivosti (podľa [1]) pre konštantný trend je:

$$\left(b_0 - t_{p,n-1} s \sqrt{1 + \frac{1}{n}}, b_0 + t_{p,n-1} s \sqrt{1 + \frac{1}{n}} \right) \quad (1.4)$$

kde $s = \sqrt{\sum_{t=1}^n \frac{(y_t - \bar{y})^2}{n-1}}$ a $t_{p,n-1}$ označuje kritickú hodnotu t-rozdeleniana hladine

významnosti p s $n-1$ stupňami voľnosti.

Za predpokladu, že trend je lineárny, opäť minimalizujeme výraz (1.1) s tým rozdielom, že trendová zložka má tvar: $Tr_t = \beta_0 + \beta_1 t$ pre $t = 1, 2, 3, \dots, n$

Metódou najmenších štvorcov dostávame odhady b_0 a b_1 parametrov β_0 a β_1 .

$$b_1 = \frac{\sum_{t=1}^n t y_t - \frac{n+1}{2} \sum_{t=1}^n y_t}{\sum_{t=1}^n t^2 - n \left(\frac{n+1}{2} \right)^2} \quad \text{a} \quad b_0 = \bar{y}_t - b_1 \frac{n+1}{2} \quad (1.2b)$$

$$\text{Pre predikciu platí vzťah } \hat{y}_T = b_0 + b_1 T \quad (1.3b)$$

Konfidenčný (1-p).100% interval (podľa [1]) je pre tento model vyjadrený vzťahom

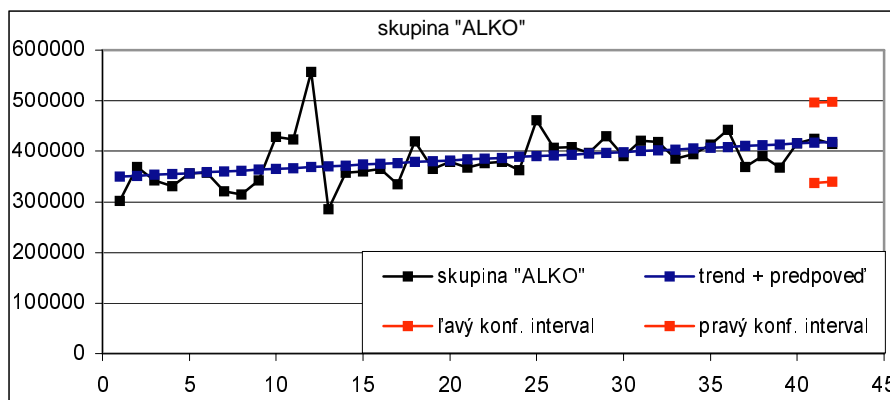
$$\left(b_0 + b_1 T - t_{p,n-2} s f_T, b_0 + b_1 T + t_{p,n-2} s f_T \right) \quad (1.4b)$$

² Skupina ALKO je skupina alkoholických nápojov. Je tvorená podskupinami PIVO, VÍNO a LIEHOVINY.

$$\text{kde } s = \sqrt{\frac{\sum_{t=1}^n y_t^2 - \sum_{t=1}^n \hat{y}_t^2}{n-2}}, \quad f_T = \sqrt{1 + \frac{1}{n} + \frac{\left(T - \frac{n+1}{2}\right)^2}{\sum_{t=1}^n t^2 - n\left(\frac{n+1}{2}\right)^2}} \text{ a } t_{p,n-2} \text{ označuje kritickú hodnotu}$$

t-rozdelenia na hladine významnosti p s $n-2$ stupňami voľnosti.

Analogicky by sme mohli postupovať k trendom vyšších rádov, z charakteru dát je však zrejmé, že lineárny trend je postačujúci. Pozrime sa, ako vyzerá predikcia budúceho vývoja na skutočných dátach. Analyzujeme týždenné dáta skupín ALKO a TABAK³ v období 7.1.2002 – 27.10.2002 (42 týždňov). Trend, posúdením grafického priebehu, volíme lineárny. Odhad urobíme bez použitia posledných 2 týždňov, tieto využijeme na kontrolu kvality predikcie. Pomocou vzorca (1.2b) určíme pre skupinu ALKO koeficienty b_0 a b_1 nasledovne: $b_0 = 348269$ a $b_1 = 1674$. Hodnota predikcie pre 41. a 42. týždeň a konfidenčné intervaly sú zobrazené v tabuľke 1.1 a na grafe.



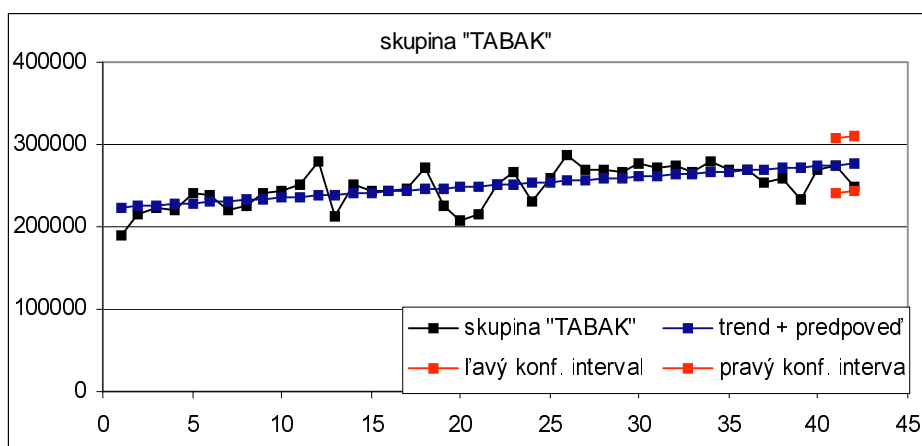
Obr. 1.2

týždeň	skutočnosť	predikcia	rozdiel	v [%]	konfidenčný interval	
					ľavý	pravý
41	423934	416889	7045	1,7	337799	495979
42	414663	418562	-3899	0,9	339191	497934

Tab. 1. 1

³ Skupina TABAK obsahuje najmä tento tovar: cigarety a iné potreby pre fajčiarov, noviny, cestovné lístky a telefónne karty.

Rovnako určíme odhady b_0 a b_1 parametrov V_{r+3} a β_1 pre skupinu TABAK. Odhady nadobúdajú hodnoty $b_0 = 221887$ a $b_1 = 1287$. Hodnota predikcie pre 41. a 42. týždeň a konfidénčné intervaly sú opäť zobrazené v tabuľke a na grafe



Obr. 1. 3

Týždeň	Skutočnosť	Predikcia	Rozdiel	v [%]	Konfidénčný interval	
					ľavý	pravý
41	274753	274663	90	0.03	241221	308104
42	248383	275950	-27567	11.1	242389	309511

Tab. 1. 2

Nebudeme sa teraz zaoberať diskusiou o adekvátnosti tohoto modelu a na tomto mieste urobíme dve poznámky. Prvá sa týka samotných dát, v 12. týždni upozorňujeme na vplyv Veľkej noci na predaj alkoholu. Druhá poznámka sa týka rastu predaja alkoholu (kladný parameter beta β_1 , hodnota t-štatistiky jeho nulovosť zamieta). Otázka, či ide o nastupujúci alkoholizmus u populácie alebo vplyv inflácie a rastúcich cien alkoholu, bude zodpovedaná neskôr.

Pristúpme teraz k trendovej zložke z inej strany, pokúsime sa aplikovať adaptívny prístup. Zástupcami adaptívneho prístupu sú napríklad kľzavé priemery. Výhodou kľzavých priemerov je schopnosť pracovať s dátami, ktoré v čase výrazne menia svoj charakter. Použitie matematických kriviek s nemennými parametrami je na popis takéhoto druhu dát obtiažne. Predpokladáme však, že na krátkych časových úsekoch je popis vhodnou matematickou krivkou (s konštatntnými parametrami) možný, i keď na odlišných intervaloch sú parametre rôzne. Časový rad po úsekoch vyrovnávame preložením polynomickej krivky.

Vychádzame z toho, že väčšina funkcií môže byť polynómom dobre aproximovaná. Ukážeme, že pri vyrovnávaní časového radu ide v podstate o vytváranie lineárnej kombinácie členov pôvodného časového radu. Samotnú konštrukciu urobíme nasledovne. Cez nepárny počet susedných členov časového radu preložíme polynóm a hodnotu prostredného člena vyrovnáme hodnotou polynómu v tomto bode. Otvorená zostáva otázka voľby dĺžky a rádu kĺzavého priemeru, vyrovnanie krajných hodnôt a tvorba prognóz.

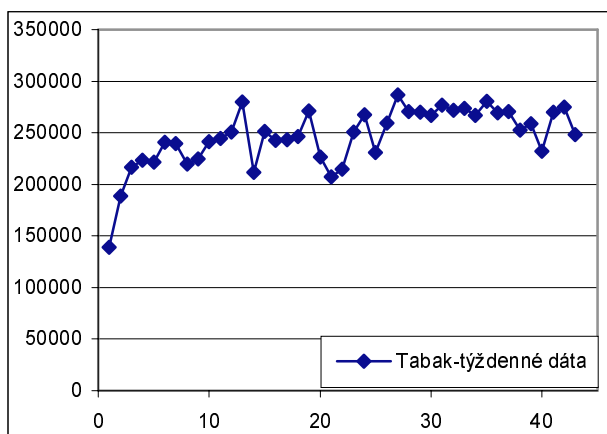
S rastúcou dĺžkou kĺzavého priemeru sa časový rad stáva vyhladenejší. Týmto spôsobom sa odfiltrujú krátkodobé fluktuácie. Pri neperiodických časových radoch sa snažíme použiť kĺzavé priemery dĺžky, ktorá zodpovedá cyklu, ktorý chceme z radu odfiltrovať. Pri periodických časových radoch volíme dĺžku kĺzavého priemeru tak, aby zodpovedala perióde sezónnych výkyvov.

Odpoveď na otázku voľby rádu poskytuje objektívne kritérium (detailne opísané v [1]). Nech časový rad y_t je tvorený trendom a náhodnou chybou, kde trend Tr_t je polynómom rádu r a kde E_t je biely šum ($y_t = Tr_t + E_t$).

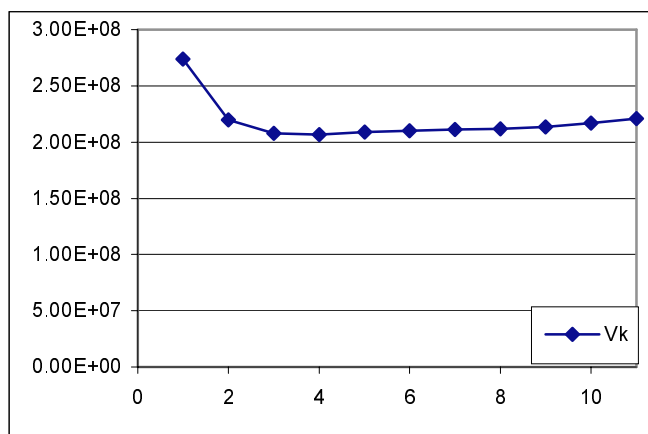
$$\text{Zavádzame výrazy } V_k = \frac{\sum_{t=k+1}^n (\Delta^k y_t)^2}{\binom{2k}{k} (n-k)} \quad (1.5)$$

kde $\Delta^k y_t$ označuje k -tu diferenciu. Pre $k \geq r+1$ je V_k odhadom rozptylu σ^2 . Vypočítame teda odhady V_k až do okamihu, kým hodnoty nezačnú konvergovať k nejakej konštante. Ak sú hodnoty $V_{r+1} V_{r+2} V_{r+3} \dots$ blízko konštanty, určíme za rád kĺzavého priemeru práve hodnotu r .

Aplikujme vyššie spomínaný postup na skupinu TABAK za obdobie od 1.1.-31.10.2002 (týždenné dáta, 43 týždňov) a určíme rád kĺzavých priemerov. Na grafe 1.4 vidíme samotné dáta, na x-ovej osi je poradové číslo týždňa roku 2002, na y-ovej osi je hodnota tržieb. Na grafe 1.5 vidíme prvých 11 hodnôt V_k , z nich určíme rád kĺzavého priemeru 3.



Obr. 1.4



Obr. 1.5

Prejdime k samotnej eliminácii trendovej zložky. Nebudeme uvažovať posledné 3 týždne, tieto použijeme na porovnanie s prognózou. Čo sa týka dĺžky kľavého priemeru, overíme niekoľko možných dĺžok. Koefficienty vyrovnávajúceho polynómu určíme metódou najmenších štvorcov, minimalizujeme výraz:

$$\sum_{\tau=-m}^m (y_{t+\tau} - (\beta_0 + \beta_1\tau + \beta_2\tau^2 + \beta_3\tau^3))^2 \rightarrow \min \quad (1.6)$$

Po príslušných úpravách dostávame systém rovníc, z ktorého určíme odhady b_0, b_1, b_2, b_3 parametrov $\beta_0, \beta_1, \beta_2, \beta_3$:

$$\sum_{\tau=-m}^m y_{t+\tau}\tau^j - b_0 \sum_{\tau=-m}^m \tau^j - b_1 \sum_{\tau=-m}^m \tau^{j+1} - b_2 \sum_{\tau=-m}^m \tau^{j+2} - b_3 \sum_{\tau=-m}^m \tau^{j+3} = 0 \quad \text{pre } j=0,1,2,3 \quad (1.7)$$

Ak uvažujeme kľavé priemery dĺžky 5 ($m = 2$), po dosadení a úprave dostávame pre odhady b_0, b_1, b_2, b_3 :

$$b_0 = \frac{1}{35} (17 \sum_{\tau=-2}^2 y_{t+\tau} - 5 \sum_{\tau=-2}^2 \tau^2 y_{t+\tau}) = \frac{1}{35} (-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}) \quad (1.9)$$

$$b_1 = \frac{1}{72} (65 \sum_{\tau=-2}^2 \tau y_{t+\tau} - 17 \sum_{\tau=-2}^2 \tau^3 y_{t+\tau})$$

$$b_2 = \frac{1}{14} (\sum_{\tau=-2}^2 \tau^2 y_{t+\tau} - 2 \sum_{\tau=-2}^2 y_{t+\tau}) \quad (1.10)$$

$$b_3 = \frac{1}{72} (5 \sum_{\tau=-2}^2 \tau^3 y_{t+\tau} - 17 \sum_{\tau=-2}^2 \tau y_{t+\tau})$$

Pre vyrovnanie hodnôt od $m+1$ až po $n-m$ je zaujímavý len odhad b_0 to je hodnota polynómu v prostrednom bode a zároveň vyrovnaná hodnota časového radu

$$\hat{y}_{t+k} = b_0 + b_1k + b_2k^2 + b_3k^3 = b_0 \quad \text{kde } m+1 \leq t \leq n-m \quad k=0 \quad (1.11)$$

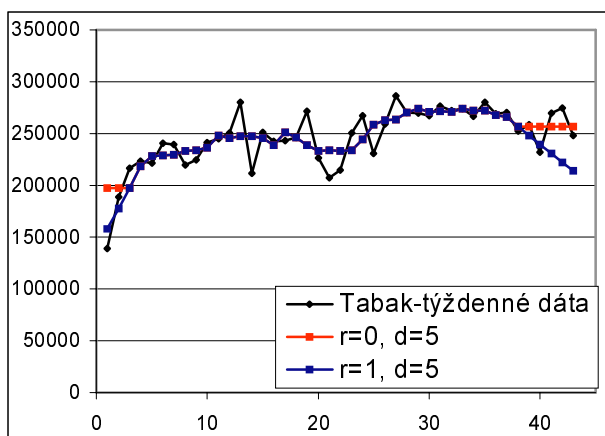
Vyrovnanie posledných m hodnôt urobíme nasledovne (postup pre prvých m hodnôt je analogický, rovnaký je postup aj pre konštrukciu predpovede). Cez posledných $2m+1$ hodnôt sa preloží polynóm (v tomto prípade 3. rádu). Vyrovnané hodnoty v posledných m bodoch (v našom prípade $m=2$) vypočítame ako

$$\hat{y}_{n-m+k} = b_0 + b_1k + b_2k^2 + b_3k^3 \quad \text{kde } k=1,2 \quad (3,4,\dots,m) \quad (1.12)$$

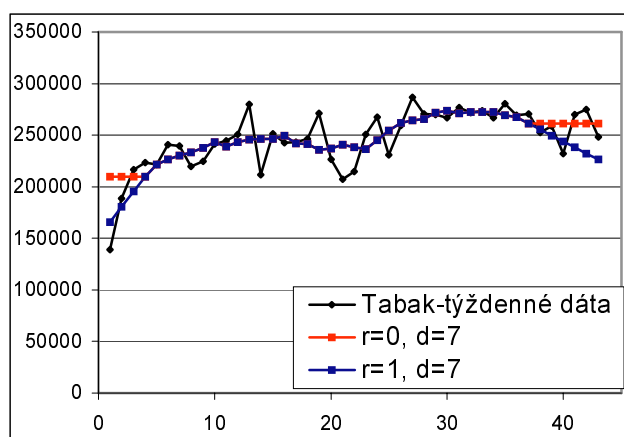
kde odhady b_0, b_1, b_2, b_3 dostaneme zo vzťahov (1.9) a (1.10).

Veľmi jednoducho vypočítame aj prognózu ďalšieho vývoja, v (1.12) položíme $k=3,4,\dots$ (predpoveď na jeden krok dopredu, dva kroky dopredu atď). Vyrovnanie prvých hodnôt urobíme analogicky.

Analyzujme opäť týždenné dáta skupiny TABAK. Použijeme kľzavé priemery rôznych rádov a dĺžok. Ku každému priemeru je pridaná predpoveď na obdobie troch týždňov dopredu (posledné tri hodnoty v grafe). Na grafoch je na x-ovej osi znázornený čas - poradové číslo týždňa v roku 2002, na y-ovej osi je celková tržba. Písmenom r označujeme rád kľzavého priemeru, písmenom d dĺžku kľzavého priemeru.

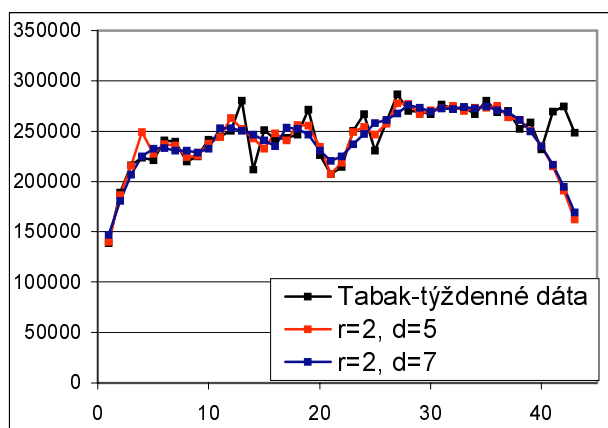


Obr. 1. 6

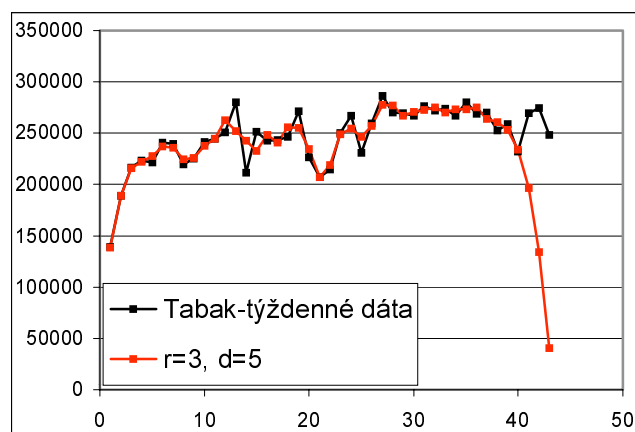


Obr. 1. 7

Pridávame tabuľky s predpoveďami na 3 týždne dopredu, spolu s hodnotami uvádzame aj absolútnu a relatívnu chybu. Vidíme pomerne dobré predpovede pri vyhladzovaní aritmetickým priemerom (percentuálny rozdiel oproti skutočnosti je na úrovni cca 5%), kvalita sa zhoršuje s rastúcim rádom polynómu, ktorým vyhladzujeme. Predpovede kľzavých priemerov väčšej dĺžky boli ku skutočným hodnotám bližšie.



Obr. 1.8



Obr. 1.9

skutočnosť	r=0,d=5		r=0, d=7		r=1,d=5		r=1, d=7	
týždeň hodnota	predpoveď	rozdiel v [%]	predpoveď	rozdiel v [%]	predpoveď	rozdiel v [%]	predpoveď	rozdiel v [%]
41 269660	256557	13103 5%	261417	8243 3%	230991	38669 15%	238046	31614 12%
42 274753	256557	18196 7%	261417	13336 5%	222469	52284 20%	232204	42549 17%
43 248383	256557	-8174 3%	261417	-13033 5%	213947	34436 13%	226361	22022 9%

skutočnosť	r=2, d=5			r=2, d=7			r=3, d=5		
týždeň hodnota	predpoveď rozdiel v [%]			predpoveď rozdiel v [%]			predpoveď rozdiel v [%]		
41 269660	215191	54469	20	216566	53094	20	196270	73390	27
42 274753	190869	83884	31	194614	80139	29	134107	140646	51
43 248383	162033	86350	35	169081	79302	32	40399	207984	84

Tab. 1.3

Môžeme konštatovať, že priemery vyšších rádoov ($r = 2, r = 3$) sa nehodia na konštrukciu predpovedí. Odpoveď na otázku, prečo pri vyrovnávaní polynómami vyšších rádoov nedosahujeme také dobré predpovede ako pri polynómoch nižších rádoov ($r = 0, r = 1$) je pomerne jednoduchá. Pracujeme s dátami, ktoré vykazujú konštantný, resp. lineárny trend. Je prirodzené, že ak cez posledných $2m + 1$ bodov preložíme polynóm vyššieho rádu a trend je konštantný, resp. lineárny, tak rozdiel medzi predikovanými a skutočnými hodnotami sa s časom (vzdialenosťou od stredného, t.j. m -tého bodu) zväčšuje úmerne rádu polynómu.

1.2 Inflácia

Takmer všetky skúmané skupiny tovarov dlhodobovo vykazujú približne lineárny trend rastu, ak skúmame celkové tržby skupín v bežných cenách. Iste je namieste otázka, do akej miery je rast tržieb spôsobený rastom spotreby a do akej miery infláciou a teda novými vyššími cenami. Túto otázku budeme z viacerých dôvodov diskutovať len okrajovo. Prvým dôvodom je skutočnosť, že primárnym cieľom je konštrukcia predpovedí za účelom optimálneho zásobenia skladu hypermarketu. Tieto predpovede sa robia len na krátke obdobie, zvyčajne nie dlhšie ako je doba medzi dvoma dodávkami tovaru. V praxi ide o obdobie niekoľkých dní až niekoľkých týždňov. Tento fakt je daný pomerne rýchlou obrátkovosťou väčšiny tovarov. Druhým dôvodom je skutočnosť, že nie je rozdiel v tom, či

predikované množstvo tovaru určíme v nových (bežných) cenách z pôvodného trendu aj za predpokladu, že je zafarbený infláciou, alebo ho určíme v starej cenovej hladine pomocou trendu, odvodeného od pôvodného pomocou cenových deflátorov pre danú skupinu. Cenové deflátoary sú však dostupné len pre skupiny „potraviny a nealkoholické nápoje“ a „alkohol a tabak“.

Analyzujeme teraz dáta skupín ALKO a TABAK, ktoré upravíme pomocou indexov spotrebiteľských cien. Keďže indexy spotrebiteľských cien sa vzťahujú na obdobie jedného mesiaca, agregujeme denné dáta na časové obdobie 1 mesiac. Pomocou cenových deflátorov prepočítame mesačné tržby k dátumu 1.1.2002. Iste by sa dala ešte vykonať ďalšia úprava dát, napríklad normalizácia na štandardný (30 denný) mesiac. Väčší vplyv na objem predaného množstva ako rozdiel medzi 30 a 31 dňovými mesiacmi má počet piatkov a sobôt v danom mesiaci. Neskôr ukážeme, že najväčší objem obchodu sa realizuje práve počas týchto dní. Dáta teda ďalej neupravujeme.

Podľa údajov ŠÚ SR ceny alkoholu a tabaku za obdobie prvých 10 mesiacov roku 2002 stúpili o 2,2% (v porovnaní s januárom 2002). Vývoj celkového mesačného obratu v bežných cenách a v stálych cenách je pre každú skupinu zobrazený v grafe 1.10 a 1.11. Uvažujme lineárny trend ($Tr_t = \beta_0 + \beta_1 t$). Odhady b_0 a b_1 parametrov β_0 a β_1 sú uvedené v tabuľke 1.4 spolu s hodnotou t-štatistiky pre odhad b_1 .

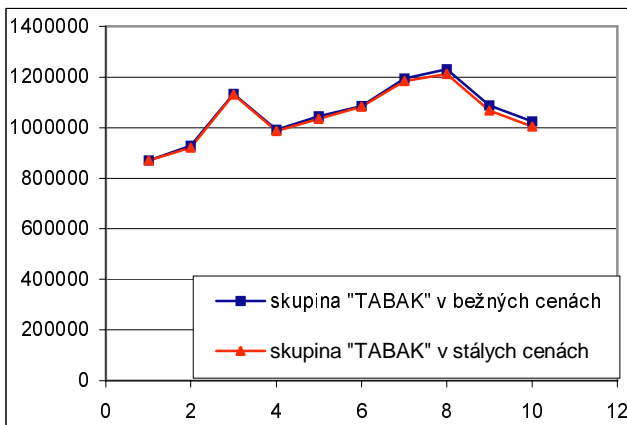
	skupina "ALKO"		skupina "TABAK"	
	bežné ceny	stále ceny	bežné ceny	stále ceny
b_0	1549763	1552727	937344	940327
b_1	18465	13661	22174	19823
t-stat	1.028	0.924	2.102	1.858

Tab. 1.4

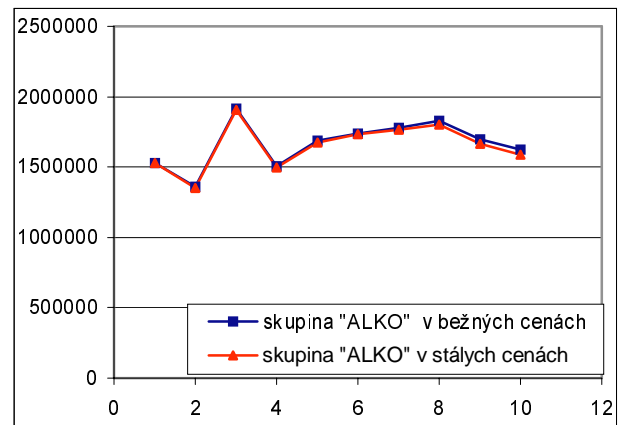
Pri stálych cenách vidíme v oboch skupinách pokles hodnoty parametra b_1 . Tento by mohol nasvedčovať tomu, že časť rastu tržieb ide na vrub inflácie a časť na vrub vyššej spotreby. Je tu však nízka hodnota t-štatistiky pre odhad b_1 parametra β_1 , ktorá nezamieta jeho nulovosť. Nízka hodnota t-štatistiky bude zrejme zapríčinená nízkym počtom hodnôt a tým vysokou disperziou odhadu. Pre veľký rozptyl odhadu b_1 je potrebné interpretovať zníženie

parametra b_1 s dávkou opatrnosti. Skúmať vplyv inflácie na rast tržieb by sme mohli v prípade, že by bol k dispozícii dlhší časový rad.

Na záver môžeme konštatovať, že nízka miera inflácie nepredstavuje pre krátku predpoveď problém, pokiaľ ide o celé skupiny. Pri jednotlivých tovaroch však už aj malá zmena ceny, ako ukážeme neskôr, môže zapríčiniť výraznú zmenu dopytu.



Obr. 1.10



Obr. 1.11

1.3 Sezónnosť

Nemenej dôležitou časťou dekompozície je určenie sezónnosti. Na trendovej zložke môže byť nabalene niekoľko druhov sezónnych vplyvov (tržba vo februárový pondelok môže vyzeráť inak ako tržba v decembrový pondelok a už úplne inak ako na Veľkonočný pondelok). Vzhľadom na krátkosť nášho časového radu, kde máme len 400 denných údajov, niektoré sezónne vplyvy nedokážeme určiť a odfiltrovať. Ide najmä o javy pravidelne sa opakujúce každý rok, teda vplyv jednotlivých mesiacov a vplyv sviatkov ako napríklad Veľká noc a Vianoce. V našom prípade sa sústreďujeme hlavne na určenie vplyvu jednotlivých dní. Sezónnu zložku v časovom rade sa pokúsime odfiltrovať pomocou kľzavých priemerov. Môžeme voliť medzi aditívnym a multiplikatívnym modelom. Na predpovede budeme používať aditívny model z toho dôvodu, že dosahoval lepšie hodnoty predpovedí. Predpokladáme aditívny model, uvažujeme teda časový rad v tvare:

$$y_t = Tr_t + Sz_t + E_t \quad (1.13)$$

kde Tr_t je trendová zložka v čase t , Sz_t je sezónna zložka v čase t a E_t je chyba v čase t . Po vyrovnaní časového radu a odfiltrovaní trendovej zložky dostávame vzťah:

$$y_t - Tr_t = Sz_t + E_t \quad (1.14)$$

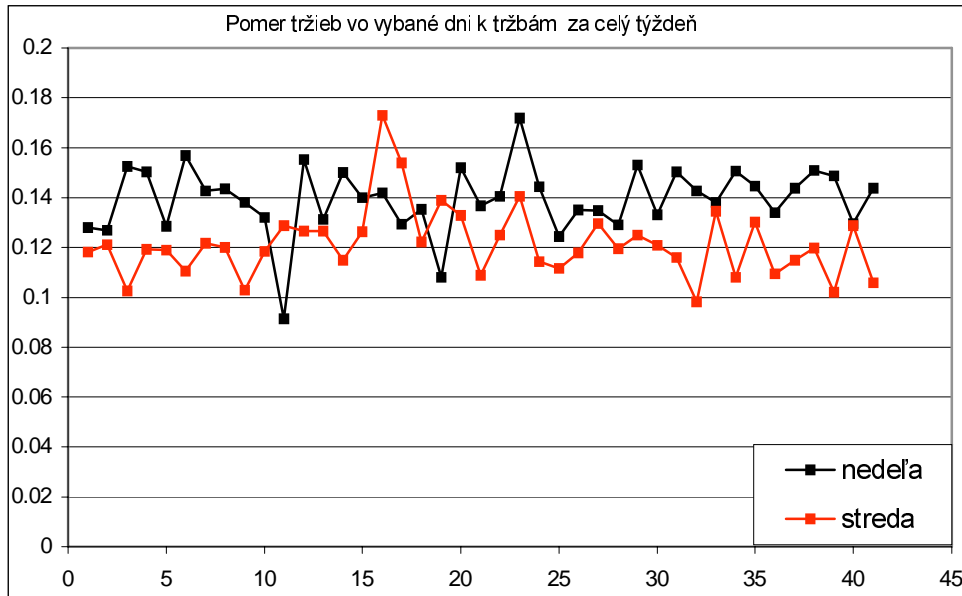
Pri multiplikatívnom modeli použijeme namiestorozdielu hodnôt y_t a Tr_t ich podiel a ďalej postupujeme analogicky ako pri aditívnych modeloch. To, čo zostalo po odfiltrovaní trendu, by mala byť sezónna zložka zafarbená už len náhodnou chybou. Vzhľadom na krátke obdobie predpovede najvýznamnejším „sezónnym faktorom“ bude vplyv jednotlivých dní. Zavedieme označenie $Sz_{1+7j}, \dots, Sz_{7+7j}$ pre „sezónny vplyv“ pondelka až nedele, kde j je číslo periódy, v našom prípade číslo týždňa v roku. Pre určenie niektorých ďalších sezónnych vplyvov, napríklad vplyv niektorých sviatkov, dní voľna a niektorých špeciálnych dní by sme potrebovali dlhší časový rad. Pod pojmom „špeciálne dni“ budeme ďalej rozumieť dni, ktoré výrazne ovplyvňujú predaj, napríklad sviatky Veľká noc a Vianoce, dni pracovného voľna pripadajúce na pondelok až piatok, dni typu Mikuláš a Sv. Valentín, ale aj začiatok prázdnin či posledný deň roka. Špeciálne dni sa môžu pre rôzne skupiny líšiť. Kým napríklad Mikuláš nemá vplyv na predaj alkoholu, na predaj cukrovín je jeho vplyv výrazný.

Elimináciu chyby zo zložky $Sz_{i+7j} + E_{i+7j}$ môžeme urobiť dvoma spôsobmi. Prvým spôsobom je obyčajný aritmetický priemer jednotlivých sezónnych vplyvov (aritmetický priemer všetkých pondelkov, utorkov atď). Pomocou neho určíme hodnotu:

$$\bar{S}z_i = \frac{1}{n} \sum_{j=0}^{n-1} (Sz_{i+7j} + E_{i+7j}) \quad (1.15)$$

kde n je počet týždňov, i je deň týždňa ($i = 1$ pondelok, $i = 7$ nedeľa) a j je číslo týždňa. Hodnotu $\bar{S}z_i$ budeme považovať za sezónny vplyv i -teho dňa. Vzťah sa opiera o predpoklad, že stredná hodnota chyby je nulová. Druhou možnosťou je preložiť hodnoty $Sz_{i+7j} + E_{i+7j}$ matematickou krivkou a sledovať tak, či sa napríklad s časom nemení vplyv jednotlivých sezón (dní). V našom prípade sa však nepreukázalo, že by popularita nákupov v niektorých

dňoch týždňa rástla na úkor iných dní, zostávame preto pri aritmetickom priemere. Na grafe 1.12 je znázornený pomer tržieb vo vybrané dni k tržbám za celý týždeň, na x-ovej osi je číslo týždňa roku 2002.



Obr. 1.12

Urobme teraz kľzavé priemery a odhadnime sezónny vplyv jednotlivých dní postupom uvedeným vyššie. Nakoľko nám ide len o odhad sezónneho vplyvu, nebudeme sa zaoberať vyrovnávaním krajných hodnôt. Dĺžku kľzavého priemeru volíme tak, aby zodpovedala perióde sezónnych výkyvov. V našom prípade bude dĺžka kľzavého priemeru rovná dĺžke týždňa, rád kľzavého priemeru odôvodníme neskôr. Predpoveď budeme konštruovať tak, že cez hodnoty radu očistené od sezónnych vplyvov ($y_t - \bar{S}_t$) preložíme vhodnú matematickú krivku, v našom prípade vystačíme s priamkou. Túto predĺžime do „budúcnosti“, kde opätovne aplikujeme vplyv sezónnych faktorov. Predpoveď nadobúda tvar:

$$\hat{y}_{t+7 \cdot j+i} = \hat{T}_{t+7 \cdot j+i} + \bar{S}_{t+i} \quad (1.16)$$

kde j označuje číslo týždňa a i označuje deň v týždni pre ktorý je predpoveď zostrojená, $\hat{T}_{t+7 \cdot j+i}$ označuje predpoveď trendovej zložky a \bar{S}_{t+i} označuje predpoveď sezónneho faktora v čase $t + 7 \cdot j + i$ urobenú v čase t .

Čo sa týka rádu kĺzavých priemerov, budeme používať len kĺzavé priemery nultého rádu. Priemery vyšších rádov do trendovej zložky zahŕňajú väčšiu časť sezónnych vplyvov ako kĺzavé priemery nultého rádu, preto je tento odhad sezónnosti v predpovediach (viď vyššie) nepoužiteľný. Pred samotným odhadom musíme ešte riešiť otázku dní, kedy bola prevádzka zatvorená a hodnota tržieb dôsledkom toho nulová. Môžeme túto „dieru“ v časovom rade istým spôsobom „zalepiť“ alebo nechať hodnotu tržieb nulovú. Zaoberajme sa možnosťou, že necháme hodnotu tržieb nulovú. Vyrovnaná hodnota trendovej zložky Tr_t pomocou kĺzavých priemerov bude určite kladná (hodnota priemeru tržieb 3 predchádzajúcich a 3 nasledujúcich dní). Sezónny vplyv plus chyba v tomto dni, reprezentovaný rozdielom $y_t - Tr_t$, bude potom vždy záporný (lebo $y_t = 0$) a jeho hodnota bude od skutočnej hodnoty vzdialená. Táto chyba sa prejaví aj na aritmetickom priemere, ktorým budeme odhadovať celkový sezónny vplyv toho ktorého dňa. Naviac ak cez hodnoty radu očisteného od sezónnych vplyvov ($y_t - Sz_t$) preložíme matematickú krivku za účelom prognózy, dni kedy bola prevádzka zatvorená budú outliermi. Preto sa skôr prikláňame k možnosti túto „dieru“ zalepiť, napríklad hodnotou $\tilde{y}_t = (y_{t-7} + y_{t+7})/2$. Iste príde aj takto k určitému vychýleniu sezónnych vplyvov, sme však toho názoru, že chyba bude menšia a rovnako sa vyhneme aj problémom s outliermi.

Analyzujeme dáta za obdobie 15.4.2002-27.10.2002 skupín ALKO, TABAK, CUKROVINKY a PEKÁRENSKÉ VÝROBKÝ⁴. Dni, počas ktorých bola prevádzka zatvorená, boli ošetrené vyššie spomenutou metódou. Posledných sedem dní nebudeme používať na odhad sezónneho vplyvu, využijeme ich len na kontrolu predpovede. Sezónne vplyvy jednotlivých dní boli pre tieto skupiny a aditívny model určené nasledovne:

ADIT	ALKO	TABAK	CUKOR	PEKÁRENSKE VÝROBKÝ
PONDELOK	-13353	-7201	-12562	-2549
UTOROK	-9910	-4998	-7812	-642
STREDA	-7688	-5595	-7012	-1851
ŠTVRTOK	-4481	-1944	-1387	-229
PIATOK	23733	8371	25331	12107
SOBOTA	21895	12610	18054	7972
NEDEĽA	-10388	-1158	-14374	-14701

⁴ Skupina CUKROVINKY je tvorená najmä podskupinami ČOKOLÁDA, LÍZATKÁ a CUKRÍKY, DEZERTY. Skupina PEKÁRENSKÉ VÝROBKÝ je tvorená najmä podskupinou CHLIEB a PEČIVO.

PRIEMER	57040	37171	68420	47617
---------	-------	-------	-------	-------

Tab. 1.5

Hodnota PRIEMER určuje priemernú hodnotu tržieb za spomínané obdobie. Hodnota je uvedená len na porovnanie so sezónnymi vplyvmi. Nakoniec by mali byť jednotlivé sezónne vplyvy normované tak, aby ich súčet bol 0. Nakoľko sa však celkový súčet všetkých sezónnych vplyvov líši od 0 zanedbateľne, túto procedúru vynecháme. Pre úplnosť ešte pridávame hodnoty multiplikatívneho modelu (normované):

MULTI	ALKO	TABAK	CUKOR	PEKÁRENSKE VÝROBKY
PONDELOK	0.767	0.805	0.819	0.947
UTOROK	0.826	0.866	0.885	0.985
STREDA	0.865	0.850	0.901	0.961
ŠTVRTOK	0.921	0.945	0.985	0.995
PIATOK	1.419	1.228	1.368	1.255
SOBOTA	1.384	1.339	1.254	1.165
NEDEĽA	0.818	0.968	0.788	0.691

Tab. 1.6

Môžeme konštatovať to, čo poznáme zo skutočného života. Najviac nákupov sa uskutočňuje v piatok a sobotu. Zaujímavá je skutočnosť, že obrat za piatok a sobotu bežne dosahuje pri niektorých skupinách 40% celkového obratu skupiny. Na druhej strane vidíme, že pri niektorých skupinách je vplyv jednotlivých dní rozložený rovnomernejšie, napríklad skupina pekárenských výrobkov. Za týmto rozdelením je potrebné hľadať dva hlavné dôvody. Prvým je rozdelenie spotreby počas týždňa. Kým spotreba niektorých výrobkov je rozdelená rovnomerne počas celého týždňa, napríklad už spomínané pekárske výrobky, iné výrobky majú spotrebu nerovnomernejšiu, napríklad alkohol počas víkendu (rôzne oslavy). Druhým dôvodom je možnosť nakúpiť výrobok do zásoby. Pri niektorých výrobkoch neexistuje možnosť kúpiť tovar na dlhšie obdobie dopredu a vyhnúť sa tak opätovnej ceste do obchodu. Ide napríklad o mliečne výrobky alebo ovocie. Iné výrobky sa dajú kúpiť v množstvách, ktoré vystačí na dlhšiu dobu, napríklad káva a drogistický tovar. Preto je prirodzené, že nákup týchto výrobkov sa ponechá na veľký piatkový alebo sobotný nákup, keď rodina ide nakupovať povedzme autom.

Čo sa týka predpovedí, urobíme krátku predpoveď postupom uvedeným vyššie. Výsledok je v tabuľke 1.7 a tiež v grafoch v 2. kapitole spolu s hodnotami predpovedí získanými z regresie.

SKUPINA	ALKO				TABAK				
	DEŇ	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]
PONDELOK		40025	46620	6595	16.5	28127	32203	4076	14.5
UTOROK		47108	50094	2986	6.3	36576	34429	-2147	-5.9
STREDA		43451	52347	8896	20.5	26280	33856	7576	28.8
ŠTVRTOK		64183	55586	-8597	-13.4	32033	37529	5496	17.2
PIATOK		79123	83832	4709	6.0	40389	47867	7478	18.5
SOBOTA		86383	82025	-4358	-5.0	49253	52130	2877	5.8
NEDEĽA		54391	49774	-4617	-8.5	35725	38384	2660	7.4
CELKOM		414664	420278	5614	1.4	248383	276399	28016	11.3

DEŇ	CUKROVINKY				PEKÁRENSKÉ VÝROBKY			
	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]
PONDELOK	56931	54569	-2362	-4.1	48525	47032	-1493	-3.1
UTOROK	63786	59320	-4465	-7.0	50027	48962	-1065	-2.1
STREDA	60998	60121	-877	-1.4	46585	47775	1190	2.6
ŠTVRTOK	77672	65747	-11925	-15.4	53012	49420	-3592	-6.8
PIATOK	102958	92466	-10492	-10.2	63546	61778	-1768	-2.8
SOBOTA	98096	85189	-12907	-13.2	61918	57666	-4252	-6.9
NEDEĽA	68003	52762	-15241	-22.4	39347	35015	-4332	-11.0
CELKOM	528443	470175	-58268	-11.0	362960	347649	-15311	-4.2

Tab. 1.7

V kolonke „SKUTOČNOSŤ“ sú uvedené skutočné tržby, v kolonke „ODHAD“ sú uvedené hodnoty predikovaných tržieb, „ROZDIEL“ je hodnota predikcie mínus skutočná hodnota, záporné znamienko označuje deficit na sklade. Kolonka „CELKOM“ obsahuje agregované hodnoty za celý týždeň.

Môžeme konštatovať, že niektoré skupiny sa nám podarilo predpovedať lepšie, iné s menším úspechom, celkovo sme však získali určitý prehľad o vplyve jednotlivých dní. Ak by sme porovnali predpoveď na týždeň dopredu zloženú zo súčtu predpovedí na 7 dní

(založenú na denných dátach) s predpoveďou založenou na týždenných dátach, vidíme len mierne zhošenie. Celkovo by sme mohli konštatovať, že tvorba prognóz pomocou kľzavých priemerov nie je najhoršou cestou.

2. KAPITOLA – regresný prístup

Venujme sa teraz prognózam a problému sezónnych faktorov z iného pohľadu. Na začiatku urobíme konštrukciu regresného modelu, ktorý bude využívať nasledujúce vystvetľujúce premenné: čas (trend) a 6 binárnych premenných pre jednotlivé dni týždňa. Neskôr urobíme ďalšie úpravy. Použijeme metódu obyčajných najmenších štvorcov. Predpokladajme model:

$$y = X\beta + \varepsilon \quad (2.1)$$

(resp. $y_t = \beta_0 + \beta_1 t + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t} + \beta_6 x_{6,t} + \beta_7 x_{7,t} + \varepsilon_t$ pre $t = 1, 2, \dots, n$)

kde y je vektor tržieb, X je matica plánu pozostávajúca z konštanty, časti reprezentujúcej čas (lineárny trend) a z časti obsahujúcej 6 binárnych premenných pre dni utorok až nedeľa, β je vektor parametrov a ε je reziduálna zložka. Zároveň predpokladáme:

$E(\varepsilon) = 0$	Stredná hodnota chýb je nulová
$Var(\varepsilon) = \sigma^2 I$ $Cov(\varepsilon_i, \varepsilon_j) = 0$	Homoskedasticita a neautokorelovanosť
$E(X^T \varepsilon) = 0$	X je nestochastická
$\varepsilon \sim N(0, \sigma^2 I)$	Normálne rozdelenie reziduálov

Minimalizujeme výraz:

$$\sum_{t=1}^n (y_t - (\beta_0 + \beta_1 t + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t} + \beta_6 x_{6,t} + \beta_7 x_{7,t}))^2 \rightarrow \min \quad (2.2)$$

Po príslušných úpravách dostávame odhad \hat{b} parametrov β v tvare:

$$\hat{b} = (X^T X)^{-1} X^T y \quad (2.3)$$

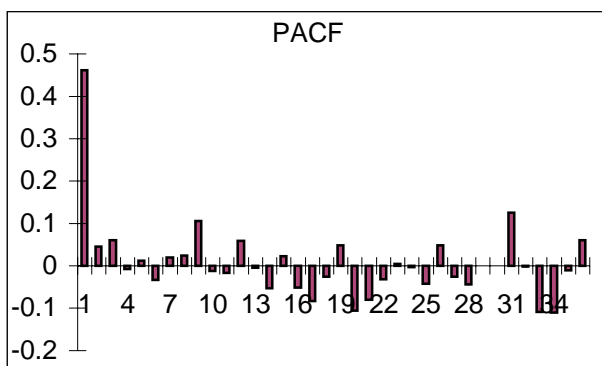
pre ktorý za platnosti predpokladov platí:

$$\hat{b} \sim N(\beta, \sigma^2 (X^T X)^{-1}) \quad (2.4)$$

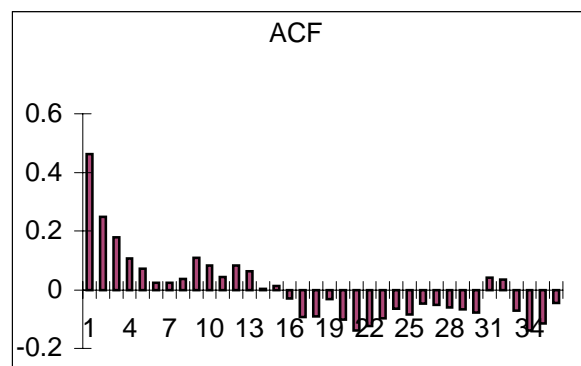
Urobme konštrukciu regresného modelu (2.1) pre skupinu ALKO za obdobie 3.1.-27.10.2002. Vektor tržieb y (ALKO02) bude pozostávať len z nenulových tržieb, dni kedy bola prevádzka zatvorená nebudeme uvažovať, patrične upravíme aj maticu plánu X . Okrem odhadu metódou najmenších štvorcov uvádzame aj grafy hodnôt parciálnej autokorelačnej funkcie PACF a autokorelačnej funkcie ACF reziduálov, rovnako ako aj graf skutočných a fitovaných hodnôt a reziduály. Na x-ovej osi je nanesený čas, na y-ovej osi hodnota tržieb.

Dependent Variable: ALKO02				
Included observations: 294				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36452.39	1681.493	21.67858	0.0000
TIME	36.51814	6.066187	6.019950	0.0000
UT	2583.676	1975.721	1.307713	0.1920
ST	5237.147	1964.733	2.665577	0.0081
SV	10636.03	1964.738	5.413457	0.0000
PI	36596.85	1964.763	18.62660	0.0000
SO	36554.22	1964.805	18.60450	0.0000
NE	3280.272	1976.139	1.659940	0.0980
R-squared	0.753843	Mean dependent var		55697.08
Adjusted R-squared	0.747818	S.D. dependent var		17572.35
S.E. of regression	8824.436	Akaike info criterion		21.03527
Sum squared resid	2.23E+10	Schwarz criterion		21.13550
Log likelihood	-3084.185	F-statistic		125.1227
Durbin-Watson stat	1.075697	Prob(F-statistic)		0.000000

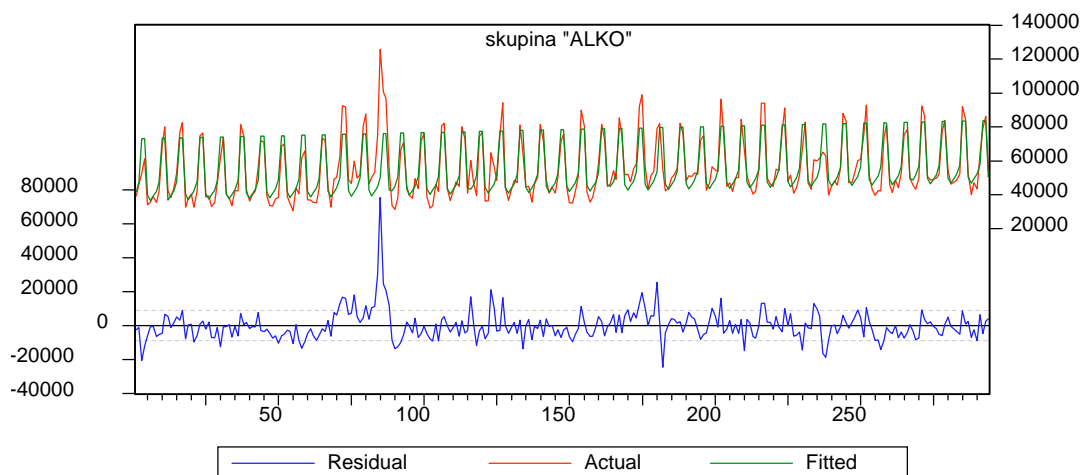
Tab. 2.1



Obr. 2. 1



Obr. 2.2



Obr. 2. 3

Vidíme pomerne dobrý fit, R^2 je približne na úrovni 75%. Hodnota t-štatistiky hovorí o nevýznamnosti niektorých premenných (utorok, nedeľa). Problémom je aj hodnota Jarque-Bera štatistiky, ktorá zamieta normálne rozdelenie reziduálov. O nevhodnosti modelu tiež hovorí hodnota Durbin-Watsonovej štatistiky, ktorá poukazuje na problém autokorelácie. Aj vizuálnym posúdením hodnôt autokorelačnej a parciálnej autokorelačnej funkcie reziduálov vidíme, že zrejme ide o autoregresný proces 1. rádu. Autokorelačná funkcia ACF a parciálna autokorelačná funkcia PACF sú diskutované v ďalšej časti tejto práce (3. KAPITOLA - modely ARIMA a SARIMA).

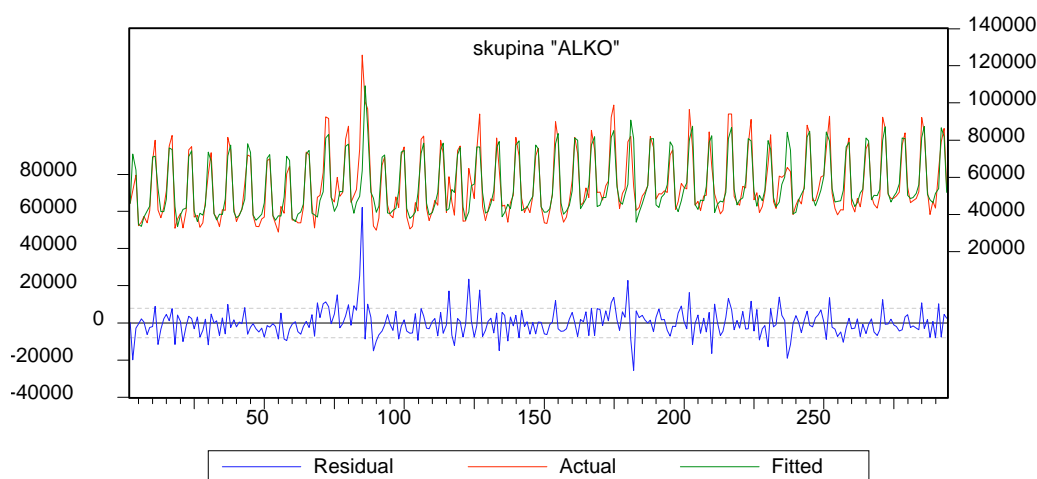
Na grafe tržieb vidíme niekoľko špeciálnych dní, ktoré by bolo možné ošetriť, napríklad pomocou dummy premenných. Najväčšia hodnota reziduálov sa objavuje pri Veľkej noci, tesne pred ňou by zvýšená hodnota tržieb mohla napovedať niečo o istom predzásobovaní sa alkoholom, alebo tiež o tom, že 19.3. slávi meniny Jozef. Ďalšie mierne zvýšené hodnoty reziduálov vidíme pri dňoch 1. a 8. máj. Neskôr vidíme na grafe koniec (stredoškolského) školského roka a začiatok prázdnin. Najst' dôvod vyššej alebo nižšej hodnoty tržieb, resp. spotreby alkoholu sa samozrejme nie vždy podarilo (najmä niektoré prepady tržieb).

Skúsme sa teraz venovať problému autokorelácie. Urobme modifikáciu prvého regresného modelu a pridajme novú vysvetľujúcu premennú – hodnotu tržieb posunutú o krok dozadu. Javí sa celkom prirodzené, že hodnota zajtrajšieho nákupu, môže závisieť aj od hodnoty minulého. Ak bol totiž nákup v minulosti väčší, domáce zásoby môžu byť dostatočné a hodnota budúceho nákupu menšia. Opačne, ak bola hodnota minulého nákupu malá, hodnota budúceho nákupu môže byť vyššia. Druhý regresný model s hodnotou tržieb

oneskorenou o jeden krok má tvar ako predošlý, navyac ALKO02(-1) predstavuje hodnotu tržieb posunutú o krok dozadu. Model dosahuje nasledujúce hodnoty.

Dependent Variable: ALKO02				
Included observations: 293 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18574.67	2566.834	7.236413	0.0000
TIME	21.32032	5.720840	3.726781	0.0002
ALKO02(-1)	0.439392	0.051042	8.608435	0.0000
UT	4103.869	1774.216	2.313060	0.0214
ST	5909.707	1766.901	3.344673	0.0009
SV	10087.76	1756.752	5.742277	0.0000
PI	33675.51	1788.123	18.83288	0.0000
SO	22225.05	2419.347	9.186382	0.0000
NE	-10964.15	2420.023	-4.530597	0.0000
R-squared	0.804252	Mean dependent var	55752.93	
Adjusted R-squared	0.798738	S.D. dependent var	17576.25	
S.E. of regression	7885.095	Akaike info criterion	20.81357	
Sum squared resid	1.77E+10	Schwarz criterion	20.92661	
Log likelihood	-3040.188	F-statistic	145.8559	
Durbin-Watson stat	2.001080	Prob(F-statistic)	0.000000	

Tab. 2.2



Obr. 2.4

Vidíme zlepšenie hodnoty R^2 , model nemá problémy s autokoreláciou, zlepšila sa hodnota AIC a model neobsahuje nevýznamné premenné. Rovnako sa nám zlepšila celková hodnota F-štatistiky. Stále však čelíme problému nie normálneho rozdelenia reziduálov

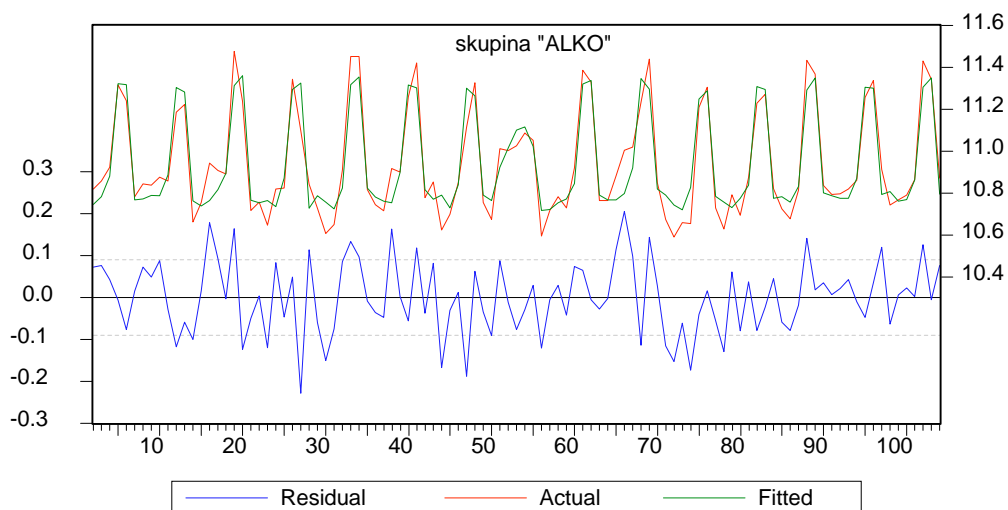
a problému heteroskedasticity, čo by mohlo mať za následok chybu v konfidenčných intervaloch pri prognózach. Nenormalita reziduálov je zrejme spôsobená pomerne veľkým počtom „špeciálnych“ dní (Veľká noc, 1. máj, 8.máj, príchod prázdnin...). Zahrnutie každého z týchto dní do modelu, by si vyžadovalo pomerne veľké množstvo binárnych premenných, tieto by však mohli zhoršiť niektoré iné vlastnosti modelu. Model sa podarilo ešte mierne vylepšiť o vplyv Veľkej noci zahrnutím binárnej premennej, ktorú by sme mohli nazvať „Easter or not“. Hodnoty R^2 aj AIC sa zlepšili, rovnako sa zlepšila aj hodnota F-štatistiky, zostáva však problém nie normálneho rozdelenia reziduálov.

Venujme sa teraz predpovediam. Analyzujeme skupiny ALKO, TABAK, PEKÁRENSKÉ VÝROBKY a CUKROVINKY. Predpovede budeme konštruovať pomocou modelu skráteného na obdobie 8.7.2002-27.10.2002. Takto sa vyhneme problémom s niektorými špeciálnymi dňami. Skrátené modely dosahujú lepšie hodnoty R^2 a AIC a netrpia poruchami ako predchádzajúce „dlhé“ modely. Pre krátke predpovede dopredu sú vhodnejšie. Na zachytenie vplyvu „špeciálnych“ dní by sme potrebovali dlhší časový rad. Posledný týždeň nebudeme brať pri tvorbe regresie do úvahy, použijeme ho len na porovnanie kvality predpovedí.

Pre skupinu „ALKO“ bol vybraný model so šiestimi binárnymi premennými a jednou časovou posunutou premennou. Štyri binárne premenné sú pre dni štvrtok, piatok, sobota a nedeľa a dve binárne premenné slúžia na ošetrenie príchodu školského roka. Premenná $USS = 1$ v dňoch 27.8.-29.8.2002 (sú to dni utorok, streda, štvrtok, na grafe hodnoty 51, 52, 53) a premenná $PS = 1$ pre dni 30.8-31.8.2002 (dni piatok a sobota, na grafe 54, 55). Ich použitie je z grafu zrejmé. Hodnota tržieb prešla logaritmickou transformáciou. Model má normálne rozdelené reziduály, netrpí problémom heteroskedasticity a autokorelácie. Na grafe je na y-ovej osi nanosený logaritmus skutočných a odhadnutých hodnôt a hodnôt reziduálov, na x-ovej osi je čas. Model vyzerá nasledovne:

Dependent Variable: LOG(ALKO)				
Included observations: 103 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.862038	0.962021	8.172421	0.0000
SV	0.088569	0.027299	3.244343	0.0016
PI	0.511209	0.029679	17.22473	0.0000
SO	0.401568	0.055553	7.228563	0.0000
NE	-0.128613	0.055382	-2.322283	0.0224
PS	-0.233142	0.066002	-3.532374	0.0006
USS	0.181032	0.054626	3.314035	0.0013
LOG(ALKO(-1))	0.269890	0.089429	3.017935	0.0033
R-squared	0.880763	Mean dependent var	10.94274	
Adjusted R-squared	0.871977	S.D. dependent var	0.251772	
S.E. of regression	0.090085	Akaike info criterion	-1.901640	
Sum squared resid	0.770953	Schwarz criterion	-1.697000	
Log likelihood	105.9344	F-statistic	100.2473	
Durbin-Watson stat	1.966608	Prob(F-statistic)	0.000000	

Tab. 2.3



Obr. 2.5

Uvádzame ešte prognózu a konfidenčné intervaly na 7 nasledujúcich dní:

SKUPINA	ALKO				konfidenčný interval ⁵		
	DEŇ	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]	ľavý	pravý
PONDELOK		40025	48832	8807	22.0	40743	58528
UTOROK		47108	47849	741	1.6	39927	57344
STREDA		43451	47587	4136	9.5	39709	57029
ŠTVRTOK		64183	51917	-12265	-19.1	43143	62476
PIATOK		79123	81109	1987	2.5	66439	99018
SOBOTA		86383	81987	-4395	-5.1	68144	98643
NEDEĽA		54391	48390	-6001	-11.0	39395	59438
CELKOM		414663	407672	-6991	-1.7		

Tab. 2.4

V stĺpci „konfidenčné intervaly“ sú uvedené dvojstranné 95% konfidenčné intervaly vypočítané pomocou vzťahu (uvedeného v [4]):

$$\hat{y}_p + t_{\alpha/2} s_{ep} < y_p < \hat{y}_p + t_{1-\alpha/2} s_{ep} \quad (2.5)$$

kde $t_{\alpha/2}$ a $t_{1-\alpha/2}$ sú kvantily Studentovho rozdelenia s $n-k$ stupňami voľnosti, \hat{y}_p je odhadovaná budúca hodnota tržieb, y_p je skutočná hodnota budúcich tržieb a pre s_{ep} platí vzťah

$$s_{ep} = \sqrt{\frac{e^T e}{n-k}} \sqrt{1 + X_p^T (X^T X)^{-1} X_p} \quad (2.6)$$

kde e je vektor reziduií, n je počet pozorovaní, k je počet vysvetľujúcich premenných, X je matica plánu a X_p je matica (vektor) plánu budúcich pozorovaní.

Rovnako ako v prípade kľzavých priemerov aj tento model indikuje, že najväčší obrat sa uskutočňuje v piatok a sobotu. Výsledky regresie sú porovnateľné s výsledkami kľzavých priemerov. Skutočnosť, že model obsahuje časovo oneskorenú premennú mierne komplikuje

⁵ Nesymetrickosť konfidenčného intervalu vyplýva z logaritmickej transformácie. Konfidenčný interval je symetrický pre model s logaritmovanými hodnotami tržieb. Spätnou transformáciou sa symetria stráca.

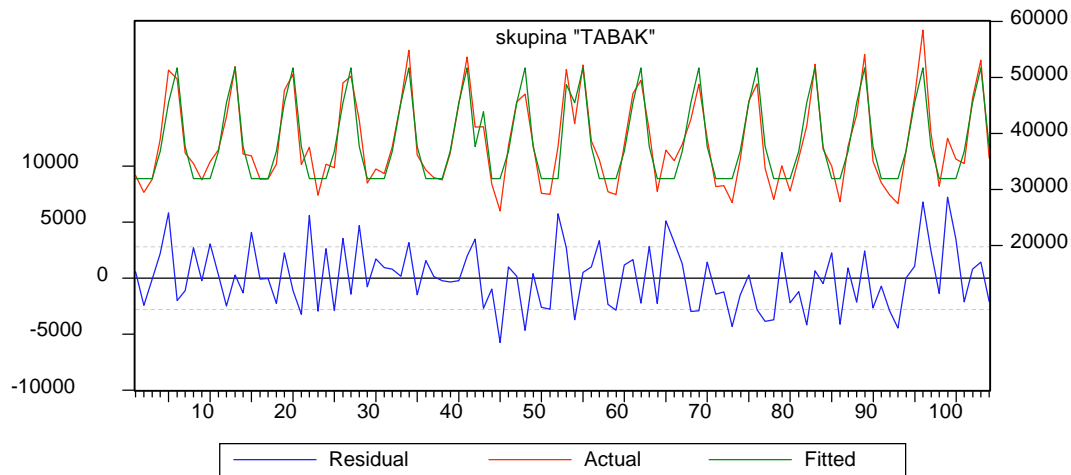
konštrukciu prognóz. Problém sa rieši rekurentne. Najskôr sa vypočíta prognóza o krok dopredu. Pre túto prognózu máme časovo oneskorenú premennú ešte k dispozícii. Na prognózu o dva kroky dopredu sa ako časovo oneskorená premenná použije predikcia o jeden krok dopredu, na predikciu o tri kroky dopredu sa ako časovo oneskorená premenná použije predikcia na dva kroky dopredu, takto postupujeme ďalej.

Pre skupinu TABAK bol zostrojený model s piatimi binárnymi premennými. Štyri binárne premenné zodpovedajú dňom štvrtok až nedeľa, binárna premenná HOLIDAY má hodnotu 1 v dňoch 18. augusta a 29. augusta. V tomto prípade nebolo potrebné siahnúť po časovo oneskorenej premennej ani po logaritmickej transformácii. Model má normálne rozdelené reziduály, netrpí autokoreláciou ani heteroskedasticitou.

Dependent Variable: TABAK				
Included observations: 104 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31944.24	424.7351	75.20980	0.0000
SV	4850.886	842.1152	5.760359	0.0000
PI	13541.91	838.6609	16.14706	0.0000
SO	19737.20	838.6609	23.53419	0.0000
NE	5729.956	838.6609	6.832268	0.0000
HOLIDAY	11933.84	2026.241	5.889646	0.0000
R-squared	0.874790	Mean dependent var	38499.69	
Adjusted R-squared	0.868402	S.D. dependent var	7720.619	
S.E. of regression	2800.765	Akaike info criterion	18.76913	
Sum squared resid	7.69E+08	Schwarz criterion	18.92169	
Log likelihood	-969.9950	F-statistic	136.9374	
Durbin-Watson stat	2.169708	Prob(F-statistic)	0.000000	

Tab. 2.5

V grafe je na x-ovej osi čas, na y-ovej osi je skutočná a fitovaná hodnota tržieb a reziduály. Graf vyzerá nasledovne:



Obr. 2. 6

Predpoveď na 7 dní dopredu a konfidenčné intervaly majú tvar:

SKUPINA	TABAK				konfidenčný interval	
	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]	ľavý	pravý
PONDELOK	28127	31944	3817	13.6	26323	37566
UTOROK	36576	31944	-4632	-12.7	26323	37566
STREDA	26280	31944	5664	21.6	26323	37566
ŠTVRTOK	32033	36795	4762	14.9	31049	42542
PIATOK	40389	45486	5097	12.6	39746	51226
SOBOTA	49253	51681	2428	4.9	45941	57422
NEDEĽA	35725	37674	1949	5.5	31934	43415
CELKOM	248383	267470	19087	7.7		

Tab. 2.6

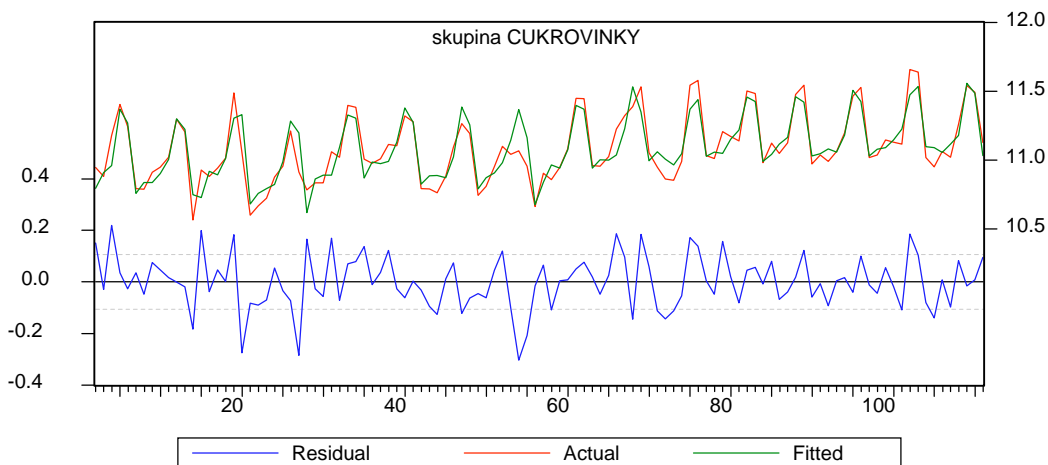
Predpovede regresného modelu pre skupinu „TABAK“ sú lepšie ako predpovede konštruované pomocou kľzavých priemerov.

Pre skupinu „CUKROVINKY“ bol vybraný model s lineárnym trendom, časovo oneskorenou premennou, ktorou sa ošetril problém autokorelácie a štyrmi binárnymi premennými zodpovedajúcimi dňom štvrtok až nedeľa. V modeli bola použitá logaritmická transformácia. Model prešiel testom na autokoreláciu, heteroskedasticitu a normalitu reziduálov. Model má tvar:

Dependent Variable: LOG(CUKOR)				
Included observations: 110 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.263721	0.921999	5.709029	0.0000
TIME	0.001405	0.000398	3.531929	0.0006
SV	0.084326	0.031972	2.637463	0.0096
PI	0.338046	0.034784	9.718516	0.0000
SO	0.118978	0.051984	2.288733	0.0241
NE	-0.310075	0.047565	-6.518909	0.0000
LOG(CUKOR(-1))	0.515332	0.085726	6.011353	0.0000
R-squared	0.823849	Mean dependent var	11.08974	
Adjusted R-squared	0.813588	S.D. dependent var	0.244413	
S.E. of regression	0.105526	Akaike info criterion	-1.598191	
Sum squared resid	1.146989	Schwarz criterion	-1.426342	
Log likelihood	94.90049	F-statistic	80.28759	
Durbin-Watson stat	1.964015	Prob(F-statistic)	0.000000	

Tab. 2.7

V grafe je na y-ovej osi logaritmus skutočných a fitovaných hodnôt a logaritmus reziduálov, na x-ovej osi je čas. Graf vyzerá nasledovne:



Obr. 2. 7

Predpoveď bola počítaná rovnakým spôsobom ako pre model skupiny „ALKO“ a má tvar:

SKUPINA	CUKROVINKY				konfidenčný interval ⁶	
	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]	ľavý	pravý
PONDELOK	56931	64895	7964	14.0	52323	80488
UTOROK	63786	66984	3198	5.0	53992	83101
STREDA	60998	68162	7164	11.7	54929	84583
ŠTVRTOK	77672	74874	-2798	-3.6	60089	93297
PIATOK	102958	102751	-207	-0.2	82030	128706
SOBOTA	98096	99510	1414	1.4	79858	123998
NEDEĽA	68003	62947	-5056	-7.4	49876	79443
CELKOM	528443	540122	11679	2.2		

Tab. 2. 8

Regresný model pre skupinu „CUKROVINKY“ dosahuje výrazne lepšiu predpoveď ako predpoveď pomocou kľzavých priemerov.

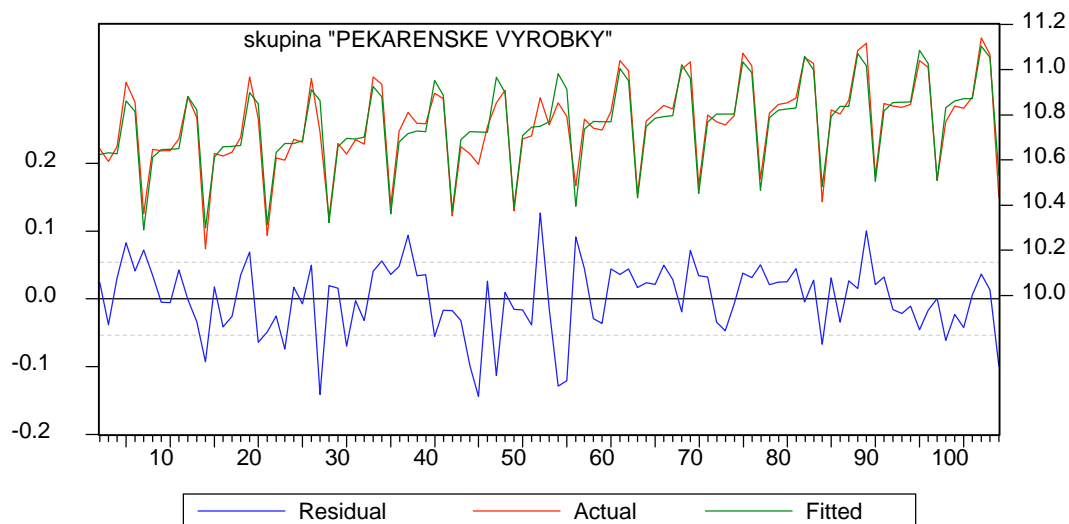
Model pre poslednú analyzovanú skupinu „PEKÁRENSKÉ VÝROBKY“ má tvar:

Dependent Variable: LOG(PEKA)				
Included observations: 103 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.482788	0.475428	19.94578	0.0000
TIME	0.002255	0.000211	10.69573	0.0000
PI	0.224169	0.016417	13.65494	0.0000
SO	0.143443	0.021409	6.700165	0.0000
NE	-0.372972	0.019710	-18.92339	0.0000
LOG(PEKA(-1))	0.107289	0.045166	2.375414	0.0195
R-squared	0.928494	Mean dependent var	10.75564	
Adjusted R-squared	0.924808	S.D. dependent var	0.196147	
S.E. of regression	0.053786	Akaike info criterion	-2.951125	
Sum squared resid	0.280613	Schwarz criterion	-2.797645	
Log likelihood	157.9829	F-statistic	251.9046	
Durbin-Watson stat	1.644622	Prob(F-statistic)	0.000000	

Tab. 2.9.

⁶ Nesymetrickosť konfidenčného intervalu vyplýva z logaritmickej transformácie. Konfidenčný interval je symetrický pre model s logaritmovanými hodnotami tržieb. Spätnou transformáciou sa symetria stráca.

Model obsahuje lineárny trend, tri binárne premenné zodpovedajúce dňom piatok až nedeľa a časovo oneskorenú premennú. V modeli bola použitá logaritmická transformácia. Model prešiel testom na autokoreláciu, heteroskedasticitu a normalitu reziduálov. V grafe je na x-ovej osi čas, na y-ovej osi logaritmus skutočných a fitovaných hodnôt a reziduálov.



Obr. 2. 8

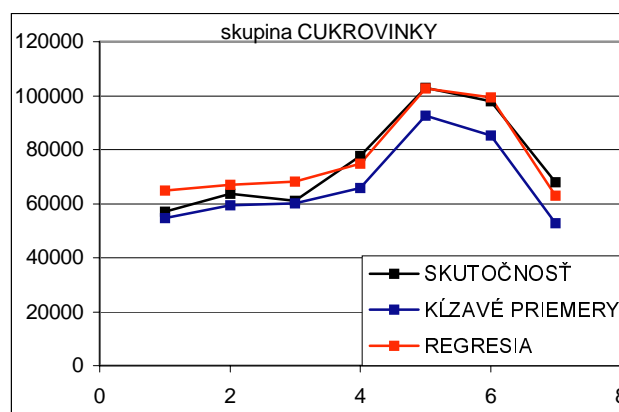
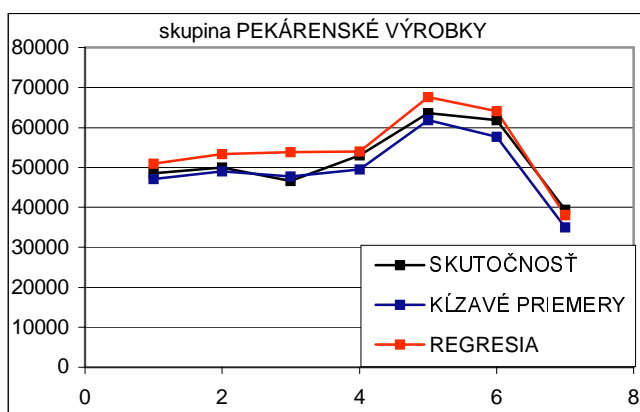
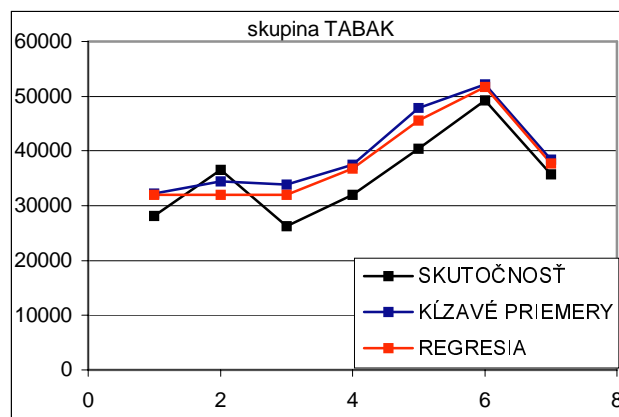
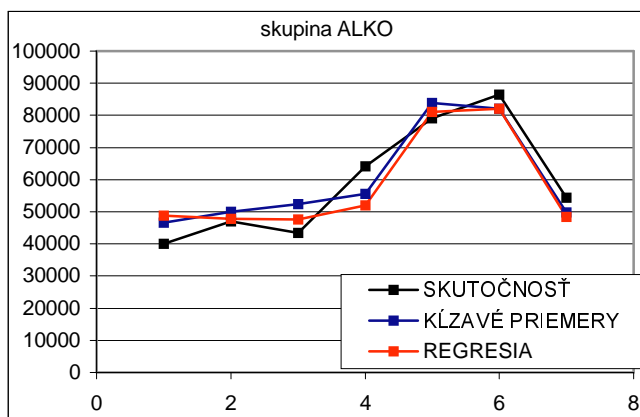
Predpoveď na týždeň dopredu vyzerá nasledovne:

SKUPINA	PEKÁRENSKÉ VÝROBKY				konfidenčný interval ⁷	
	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]	ľavý	pravý
PONDELOK	48525	50958	2433	5.0	45830	56660
UTOROK	50027	53351	3324	6.6	47968	59339
STREDA	46585	53736	7150	15.3	48309	59772
ŠTVRTOK	53012	53898	887	1.7	48452	59957
PIATOK	63546	67616	4070	6.4	60566	75487
SOBOTA	61918	64053	2134	3.4	57459	71403
NEDEĽA	39347	38082	-1265	-3.2	33860	42829
CELKOM	362960	381694	18734	5.2		

Tab. 2.10

Predpoveď získaná pomocou regresie je porovnateľná s predpoveďou získanou pomocou kľzavých priemerov. Na záver uvádzame pre každú skupinu graf so skutočnými hodnotami, hodnotami predpovedí získanými pomocou kľzavých priemerov a hodnotami predpovedí získanými pomocou regresie.

⁷ Nesymetrickosť konfidenčného intervalu vyplýva z logaritmickkej transformácie. Konfidenčný interval je symetrický pre model s logaritmovanými hodnotami tržieb. Spätnou transformáciou sa symetria stráca.



Počas celého roka vidíme množstvo „špeciálnych dní“, pre ktoré je potrebné isté ošetrenie. Jednou z možností sú binárne premenné, tieto sú skôr vhodné na ošetrenie jednorázových vplyvov ako napríklad Mikuláš, Sv.Valentín, niektoré jednodňové sviatky. Binárne premenné však nie vždy môžu vystihnúť celý efekt. Ako príklad môžeme spomenúť rast tržieb už niekoľko týždňov pred Veľkou nocou alebo Vianocami, ktorý môžeme premennými ťažko ošetriť. Pre modelovanie Veľkej noci a Vianoc by bolo vhodné zaviesť zvláštny režim, stále je však problémom krátky časový rad, kvôli ktorému je možné len ťažko navrhnúť správny model.

Čo sa týka vzájomného porovnania regresného prístupu a prístupu cez kľzavé priemery, môžeme konštatovať, že oba spôsoby dosahujú pomerne dobré výsledky prognóz, mierne lepšie sa zdajú byť pri použití regresného prístupu. Výhodou regresného prístupu je aj pomerne jednoduchý výpočet konfidenčných intervalov. Pri oboch modeloch nie je výnimočná chyba dennej prognózy na úrovni vyše 15%, skutočná hodnota sa však väčšinou drží v hraniciach konfidenčného intervalu. Akumulovaná chyba za celý týždeň (daná ako pomer súčtu chýb predikcie a súčtu skutočného dopytu za celý týždeň) však už dosahuje lepšie hodnoty, približne na úrovni 5% skutočného dopytu, čo je akceptovateľná hodnota.

3. KAPITOLA – Box-Jenkinsova metodológia

V tejto časti práce sa budeme zaoberať Box - Jenkinsovou metodológiou. Ide o pomerne nový prístup k analýze časových radov. Časový rad je chápaný ako rad stochastického charakteru, avšak aj v tejto „náhode“ sa snažíme zachytiť isté zákonitosti. V zásade môžeme postupovať dvomi spôsobmi. Prvým spôsobom je klasická dekompozícia časového radu a následná aplikácia Box – Jenkinsovho prístupu na reziduálnu časť časového radu. Druhá možnosť je chápať Box – Jenkinsovú metodológiu ako samostatný prístup k trendovej a sezónnej zložke. Výhodou v prípade stochastického modelovania trendu a sezónnosti je, že sa model prispôbuje zmenám časového radu.

3.1. Základné pojmy

Stacionarita

Stacionarita časového radu znamená v istom zmysle ustálené správanie sa časového radu. Stacionaritu rozlišujeme na silnú a slabú. Pod silnou stacionaritou rozumieme invariantnosť pravdepodobnostného rozdelenia stochastického procesu vzhľadom na posunutie v čase, platí teda, že $\{y_t, y_{t-1}, y_{t-2}, \dots\}$ a $\{y_{t+k}, y_{t-1+k}, y_{t-2+k}, \dots\}$ majú rovnaké rozdelenie. Pod slabou stacionaritou rozumieme konštantnú strednú hodnotu a konštantnú disperziu stochastického procesu a kovariančnú maticu nezávislú na čase. V praxi uvažujeme slabú stacionaritu.

Autokovariancia , autokorelácia, parciálna autokorelácia

Autokovariančná funkcia je definovaná ako:

$$\gamma_k = \text{cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] \quad (3.1)$$

Autokorelačná funkcia je definovaná ako:

$$\rho_k = \frac{\text{cov}(X_t, X_{t+k})}{\sqrt{D(X_t)}\sqrt{D(X_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (3.2)$$

Parciálna autokorelačná funkcia je definovaná ako:

$$\rho_{kk} = \text{corr}(X_t, X_{t+k} \mid X_{t+1}, X_{t+2}, \dots, X_{t+k-1}) \quad (3.3)$$

Je to korelácia náhodných veličín X_t a X_{t+k} pri konštatntných (pevných hodnotách) $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$. Je počítaná rekurentne, vplyv $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ sa dá odstrániť aj pomocou metódy najmenších štvorcov.

Doteraz spomínané funkcie boli teoretické funkcie, v praxi sa však musíme uspokojiť s ich odhadmi. Odhad strednej hodnoty časového radu má tvar výberového priemeru. Odhad autokovariančnej a autokorelačnej funkcie urobíme pomocou výberovej autokovariančnej funkcie a výberovej autokorelačnej funkcie (podľa [1]).

Výberové funkcie

$$\text{Výberový priemer: } \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \quad (3.4)$$

$$\text{Výberová autokovariančná funkcia: } c_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (3.5)$$

$$\text{Výberová autokorelačná funkcia: } r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (3.6)$$

Výberová parciálna autokorelačnú funkcia sa počíta rekurentne:

$$r_{k,k} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \quad \text{pre } k > 1 \quad (3.7)$$

$$r_{k,j} = r_{k-1,j} - r_{k,k} r_{k-1,k-j} \quad \text{pre } j = 1, 2, 3, \dots, k-1 \quad (3.8.)$$

kde počiatočný odhad je: $r_{1,1} = r_1$

Správanie sa autokorelačnej a parciálnej autokorelačnej funkcie má dôležitú vypovedaciu schopnosť, výber a konštrukcia vhodného modelu je založená práve na ich priebehu. Nakoľko stále pracujeme len s odhadmi skutočných hodnôt, dôležité je zodpovedať otázku, kedy môžeme tvrdiť, že hodnota $\rho_k = 0$. Na jej zodpovedanie najprv vypočítame

smerodajnú odchýľku odhadu r_k autokorelačnej funkcie ρ_k . Používa sa Bartlettova aproximácia:

Ak je $\rho_k = 0$ pre $k > k_0$ potom platí:

$$\sigma(r_k) = \sqrt{\text{var}(r_k)} \approx \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2\right)} \quad \text{pre } k > k_0 \quad (3.9)$$

Rozhodovanie o tom, či $\rho_k = 0$ je založené na porovnaní hodnoty $|r_k|$ s hodnotou $2\sigma(r_k)$. Ak hodnota $|r_k|$ prekračuje dvojsigmový interval, máme voči nulovosti ρ_k vážne výhrady. Rozhodovanie sa opiera o asymptotickú normalitu odhadu r_k s asymptotickou strednou hodnotou ρ_k a fakt, že náhodná veličina s normálnym rozdelením prekročí dvojnásobok strednej hodnoty (správnejšie 1.96-násobok) s pravdepodobnosťou 5%. Bod k_0 má významnú úlohu pri identifikácii modelu. Rozhodovanie o tom, či hodnota parciálnej autokorelačnej funkcie ρ_{kk} je nulová je analogické, porovnáva sa hodnota $r_{k,k}$ s dvojnásobkom hodnoty smerodajnej odchýľky odhadu, teda s hodnotou $2\sigma(r_{k,k})$. Odhad $\sigma(r_{k,k})$ je konštruovaný pomocou Quenouillovej aproximácie.

Ak je $\rho_{kk} = 0$ pre $k > k_0$ potom platí:

$$\sigma(r_{k,k}) \approx \sqrt{\frac{1}{n}} \quad \text{pre } k > k_0 \quad (3.10)$$

Proces bieleho šumu

Biely šum by sme mohli nazvať aj „čistou náhodou“. Pre biely šum $\{a_t\}$ platí:

autokovariančná funkcia má tvar:
$$\gamma_k = \begin{cases} \sigma_a^2 & \text{pre } k = 0 \\ 0 & \text{pre } k \neq 0 \end{cases}$$

autokorelačná funkcia má tvar:
$$\rho_k = \begin{cases} 1 & \text{pre } k = 0 \\ 0 & \text{pre } k \neq 0 \end{cases}$$

parciálna autokorelačná funkcia má tvar: $r_{k,k} = \begin{matrix} 1 & \text{pre } k = 0 \\ 0 & \text{pre } k \neq 0 \end{matrix}$

Operátor spätného posunutia

Operátorom spätného posunutia nazveme operátor B , pre ktorý platí: $By_t = y_{t-1}$. Po viacnásobnom aplikovaní: $B^j y_t = y_{t-j}$.

Lineárny proces

Lineárnym procesom nazývame časový rad v tvare:

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

(3.11)

kde y_t je časový rad, ψ_i sú parametre, $\psi_i = 1$, μ je stredná hodnota časového radu. Kvôli zjednodušeniu predpokladáme $\mu = 0$, v opačnom prípade prejdeme k novému časovému radu centrovaním ($y'_t = y_t - \mu$). Hodnota ε_t predstavuje proces bieleho šumu. Zapísané pomocou operátora spätného posunutia dostávame:

$$y_t = \psi(B)\varepsilon_t$$

$$\text{kde } \psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots = 1 + \sum_{j=1}^{\infty} \psi_j B^j \quad (3.12)$$

Podmienka pre existenciu lineárneho procesu má tvar:

$$\psi(B) \text{ konverguje pre } |B| < 1 \quad (3.13)$$

v opačnom prípade rad (suma náhodných veličín) nekonverguje. Podmienka rovnako zabezpečí aj stacionaritu procesu. Proces je možné zapísať aj pomocou minulých hodnôt:

$$y_t = \mu + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \dots + \varepsilon_t \quad (3.14)$$

kde y_t je časový rad, π_i sú parametre, μ je jeho stredná hodnota (pre jednoduchosť opäť predpokladajme $\mu = 0$, inak prejdeme k novému časovému radu centrovaním) a ε_t je biely šum. Zapísané pomocou operátora spätného posunutia dostávame:

$$\pi(B)y_t = \varepsilon_t$$

$$\text{kde } \pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = 1 - \sum_{j=1}^{\infty} \pi_j B^j \quad (3.15)$$

Proces zapísaný v tomto tvare sa nazýva invertibilný. Postačujúca podmienka pre invertibilitu lineárneho procesu je v tvare:

$$\pi(B) \text{ konverguje pre } |B| < 1 \quad (3.16)$$

Aby bola splnená podmienka stacionarity procesu zapísaného v invertibilnom tvare, musia ležať všetky korene polynómu mimo jednotkového kruhu, t.j. $|\pi_i| < 1 \quad \forall i$.

3.2. Procesy MA, AR, ARMA

Proces kĺzavých priemerov

Proces kĺzavých priemerov rádu q (ďalej len MA(q)) má tvar:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.17)$$

Pomocou operátora spätného posunutia ho môžeme zapísať ako:

$$y_t = \theta(B)\varepsilon_t \quad (3.18)$$

$$\text{kde } \theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q = 1 + \sum_{j=1}^q \theta_j B^j$$

Proces je stacionárny bez nutnosti splnenia ďalších podmienok. Hodnota autokorelačnej funkcie má tvar:

$$\rho_k = \frac{\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad \text{pre } k=1,2,\dots,q \quad (3.19)$$

$$\rho_k = 0 \quad \text{pre } k > q$$

Získali sme q rovníc o q neznámych, pomocou hodnôt ρ_1, \dots, ρ_q môžeme teda vypočítať hodnoty $\theta_1, \dots, \theta_q$, postup je iteračný. Ide len o prvotný odhad parametrov $\theta_1, \dots, \theta_q$.

Autoregresný proces

Autoregresným procesom rádu p (ďalej len AR(p)) nazývame proces v tvare:

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (3.20)$$

Proces je invertibilný bez nutnosti splnenia ďalších podmienok. Proces je stacionárny, ak hodnota $|\varphi_i| < 1 \quad \forall i$. Proces MA(q) je invertibilný, ak všetky korene polynómu ležia mimo jednotkového kruhu. Pomocou operátora spätného posunutia môžeme proces zapísať ako:

$$\varphi(B)y_t = \varepsilon_t \quad (3.21)$$

$$\text{kde } \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p = 1 - \sum_{j=1}^p \varphi_j B^j$$

Pre autokorelačnú funkciu vzniká sústava rovníc (Yule-Walkerova sústava rovníc):

$$\begin{aligned} \rho_1 &= \varphi_1 + \varphi_2 \rho_1 + \dots + \varphi_p \rho_{p-1} \\ \rho_2 &= \varphi_1 \rho_1 + \varphi_2 + \dots + \varphi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \dots + \varphi_p \end{aligned} \quad (3.22)$$

Riešením sústavy dostávame odhad hodnoty parametrov $\varphi_1, \dots, \varphi_p$ vyjadrený pomocou hodnôt autokorelačnej funkcie ρ_1, \dots, ρ_p . Aj tu ide len o akýsi prvotný odhad. Ďalšie upresnenie sa vykoná napríklad metódou maximálnej vierohodnosti, prvotné odhady parametrov $\varphi_1, \dots, \varphi_p$, resp. $\theta_1, \dots, \theta_q$ sa použijú ako štartovací bod iteračného procesu. Podrobnejšie sa nebudeme zaoberať metódou odhadu parametrov, hodnoty $\varphi_1, \dots, \varphi_p$ a $\theta_1, \dots, \theta_q$ dostaneme ako softvérový výstup.

Zmiešaný proces

Zmiešaným procesom rádu p a q (ďalej len ARMA(p, q)) rozumieme proces v tvare:

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.23)$$

Pomocou operátora spätného posunutia môžeme proces zapísať ako

$$\varphi(B)y_t = \theta(B)\varepsilon_t \quad (3.24)$$

kde výrazy $\varphi(B)y_t$ a $\theta(B)\varepsilon_t$ sú definované ako vyššie.

Proces ARMA(p,q) je stacionárny, ak spĺňa podmienku stacionarity ako AR(p) proces a je invertibilný, ak spĺňa podmienku invertibility ako MA(q) proces. Pre autokorelačnú funkciu dostávame rovnakú sústavu rovníc ako pre AR(p) proces.

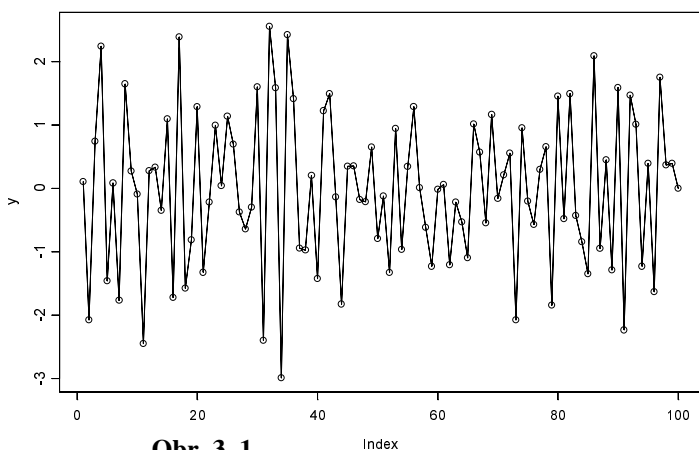
3.3 Identifikácia modelu

Ako už bolo spomenuté, vhodný model sa snažíme vybrať podľa správania sa autokorelačnej a parciálnej autokorelačnej funkcie. Uvádzame tabuľku, z ktorej je ďalšia identifikácia jasná. Kritéria sa opierajú o odvodené vzťahy pre autokorelačnú funkciu a parciálnu autokorelačnú funkciu. Krivkou U nazývame geometricky klesajúcu postupnosť alebo sínusoidu s geometricky klesajúcou amplitúdou, resp. ich kombináciu.

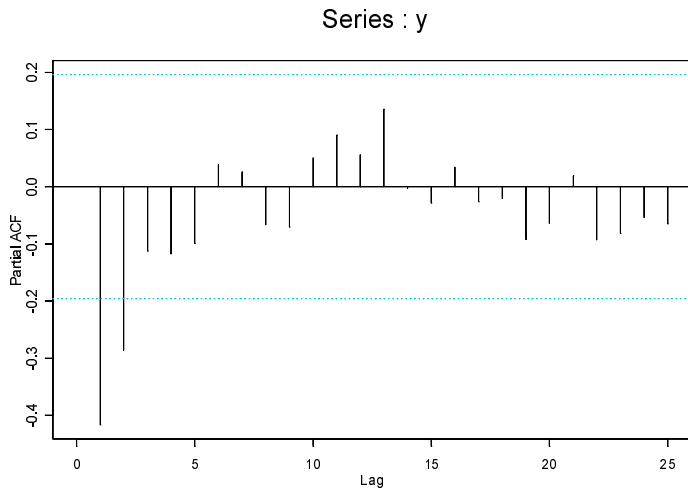
	Model identifikujeme ako AR(p) ak	Model identifikujeme ako MA(q) ak	Model identifikujeme ako ARMA(p,q) ak
ρ_k	neexistuje k_0 , ρ_k je ohraničená krivkou U	$k_0 = q$	neexistuje k_0 , ρ_k je ohraničená po prvých q-p hodnotách krivkou U
$\rho_{k,k}$	$k_0 = p$	neexistuje k_0 , $\rho_{k,k}$ je ohraničená krivkou U	neexistuje k_0 , $\rho_{k,k}$ je ohraničená po prvých p-q hodnotách krivkou U

Tab. 3.1

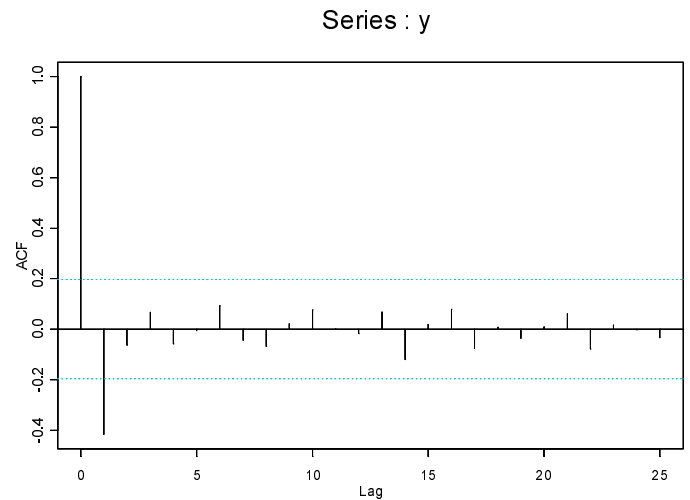
Na nasledujúcich grafoch je zobrazená realizácia procesu MA(1) $y_t = \varepsilon_t - 0.6\varepsilon_{t-1}$, hodnoty parciálnej autokorelačnej funkcie PACF a hodnoty autokorelačnej funkcie ACF.



Obr. 3.1



Obr. 3. 2



Obr. 3. 2

3.4 Nestacionárne procesy

Ďalej sa budeme venovať nestacionárnym procesom. Doteraz diskutované procesy sa vyznačovali konštantnou strednou hodnotou a konštantným rozptylom. Kovariancia (korelácia) závisela len na vzdialenosti $t - k$. V praxi sa však stacionárne procesy vyskytujú zriedkavo, týka sa to najmä ekonomických časových radov. Budeme rozlišovať dva druhy nestacionarity, nestacionarita procesu v strednej hodnote a nestacionarita procesu v disperzii.

Procesy nestacionárne v strednej hodnote

Procesy nestacionárne v strednej hodnote sa vyznačujú trendom. Zavedieme pojem deterministického a stochastického trendu. O deterministickom trende hovoríme, ak môžeme strednú hodnotu nestacionárneho procesu chápať ako funkciu času. Napríklad vývoj strednej hodnoty môže byť lineárne rastúci. Proces obsahujúci deterministický trend môže mať napríklad tvar:

$$X_t = \alpha_0 + \alpha_1 t + \varepsilon_t$$

kde α_0 a α_1 sú parametre, t je čas a ε_t je biely šum.

Vieme už, že ak všetky korene polynómu $\varphi(B)$ ležia mimo jednotkového kruhu, procesy AR a ARMA sú stacionárne. Ak neležia všetky korene mimo jednotkového kruhu, proces je nestacionárny. Platí: ak jeden koreň polynómu leží na jednotkovej kružnici a ostatné mimo, proces nie je stacionárny, ale obsahuje tzv. lineárny stochastický trend. Obdobne ak ležia na jednotkovom kruhu dva korene a ostatné ležia mimo, proces obsahuje tzv. parabolický stochastický trend. Ak aspoň jeden bod leží vnútri jednotkového kruhu, hovoríme

o nestacionárnom procese explozívneho typu. Nestacionárne procesy so stochastickým trendom je však možné diferenciaciou pretransformovať na procesy stacionárne. Ak proces po d-tej diferencii neobsahuje žiadnu deterministickú zložku a je ho možné popísať stacionárnym a invertibilným modelom ARMA(p,q), hovoríme, že tento proces je integrovaným procesom d-teho rádu. Takýto proces nazveme ARIMA(p,d,q) a zapíšeme ho ako:

$$\varphi(B)\Delta^d y_t = \theta(B)\varepsilon_t \quad (3.25)$$

kde $\Delta^d y_t$ je d-ta diferencia časového radu y_t .

Diferenčný operátor Δ nahradíme pomocou operátora spätného posunutia ako $(1-B)$, platí totiž $\Delta y_t = y_t - y_{t-1} = (1-B)y_t$. Analogicky môžeme zapísať $\Delta^d = (1-B)^d$. Model ARIMA(p,d,q) teda všeobecne zapíšeme ako:

$$\varphi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \quad (3.26)$$

Rád diferencovania, ktorý býva len zriedkavo vyšší ako 2, môžeme určiť viacerými metódami. Jednou z nich je posúdenie grafického záznamu časového radu spolu s jeho prvou a druhou diferenciou a následné optické posúdenie stacionarity. Objektívnejšie kritérium spočíva v posúdení hodnôt rozptylu časového radu a rozptylu jeho diferencií $\hat{\sigma}_y^2, \hat{\sigma}_{\Delta y}^2, \hat{\sigma}_{\Delta^2 y}^2$. Za rád diferencie d potom volíme tú hodnotu, pre ktorú dosahuje rad najmenší rozptyl.

Procesy nestacionárne v disperzii

Ak rozptyl časového radu v čase nie je nemenný, hovoríme o procese nestacionárnom v rozptyle (disperzii). Vhodnou transformáciou opäť prejdeme k procesu, ktorý je vo svojej podstate lineárny a vieme ho popísať už niektorým zo známych modelov AR, MA a ARIMA.

Box a Cox navrhli transformáciu v tvare:

$$y_t^{(\lambda)} = \frac{y_t^\lambda - 1}{\lambda} \quad \text{pre } \lambda \neq 0 \quad (3.27)$$

$$y_t^{(\lambda)} = \ln y_t \quad \text{pre } \lambda = 0$$

kde $y_t^{(\lambda)}$ označuje transformovanú hodnotu. Parameter λ sa určí metódou maximálnej vierohodnosti.

Iná používaná transformácia má tvar:

$$\begin{aligned} y_t^{(\lambda)} &= y_t^\lambda && \text{pre } \lambda \neq 0 \\ y_t^{(\lambda)} &= \ln y_t && \text{pre } \lambda = 0 \end{aligned} \quad (3.28)$$

kde parameter λ určíme nasledovne. Pozorovania sa rozdelia na krátke úseky (4-12 pozorovaní), na každom úseku sa vypočíta aritmetický priemer, ktorý označíme m a rozdiel medzi maximálnou a minimálnou hodnotou, označíme r , niekedy sa namiesto rozdielu maximálnej a minimálnej hodnoty počíta rozptyl hodnôt. Hodnoty m a r sa nanesú do grafu (na x-ovú os m). Ak sú takýmto spôsobom skonštruované body približne konštantné, znamená to, že s rastúcou strednou hodnotou sa rozptyl nemení, transformáciu preto nie je potrebné vykonať. Ak rozptyl rastie lineárne (body približne ležia na lineárne rastúcej priamke) volíme $\lambda = 0$ a robíme logaritmickú transformáciu, poprípade volíme $\lambda = 0.5$. Ak rozptyl rastie exponenciálne, volíme $\lambda = -0.5$. Podotýkame však, že ide len o hrubý odhad parametra λ , presnosť je do istej miery vyvážená jednoduchosťou tejto metódy. V ekonomickej praxi sa často vyskytuje práve logaritmická transformácia, napríklad rozptyl predaja sa zvyšuje úmerne s jeho rastom. Preto sa niekedy rozlišuje len medzi možnosťami $\lambda = 0$ a $\lambda = 1$.

3.5 Sezónne procesy

Rovnako ako trend aj sezónna zložka je v Box-Jenkinsovej metodológii modelovaná stochasticky. V modeli ARIMA sme predpokladali vzájomnú závislosť (koreláciu) medzi veličinami $\dots X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2} \dots$. Pretože v časovom rade sú obsiahnuté aj sezónne výkyvy, môžeme predpokladať určitú závislosť (koreláciu) aj medzi premennými $\dots X_{t-2p}, X_{t-p}, X_t, X_{t+p}, X_{t+2p} \dots$, kde p je dĺžka sezóny. Predpokladáme, že časový rad obsahuje oba typy závislosti. Na zachytenie závislosti vnútri periódy použijeme ARIMA model. Rovnako skonštruujeme ARIMA model aj pre zachytenie závislosti medzi jednotlivými periódami. Predpokladajme, že závislosť vo vnútri periódy je popísaná ARIMA modelom:

$$\varphi(B)(1-B)^d y_t = \theta(B)\eta_t \quad (3.29)$$

kde $\varphi(B)$ a $\theta(B)$ sú autoregresný operátor a operátor kĺzavých priemerov. Predpokladáme, že v procese η_t je zachytená závislosť aj medzi jednotlivými periódami, preto aj tento proces popíšeme modelom ARIMA:

$$\Phi(B^L)(1-B^L)^D \eta_t = \Theta(B^L) \varepsilon_t \quad (3.30)$$

kde $\Phi(B^L) = 1 - \Phi_1 B^L - \Phi_2 B^{2L} - \dots - \Phi_p B^{pL}$ je sezónny autoregresný operátor rádu P a $\Theta(B^L) = 1 + \Theta_1 B^L + \Theta_2 B^{2L} + \dots + \Theta_q B^{qL}$ je sezónny operátor kĺzavého súčtu rádu Q.

Dĺžku periódy označíme L ($L = 4$ štvrt'ročné dáta, $L = 12$ mesačné dáta, $L = 7$ týždenné dáta). Znakom B^L označíme operátor sezónneho posunutia s dĺžkou sezóny L, platí $B^L y_t = y_{t-L}$, znakom $\Delta^D = (1 - B^L)^D$ potom označíme D-tu sezónnu diferenciu. Spojením týchto dvoch modelov vzniká sezónny model ARIMA, označujeme SARIMA(p,d,q) x (P,D,Q)_s. Model má tvar :

$$\varphi(B)\Phi(B_L^S)\Delta^d \Delta^D y_t = \theta(B)\Theta(B_L^S)\varepsilon_t \quad (3.31)$$

pomocou operátora sezónneho posunutia ho môžeme tiež zapísať ako:

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 - \Phi_1 B^L - \dots - \Phi_p B^{pL})(1 - B)^d (1 - B^L)^D = \\ = (1 + \theta_1 B + \dots + \theta_q B^q)(1 + \Theta_1 B^L + \dots + \Theta_q B^{qL}) \quad (3.32)$$

Pri identifikácii modelu sa opierame hlavne o priebeh autokorelačnej funkcie ACF a parciálnej autokorelačnej funkcie PACF. Ak je to nutné, najskôr sa vykoná stacionarizácia časového radu, napríklad logaritmicou transformáciou. Následne sa posúdi nutnosť sezónneho diferencovania. Ak hodnota ACF periodicky nadobúda lokálne extrémny, je potrebné zvážiť sezónne diferencovanie. Ak po sezónnom diferencovaní klesajú hodnoty ACF príliš pomaly (lineárne) dáta diferencujeme aj obyčajne (nesezónne). Po sezónnej aj obyčajnej diferencii nadobúdajú ACF a PACF v prvej perióde rovnaké tvary ako pri nesezónnom procese. V ostatných periódách sú tvary autokorelačnej a parciálnej autokorelačnej funkcie pomerne komplikované a opis ich priebehu je zložitý. Výber vhodného modelu bude ešte diskutovaný v časti s praktickými aplikáciami.

3.6 Prognózy

Po identifikácii modelu prejdeme k prognózam. Pre jednoduchosť uvažujme proces ARMA v tvare:

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.33)$$

Predpokladáme stacionárnosť a invertovateľnosť. V čase t sa snažíme urobiť predpoveď o k krokov dopredu, odhadovanú hodnotu označíme \hat{y}_{t+k} . Predpoveď hľadáme ako lineárnu kombináciu hodnôt $y_t, y_{t-1}, y_{t-2}, \dots$ resp. lineárnu kombináciu hodnôt $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ (proces je invertovateľný). Navyše požadujeme, aby predpoveď mala v triede lineárnych predpovedí najmenšiu strednú štvorcovú chybu. Minimalizujeme teda výraz

$$E[(y_{t+k} - \hat{y}_{t+k})^2] \rightarrow \min \quad (3.34)$$

Predpoveď hľadáme v tvare:

$$\hat{y}_{t+k} = \psi_k^* \varepsilon_t + \psi_{k+1}^* \varepsilon_{t-1} + \psi_{k+2}^* \varepsilon_{t-2} + \dots \quad (3.35)$$

Skutočnú hodnotu y_{t+k} však môžeme zapísať (keďže proces je invertovateľný) aj ako:

$$y_{t+k} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \psi_2 \varepsilon_{t+k-2} + \dots + \psi_k \varepsilon_t + \dots \quad (3.36)$$

Minimalizácia chyby potom nadobúda tvar:

$$(3.37)$$

$$E[(y_{t+k} - \hat{y}_{t+k})^2] = E\left[\left((\varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \psi_2 \varepsilon_{t+k-2} + \dots + \psi_k \varepsilon_t + \dots) - (\psi_k^* \varepsilon_t + \psi_{k+1}^* \varepsilon_{t-1} + \psi_{k+2}^* \varepsilon_{t-2} + \dots)\right)^2\right]$$

$$E[(y_{t+k} - \hat{y}_{t+k})^2] = E\left[\left(\varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \psi_2 \varepsilon_{t+k-2} + \dots + \psi_{k-1} \varepsilon_{t-1} + (\psi_k - \psi_k^*) \varepsilon_t + (\psi_{k+1} - \psi_{k+1}^*) \varepsilon_{t-1} + \dots\right)^2\right]$$

Predpoklad nulovej korelácií chýb nám dovoľuje písať pre všetky $i \neq j$ $E[(\varepsilon_i \varepsilon_j)^2] = 0$.

Vzťah pre strednú štvorcovú chybu potom po úprave vyzerá nasledovne:

$$E[(y_{t+k} - \hat{y}_{t+k})^2] = \left(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{k-1}^2 + \sum_{j=k}^{\infty} (\psi_j - \psi_j^*)^2\right) \sigma_\varepsilon^2 \quad (3.38)$$

a minimum nadobúda pre $\psi_j = \psi_j^*$ $j = k, k+1, k+2, \dots$

Potom pre predpoveď platí vzťah:

$$\hat{y}_{t+k} = \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \psi_{k+2} \varepsilon_{t-2} + \dots \quad (3.39)$$

Analogicky môžeme zapísať:

$$\hat{y}_{t+k} = [y_{t+k}] \quad (3.40)$$

kde hranatými zátvorkami označíme podmienenú strednú hodnotu pri pevných hodnotách $y_t, y_{t-1}, y_{t-2}, \dots$. Chyba predpovede a jej rozptyl majú tvar:

$$e_{t+k}(t) = y_{t+k} - \hat{y}_{t+k}(t) = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \dots + \psi_{k-1} \varepsilon_{t+1} \quad (3.41)$$

$$\text{var}(e_{t+k}(t)) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{k-1}^2) \sigma_\varepsilon^2 \quad (3.42)$$

Praktický výpočet predpovedí robíme iteračne. Najprv vypočítame predpoveď o jeden krok dopredu, pomocou nej vypočítame predpoveď o dva kroky dopredu atď. Keďže platí vzťah:

$$y_{t+k} = \varphi_1 y_{t+k-1} + \varphi_2 y_{t+k-2} + \dots + \varphi_p y_{t+k-p} + \varepsilon_{t+k} + \theta_1 \varepsilon_{t+k-1} + \theta_2 \varepsilon_{t+k-2} + \dots + \theta_q \varepsilon_{k+t-q} \quad (3.43)$$

môžeme písať:

$$(3.44)$$

$$\hat{y}_{t+k}(t) = \varphi_1 [y_{t+k-1}] + \varphi_2 [y_{t+k-2}] + \dots + \varphi_p [y_{t+k-p}] + [\varepsilon_{t+k}] + \theta_1 [\varepsilon_{t+k-1}] + \theta_2 [\varepsilon_{t+k-2}] + \dots + \theta_q [\varepsilon_{k+t-q}]$$

kde hranatými zátvorkami opäť označíme podmienenú strednú hodnotu. Vieme tiež, že platí:

$$[y_{t+j}] = \hat{y}_{t+j}(t) \quad \text{pre } j > 0$$

$$[y_{t+j}] = y_{t+j} \quad \text{pre } j \leq 0$$

$$(3.45)$$

$$[\varepsilon_{t+j}] = 0 \quad \text{pre } j > 0$$

$$[\varepsilon_{t+j}] = \varepsilon_{t+j} = y_{t+j} - \hat{y}_{t+j}(t+j-1) \quad \text{pre } j \leq 0$$

Pre začatie iteračného postupu volíme $\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_q = 0$. Konfidenčný interval pre predpoveď na k krokov dopredu $\hat{y}_{t+k}(t)$ môžeme pomocou (3.42) zapísať ako:

$$(\hat{y}_{t+k}(t) - 2\sigma(\varepsilon_{t+k}(t)), \hat{y}_{t+k}(t) + 2\sigma(\varepsilon_{t+k}(t))) \quad (3.46)$$

Vidíme, že konfidenčný interval sa pre rastúce k (horizont predpovede) zväčšuje. Na záver už len spomenieme, že vo vzorcoch pre predpovede pracujeme s odhadmi parametrov $\varphi_1, \dots, \varphi_p$ a $\theta_1, \dots, \theta_q$, nie so skutočnými parametrami. Vzorce však nie sú príliš citlivé na chyby, ktoré vznikli pri odhade parametrov, obzvlášť pri dlhších časových radoch.

3.7 Aplikácie

Venujme sa teraz aplikáciám sezónnych modelov SARIMA. Aplikovali sme vyššie spomínaný postup na vybrané skupiny tovarov: ALKO, CUKROVINKY, TABAK a PEKÁRENSKÉ VÝROBKY. Niektoré modely si vyžiadali transformáciou za účelom stacionarizovania časového radu. Použili sme transformáciu v tvare:

$$y_t^{(\lambda)} = y_t \quad \text{pre } \lambda = 1$$

$$y_t^{(\lambda)} = \ln y_t \quad \text{pre } \lambda = 0$$

Rad sme rozdelili na úseky dĺžky 6. Na každom úseku sme vypočítali hodnotu aritmetického priemeru (označíme m) a rozdielu maximálnej a minimálnej hodnoty (označíme r) a naniesli sme ich do grafu (na x-ovú os m). Následne sme skonštruovali jednoduchý lineárny regresný model v tvare $r = \beta_0 + \beta_1 m + \varepsilon$ a testovali sme hypotézu o nulovosti parametra β_1 . Hypotéza nebola zamietnutá pri skupine „TABAK“.

Vhodný model sa vyberal najmä podľa správania sa autokorelačnej funkcie ACF a parciálnej autokorelačnej funkcie PACF, tiež sa pozorovala hodnota Akaikeho informačného kritéria. Dôraz sa rovnako kládol na korelogram reziduálov a na hodnotu Q-štatistiky (Portmanteau test). Posledným overovacím krokom pre prijatie modelu bolo porovnanie predikovaných hodnôt so skutočnými. Model sa odhadol bez použitia dát za posledné tri týždne. Následne sa zostrojila prognóza na dva týždne dopredu, ktorá bola porovnaná so skutočnými hodnotami. V prípade uspokojivej prognózy bol model prijatý a parametre modelu nanovo odhadnuté už

s použitím všetkých dát s výnimkou dát za posledný týždeň. Tieto boli použité už len na porovnanie kvality predpovede.

Pre skupinu ALKO bola použitá logaritmická transformácia a následne vybraný model SARIMA(0,0,1) x (0,1,1)₇ s parametrami

Dependent Variable: DLOG(ALKO,0,7)				
Included observations: 98 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000537	0.002528	0.212633	0.8321
MA(1)	0.319752	0.096056	3.328801	0.0012
SMA(7)	-0.918041	0.027256	-33.68203	0.0000
R-squared	0.552081	Mean dependent var		0.001867
Adjusted R-squared	0.542651	S.D. dependent var		0.152640
S.E. of regression	0.103227	Akaike info criterion		-1.673645
Sum squared resid	1.012296	Schwarz criterion		-1.594514
Log likelihood	85.00863	F-statistic		58.54598
Durbin-Watson stat	1.901510	Prob(F-statistic)		0.000000

Tab. 3.1

Zároveň bol otestovaný aj regresný model, kde na reziduálnu časť bola aplikovaná Box-Jenkinsova metodológia. Tento model dosahoval mierne lepšiu hodnotu AIC, predpoveď modelu bola takmer identická s predpoveďou modelu SARIMA.

Predikcia zostrojená pomocou modelu SARIMA na týždeň dopredu je v tabuľke 3.2 a v grafe na konci kapitoly.

SKUPINA	ALKO			
	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]
PONDELOK	40025	47272	7247	18.1
UTOROK	47108	47503	395	0.8
STREDA	43451	48685	5234	12.0
ŠTVRTOK	64183	52847	-11336	-17.7
PIATOK	79123	81759	2636	3.3
SOBOTA	86383	81756	-4627	-5.4
NEDEĽA	54391	49005	-5386	-9.9
CELKOM	414664	408827	-5837	-1.4

Tab. 3.2

Pre skupinu CUKROVINKY bola použitá logaritmická transformácia a z niekoľkých testovaných modelov bol vybraný model SARIMA(1,0,0) x (0,1,1)₇. Bol uvažovaný aj model s obyčajnou diferenciou, ten však dosahoval mierne horšie výsledky predpovede.

Dependent Variable: DLOG(CUKOR,0,7)				
Included observations: 83 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.024765	0.007244	3.418541	0.0010
AR(1)	0.505449	0.096396	5.243477	0.0000
MA(7)	-0.915853	0.028714	-31.89614	0.0000
R-squared	0.615002	Mean dependent var		0.013008
Adjusted R-squared	0.605377	S.D. dependent var		0.176044
S.E. of regression	0.110589	Akaike info criterion		-1.530513
Sum squared resid	0.978398	Schwarz criterion		-1.443085
Log likelihood	66.51628	F-statistic		63.89654
Durbin-Watson stat	1.932110	Prob(F-statistic)		0.000000

Tab. 3.3

Predikcia na týždeň dopredu je v tabuľke 3.4, graf je na konci kapitoly.

SKUPINA	CUKROVINKY			
DEŇ	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]
PONDELOK	56931	64390	7459	13.1
UTOROK	63786	68397	4612	7.2
STREDA	60998	72387	11388	18.7
ŠTVRTOK	77672	77606	-66	-0.1
PIATOK	102958	107761	4804	4.7
SOBOTA	98096	103460	5364	5.5
NEDEĽA	68003	63452	-4551	-6.7
CELKOM	528443	557452	29010	5.5

Tab. 3.4

Pre skupiny pekárenských výrobkov bola opäť použitá logaritmická transformácia a následne bol vybraný model SARIMA(1,0,0) x (0,1,1)₇ uvedený v tabuľke 3.5.

Dependent Variable: DLOG(PEKA,0,7)				
Included observations: 83 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.021656	0.002357	9.188592	0.0000
AR(1)	0.248445	0.108284	2.294395	0.0244
MA(7)	-0.920283	0.029301	-31.40837	0.0000
R-squared	0.567454	Mean dependent var		0.016110

Adjusted R-squared	0.556640	S.D. dependent var	0.082289
S.E. of regression	0.054792	Akaike info criterion	-2.935060
Sum squared resid	0.240176	Schwarz criterion	-2.847632
Log likelihood	124.8050	F-statistic	52.47574
Durbin-Watson stat	1.988473	Prob(F-statistic)	0.000000

Tab. 3.5

Predikciu na týždeň dopredu uvádzame v tabuľke 3.6, graf je zobrazený na konci kapitoly.

SKUPINA	PEKÁRENSKÉ VÝROBKY			
DEŇ	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]
PONDELOK	48525	50725	2200	4.5
UTOROK	50027	52491	2464	4.9
STREDA	46585	53248	6663	14.3
ŠTVRTOK	53012	55025	2014	3.8
PIATOK	63546	68311	4765	7.5
SOBOTA	61918	64968	3050	4.9
NEDEĽA	39347	38055	-1291	-3.3
CELKOM	362960	382824	19864	5.5

Tab. 3.6

Pri skupine TABAK nebolo potrebné siahnuť po logaritmickej transformácii. Boli otestované viaceré modely (vrátane modelov s logaritmickej transformáciou), následne bol ako vhodný model vybraný model SARIMA(0,1,1) x (1,1,1)₇. Model je uvedený v tabuľke 3.7, hodnota predikcie na týždeň dopredu je v tabuľke 3.8.

Dependent Variable: D(TABAK,1,7)					
Included observations: 90 after adjusting endpoints					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	5.436662	21.61363	0.251539	0.8020	
AR(7)	-0.472033	0.120094	-3.930540	0.0002	
MA(1)	-0.895042	0.058299	-15.35265	0.0000	
SMA(7)	-0.349207	0.108552	-3.216952	0.0018	
R-squared	0.664503	Mean dependent var	-65.31151		
Adjusted R-squared	0.652800	S.D. dependent var	6359.547		
S.E. of regression	3747.281	Akaike info criterion	19.33888		
Sum squared resid	1.21E+09	Schwarz criterion	19.44998		
Log likelihood	-866.2494	F-statistic	56.77871		
Durbin-Watson stat	2.047637	Prob(F-statistic)	0.000000		

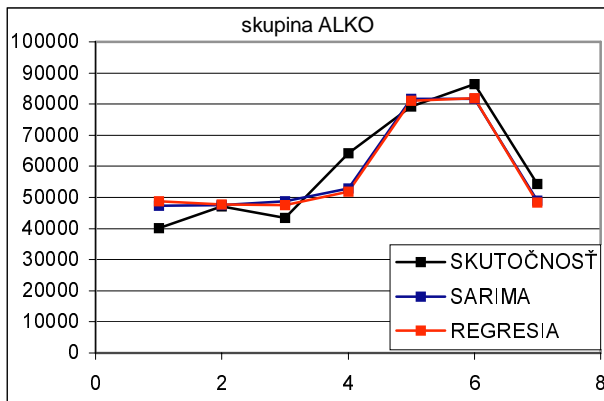
Tab. 3.7

SKUPINA	TABAK			
DEŇ	SKUTOČNOSŤ	ODHAD	ROZDIEL	v [%]
PONDELOK	28127	31909	3782	13.4

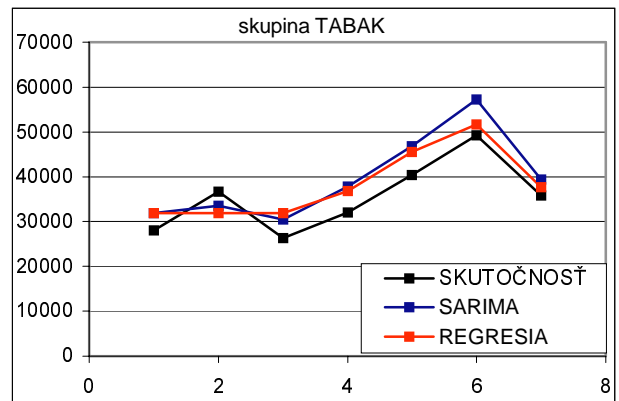
UTOROK	36576	33562	-3014	-8.2
STREDA	26280	30511	4231	16.1
ŠTVRTOK	32033	37830	5796	18.1
PIATOK	40389	46841	6452	16.0
SOBOTA	49253	57310	8058	16.4
NEDELA	35725	39375	3651	10.2
CELKOM	248383	277338	28955	11.7

Tab. 3.8

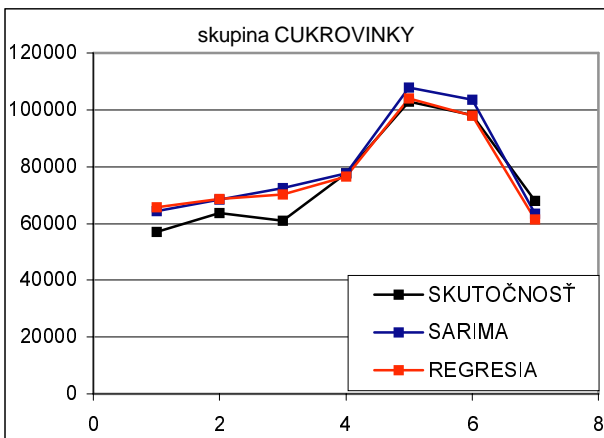
Na záver uvádzame grafy skutočných hodnôt a predikcie na týždeň dopredu. Pre porovnanie uvádzame aj výsledky regresného prístupu.



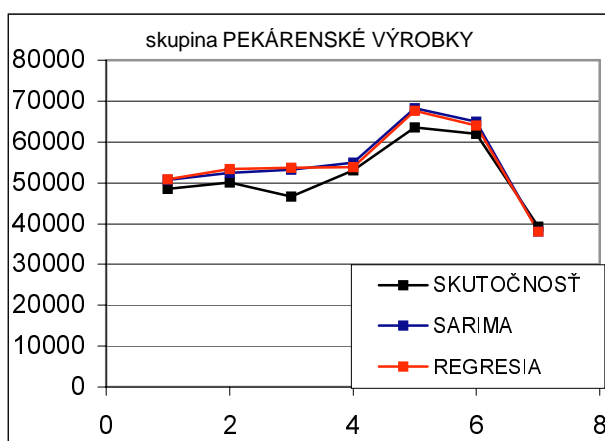
Obr. 3.4



Obr. 3.5



Obr. 3.6



Obr. 3.7

Záverom môžeme konštatovať, že prognózy modelov SARIMA sú dobrou konkurenciou prognózam regresného prístupu a prístupu cez kľzavé priemery. Menšia nevýhoda je v pomerne zložitej identifikácii vhodného modelu. Aplikácia modelov

SARIMA na reziduálnu časť regresných modelov prinášala prognózu takmer na rovnakej úrovni ako aplikácia modelu SARIMA na hodnotu tržieb. Hodnoty prognóz určené týmto spôsobom sú na akceptovateľnej úrovni.

4. KAPITOLA – mikroekonomický prístup

Doteraz sme zaoberali len celými skupinami tovarov, v tejto časti sa však pokúsime načrtnúť mikroekonomický prístup a predpovedať správanie sa jednotlivých tovarov. Skupina tovarov bola pomerne necitlivá na zmenu ceny jedného tovaru, resp. niekoľko málo tovarov. Zrejme to bolo spôsobené možnosťou substitúcie medzi jednotlivými produktami. Ak sa daný výrobok nachádza v cenovej akcii (a pravdepodobne aj v letákovej ponuke) je prirodzené, že dopyt po tomto výrobku je zvýšený. Tento zvýšený dopyt sa však iste prejaví poklesom dopytu po konkurenčných výrobkoch. Mohli by sme teda konštatovať, že dopyt po tovare závisí od ceny tovaru, od ceny konkurenčných tovarov a od toho, či je výrobok uvedený v letákovej ponuke. Posledný faktor sa javí ako najsilnejší akcelerátor predaja.

Sú tovary, ktoré menia cenu veľmi zriedkavo, je ich však minimum. Väčšina výrobkov mení cenu pomerne často, v hypermarketoch by sme mohli dokonca hovoriť o „plávajúcich cenách“. Tvary môžu meniť cenu v závislosti od dodávateľov, prípadných cenových akcií, ale aj s blížiacim sa koncom doby trvanlivosti. Ak sa dlhodobo nemení cena tovaru a zároveň zostáva nemenná aj cena substitučných (konkurenčných) tovarov, môžeme väčšinou prognózu určiť pomerne jednoducho. Príkladom by mohol byť tovar, ktorý nesie oficiálny názov „rožok štandard“. Ak sa však cena mení, mení sa aj dopyt po tovare.

Začnime teda opisom modelu. Budeme predpokladať, že dopyt po výrobku je závislý len od jeho ceny a pokúsime sa odhadnúť dopytovú funkciu. Predpokladajme, že dopyt po výrobku má konštantnú cenovú elasticitu. Ukážeme, že tento predpoklad vedie na obyčajnú diferenciálnu rovnicu. Označme $D_1 = D(p_1)$ a $D_2 = D(p_2)$ dopyt po tovare pri cene p_1 a pri cene p_2 . Predpoklad konštantnej cenovej elasticity dopytu znamená:

$$-\varepsilon_{D,p} = \frac{\frac{\Delta D}{D}}{\frac{\Delta p}{p}} = \frac{\Delta D}{\Delta p} \frac{p}{D} = \text{konšt} \quad (4.1)$$

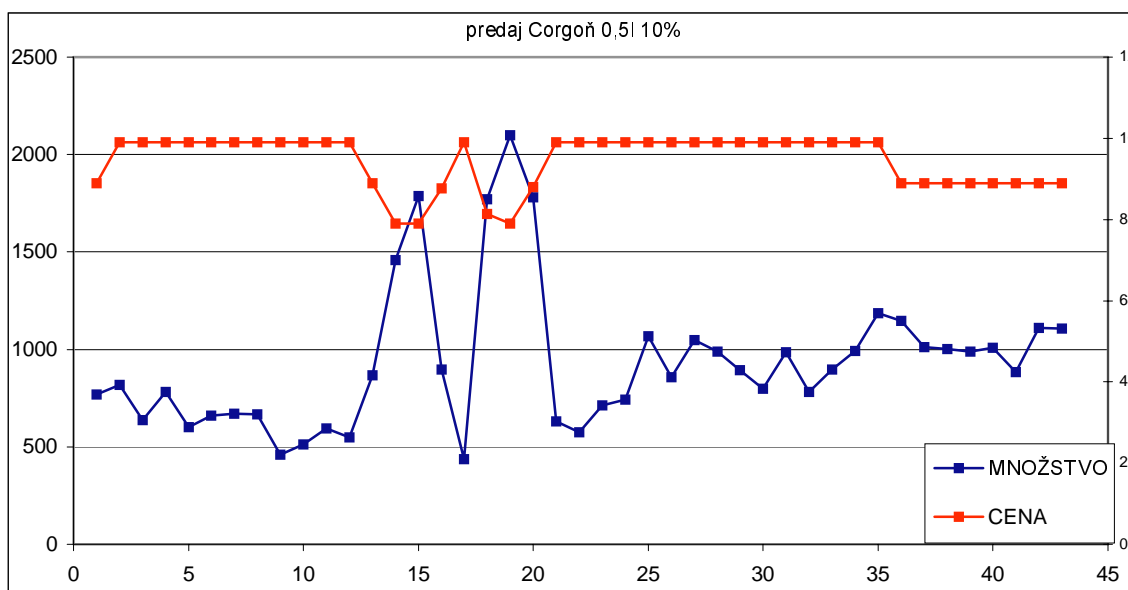
Prejdeme k derivácii a riešime obyčajnú diferenciálnu rovnicu:

$$\frac{dD}{dp} = -\varepsilon \frac{D}{p} \quad \Rightarrow \quad \frac{dD}{D} = -\varepsilon \frac{dp}{p} \quad \Rightarrow \quad D(p) = k_1 p^{-\varepsilon} \quad (4.2)$$

kde k_1 je konštanta a ε je elasticita, o ktorej predpokladáme, že je tiež konštantná. Parametre modelu k_1 a ε vieme určiť, ak poznáme dve ceny a k nim prislúchajúci dopyt. Po úprave dostávame pre k_1 a ε nasledujúce vzťahy:

$$\varepsilon = \frac{\ln\left(\frac{D_2}{D_1}\right)}{\ln\left(\frac{p_1}{p_2}\right)} \quad k_1 = D_1 p_1^\varepsilon \quad (4.3)$$

Skúsime model nakalibrovať na tovar, ktorý mení cenu pomerne často. Ako príklad budeme uvažovať konkrétny tovar, fľaškové pivo Corgoň. V nasledujúcom grafe sú uvedené týždenné dáta predaja piva Corgoň, na x-ovej osi je poradové číslo týždňa, na ľavej y-ovej osi je množstvo (počet fliaš), na pravej y-ovej osi je cena. Tá je počítaná ako pomer týždenných tržieb a množstva predaného za týždeň. Preto ak sa cena menila počas týždňa (napríklad cenová akcia sa začala v piatok) nemá cena „baťovský tvar“.



Obr. 4. 1

Vidíme, že s poklesom ceny rastie dopyt, tovar je teda obyčajným tovarom (v zmysle mikroekonomickej definície). V období 5.4.-18.4.2002 (14.-16. týždeň) sa uskutočnila akcia, výrobok bol uvedený v letákovvej ponuke. Rozdelíme časový rad na niekoľko období podľa výšky ceny a v každom období spočítame priemerný dopyt za týždeň. V tabuľke 4.1 je časové

obdobie udávané v týždňoch roka 2002, cena v danom období a priemerný dopyt za týždeň v tomto období. Odhad parametrov k_1 a ε je v tabuľke 4.2.

Obdobie	Cena	Dopyt
2-12	9.9	632
36-43	8.9	1032

Tab. 4.1

Parameter	Odhad
k_1	24786859
ε	4.613893119

Tab. 4.2

Pomocou týchto hodnôt urobíme odhad dopytu pri zmene ceny. Výsledky sú v tabuľke 4.3.

Obdobie	Cena	Dopyt	
		odhadovaný	skutočný
14-15	7.9	1789	1623
18-20	8.28	1443	1884

Tab. 4.3

Obdobie je časové obdobie v týždňoch, cena je priemerná cena tovaru za toto obdobie (počítaná ako pomer tržby a počtu predaných výrobkov). Odhadovaný dopyt je týždenný dopyt odhadnutý modelom a skutočný dopyt je priemerný týždenný dopyt v tomto období.

Môžeme konštatovať pomerne dobrý odhad dopytu. Bolo by zaujímave sledovať dopyt po

Obdobie	Cena	Dopyt
34-44	11.9	488
26-31	10.9	674

tomto výrobku pri avšak bez uvedenia v letákovej ponuke.

Parameter	Odhad
k	4539969
eps	3.690175

akciovej cene, výrobku Model však

nedáva rovnako dobré výsledky pre niektoré ďalšie tovary. Uvádžame ešte jednu aplikáciu, tentokrát na pivo Zlatý bažant. V tabuľke 4.4 je cena a dopyt počas rôznych období, v tabuľke 4.5. odhad parametrov k_1 a ε .

Tab. 4.4

Tab. 4.5

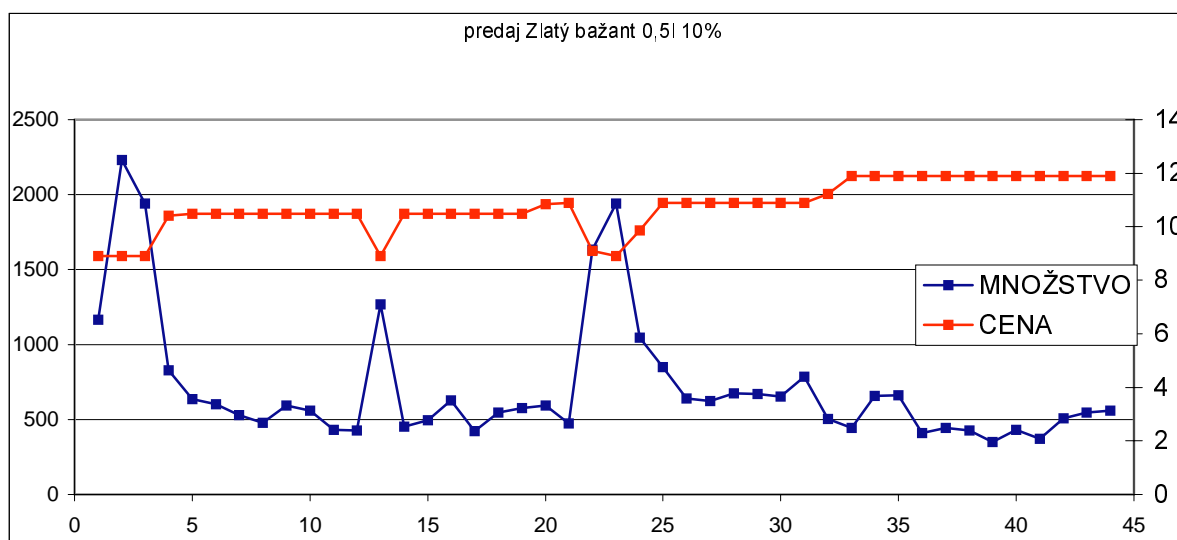
Pomocou odhadnutých parametrov k_1 a ε bol zostrojený odhad dopytu pri zmene ceny. Výsledky sú uvedené v poslednej tabuľke 4.6.

Obdobie	Cena	Dopyt	
		odhadovaný	skutočný
1-3	8.9	1424	1778
22-24	9.3	1217	1540
14-19	10.5	774	520

Pre úplnosť uvádzame ešte graf vývoja dopytu a ceny piva Zlatý bažant. Na x-ovej osi je čas

Tab. 4.6

v týždňoch, na pravej y-ovej osi je počet predaných fliaš piva a na pravej y-ovej osi je cena.



Obr. 4.2

Aj tu vidíme relatívne dobrú prognózu. Cesta k zlepšeniu by mohla viesť cez modely, ktoré by rovnako zahŕňali aj ceny konkurenčných výrobkov a skutočnosť, či je model v letákovvej ponuke alebo nie. Nakoľko však každá podskupina výrobkov obsahuje niekoľko desiatok navzájom substitučných výrobkov (napríklad podskupina PIVO skupiny ALKO obsahuje vyše 50 výrobkov) a v podskupine takmer neustále dochádza k zmene ceny niektorého z výrobkov, k cenovým a letákovým akciám a k dispozícii máme len obmedzený počet pozorovaní, bolo by takéto modelovanie náročné.

Záver

V diplomovej práci sme sa zaoberali niekoľkými metódami. Pomocou kľzavých priemerov a regresie sme odhadli vplyv jednotlivých dní v týždni na hodnotu tržieb. Konštatujeme výrazný nárast tržieb v piatok a sobotu, kde tržby počas týchto dvoch dní môžu dosahovať až 40% celkovej týždennej tržby. Na odhad budúceho vývoja spotreby sme použili okrem dvoch vyššie spomenutých metód aj modely SARIMA. Pre celé skupiny tovarov sme počas bežných dní dosahovali spomenutými metódami porovnateľné prognózy na dobrej úrovni. Počas roka sa však vyskytuje veľké množstvo dní, ktoré musia byť špeciálne ošetrené. Ide predovšetkým o sviatky, dni pracovného voľna pripadajúce na pondelok až piatok a dni typu Mikuláš alebo Sv. Valentín. Tieto dni výrazne ovplyvňujú hodnotu tržieb v jednotlivých skupinách. Niektoré dni ovplyvňujú hodnotu tržieb len krátku dobu (Mikuláš spôsobí výrazný nárast tržieb v skupine CUKROVINKY len 5.12.). Na ich ošetrenie sú binárne premenné postačujúce. Niektoré sviatky však ovplyvňujú hodnotu tržieb dlhšie, (nákup na Vianoce sa začínajú už v novembri). Na ich ošetrenie by bolo vhodné zaviesť zvláštny model. Vzhľadom na krátkosť časového radu (400 denných údajov) sme tento nemohli skonštruovať. Pokiaľ ide o jednotlivé tovary, ich dopyt je vysoko citlivý na viacero faktorov. Je to cena tovaru, cena konkurenčných tovarov, najväčším akcelerátorm predaja je letáková akcia. Správanie sa tovarov pri zmene ceny sme popísali pomocou mikroekonomického prístupu. Prognózy dávajú len hrubý odhad. Dopyt pri cenových akciách je všeobecne ťažko predpovedateľný, nárast predaja počas akcie sa v praxi pohybuje od 50% (nevydarené akcie) do rádovo niekoľko 100% až 1000% (obzvlášť vydarené). Celkovo môžeme konštatovať, že výsledky v tejto práci by boli použiteľné na tvorbu prognóz za účelom optimálneho zásobenia skladu.

Zoznam použitej literatúry:

- [1] Cipra, T., Analýza časových řad s aplikacemi v ekonomii, SNTL – Nakladatelství technické literatury, 1986
- [2] Makridakis, S., Wheelwright, S.C., Forecasting methods for management, John Wiley & Sons, 1989
- [3] Kozák J., Hindls, R., Arlt, J., Úvod do analýzy ekonomických časových řad, Vysoká škola ekonomická v Praze, 1994
- [4] Hušek, R. a kol, Ekonometrické modely řízení a plánování, SNTL – Nakladatelství technické literatury, 1987
- [5] Rublíková, E., Prognostická statistika, Edičné stredisko EU Bratislava, 1984