

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



PREDIKCIA UKAZOVATEĽOV KVALITY AKTÍV
RETAILOVÉHO PORTFÓLIA KOMERČNEJ BANKY

DIPLOMOVÁ PRÁCA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

**PREDIKCIA UKAZOVATEĽOV KVALITY AKTÍV
RETAILOVÉHO PORTFÓLIA KOMERČNEJ BANKY**

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomicko-finančná matematika a modelovanie
Študijný odbor: Aplikovaná matematika
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci práce: doc. Mgr. Marián Grendár, CSc.

Bratislava 2016

Bc. Martin Oberuč



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Martin Oberuč
Študijný program: ekonomicko-finančná matematika a modelovanie
(Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Predikcia ukazovateľov kvality aktív retailového portfólia komerčnej banky
Forecasting the asset quality indicators for the retail portfolio of a commercial bank

Cieľ: Pomocou vhodných štatistických modelov predpovedať hodnoty ukazovateľov kvality aktív retailového portfólia komerčnej banky so zohľadnením makroekonomických ukazovateľov a iných vysvetľujúcich premenných.

Vedúci: doc. Mgr. Marián Grendár, PhD.
Konzultant: Mgr. Marek Mikoška
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.
Dátum zadania: 10.02.2015

Dátum schválenia: 11.02.2015
prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Pod'akovanie

Rád by som využil tento priestor a poďakoval sa vedúcemu svojej diplomovej práce, doc. Mgr. Mariánovi Grendárovi, CSc., za jeho ľudský prístup, neoceniteľné rady a podnetné pripomienky, ktoré mi pri písaní tejto práce veľmi pomohli. Moje poďakovanie patrí aj konzultantovi práce, Mgr. Marekovi Mikoškovi, za námet na spracovanie danej témy, poskytnutie užitočnej literatúry a ochotu diskutovať o praktických problémoch, ktoré sa počas písania práce vyskytli. V neposlednom rade ďakujem svojim rodičom za ich obetavosť, trpezlivosť a podporu počas celého štúdia.

Abstrakt

OBERUČ, Martin: *Predikcia ukazovateľov kvality aktív retailového portfólia komerčnej banky*. [Diplomová práca]. Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky. Katedra aplikovanej matematiky a štatistiky. Školiteľ práce: doc. Mgr. Marián Grendár, CSc. Bratislava: Fakulta matematiky, fyziky a informatiky UK, 2016. 92 s.

Cieľom tejto diplomovej práce je predpovedať hodnoty ukazovateľov kvality aktív retailového portfólia komerčnej banky pomocou vhodných štatistických modelov. Práca je rozdelená na sedem kapitol. Prvé štyri kapitoly tvoria teoretickú a zvyšné tri praktickú časť práce. V prvej kapitole charakterizujeme retailové portfólio komerčnej banky, popisujeme ukazovatele kvality aktív a rozoberáme možnosti ich predikcie. Druhá kapitola sa venuje modelu lineárnej regresie, so zameraním na najznámejšie problémy vyskytujúce sa v praxi a možnosti ich riešenia. V tretej kapitole definujeme dekompozičné metódy analýzy časových radov a stručne sa zmiňujeme aj o teórii Kalmanovho filtra. Štvrtá kapitola sa zaoberá najmä modelmi Boxovej – Jenkinsovej metodológie. Okrem toho tiež približuje problematiku kointegrácie a kauzálneho pôsobenia časových radov. Piata kapitola obsahuje charakteristiku reálnych dát, s ktorými sme v diplomovej práci pracovali. V šiestej a siedmej kapitole sú pomocou metód popísaných v teoretickej časti vytvorené predikčné modely pre dve rozdielne časti retailového portfólia, s následným porovnaním a vyhodnotením výsledkov.

Kľúčové slová: predpovedanie, ukazovatele kvality aktív, retailové portfólio komerčnej banky, lineárna regresia, dekompozičné metódy analýzy časových radov, Kalmanov filter, Boxova – Jenkinsova metodológia

Abstract

OBERUČ, Martin: *Forecasting the asset quality indicators for the retail portfolio of a commercial bank*. [Diploma thesis]. Comenius University in Bratislava. Faculty of Mathematics, Physics and Informatics. Department of Applied Mathematics and Statistics. Supervisor: doc. Mgr. Marián Grendár, CSc. Bratislava: Faculty of Mathematics, Physics and Informatics UK, 2016. 92 p.

The aim of this diploma thesis is to forecast the asset quality indicators for the retail portfolio of a commercial bank using appropriate statistical models. The thesis is divided into seven chapters. The first four chapters comprise the theoretical part and remaining three chapters form the practical part of the thesis. In the first chapter, we characterise the retail portfolio of a commercial bank, describe asset quality indicators and discuss the possibilities of their forecasting. The second chapter deals with a model of linear regression with focus on the most common problems in practice and the possibilities of their solutions. In the third chapter, we define the decomposition methods of time series analysis and briefly mention Kalman filter, too. The fourth chapter handles mostly with the models of the Box – Jenkins methodology. Moreover, it also describes the issues of cointegration and Granger causality of time series. The fifth chapter includes characteristics of the real data dealt with in the diploma thesis. In the sixth and the seventh chapter, the forecasting models for two different parts of the retail portfolio are built with consequent comparison and evaluation of results.

Key words: forecasting, asset quality indicators, retail portfolio of a commercial bank, linear regression, decomposition methods of time series analysis, Kalman filter, Box – Jenkins methodology

Obsah

Úvod	9
1 Ukazovatele kvality aktív a ich predpovedanie	11
1.1 Komerčná banka a jej aktíva	11
1.2 Retailové portfólio komerčnej banky	11
1.2.1 Hypotekárny úver.....	12
1.2.2 Spotrebný úver	12
1.2.3 Kontokorentný úver	12
1.2.4 Kreditná karta.....	13
1.3 Finančné riziko a jeho riadenie.....	13
1.4 Ukazovatele kvality aktív	15
1.4.1 Miera omeškania	15
1.4.2 Miera zlyhania	16
1.5 Predpovedanie	16
1.5.1 Bodová a intervalová predpoveď	16
1.5.2 Jednokroková a viackroková predpoveď	17
1.5.3 Statická a dynamická predpoveď	17
1.5.4 Predpoveď vo vzorke a mimo vzorky.....	17
1.5.5 Predpoveď v ekonometrickom modeli a v modeli časového radu.....	18
1.6 Miery presnosti predpovede	18
1.7 Možnosti predpovedania ukazovateľov kvality aktív.....	19
1.8 Výber vysvetľovanej premennej.....	21
2 Model lineárnej regresie v praxi	22
2.1 Metóda najmenších štvorcov	22
2.2 Metóda maximálnej vierohodnosti	23
2.3 Kritériá výberu modelu.....	24
2.3.1 Koeficient determinácie	25
2.3.2 Informačné kritériá.....	25
2.3.3 Iteračné výberové metódy	26
2.4 Predpovedanie	26
2.5 Špecifiká lineárnej regresie v praxi	27

2.5.1	Heteroskedasticita	27
2.5.2	Multikolinearita.....	28
2.5.3	Autokorelácia rezíduí.....	29
3	Dekompozičné metódy analýzy časových radov.....	32
3.1	Dekompozícia časového radu.....	32
3.2	Popis trendovej zložky matematickými krivkami	33
3.3	Metóda kľzavých priemerov a kubický vyhladzovací splajn.....	34
3.4	Exponenciálne vyhladzovanie	36
3.4.1	Jednoduché exponenciálne vyhladzovanie	37
3.4.2	Dvojité exponenciálne vyhladzovanie	38
3.5	Holtova metóda	39
3.6	Kalmanov filter.....	40
4	Autokorelačné metódy analýzy časových radov.....	42
4.1	Stacionárne časové rady	42
4.2	Autokorelačná funkcia a jej odhad	42
4.3	Parciálna autokorelačná funkcia a jej odhad	43
4.4	Proces kľzavých súčtov MA	44
4.5	Autoregresný proces AR	45
4.6	Zmiešaný ARMA proces	45
4.7	Testy jednotkového koreňa.....	46
4.7.1	Dickeyov – Fullerov test.....	46
4.7.2	Rozšírený Dickeyov – Fullerov test.....	47
4.8	Proces ARIMA	47
4.9	Predpovedanie	48
4.10	Kointegrácia	49
4.10.1	Test založený na rezíduách	49
4.10.2	Johansenov test	50
4.11	Grangerova kauzalita.....	51
5	Dáta.....	52
5.1	Vnútrobankové vysvetľujúce premenné.....	52
5.1.1	Údaje zo žiadostí.....	52
5.1.2	Údaje o portfóliu	53

5.2	Makroekonomické vysvetľujúce premenné	53
5.2.1	Hrubý domáci produkt	53
5.2.2	Nezamestnanosť	54
5.2.3	Inflácia	54
5.2.4	Finančná kríza	54
6	Predikčné modely pre prvú časť portfólia.....	55
6.1	Model lineárnej regresie	55
6.2	Modely časových radov	62
6.2.1	Kubický vyhladzovací splajn	62
6.2.2	Holtova metóda	64
6.2.3	Kalmanov filter	66
6.2.4	ARIMA model	67
6.3	Vyhodnotenie výsledkov	69
6.3.1	Jednokrokové predpovede	69
6.3.2	Viac krokové predpovede	70
7	Predikčné modely pre druhú časť portfólia	72
7.1	Model časových radov	72
7.1.1	Kubický vyhladzovací splajn	72
7.1.2	Holtova metóda	73
7.1.3	Kalmanov filter	74
7.1.4	ARIMA model	75
7.2	Modely lineárnej regresie	77
7.2.1	Model lineárnej regresie s ARIMA rezíduami a nezamestnanosťou	78
7.2.2	Model lineárnej regresie s ARIMA rezíduami a zmenou HDP	80
7.2.3	Model lineárnej regresie s ARIMA rezíduami, zmenou HDP a nezamestnanosťou	82
7.3	Vyhodnotenie výsledkov	84
7.3.1	Jednokrokové predpovede	84
7.3.2	Viac kroková predpoveď	85
	Záver	87
	Zoznam použitej literatúry	89

Úvod

Komerčné banky sú najdôležitejšími finančnými inštitúciami, ktorých funkciou je okrem iného aj poskytovanie úverov. Poskytnuté úverové produkty tvoria úverové portfólio a predstavujú najpodstatnejšiu časť aktív väčšiny komerčných bánk. Portfólio všetkých úverov poskytnutých domácnostiam nazývame retailovým portfóliom komerčnej banky. Skladá sa z hypotekárnych úverov, spotrebných úverov, kontokorentných úverov a kreditných kariet.

Úverové produkty sa poskytujú na základe zmluvy, v ktorej sa klient zaväzuje vrátiť komerčnej banke poskytnuté peňažné prostriedky za vopred dohodnutých podmienok. Riziko nesplnenia týchto podmienok sa nazýva kreditným rizikom. Ukazovatele kvality aktív sú indikátory, ktoré banka na pravidelnej báze sleduje a vyhodnocuje v rámci jeho riadenia. Medzi tieto ukazovatele zaraďujeme napríklad mieru omeškania a mieru zlyhania.

Výskum v oblasti predpovedania kreditného rizika začal výrazne naberať na intenzite približne pred štvrtstoročím. Toto predpovedanie môže prebiehať na úrovni jednotlivých klientov, úverov alebo na úrovni celého úverového portfólia. Poslednému prístupu sa doteraz venovalo najmenej pozornosti, a to aj napriek tomu, že je najvhodnejší pre zohľadnenie makroekonomických vplyvov. V diplomovej práci sa preto venujeme predpovedaniu na úrovni portfólia.

Cieľom diplomovej práce je pomocou vhodných štatistických modelov predpovedať hodnoty miery omeškania dvoch odlišných častí retailového portfólia Všeobecnej úverovej banky, a.s.. Pri predpovedaní pritom chceme okrem iného využiť aj už spomínané makroekonomické ukazovatele. Diplomová práca je rozdelená na dve hlavné časti, teoretickú a praktickú. Teoretická časť diplomovej práce sa skladá zo štyroch kapitol, praktickú časť tvoria tri ďalšie kapitoly.

Prvá kapitola začína definíciou komerčnej banky a jej retailového portfólia. Stručne popísané sú v nej jednotlivé produkty, z ktorých sa toto portfólio skladá. Ďalej je v prvej kapitole definované finančné riziko a proces jeho riadenia. Podrobnejšie sa pritom venujeme kreditnému riziku a ukazovateľom kvality aktív. Prvá kapitola je zakončená vysvetlením teoretického pozadia predpovedania, zhrnutím možností predikcie kvality aktív a odôvodnením výberu vysvetľovanej premennej.

ÚVOD

V druhej kapitole je popísaná teória lineárnej regresie a dvoch typov odhadu jej parametrov, metódy najmenších štvorcov a metódy maximálnej vierohodnosti. Venujeme sa v nej kritériám výberu modelu, konštrukcii predpovedí a problémom, ktoré sa v praxi vyskytujú najčastejšie, konkrétne multikolinearite, heteroskedasticite a autokorelácii rezíduí. Pri každom z nich sú stručne popísané metódy na identifikáciu, následky v prípade ignorovania a tiež možnosti riešenia.

Tretia kapitola je venovaná najmä dekompozičným metódam analýzy časových radov. Popisujeme v nej jednotlivé zložky rozkladu časového radu. Okrem eliminácie matematickými krivkami v tretej kapitole definujeme aj adaptívne prístupy k trendovej zložke. Detailnejšie sa pritom venujeme metóde kubického vyhladzovacieho splajnu, exponenciálnemu vyhladzovaniu a Holtovej metóde. Tretia kapitola je zakončená popisom teoretického základu Kalmanovho filtra.

Vo štvrtej kapitole sa zaoberáme hlavne Boxovou – Jenkinsovou metodológiou. Sú v nej definované pojmy stacionarity, autokorelačnej a parciálnej autokorelačnej funkcie. Ďalej je v nej popísaný proces kĺzavých súčtov, autoregresný proces a tiež zmiešaný proces. V tejto kapitole sa venujeme aj testom jednotkového koreňa, ARIMA procesom a konštrukcii predpovedí. Na záver sa stručne zmienujeme o kointegrácii časových radov, Grangerovej kauzalite a o ich testovaní.

Piata kapitola popisuje reálne dáta, ktoré sme použili pri vývoji predikčných modelov. Šiestu kapitolu začíname modelom lineárnej regresie pre prvú časť portfólia, pre ktorú sú v tejto kapitole vyvinuté aj štyri ďalšie modely časových radov. Pri druhej časti portfólia postupujeme podobne, keď po vytvorení štyroch modelov časových radov pristupujeme v siedmej kapitole aj ku konštrukcii troch modelov lineárnej regresie, s využitím makroekonomických vysvetľujúcich premenných.

Vyvinuté predikčné modely pre obidve časti portfólia validujeme viackrokovou predpoveďou a jednokrokovými predpoveďami na testovacej vzorke dát, keď kvalitu týchto predpovedí vzájomne porovnávame pomocou niekoľkých mier. Pre každú časť portfólia si na základe porovnania vyberáme jeden model na dlhodobé a jeden model na krátkodobé predpovede. Ich presnosť v praxi testujeme pomocou predikcie pre tri mesiace, ktoré neboli súčasťou vývojovej vzorky dát.

1 Ukazovatele kvality aktív a ich predpovedanie

1.1 Komerčná banka a jej aktíva

Komerčná banka môže byť podľa [60] charakterizovaná ako najdôležitejšia finančná inštitúcia v ekonomike, ktorá prijíma vklady, poskytuje úvery a spomedzi všetkých finančných inštitúcií ponúka najširšiu škálu služieb. Komerčná banka je však v prvom rade podnikateľský subjekt, ktorého základným cieľom je snaha o dosahovanie maximálneho zisku.

Ako sa uvádza v [8], každá komerčná banka pôsobiaca na území Slovenskej republiky je súčasťou dvojstupňového bankového systému, ktorý štandardnou formou usporiadania bankovníctva v trhových ekonomikách. Jeho základnou charakteristikou je, že komerčné banky a centrálna banka, ktorej funkciu zastáva Národná banka Slovenska, sú oddelené inštitúcie a majú navzájom odlišné funkcie a ciele.

Operácie, ktoré komerčná banka vykonáva, môžeme rozdeliť na aktívne a pasívne. Podľa [23] je pri aktívnych operáciách banka v postavení veriteľa, to znamená, že jej vznikajú pohľadávky, prípadne vlastnícke práva. Najpodstatnejšiu časť aktív predstavujú vo väčšine komerčných bánk úverové produkty, ktoré majú významný podiel na výnosoch, no sú veľmi podstatné aj z hľadiska bankových rizík.

Za úverové produkty komerčnej banky môžeme podľa [22] považovať tie produkty, ktoré umožňujú klientom určitý spôsob financovania. Poskytujú sa na základe úverovej zmluvy, v ktorej sa banka zaväzuje v dohodnutý termín poskytnúť klientovi peňažné prostriedky. Klient sa, naopak, zaväzuje poskytnuté peňažné prostriedky vrátiť za dohodnutých podmienok, vrátane úrokov.

1.2 Retailové portfólio komerčnej banky

Portfólio úverových produktov, ktoré ponúkajú komerčné banky svojim klientom, je veľmi široké. Podľa [57] je zo všetkých finančných inštitúcií práve ich portfólio najviac diverzifikované. Množstvo úverov sa okrem spoločností, podnikateľov či iných bánk poskytuje aj domácnostiam. Portfólio úverov poskytnutých domácnostiam nazývame retailovým portfóliom komerčnej banky.

Retailové portfólio komerčnej banky môžeme rozdeliť na dve základné časti, na časť zabezpečenú a časť nezabezpečenú. Do zabezpečenej časti retailového portfólia

patria hypotekárne úvery, ktoré sú zvyčajne kryté nehnuteľnosťou. Nezabezpečená časť sa skladá z troch typov bankových produktov, z ktorých ani jeden nebýva týmto spôsobom krytý. Patria do nej spotrebné úvery, kontokorentné úvery a kreditné karty.

1.2.1 Hypotekárny úver

Za základný charakteristický znak hypotekárneho úveru možno považovať spôsob jeho zabezpečenia, ktorým je záložné právo na nehnuteľnosť. Z povahy tohto zabezpečenia vyplýva relatívne nízke riziko nesplatenia a možnosť rozloženia splátok na dlhšie časové obdobie. Spravidla preto ide o dlhodobý úver s dobou splatnosti na viac rokov.

Podľa [22] však musí platiť zásada, že splatnosť hypotekárneho úveru nemôže byť dlhšia ako životnosť založenej nehnuteľnosti. Banka tiež musí brať do úvahy schopnosť klienta splácať úver po celú dobu jeho splatnosti. Doba splatnosti úveru je často ovplyvnená aj ekonomickou výhodnosťou pre klientov. Ak je vyššia, klesá výška splátok hypotekárneho úveru, ale zároveň sa zvyšuje celkový objem úrokov.

1.2.2 Spotrebný úver

Spotrebný úver slúži na nákup tovarov a služieb. Ako sa uvádza v [25], rozlišujeme dva typy spotrebných úverov. Účelový spotrebný úver je určený na presne špecifikovaný účel, ktorý je dohodnutý v úverovej zmluve. S tým je spojená povinnosť klienta v stanovenej lehote zdokladovať použitie úveru. Ak si dlžník túto povinnosť nesplní, môže banka pristúpiť k sankciám, napríklad k navýšeniu úrokovej sadzby.

Naopak, pri bezúčelovom spotrebnom úvere banka dokladovanie použitia finančných prostriedkov od klienta nepožaduje. Klient môže peniaze použiť prakticky na čokoľvek, aj na rovnaký účel ako pri účelovom spotrebnom úvere, ale v tomto prípade nemusí banke použitie úveru dokladovať. V porovnaní s účelovým spotrebným úverom je bezúčelový spotrebný úver zvyčajne drahší.

1.2.3 Kontokorentný úver

Kontokorentný úver sa používa na krytie opakujúcich sa krátkodobých potrieb. Jeho účelom je podľa [58] preklopenie časového nesúladu medzi príjmami a výdavkami klienta. Umožňuje opakované čerpanie a splácanie peňažných prostriedkov podľa aktuálnej potreby, až do výšky stanoveného úverového limitu stanoveného komerčnou bankou.

1 UKAZOVATELE KVALITY AKTÍV A ICH PREDPOVEDANIE

Tento úver banka poskytuje iba klientom s výbornou platobnou disciplínou, ktorí si v minulosti plnili všetky predchádzajúce záväzky v stanovených termínoch a výške. Kontokorentný úver sa v praxi zvykne nazývať aj povolené prečerpanie na účte, keďže sa poskytuje na bežnom účte klienta. Tento účet sa vzápätí stáva debetno-kreditným, teda kontokorentným účtom.

1.2.4 Kreditná karta

Ako sa uvádza v [25], pri čerpaní prostriedkov z kreditnej karty je čerpaná pôžička, ktorú neskôr držiteľ kreditnej karty neskôr spláca. Môže sa jednať o platbu za tovary a služby, prípadne pri výber z bankomatu. Kreditné karty nie sú viazané na bežné účty. Čerpaný úver sa nazýva revolvingový, teda opakujúci sa, pretože stanovený úverový limit sa s každou uskutočnenou splátkou úveru automaticky obnovuje.

Kreditné karty sú zaujímavé predovšetkým vďaka tzv. bezúročnému obdobiu, počas ktorého klient za čerpané prostriedky neplatí úrok. Toto obdobie obvykle trvá jeden až dva mesiace. Ak však počas nich klient svoj dlh neuhradí, začínajú sa dlžné prostriedky úročiť. Úrokové sadzby pri kreditných kartách sú podstatne vyššie ako pri spotrebných alebo kontokorentných úveroch.

1.3 Finančné riziko a jeho riadenie

Podľa [42] môže byť finančné riziko vo všeobecnosti definované ako potenciálna budúca strata, ktorá sa viaže k určitému aktívu alebo portfóliu aktív. Je pri ňom nevyhnutné rozlišovať pravdepodobnosť toho, že udalosť nastane, a tiež veľkosť jej potenciálneho vplyvu. Finančné riziko môžeme rozdeliť na štyri základné typy, v závislosti od druhu neistoty, ktorý vyjadrujú.

Prvým typom finančného rizika, ktorému je komerčná banka vystavená, je trhové riziko. Vo všeobecnosti zahŕňa riziko straty v dôsledku poklesu trhových cien aktív, ktoré má banka v portfóliu. K tomuto poklesu môže podľa [16] dôjsť z viacerých príčin, napríklad kvôli pohybu úrokových sadzieb (úrokové riziko), výmenných kurzov (menové riziko), cien akcií (akciové riziko) či komodít (komoditné riziko).

Komerčná banka je ďalej vystavená riziku likvidity. Je to riziko straty v dôsledku momentálneho nedostatku hotových finančných prostriedkov. Ako sa uvádza v [16], ďalej sa toto riziko delí na riziko trhovej likvidity a riziko financovania. Prvý typ rizika vyjadruje riziko straty v prípade nedostatočnej aktivity trhu na potrebný

1 UKAZOVATELE KVALITY AKTÍV A ICH PREDPOVEDANIE

prístup k hotovosti a druhý riziko straty v prípade momentálnej platobnej neschopnosti v dôsledku nesúladu vo finančných tokoch.

Operačné riziko je spomedzi štyroch základných typov rizika najťažšie merateľné. Podľa [42] zahŕňa najmä riziko straty v dôsledku zlyhania ľudského faktora alebo interných systémov banky, prípadne v dôsledku nesprávne nastavených bankových procesov. Okrem toho k operačnému riziku zaraďujeme aj riziko straty v dôsledku interného alebo externého podvodu, či živeľnej pohromy.

Kreditné riziko by sme vo všeobecnosti mohli charakterizovať ako riziko straty v dôsledku neschopnosti alebo neochoty zmluvného partnera splatiť svoje pohľadávky načas a v plnej výške, podľa vopred dohodnutých podmienok. Ako sa uvádza v [43], pod pojem kreditného rizika patria aj riziko suverenity, riziko koncentrácie, riziko vypořádania a riziko protistrany.

Ako sa uvádza v [42], okrem štyroch základných typov rizika musí komerčná banka brať do úvahy aj ďalšie typy rizík. Patria sem riziko straty v dôsledku strategicky zlého obchodného rozhodnutia (strategické riziko), právneho problému (právne riziko), zlej reputácie (reputačné riziko), nevhodného použitia modelu (riziko modelu), kolapsu celého finančného systému alebo celého trhu (systémové riziko).

Pri podnikaní musí komerčná banka dbať o dlhodobú stabilitu a bezpečnosť. Špeciálne pravidlá, ktoré sa na jej podnikanie vzťahujú, podľa [37] vyplývajú aj z toho, že komerčné banky vo svojej podstate obchodujú so zverenými peňažnými prostriedkami. Vyžadujú si preto prísnejšie pravidlá a dohľad nad svojou činnosťou. Neoddeliteľnou súčasťou fungovania komerčnej banky preto musí byť riadenie rizík.

Proces riadenia rizík sa podľa [42] skladá z troch hlavných častí, konkrétne z identifikácie, merania a samotného riadenia rizík. Identifikácia rizík sa zameriava na rozpoznanie rizikových faktorov, ktorým je banka vystavená. Každá banková transakcia totiž so sebou nesie niekoľko typov rizika. Mnohé z nich, ktoré za normálnych okolností považujeme za vedľajšie, sa napríklad v období krízy môžu ukázať ako kľúčové.

Riadenie rizík vo všeobecnosti zahŕňa nastavenie rizikových limitov, ktoré je založené na určenom rizikovom apetíte, spätnom testovaní (testovaní kvality modelov na historických dátach) a stresovom testovaní (hodnotení vplyvu niekoľkých veľmi vážnych, no do určitej miery pravdepodobných scenárov). S meraním kreditného rizika retailového portfólia komerčnej banky úzko súvisia ukazovatele kvality aktív.

1.4 Ukazovatele kvality aktív

Ukazovatele kvality aktív sú indikátory, ktoré banka sleduje a vyhodnocuje v rámci sledovania kvality portfólia na pravidelnej báze, najčastejšie mesačne. Sledovaním a vyhodnocovaním kvality portfólia sa obvykle zaoberá samostatné oddelenie v rámci riadenia rizík. Ukazovatele sú merané na úrovni celého portfólia, ale aj na úrovni jednotlivých aktív, a môžu veľa napovedať o ich kvalite.

1.4.1 Miera omeškania

Stav, keď má klient voči banke finančný záväzok (istinu, úroky alebo iné poplatky) po lehote splatnosti, sa nazýva delikvencia. Jedná sa o porušenie zmluvných podmienok klientom, napríklad kvôli nepovolenému prekročeniu limitu povoleného prečerpania na bežnom účte alebo nedodržaniu splátkového kalendára (napríklad nezaplatenie splátky istiny, nezaplatenie úrokov v stanovený termín alebo iné).

Miera omeškania predstavuje percentuálny podiel sumy zostatkov v delikvencii k sume celkového zostatku daného portfólia. Pre vyhodnocovanie kvality retailového portfólia sa najčastejšie používa miera omeškania 90–360 (DLQ_{90-360}), ktorá hovorí o úveroch v delikvencii 90 až 360 dní. Najväčšiu výpovednú hodnotu má miera omeškania jednotlivých častí portfólia. Vypočítať sa dá pomocou vzorca

$$DLQ_{90-360} = \frac{\sum ZOSTATOK(90+) - \sum ZOSTATOK(360+)}{\sum ZOSTATOK(PORTFÓLIO) - \sum ZOSTATOK(360+)} \cdot 100\% .$$

Okrem miery omeškania 90–360 sa pri sledovaní kvality aktív sústreďuje značná pozornosť aj na mieru omeškania 30–360 (DLQ_{30-360}), ktorá v sebe nesie aj informáciu o klientoch, ktorí sú v delikvencii 30 až 89 dní. Nezanedbateľná časť týchto klientov totiž v nasledujúcich mesiacoch môže prejsť do delikvencie viac ako 90 dní. Vyjadriť sa dá ako

$$DLQ_{30-360} = \frac{\sum ZOSTATOK(30+) - \sum ZOSTATOK(360+)}{\sum ZOSTATOK(PORTFÓLIO) - \sum ZOSTATOK(360+)} \cdot 100\% .$$

Miera omeškania 90+ (DLQ_{90+}) hovorí aj o úveroch v omeškani viac ako rok. Jej hodnota sa počíta vzorcom

$$DLQ_{90+} = \frac{\sum ZOSTATOK(90+)}{\sum ZOSTATOK(PORTFÓLIO)} \cdot 100\% .$$

1.4.2 Miera zlyhania

Za vznik zlyhania v súvislosti s konkrétnym úverom sa považuje, ak konkrétny úver dlžníka je v omeškaní v súvislosti s plnením záväzku voči banke viac ako 90 dní, prípadne banka usúdi, že dlžník si na konkrétnom úvere pravdepodobne nesplní svoj záväzok voči nej bez toho, aby prišlo napríklad k realizácii zabezpečenia alebo inej významnej komplikácii.

Miera zlyhania (*DEFAULT*) je definovaná ako podiel sumy zostatkov zlyhaných zmlúv k celkovej sume zostatkov daného portfólia. Pre vyhodnocovanie kvality retailového portfólia je rovnako dôležitá ako miera omeškania. Miera zlyhania sa vyjadruje aj pre celé portfólio, no väčšiu výpovednú hodnotu má zvyčajne miera zlyhania jeho jednotlivých častí. Počíta sa vzorcom

$$DEFAULT = \frac{\sum ZOSTATOK(DEFAULT)}{\sum ZOSTATOK(PORTFÓLIO)} \cdot 100\% .$$

1.5 Predpovedanie

Predpovedanie je odhadovanie budúcej hodnoty premennej pre určitý časový okamih v budúcnosti. Podľa [4] sa zvyčajne vykonáva s úmyslom pomôcť budúcemu rozhodovaniu a plánovaniu. Vychádza z predpokladu, že ak sme schopní predpovedať, čo sa udeje v budúcnosti, môžeme tomu prispôsobiť naše súčasné správanie s cieľom byť v budúcnosti v čo najlepšej pozícii.

Podľa [13] takmer každá aktivita, ktorá sa týka vypožičiavania peňazí, v sebe zahŕňa určitú formu predpovedania. Pri retailovom portfóliu komerčnej banky je cieľom predpovedať mesačný vývoj portfólia na niekoľko mesiacov vopred. Predpovede pritom môžeme rozdeliť do niekoľkých hlavných kategórií. V nasledujúcich častiach podkapitoly si uvedieme delenie podľa [16].

1.5.1 Bodová a intervalová predpoveď

Bodová predpoveď je hodnota, ktorá predstavuje odhad budúcej hodnoty uvažovaného časového radu v určitom, presne stanovenom, časovom okamihu. Bodová predpoveď je vždy zaťažená chybou, preto je potrebné brať konkrétnu hodnotu, ktorú poskytuje, s rezervou. Z praktického pohľadu sú preto pre používateľa zrejme zaujímavejšie intervalové ako bodové predpovede.

1 UKAZOVATELE KVALITY AKTÍV A ICH PREDPOVEDANIE

Intervalová predpoveď je interval, ktorý je analógiou intervalu spoľahlivosti zo štatistiky. Rozdiel je v tom, že namiesto neznámeho parametra sa pri ňom odhaduje budúca hodnota časového radu. Napríklad 95-percentný predikčný interval udáva dolnú a hornú hranicu, medzi ktorými bude ležať budúca hodnota časového radu s pravdepodobnosťou 95 percent.

1.5.2 Jednokroková a viackroková predpoveď

Jednokroková predpoveď je skonštruovaná pre jedinú, zvyčajne nasledujúce, pozorovanie. Využiť sa dá, pokiaľ našim cieľom nie je dlhodobá, ale čo najpresnejšia predpoveď, a tiež na validáciu predikčného modelu. V tomto prípade sa postupuje konštrukciou postupnosti predpovedí pre viac pozorovaní. Tieto predpovede sa konštruujú vždy o jeden krok dopredu, na základe skutočných hodnôt.

Viackroková predpoveď je predpoveď pre viac, zvyčajne nasledujúcich, pozorovaní súčasne. Jedná sa o množinu predpovedaných hodnôt, ktoré sú všetky skonštruované v tom istom čase. Viackrokovou predpoveďou preto nie je napríklad spomínaná postupnosť predpovedí pre viac pozorovaní konštruovaných postupne v čase o jeden krok dopredu na základe skutočných hodnôt.

1.5.3 Statická a dynamická predpoveď

Toto rozdelenie sa používa v prípadoch, kedy medzi vysvetľujúcimi premennými pre predpovedanú vysvetľovanú premennú figurujú aj oneskorené hodnoty tejto vysvetľovanej premennej. Táto situácia nastáva napríklad v autoregresných modeloch. Pokiaľ sú skutočné oneskorené hodnoty k dispozícii, statická predpoveď ich využíva ako oneskorené hodnoty predpovedanej vysvetľovanej premennej.

Naopak, dynamická predpoveď nevyužíva známe oneskorené hodnoty vysvetľovanej premennej ani v prípade, ak sú už známe, ale dosadí namiesto nich predchádzajúce výsledky predpovede. Dynamická predpoveď by preto v porovnaní so statickou predpoveďou mala byť vo väčšine prípadov menej presná. Na druhej strane však umožňuje konštruovať predpovede na dlhšie časové obdobie.

1.5.4 Predpoveď vo vzorke a mimo vzorky

Predpoveď vo vzorke je predpoveď pre dáta, ktoré boli použité pri konštrukcii predikčného modelu. V tomto prípade sa preto dajú očakávať dobré predikčné výsledky, keďže v predpovedanom horizonte zostávajú v platnosti predpoklady modelu. Môže sa

používať pri spätnom testovaní odhadnutých parametrov modelu, ale zvyčajne je vhodnejšie použiť pre tento účel predpoveď mimo vzorky.

Predpoveď mimo vzorky je predpoveďou dát, ktoré sa na konštrukcii predikčného modelu nezúčastnili. V dobe konštrukcie modelu totiž buď neboli k dispozícii, alebo boli z pôvodnej vzorky zámerne odstránené za účelom posúdenia predikčnej schopnosti modelu. Takto odstránené dáta sa označujú ako utajená vzorka. Predpoveď mimo vzorky je zvyčajne menej presná ako predpoveď vo vzorke.

1.5.5 Predpoveď v ekonometrickom modeli a v modeli časového radu

Predpoveď v ekonometrickom modeli dáva do vzťahu budúcu hodnotu vysvetľovanej premennej a budúce hodnoty vysvetľujúcich premenných. Takéto modely môžu pri existencii dlhodobých vzťahov medzi premennými fungovať uspokojivo aj pre dlhodobé predpovede. Na druhej strane pri nich môže vzniknúť problém s dostupnosťou predpovedí príslušných vysvetľujúcich premenných.

Predpoveď v modeli časového radu konštruuje predpoveď ako autoprojekciu minulých a súčasných hodnôt časového radu do budúcnosti. Môže byť založená na dekompozičných alebo autokorelačných vlastnostiach časového radu. Jej hlavnou výhodou je, že pre konštrukciu predpovedí využíva iba minulé hodnoty vysvetľovanej premennej a nevyžaduje si predpovede iných vysvetľujúcich premenných.

1.6 Miery presnosti predpovede

Chyba predpovede \hat{y}_t skutočnej hodnoty y_t je definovaná ako $e_t = y_t - \hat{y}_t$. Ďalej si uvedieme si miery, ktoré sa podľa [16] najčastejšie používajú na ohodnotenie kvality skonštruovaných predpovedí. Sú vhodné na súhrnné posúdenie presnosti predpovedaných hodnôt, pričom nezáleží na tom, či sa jedná o viackrokové alebo jednokrokové, statické alebo dynamické predpovede.

Súčet štvorcových chýb *SSE*

$$SSE = \sum_{t=T+1}^{T+h} (y_t - \hat{y}_t)^2 = \sum_{t=T+1}^{T+h} (e_t)^2$$

je analógiou ku kritériu najmenších štvorcov v regresných modeloch. Stredná štvorcová chyba *MSE*

$$MSE = \frac{1}{h} \cdot \sum_{t=T+1}^{T+h} (y_t - \hat{y}_t)^2 = \frac{1}{h} \cdot \sum_{t=T+1}^{T+h} (e_t)^2$$

je často používanou kvadratickou chybovou funkciou. Odmocninová priemerná štvorcová chyba *RMSE*

$$RMSE = \sqrt{\frac{1}{h} \cdot \sum_{t=T+1}^{T+h} (y_t - \hat{y}_t)^2} = \sqrt{\frac{1}{h} \cdot \sum_{t=T+1}^{T+h} (e_t)^2}$$

vzniká modifikáciou *MSE*, aby bola meraná v rovnakých jednotkách ako daný časový rad. Ako sa uvádza napríklad v [31] a [32], veľmi často sa využíva aj v bankovom prostredí.

Priemerná absolútna chyba *MAE*

$$MAE = \frac{1}{h} \cdot \sum_{t=T+1}^{T+h} |y_t - \hat{y}_t| = \frac{1}{h} \cdot \sum_{t=T+1}^{T+h} |e_t|$$

nepenalizuje veľké chyby tak výrazným spôsobom ako *MSE*, preto ju niektorí autori odporúčajú pre ohodnocovanie chýb predpovedí v časových radoch s odľahlými pozorovaniami. Priemerná absolútna percentuálna chyba *MAPE*

$$MAPE = \frac{100}{h} \cdot \sum_{t=T+1}^{T+h} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

obvykle nadobúda hodnoty od 0 do 100 percent. Výsledok menší ako 100 percent znamená, že daný model predpovedá lepšie ako model s konštantnými predpoveďami na úrovni nuly, ktorý má *MAPE* rovné 100 percent.

Percento správnych jednokrokových predpovedí zmien v smere vývoja *PSP*

$$PSP = \frac{100}{h} \cdot \sum_{t=T+1}^{T+h} z_t,$$

kde $z_t = 1$ ak $(y_t - y_{t-1})(\hat{y}_t - \hat{y}_{t-1}) > 0$, inak $z_t = 0$, hovorí o úspešnosti predpovedania zmien vo vývoji vysvetľovanej premennej.

1.7 Možnosti predpovedania ukazovateľov kvality aktív

Výskum v oblasti predpovedania kreditného rizika začal podľa [18] výrazne nabrať na intenzite od začiatku deväťdesiatych rokov minulého storočia. Z hľadiska významu je to podľa autorov článku jedna z najdôležitejších aplikácií predpovedania v praxi. Modelovanie kreditného rizika a jeho predpovedanie môže prebiehať na úrovni jednotlivých klientov, úverov alebo na úrovni portfólia úverov.

1 UKAZOVATELE KVALITY AKTÍV A ICH PREDPOVEDANIE

Ako sa uvádza v [18], pri predpovedaní platobného zlyhania klienta sa sprvoti využívala lineárna diskriminačná analýza. Pokroky v oblasti výpočtovej techniky však neskôr viedli k používaniu logistickej regresie, ktorá je v praxi aj dnes jedným z najpoužívanejších nástrojov modelovania kreditného rizika na úrovni klientov. Viac o vývoji modelovania kreditného rizika sa dá dozvedieť napríklad z [11], [19] či [66].

Vďaka spomínaným pokrokom v oblasti výpočtovej techniky sa postupne, spolu s logistickou regresiou, dostávali do popredia aj ďalšie klasifikačné metódy, napríklad neurónové siete, klasifikačné stromy, či zhlukové analýzy. Používaniu týchto metód a porovnávaniu kvality ich predpovedí sa venuje množstvo článkov, okrem iných napríklad [2], [7], [34] a [62].

Čo sa týka predpovedania platobného zlyhania klientov, značná pozornosť sa v poslednej dobe venuje aj modelom prežitia, ktoré majú oproti iným modelom tú výhodu, že pri nich nemusí byť striktno určená časová perióda, v ktorej k javu dôjde. Vychádzal z nich aj článok [47]. Taktiež v nich môžu byť zohľadnené premenné závislé na čase, teda napríklad makroekonomické ukazovatele.

Aplikácia modelov prežitia do praxe pri predpovedaní platobného zlyhania klientov bola bližšie popísaná v [3], [54] a [64]. V týchto článkoch bola zároveň schopnosť modelov prežitia predpovedať platobné zlyhanie klientov porovnaná s dovedty používanými metódami. Zahrnutie makroekonomických premenných do modelu bolo predmetom článkov [5] a [6].

Ďalšou metódou, ktorá sa dá použiť na predpovedanie na úrovni jednotlivých klientov, sú Markovove modely. Oproti modelom, ktoré sme si spomínali vyššie, majú tú výhodu, že môžu fungovať viacstavovo, teda okrem budúceho platobného zlyhania klienta dokážu predpovedať aj to, v ktorej kategórii delikvencie sa klient v niektorom časovom okamihu v budúcnosti ocitne.

Markovovým modelom a ich aplikácii v oblasti kreditného rizika je venovaná pozornosť v [9], [63], [50], ale predovšetkým v [30]. Pri odhadovaní pravdepodobností prechodu medzi jednotlivými stavmi boli v tomto prípade použité najmä údaje o úvere a údaje o klientovi. O zahrnutie makroekonomických ukazovateľov sa pokúsili v článkoch [53] a [26], v ktorých pracovali s viacstavovými Markovovými modelmi.

Podľa [18] sa oveľa menej pozornosti venovalo predpovedaniu miery omeškania na úrovni portfólií. Výnimkou bol článok [17], v ktorom autori odhadli modely určené

1 UKAZOVATELE KVALITY AKTÍV A ICH PREDPOVEDANIE

na predpovedanie miery omeškania úverového portfólia komerčných bánk pomocou kointegračnej analýzy a VAR modelov. Tieto modely následne porovnali s *ARIMA* modelmi. Zistili, že ich predikčné schopnosti boli takmer identické.

Alternatívou k tomuto postupu by mohol byť model lineárnej regresie, ktorého najväčšia výhoda spočíva v jednoduchom zahrnutí makroekonomických premenných a tiež veľmi zrozumiteľnej interpretácii výsledkov. Jeho použitie má však v prípade predpovedania ukazovateľov kvality aktív na úrovni portfólia a dát v tvare časových radov isté špecifiká, ktorým sa venuje napríklad [21].

Čo sa týka modelov časových radov, okrem *ARIMA* modelov, ktoré patria do skupiny modelov využívajúce autokorelačné vlastnosti časových radov, sa pre účely predpovedania ukazovateľov kvality aktív dajú využiť aj dekompozičné metódy analýzy časových radov, medzi ktoré patrí napríklad metóda vyhladzovacieho splajnu, ktorej je venovaná pozornosť v [67].

1.8 Výber vysvetľovanej premennej

Naším cieľom bolo zvoliť si ako vysvetľovanú premennú predikčného modelu na úrovni portfólia ten ukazovateľ, ktorý má pri sledovaní kvality aktív retailového portfólia najväčší význam. Nakoniec sme sa rozhodli pre mieru omeškania, ktorá vyjadruje aktuálny podiel zostatkov v omeškaní, a pri sledovaní kvality portfólia sa na ňu v banke sústreďuje veľká pozornosť.

Ako sme už v tejto kapitole uviedli, rozlišujeme rôzne typy miery omeškania. Pre účely sledovania kvality aktív retailového portfólia je najzaujímavejším typom miera omeškania v rozmedzí 90 až 360 dní. V prípade omeškania pod 90 dní sa totiž klient ešte často krát vráti k riadnemu splácaniu, no a v prípade omeškania nad 360 dní už banka počíta s definitívnym zlyhaním splácania klientovho úveru.

2 Model lineárnej regresie v praxi

Ako sa uvádza v [41], lineárna regresia vysvetľuje zmeny hodnôt vysvetľovanej premennej zmenami hodnôt vysvetľujúcich premenných. Vysvetľovaná premenná sa obvykle označuje písmenom y a vysvetľujúce premenné písmenami x_1, x_2, \dots, x_k . Model lineárnej regresie sa dá formálne zapísať v tvare

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t,$$

kde $t=1, \dots, T$ vyjadruje čas, y_t je hodnota vysvetľovanej premennej y v čase t , $x_{t1}, x_{t2}, \dots, x_{tk}$ sú hodnoty vysvetľujúcich premenných v čase t , $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ sú neznáme parametre modelu a ε_t je jeho reziduálna zložka.

2.1 Metóda najmenších štvorcov

Zrejme najčastejšie používanou metódou odhadu neznámych parametrov modelu je postup založený na metóde najmenších štvorcov (*MNS*). Ako sa uvádza v [29], táto metóda hľadá odhady parametrov $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ tak, že vzhľadom na ne minimalizuje súčet štvorcov

$$\sum_{i=1}^T (y_t - (\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk}))^2.$$

Riešením optimalizačnej úlohy sú odhady $b_0, b_1, b_2, \dots, b_k$ parametrov $\beta_0, \beta_1, \beta_2, \dots, \beta_k$, ktoré sa zvyknú nazývať aj *MNS*-odhady $b = (X^T X)^{-1} X^T y$, kde X je matica vysvetľujúcich premenných, vrátane konštantného člena.

Na základe *MNS*-odhadov vieme vypočítať *MNS*-hodnoty

$$\hat{y}_t = b_0 + b_1 x_{t1} + b_2 x_{t2} + \dots + b_k x_{tk},$$

a pomocou *MNS*-hodnôt sa vieme dopracovať k *MNS*-rezíduám

$$\hat{\varepsilon}_t = y_t - \hat{y}_t = y_t - (b_0 + b_1 x_{t1} + b_2 x_{t2} + \dots + b_k x_{tk}).$$

Okrem parametrov $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ obsahuje model lineárnej regresie aj neznámy parameter $\sigma^2 > 0$. Za jeho *MNS*-odhad sa podľa [29] obvykle berie výraz

$$s^2 = \frac{\sum_{i=1}^T \hat{\varepsilon}_t^2}{T - k},$$

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

pričom za $MN\check{S}$ -odhad štandardnej odchýlky $\sigma > 0$ sa potom považuje odmocnina z tohto odhadu v tvare

$$s = \sqrt{\frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T-k}}.$$

Podľa [1] sa dá ukázať, že $MN\check{S}$ -odhady b_i sú nestrannými odhadmi parametrov β_i a $MN\check{S}$ -odhad s^2 je nestranným odhadom parametra σ^2 . $MN\check{S}$ -odhad s sa navyše aproximatívne považuje za nestranný odhad parametra σ a s_{b_i} za nestranný odhad štandardnej odchýlky odhadu b_i parametra β_i .

Ako sa uvádza v [16], $MN\check{S}$ -odhady b_i sú v modeli lineárnej regresie všeobecne najlepšimi nestrannými lineárnymi odhadmi parametrov β_i , pokiaľ je stredná hodnota reziduálnej zložky nulová pre všetky t

$$(I) E(\varepsilon_t) = 0,$$

rozptyl reziduálnej zložky je konštantný a konečný pre všetky t

$$(II) \text{var}(\varepsilon_t) = \sigma^2 < \infty,$$

reziduálne zložky sú navzájom nekorelované všetky $s \neq t$

$$(III) \text{cov}(\varepsilon_s, \varepsilon_t) = 0,$$

a vysvetľujúce premenné sú navzájom tiež korelované

$$(IV) h(X) = k.$$

O normálnom modeli môžeme podľa [16] hovoriť v situácií, keď navyše platí aj predpoklad normálneho rozdelenia reziduálnych zložiek pre všetky t

$$(V) \varepsilon_t \sim N(0, \sigma^2).$$

2.2 Metóda maximálnej vierohodnosti

Ako sa uvádza v [41], k odvodeniu odhadu parametrov modelu lineárnej regresie metódou maximálnej vierohodnosti MMV je potrebné využitie vierohodnostnej funkcie

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{T}{2}} \cdot \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2}\right\},$$

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

prípadne logaritmickéj vierohodnostnej funkcie, ktorá má jednoduchší tvar

$$l(\beta, \sigma^2) = \ln L(\beta, \sigma^2) = -\frac{T}{2} \cdot \ln(2\pi) - \frac{T}{2} \cdot \ln(\sigma^2) - \frac{1}{2\sigma^2} \cdot (y - X\beta)^T (y - X\beta).$$

MMV -odhady následne nájdeme minimalizáciou $l(\beta, \sigma^2)$. Keďže je táto funkcia konkávna, pri jej maximalizácii podľa [41] stačí nájsť nulový bod príslušných parciálnych derivácií

$$\frac{\partial}{\partial \beta} l(\beta, \sigma^2) = -\frac{1}{\sigma^2} \cdot (-X^T y + X^T X \beta),$$

$$\frac{\partial}{\partial \sigma^2} l(\beta, \sigma^2) = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \cdot (y - X\beta)^T (y - X\beta).$$

Ako sa uvádza v [16], *MMV* -odhad parametrov β

$$b = (X^T X)^{-1} X^T y$$

je zhodný s *MNS'* -odhadmi parametrov β a *MMV* a *MNS'* -odhady parametra σ^2 sú asymptoticky zhodné, keďže *MMV* -odhad σ^2 má tvar

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T} = \frac{T-k}{T} \cdot s^2.$$

V [36] je uvedené, že *MMV* -odhad parametra σ^2 je asymptoticky nestranný a *MMV* -odhady všetkých parametrov modelu sú asymptoticky efektívne, čo znamená, že ich rozptyly sa pre veľké hodnoty počtu pozorovaní T blížia k dolným hraniciam pre rozptyly daných parametrov.

2.3 Kritériá výberu modelu

Pri vytváraní modelu lineárnej regresie autor zažije veľa situácií, pri ktorých musí urobiť dôležité rozhodnutia. Či už pri pridaní vysvetľujúcich premenných do modelu, vynechaní niektorých vysvetľujúcich premenných, ktoré v modeli vystupujú, alebo pri ich prípadnej transformácii. Aby sa autor nemusel rozhodovať iba na základe intuície, môže si pomôcť niektorým z nasledujúcich kritérií.

2.3.1 Koeficient determinácie

Keď model lineárnej regresie odhadneme, je nutné posúdiť, či je skutočne kompatibilný s použitými dátami. Toto sa dá vykonať pomocou štatistických testov alebo na základe koeficientu determinácie, ktorý sa podľa [41] definuje ako

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS},$$

kde $RSS = \sum_{i=1}^T \hat{\varepsilon}_i^2 = \sum_{i=1}^T (y_i - \hat{y}_i)^2$, $TSS = \sum_{i=1}^T (y_i - \bar{y})^2$ a $ESS = \sum_{i=1}^T (\hat{y}_i - \bar{y})^2$.

Ako sa uvádza v [41], v praxi sa niekedy dáva prednosť takzvanému korigovanému koeficientu determinácie v tvare

$$\bar{R}^2 = 1 - \left[\frac{T-1}{T-k} \cdot (1 - R^2) \right],$$

ktorý by mal penalizovať nadmerný počet vysvetľujúcich premenných k . Pravidlo konštrukcie modelu s maximálnym koeficientom determinácie je však, aj pri využití korigovaného koeficientu determinácie, priveľmi jednoduché a podľa [16] môže viesť k neúčinným modelom s priveľkým počtom vysvetľujúcich premenných.

2.3.2 Informačné kritériá

Ako sa uvádza v [16], informačné kritériá do určitej miery riešia problém nedostatočnej penalizácie rastúceho počtu vysvetľujúcich premenných pri korigovanom koeficiente determinácie. Ich hodnoty sú funkciou počtu vysvetľujúcich premenných k , počtu pozorovaní T a MNS -odhadu rozptylu rezíduí s_k^2 . Optimálny počet vysvetľujúcich premenných sa hľadá minimalizáciou informačného kritéria cez k . Najpoužívanejšími kritériami sú podľa [16] Akaikeho informačné kritérium AIC

$$AIC(k) = \ln s_k^2 + \frac{2k}{T}$$

a Bayesovo informačné kritérium BIC

$$BIC(k) = \ln s_k^2 + \frac{k \cdot \ln T}{T}.$$

Keďže penalizačný člen v Bayesovom informačnom kritériu má pre vyššie T väčšiu hodnotu než v Akaikeho informačnom kritériu, Bayesovo informačné kritérium viac inklinuje k výberu menších modelov. Podobne je na tom podľa [29] aj korigované

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

Akaikeho informačné kritérium $AICc$, ktoré v porovnaní s Akaikeho informačným kritériom používa ešte dodatočný penalizačný člen.

2.3.3 Iteračné výberové metódy

Ako sa uvádza v [29], pri výbere vysvetľujúcich premenných môžeme použiť aj niektorú z iteračných metód. Metóda výberu vpred začína s najmenším modelom, ktorý obsahuje iba konštantný člen, a priradí k nemu potenciálnu vysvetľujúcu premennú s najväčším t -pomerom. Tento postup opakuje dovtedy, kým žiadna z nezaradených potenciálnych vysvetľujúcich premenných nie je významná.

Metóda eliminácie späť využíva podľa [29] presne opačný postup ako metóda výberu vpred. Začína s najväčším modelom, ktorý obsahuje všetkých k potenciálnych vysvetľujúcich premenných, a vyradí z nich tú, ktorá má najmenší t -pomer. Tento postup opakuje tak dlho, až sú všetky vysvetľujúce premenné, ktoré z modelu neboli vyradené, významné.

Kroková regresia je podľa [16] kombináciou metódy výberu vpred a metódy eliminácie späť, keďže pri každej pridanej vysvetľujúcej premennej vyskúša, či sa nedá niektorá zo zostávajúcich vysvetľujúcich premenných vyradiť. Kroková regresia je preto najuniverzálnejšou metódou, ktorá sa pri výstavbe modelu lineárnej regresie v praxi používa.

2.4 Predpovedanie

V modeli lineárnej regresie sa za predpoveď považuje odhad hodnoty vysvetľovanej premennej y^* pri daných hodnotách vysvetľujúcich premenných $x_1^*, x_2^*, \dots, x_k^*$.

Ako sa uvádza v [29], o ex ante predpovedi hovoríme v prípade, že hodnoty vysvetľujúcich premenných $x_1^*, x_2^*, \dots, x_k^*$ musíme najprv nájsť. O ex post predpoveď sa jedná vtedy, sú hodnoty vysvetľujúcich premenných $x_1^*, x_2^*, \dots, x_k^*$ známe.

Bodová predpoveď má tvar

$$y^* = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \dots + \beta_k x_k^* + \varepsilon^*,$$

pričom v odhadnutom modeli lineárnej regresie platí

$$\hat{y}^* = b_0 + b_1 x_1^* + b_2 x_2^* + \dots + b_k x_k^*.$$

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

Ak sú podľa [16] splnené aj dodatočné predpoklady

$$(VI) E(\varepsilon^*) = 0,$$

$$(VII) \text{var}(\varepsilon^*) = \sigma^2 < \infty,$$

$$(VIII) \text{cov}(\varepsilon_t, \varepsilon^*) = 0,$$

potom je predpoveď \hat{y}^* nestranná, keď platí $E(y^* - \hat{y}^*) = 0$. Predpoveď \hat{y}^* je zároveň najlepšia spomedzi nestranných lineárnych predpovedí v zmysle, že jej chyba $e = y^* - \hat{y}^*$ má minimálnu štandardnú odchýlku

$$s_e = \sqrt{\text{var}(e)} = \sqrt{E(y^* - \hat{y}^*)^2}.$$

2.5 Špecifiká lineárnej regresie v praxi

Ako sa uvádza v [16], v praxi často dochádza k porušeniu predpokladov (I) – (V) pre model lineárnej regresie. Porušenie predpokladu $E(\varepsilon_t) = 0$ môže viesť k nestabilite modelu, kým porušenie predpokladu $\varepsilon_t \sim N(\mu, \sigma^2)$ sa dá eliminovať napríklad transformáciou premenných, použitím robustných metód, prípadne modelovaním odľahlých pozorovaní umelo vytvorenými premennými.

2.5.1 Heteroskedasticita

O heteroskedasticite sa hovorí, ak je porušený predpoklad homoskedasticity

$$(II) \text{var}(\varepsilon_t) = \sigma^2 < \infty.$$

Keďže grafické rozpoznanie heteroskedasticity je použiteľné najmä vtedy, ak poznáme jej príčinu, v praxi sa jej prítomnosť v modeli testuje napríklad Goldfeldovým – Quandtovým, Whiteovým alebo Breuschovým – Paganovým testom.

Ako je uvedené v [29], Breuschov – Paganov test predpokladá testovaný model lineárnej regresie v obvyklom tvare

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t,$$

pričom jeho nulovou hypotézou je

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_T^2.$$

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

Alternatívna hypotéza má pre všetky $t = 1, \dots, T$ tvar

$$H_1 : \sigma_i^2 = \delta_0 + \delta_1 z_{t1} + \delta_2 z_{t2} + \dots + \delta_p z_{tp},$$

kde niektoré vysvetľujúce premenné pomocného modelu z_1, z_2, \dots, z_p môžu byť podmnožinou vysvetľujúcich premenných testovaného modelu a $\delta_0, \delta_1, \delta_2, \dots, \delta_p$ sú neznáme parametre.

Testovacia štatistika Breuschovho – Paganovho testu má podľa [29] tvar

$$\lambda_{BP} = \frac{\sum_{t=1}^T \left(\hat{\varepsilon}_t^2 - (\hat{\delta}_0 + \hat{\delta}_1 z_{t1} + \hat{\delta}_2 z_{t2} + \dots + \hat{\delta}_p z_{tp}) \right)^2}{2s^4},$$

kde $\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_p$ sú *MNS*-odhady parametrov modelu z alternatívnej hypotézy, $\hat{\varepsilon}_t$ sú *MNS*-rezíduá testovaného modelu a s^2 je ich rozptyl. Ak platí nulová hypotéza, testovacia štatistika má asymptotické χ_p^2 -rozdelenie, v ktorom je počet stupňov voľnosti rovný počtu vysvetľujúcich premenných v modeli z alternatívnej hypotézy.

Ak v modeli nie je splnený predpoklad homoskedasticity, *MNS*-odhad b zostáva podľa [16] nestranným odhadom parametra β , ale už nie je všeobecne najlepší medzi nestrannými lineárnymi odhadmi parametra β . *MNS*-odhad s^2 v tomto prípade nie je nestranným odhadom parametra σ^2 . Heteroskedasticita sa dá odstrániť napríklad logaritmickou transformáciou premenných.

2.5.2 Multikolarita

Ďalším predpokladom modelu lineárnej regresie je predpoklad

$$(IV) h(X) = k,$$

ktorého porušenie sa označuje ako multikolarita. Ako sa uvádza v [39], jedná sa o vysokú vzájomnú korelovanosť vysvetľujúcich premenných. Multikolarita sa dá podľa [29] najlepšie rozpoznať pomocou variančného inflačného faktora

$$VIF_i = \frac{1}{1 - R_i^2},$$

ktorý ukazuje, koľkokrát sa zväčšil rozptyl b_i zásluhou multikolarity. Tento faktor ohodnocuje vyššiu korelovanosť vysvetľujúcej premennej x_i so zvyšnými $k-1$

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

vysvetľujúcimi premennými, pričom pri vysokom R_i^2 môže vystreliť k vysokým hodnotám.

Čo sa týka maximálnej odporúčanej tolerancie na veľkosť variančného inflačného faktora, odporúčania autorov odbornej literatúry sa zvyknú líšiť. V [44] sa ako maximálna odporúčaná hodnota VIF_i uvádza hodnota 10. Odporúčanie VIF_i menej ako 5 môžeme nájsť v [59]. Neprekročiť v modeli lineárnej regresie hodnotu 4 odporúčajú napríklad autori [55].

Pridanie alebo odstránenie vysvetľujúcich premenných môže podľa [68] pri modeli s multikolinearitou viesť k výrazným zmenám vo veľkosti alebo signifikantnosti pôvodných odhadov jednotlivých parametrov. Multikolinearita však neznehodnocuje vlastnosti MNS -odhadu. Ak sa navyše korelovanosť medzi vysvetľujúcimi premennými udrží aj v budúcnosti, multikolinearita nemusí znehodnotiť ani predpoveď.

Pokiaľ však vynechanie vysvetľujúcich premenných, ktoré multikolinearitu spôsobujú, nenaruší praktickú interpretáciu modelu, je podľa [16] vhodnejšie pristúpiť k takému kroku. Pomôcť môže aj transformácia niektorých vysvetľujúcich premenných, rozšírenie dátovej vzorky o ďalšie pozorovania, prípadne použitie metódy hlavných komponentov.

2.5.3 Autokorelácia rezíduí

K porušeniu predpokladu vzájomne nekorelovaných rezíduí pre $s \neq t$

$$(III) \text{cov}(\varepsilon_s, \varepsilon_t) = 0$$

dochádza podľa [29] najčastejšie tak, že lineárny regresný model pracuje s dátami v tvare časových radov. Reziduálna zložka je vtedy korelovaná so svojimi oneskorenými a tiež budúcimi hodnotami. Príčinou autokorelácie však môžu byť chýbajúce oneskorené hodnoty vysvetľovanej alebo niektorej z vysvetľujúcich premenných.

Najjednoduchším typom autokorelácie, ktorý je v praxi podľa [41] schopný pokryť väčšinu jej prípadov, je reziduálna zložka ε_t v tvare autoregresného modelu prvého radu $AR(1)$

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t,$$

kde $\rho \in (-1,1)$ je parameter a u_t je biely šum, teda časový rad navzájom nekorelovaných veličín s nulovou strednou hodnotou a konštantným rozptylom.

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

Ako sa uvádza v [39], model $AR(1)$ umožňuje použiť Durbinov – Watsonov test autokorelácie rezíduí. V modeli testuje prítomnosť autokorelácie prvého stupňa pri nulovej hypotéze $H_0 : \rho = 0$. Testovacia štatistika má tvar

$$DW = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2}.$$

Pri pozitívnej autokorelácií $\rho > 0$ sú diferencie v čitateli malé, zatiaľ čo pri negatívnej autokorelácií $\rho < 0$ sú tieto diferencie veľké. Testovacia štatistika sa dá aproximovať ako $DW = 2(1 - \hat{\rho})$, pričom v nej figuruje odhad parametra $\hat{\rho}$ z $AR(1)$ modelu

$$\hat{\rho} = \frac{\sum_{t=2}^T \hat{\varepsilon}_t \cdot \hat{\varepsilon}_{t-1}}{\sum_{t=1}^T \hat{\varepsilon}_t^2}.$$

Štatistika DW sa podľa [29] neriadi žiadnym rozdelením, ale za predpokladu normality bieleho šumu u_t má dve kritické hodnoty, dolnú d_L a hornú d_H . Tieto závisia na počte pozorovaní T a počte vysvetľujúcich premenných k . Ich hodnoty sa dajú nájsť v štatistických tabuľkách alebo sa určujú simulačne. Na testovanie sa dajú použiť aj Boxov – Piercev, Ljungov – Boxov alebo Breuschov – Godfreyov test.

Dôsledky ignorovania autokorelácie rezíduí sú podľa [16] podobné ako v prípade ignorovania heteroskedasticity, keď MNS -odhad parametra b zostáva nestranným odhadom parametra β , ale nie je všeobecne najlepší medzi nestrannými lineárnymi odhadmi parametra β . V prípade pozitívnej autokorelácie môžu byť MNS -odhady štandardných odchýlok odhadnutých parametrov podhodnotené.

Riešením problému autokorelovaných rezíduí môže byť podľa [39] napríklad pridanie oneskorenej vysvetľovanej premennej na stranu vysvetľujúcich premenných, vytvorenie nového modelu lineárnej regresie s využitím vhodne zvolených diferencií vysvetľovanej aj vysvetľujúcich premenných, či aplikovanie Cochraneovej – Orcuttovej alebo Praisovej – Winstenovej iteračnej metódy.

Iným spôsobom riešenia problému autokorelovaných rezíduí môže byť podľa [38] lineárny regresný model s autokorelovanými rezíduami, ktoré sú $ARMA$ procesom. Ak biely šum označíme u_t , lineárny regresný model s predpisom

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t$$

2 MODEL LINEÁRNEJ REGRESIE V PRAXI

môže mať reziduálnu zložku v tvare autoregresného procesu $AR(p)$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots + \varphi_p \varepsilon_{t-p} + u_t,$$

v tvare autoregresného procesu $MA(q)$

$$\varepsilon_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q}$$

alebo v tvare autoregresného procesu $ARMA(p, q)$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots + \varphi_p \varepsilon_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q}.$$

Ako sa uvádza v [41], lineárny regresný model s autokorelovanými chybami je špeciálnym prípadom autoregresného modelu rozložených časových rezíduí, ktorý sa dá vo všeobecnosti zapísať v tvare

$$\varphi(B) \cdot y_t = \alpha + \sum_{i=0}^{\infty} \beta(B) \cdot x_t + \varepsilon_t,$$

kde $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ je autoregresný operátor, $\beta(B) = \beta_0 + \beta_1 B - \dots - \beta_k B^k$ je operátor rozložených časových oneskorení vysvetľujúcej premennej x a ε_t je reziduálna zložka v tvare stacionárneho procesu $ARMA(r, s)$.

3 Dekompozičné metódy analýzy časových radov

Ako sa uvádza v [15], niektoré časové rady y_t môžu byť rozložené na štyri základné zložky, ktorými sú trend T_t , sezónna zložka S_t , cyklická zložka C_t a reziduálna zložka E_t . Rozklad na tieto zložky je založený na predpoklade, že sa v jednotlivých zložkách rozkladu podarí rozpoznať pravidelné správanie daného časového radu, a nazýva sa tiež dekompozíciou časového radu.

3.1 Dekompozícia časového radu

Trend T_t reprezentuje dlhodobé zmeny v časovom rade. Sezónna zložka S_t popisuje zmeny v časovom rade pravidelne opakujúce sa v priebehu jedného kalendárneho roka. Cyklická zložka C_t hovorí o nepravidelných pohyboch okolo trendu. Ako sa uvádza v [15], sezónna a cyklická zložka sa niekedy súhrnne označujú aj ako periodické zložky časového radu.

Poslednou zložkou časového radu je podľa [15] reziduálna zložka E_t . Táto zostáva v časovom rade po odstránení trendu a periodických zložiek. Je tvorená náhodnými zmenami v priebehu časového radu, ktoré nemajú systematický charakter. Z toho dôvodu sa nezaraďuje k predchádzajúcim zložkám časového radu. Štandardným predpokladom je, že reziduálnu zložku tvorí biely šum.

Ako sa uvádza v [16], časový rad si na základe dekompozície môžeme predstaviť ako trend, na ktorý sú naviazané sezónna zložka, cyklická zložka a biely šum. Rozklad pritom môže mať dve podoby. Pri aditívnej dekompozícii

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

sa všetky zložky časového radu vyjadrujú vo svojich absolútnych hodnotách, zhodných s jednotkami časového radu y_t . Naopak, pri multiplikatívnej dekompozícii

$$y_t = T_t \cdot C_t \cdot S_t \cdot E_t$$

sa vo svojej absolútnej hodnote vyjadruje iba trendová zložka, ktorá je meraná v jednotkách časového radu y_t . Ostatné zložky sú vyjadrené v relatívnych hodnotách voči trendovej zložke, preto sú bezrozmerné.

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Ak sú k dispozícii pozorovania y_1, y_2, \dots, y_n , potom

$$\hat{y}_t = \hat{T}_t + \hat{C}_t + \hat{S}_t \text{ alebo } \hat{y}_t = \hat{T}_t \cdot \hat{C}_t \cdot \hat{S}_t$$

podľa [16] predstavuje vyrovnaný časový rad, ak $t \leq n$, alebo predpoveď časového radu na základe extrapolovaných hodnôt systematických zložiek, ak $t > n$. U niektorých časových radov môžu v rozklade niektoré zložky absentovať, ako napríklad v prípade časového radu v tvare $y_t = T_t + C_t + E_t$, ktorý neobsahuje sezónnu zložku.

3.2 Popis trendovej zložky matematickými krivkami

Ako sme už spomínali, trendová zložka T_t vyjadruje dlhodobé zmeny časového radu. V tejto časti sa budeme zaoberať metódami, ktorými sa ju snažíme popísať jednoduchými krivkami. Vypočítať budúce hodnoty týchto kriviek je pomerne jednoduché, čo umožňuje konštrukciu predpovede trendovej zložky. Obvykle sa pritom podľa [46] predpokladá, že analyzovaný časový rad má tvar

$$y_t = T_t + E_t.$$

Tento predpoklad nám umožní stotožniť predpoveď budúceho vývoja trendu T_t priamo s predpoveďou budúcich hodnôt časového radu y_t .

Ako sa uvádza v [16], najpoužívanejšími matematickými krivkami sú konštantný trend (odporúča sa použiť, ak je časový rad y približne konštantný), pri ktorom pre $t = 1, \dots, T$ platí

$$T_t = \beta_0,$$

lineárny trend (odporúča sa použiť, ak sú prvé diferencie $y_{t+1} - y_t$ približne konštantné), pri ktorom pre $t = 1, \dots, T$ platí

$$T_t = \beta_0 + \beta_1 t,$$

kvadratický trend (odporúča sa použiť, ak sú druhé diferencie $y_{t+2} - 2y_{t+1} + y_t$ približne konštantné), pri ktorom pre $t = 1, \dots, T$ platí

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2.$$

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Ďalej sú to exponenciálny trend (odporúča sa použiť, ak sú podiely susedných hodnôt y_{t+1} / y_t približne konštantné), pri ktorom pre $t = 1, \dots, T$ a $\beta > 0$ platí

$$T_t = \alpha \cdot \beta^t,$$

modifikovaný exponenciálny trend (odporúča sa použiť, ak sú podiely susedných prvých diferencií $(y_{t+2} - y_{t+1}) / (y_{t+1} - y_t)$ približne konštantné), pri ktorom pre $t = 1, \dots, T$ a $\beta > 0$ platí

$$T_t = \gamma + \alpha \cdot \beta^t,$$

logistický trend (odporúča sa použiť, ak má histogram prvých diferencií $y_{t+1} - y_t$ tvar hustoty $N(0,1)$ alebo ak sú podiely susedných prvých diferencií prevrátených hodnôt $(1/y_{t+2} - 1/y_{t+1}) / (1/y_{t+1} - 1/y_t)$ približne konštantné), pri ktorom pre $t = 1, \dots, T$, $\beta > 0$ a $\gamma > 0$ platí

$$T_t = \frac{\gamma}{1 + \alpha \cdot \beta^t},$$

Gompertzov trend (odporúča sa použiť, ak sú podiely susedných prvých diferencií zlogaritmovaných hodnôt $(\ln y_{t+2} - \ln y_{t+1}) / (\ln y_{t+1} - \ln y_t)$ približne konštantné), pri ktorom pre $t = 1, \dots, T$ a $\beta > 0$ platí

$$T_t = \exp(\gamma + \alpha \cdot \beta^t).$$

3.3 Metóda kľzavých priemerov a kubický vyhladzovací splajn

Metóda kľzavých priemerov sa zaraďuje medzi adaptívne prístupy k trendovej zložke. Ako sa uvádza v [15], adaptívne prístupy sú schopné pracovať so systematickými zložkami, ktoré v čase menia svoj charakter. Predpokladá sa však, že v krátkych úsekoch je možné popísanie trendovej zložky časového radu jednoduchou krivkou s nekonštantnými parametrami.

Ide o situáciu, keď sa časový rad nedá vyrovnať pomocou priamky

$$\beta_0 + \beta_1 \tau,$$

kde $t = 1, \dots, T$, ale pre krátke úseky časového radu so stredmi v jednotlivých časových okamihoch t sa dá použiť vyrovnanie pomocou lokálnych trendov

$$\beta_0(t) + \beta_1(t) \tau,$$

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

kde $\tau = \dots, t-1, t, t+1, \dots$. Proces eliminácie trendovej zložky sa prispôsobuje lokálnemu priebehu časového radu, pričom stupeň prispôsobenia sa dá cielene riadiť. Ďalšou výhodou adaptívnych metód je podľa [15] okrem výpočtovej nenáročnosti aj konštrukcia predpovedí, ktoré pružne reagujú na zmeny v charaktere časových radov.

Ako sa uvádza v [15], metóda konštrukcie kľzavých priemerov vyhladzovacím kubickým splajnom postupuje tak, že najprv vhodným polynómom tretieho stupňa vyrovná prvých $2m+1$ členov časového radu. Následne použije hodnotu tohto polynómu v bode $t = m+1$, teda v strede uvažovaného intervalu, ako vyrovnanú hodnotu \hat{y}_{m+1} daného časového radu v tomto bode.

Predpokladajme, že chceme kubickou parabolou vyhladiť $2m+1=5$ hodnôt časového radu, ktoré označíme ako $y_{t+\tau}$, kde $\tau = -2, -1, 0, 1, 2$. Parametre vyhladzujúceho polynómu odhadneme metódou najmenších štvorcov tak, že minimalizujeme výraz

$$\sum_{\tau=-2}^2 \left(y_{t+\tau} - (\beta_0 + \beta_1\tau + \beta_2\tau^2 + \beta_3\tau^3) \right)^2.$$

Derivovaním podľa jednotlivých koeficientov polynómu získame pre hľadané odhady b_0, b_1, b_2 a b_3 parametrov $\beta_0, \beta_1, \beta_2$ a β_3 sústavu štyroch rovníc, ktorá sa dá podľa [16] zapísať pre $j = 0, 1, 2, 3$ ako

$$\sum_{\tau=-2}^2 y_{t+\tau} \tau^j - b_0 \sum_{\tau=-2}^2 \tau^j - b_1 \sum_{\tau=-2}^2 \tau^{j+1} - b_2 \sum_{\tau=-2}^2 \tau^{j+2} - b_3 \sum_{\tau=-2}^2 \tau^{j+3} = 0.$$

Vzhľadom na to, že pre nepárne i platí

$$\sum_{\tau=-2}^2 \tau^i = 0,$$

zjednoduší sa sústava štyroch rovníc na tvar

$$\begin{aligned} 5b_0 + 10b_2 &= \sum_{\tau=-2}^2 y_{t+\tau} \\ 10b_1 + 34b_3 &= \sum_{\tau=-2}^2 \tau y_{t+\tau} \\ 10b_0 + 34b_2 &= \sum_{\tau=-2}^2 \tau^2 y_{t+\tau} \\ 34b_1 + 130b_3 &= \sum_{\tau=-2}^2 \tau^3 y_{t+\tau} \end{aligned}$$

Použitie kľzavých priemerov dĺžky $2m+1$ však zatiaľ neposkytuje predpovede budúcich hodnôt časového radu. Pri vyrovnávaní kubickou parabolou sme vyrovnávali

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

vždy 5 susedných hodnôt časového radu. Teraz za tieto hodnoty vezmeme hodnoty y_{t-4}, \dots, y_t . Chceli by sme poznať hodnoty kubickej paraboly vyrovnávajúcej posledný úsek pre $\tau = 1, 2$. K tomu potrebujeme poznať odhady parametrov $\beta_0, \beta_1, \beta_2$ a β_3 .

Ako sa uvádza v [16], pri vyhladzovaní by nám stačilo poznať iba odhad b_0 , keďže to je hodnota vyhladzovacieho polynómu $b_0 + b_1\tau + b_2\tau^2 + b_3\tau^3$ v bode $\tau = 0$, čo zodpovedá vyhladenej hodnote v strede skúmaného úseku y_{t-2}, \dots, y_{t+2} . V tom prípade by stačilo využiť prvú a tretiu rovnicu sústavy, pomocou ktorých by sme dostali odhad parametra β_0 v tvare

$$b_0 = \frac{1}{35} \left(17 \sum_{\tau=-2}^2 y_{t+\tau} - 5 \sum_{\tau=-2}^2 \tau^2 y_{t+\tau} \right).$$

Ďalšie odhady majú podľa [16] tvar

$$\begin{aligned} b_1 &= \frac{1}{72} \left(65 \sum_{\tau=-2}^2 \tau y_{t+\tau} - 17 \sum_{\tau=-2}^2 \tau^3 y_{t+\tau} \right) \\ b_2 &= \frac{1}{14} \left(\sum_{\tau=-2}^2 \tau^2 y_{t+\tau} - 2 \sum_{\tau=-2}^2 y_{t+\tau} \right) \\ b_3 &= \frac{1}{72} \left(5 \sum_{\tau=-2}^2 \tau^3 y_{t+\tau} - 17 \sum_{\tau=-2}^2 \tau y_{t+\tau} \right) \end{aligned} .$$

S ich pomocou môžeme pre posledné dve pozorovania y_{t-1} a y_t dostať vyrovnané hodnoty, keď pre $k = 1, 2$ platí

$$\hat{y}_{t-2+k} = b_0 + b_1 k + b_2 k^2 + b_3 k^3.$$

Ako sa uvádza v [16], tento postup nám ďalej umožňuje konštruovať predpovede v danom časovom rade tak, že napríklad predpoveď hodnoty y_{t+1} vytvoríme jednoducho tak, že do predchádzajúceho vzorca dosadíme $k = 3$, pričom po úpravách dostaneme predpoveď v tvare

$$\hat{y}_{t+1} = \frac{1}{5} (-4y_{t-4} + 11y_{t-3} - 4y_{t-2} - 14y_{t-1} + 16y_t).$$

3.4 Exponenciálne vyhladzovanie

V praxi sa ako adaptívny prístup k trendovej zložke často používa aj exponenciálne vyhladzovanie. Vyhladzovaná hodnota je v tejto metóde špeciálnym prípadom kĺzavého priemeru, v ktorom všetky doteraz pozorované hodnoty vážime do

minulosti exponenciálne klesajúcimi váhami. Ako sa uvádza v [15], vyrovnaný časový rad \hat{y}_t minimalizuje výrazy typu

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

kde $\beta \in (0,1)$ je vopred zvolená diskontná konštanta. Hlavnou výhodou tohto postupu je, že pri využití rekurentných vzorcov je po výpočtovej stránke veľmi jednoduchý.

3.4.1 Jednoduché exponenciálne vyhladzovanie

Jednoduché exponenciálne vyhladzovanie podľa [15] používame pre časové rady, u ktorých sa dá trend považovať za lokálne konštantný v tvare

$$T_t = \beta_0,$$

pričom úlohou je nájsť odhad parametra β_0 . Pretože sa jedná o adaptívny prístup k vyhladzovaniu časového radu, bude tento odhad závisieť aj od časového okamihu, v ktorom bol vykonaný.

Označme $b_0(t)$ odhad parametra β_0 vykonaný v čase t , ktorý predstavuje odhadnutú hodnotu trendu v čase t a zároveň vyrovnanú hodnotu \hat{y}_t daného časového radu. Hodnotu získame minimalizáciou výrazu

$$\sum_{j=0}^{\infty} (y_{t-j} - \beta_0)^2 \cdot \beta^j,$$

kde $\beta \in (0,1)$ je vopred zvolená diskontná konštanta. Ako sa uvádza v [15], ak tento výraz zderivujeme podľa β_0 a získanú deriváciu položíme rovnú nule, tak vzhľadom na konvexnosť minimalizovanej funkcie dostaneme odhad $b_0(t)$ parametra β_0 v čase t v tvare

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

Vyrovnaná hodnota časového radu v čase t je teda váženým priemerom hodnôt tohto časového radu do času t pri exponenciálne klesajúcich váhach. Získaný výraz je však v praxi takmer nepoužiteľný, preto sa podľa [15] prevádza na rekurentný tvar

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

kde $\alpha = (1 - \beta) \in (0,1)$ sa nazýva vyhladzovacia konštanta.

Ako sa ďalej uvádza v [15], rekurentný tvar vzorca demonštruje výhody exponenciálneho vyhladzovania, ktorými sú najmä výpočtová jednoduchosť, úspornosť vzhľadom k potrebnému objemu skladovaných dát a adaptívnosť metódy, keď pri

väčšom α metóda rýchlejšie reaguje na zmeny v charaktere dát. Pri menšom α sa, naopak, zosilní vyhladzovania schopnosť metódy.

Realizácia rekurentnej formule vyžaduje voľbu počiatocnej hodnoty \hat{y}_0 a vyhladzovacej konštanty α . Za počiatočnú hodnotu \hat{y}_0 sa podľa [16] najčastejšie berie aritmetický priemer krátkého počiatočného úseku časového radu. Pri voľbe vyhladzovacej konštanty α sa odporúča začať intervalom $0 < \alpha \leq 0,3$. Takmer všetky softvéry však vedia vybrať α aj automaticky, minimalizáciou SSE chýb predpovedí.

3.4.2 Dvojité exponenciálne vyhladzovanie

Ako sa uvádza v [15], dvojité exponenciálne vyhladzovanie používame pre časové rady, u ktorých sa dá trend považovať za lokálne lineárny s predpisom v tvare

$$T_{t-j} = \beta_0 - j\beta_1.$$

Odhady parametrov β_0 a β_1 v čase t , označované ako b_0 a b_1 , získame minimalizáciou výrazu

$$\sum_{j=0}^{\infty} [y_{t-j} - (\beta_0 - j\beta_1)]^2 \cdot \beta^j,$$

kde $\beta \in (0,1)$ je vopred zvolená diskontná konštantá.

Ak položíme parciálne derivácie tohto výrazu podľa β_0 a β_1 rovné nule, dostaneme sústavu rovníc v tvare

$$\sum_{j=0}^{\infty} \beta^j y_t - \beta_0 \sum_{j=0}^{\infty} \beta^j + \beta_1 \sum_{j=0}^{\infty} j\beta^j = 0,$$

$$\sum_{j=0}^{\infty} j\beta^j y_{t-j} - \beta_0 \sum_{j=0}^{\infty} j\beta^j + \beta_1 \sum_{j=0}^{\infty} j^2\beta^j = 0.$$

Táto sústava sa dá podľa [15] zjednodušiť pomocou vzorcov

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

a vyrovnávacích štatistík

$$S_t = (1-\beta) \cdot \sum_{j=0}^{\infty} \beta^j y_{t-j}, \quad S_t^{[2]} = (1-\beta) \cdot \sum_{j=0}^{\infty} \beta^j S_{t-j}.$$

na tvar

$$\beta_0 - \frac{\beta}{1-\beta} \cdot \beta_1 = S_t,$$

$$\beta \cdot \beta_0 - \frac{\beta(1+\beta)}{1-\beta} \cdot \beta_1 = S_t^{[2]} - (1-\beta)S_t.$$

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Ako sa uvádza v [15], zo sústavy rovníc v zjednodušenom tvare dostaneme hľadané odhady ako

$$b_0(t) = 2S_t - S_t^{[2]},$$
$$b_1(t) = \frac{1-\beta}{\beta} \cdot (S_t - S_t^{[2]}).$$

Predpoveď hodnoty $y_{t+\tau}$ v čase t má potom tvar

$$\hat{y}_{t+\tau} = b_0(t) + b_1(t)\tau = \left(2 + \frac{\alpha\tau}{1-\alpha}\right) \cdot S_t - \left(1 + \frac{\alpha\tau}{1-\alpha}\right) \cdot S_t^{[2]}.$$

Špeciálne pre $\tau = 0$ dostaneme vyrovnanú hodnotu časového radu ako

$$\hat{y}_t = 2S_t - S_t^{[2]}.$$

Realizácia rekurentných formúl dvojitého exponenciálneho vyhladzovania opäť vyžaduje voľbu počiatočných hodnôt S_0 a $S_0^{[2]}$ a vyhladzovacej konštanty α . Počiatočné hodnoty S_0 a $S_0^{[2]}$ sa podľa [16] obvykle určia ako

$$S_0 = b_0(0) - \frac{1-\alpha}{\alpha} \cdot b_1(0),$$
$$S_0^{[2]} = b_0(0) - \frac{2(1-\alpha)}{\alpha} \cdot b_1(0).$$

Pri voľbe vyhladzovacej konštanty α sa opäť odporúča zvoliť jej počiatočnú hodnotu z intervalu $0 < \alpha \leq 0,3$. Takmer všetky softvéry vedia aj v tomto prípade vybrať α automaticky tak, že v danom časovom rade minimalizujú SSE chýb predpovedí.

3.5 Holtova metóda

Zovšeobecnením dvojitého exponenciálneho vyhladzovania je Holtova metóda, ktorá používa až dve vyhladzovacie konštanty. Prvou vyhladzovacou konštantou je α pre vyhladzovanie úrovne L_t daného časového radu, druhou vyhladzovacou konštantou je β pre vyhladzovanie smernice B_t , pričom $0 < \alpha$ a $\beta < 1$.

Ako sa uvádza v [38], sústavu rovníc, ktorá charakterizuje Holtovu metódu, môžeme zapísať v tvare

$$\hat{y}_{t+h} = L_t + h \cdot B_t,$$
$$L_t = \alpha \cdot y_t + (1-\alpha) \cdot (L_{t-1} + B_{t-1}),$$
$$B_t = \beta \cdot (L_t - L_{t-1}) + (1-\beta) \cdot B_{t-1},$$

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

pričom prvú rovnicu nazývame rovnicou predpovede, druhú rovnicu voláme rovnicou úrovne a tretia rovnica sa zvykne nazývať rovnicou smernice.

Rovnica úrovne je váženým priemerom hodnoty y_t a jej jedнокrokovej predpovede $(L_{t-1} + B_{t-1})$. Rovnica smernice je zas váženým priemerom smernice v čase t , ktorá má tvar $(L_t - L_{t-1})$, a jej predchádzajúceho odhadu B_{t-1} . Z rovnice predpovede môžeme vidieť, že viackroková predpoveď Holtovou metódou je vlastne lineárnou funkciou, ktorá závisí od h .

Ako sa uvádza v [15], ako počiatočné hodnoty sa odporúčajú zvoliť $L_0 = y_1$ a $B_0 = y_2 - y_1$. Navyše sa podľa [16] dá ukázať, že dvojité exponenciálne vyhladzovanie s vyhladzovacou konštantou α je špeciálnym prípadom Holtovej metódy s vyhladzovacími konštantami α_H a β_H v tvare

$$\alpha_H = \alpha \cdot (2 - \alpha),$$

$$\beta_H = \frac{\alpha}{2 - \alpha}.$$

3.6 Kalmanov filter

Ako sa uvádza v [35], Kalmanov filter sa využíva sa na analýzu časových radov, ktoré sa dajú vyjadriť pomocou špeciálnej sústavy dvoch rovníc, ktorá tvorí lineárny dynamický systém. Táto sústava pozostáva zo stavovej rovnice a rovnice pozorovaní. Kalmanov filter je rekurzívny postup, ktorý umožňuje prepočítavanie odhadovanej premennej, ak je o nej dostupná nová informácia.

Stavová rovnica lineárneho dynamického systému má podľa [60] tvar

$$x_{t+1} = A_t x_t + u_t,$$

kde x_{t+1} je stavový vektor, ktorý reprezentuje správanie lineárneho dynamického systému v čase $t+1$, A_t predstavuje maticu parametrov stavovej rovnice v čase t a u_t je biely šum.

Ako sa ďalej uvádza v [10], druhá rovnica, ktorou je lineárny dynamický systém reprezentovaný, je rovnica pozorovaní v tvare

$$y_t = B_t x_t + v_t,$$

3 DEKOMPOZIČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

kde y_t predstavuje vektor pozorovaní v čase t , B_t je matica parametrov stavovej rovnice v čase t a v_t je biely šum, ktorý je nezávislý od bieleho šumu u_t .

Úlohou Kalmanovho filtra je nájsť optimálny stav, ktorý bude riešením lineárneho dynamického systému. Algoritmus má podľa [51] dva kroky, pričom v prvom (predikčnom) kroku sa snaží nájsť optimálny odhad stavovej premennej x_{t+1} na základe pozorovaných hodnôt y_t minimalizáciou MSE , kým v druhom (obnovovacom) kroku upravuje odhad x_{t+1} z minulého kroku o novú informáciu y_{t+1} .

4 Autokorelačné metódy analýzy časových radov

Z autokorelačných vlastností časových radov vychádza Boxova – Jenkinsova metodológia, ktorá bola prvýkrát formálne spísaná v [12]. Jej základom sú *ARMA* a *ARIMA* procesy. Tieto procesy umožňujú uspokojivo modelovať aj časové rady veľmi všeobecných priebehov, ktoré by sa v rámci dekompozičného prístupu modelovať nedali.

4.1 Stacionárne časové rady

Pri popise Boxovej – Jenkinsovej metodológie začneme modelovaním stacionárnych časových radov. Stacionárnym časovým radom nazývame vo všeobecnosti taký časový rad, ktorý vykazuje známky stochasticky ustáleného správania. O striktnej stacionarite podľa [16] hovoríme v prípade, keď je pravdepodobnostné správanie časového radu invariantné voči posunom v čase.

Overovanie splnenia podmienky striktnej stacionarity je pomerne náročné, preto sa podľa [16] v praxi pod pojmom stacionarita zvyčajne myslí slabá stacionarita, ktorej definícia je menej obmedzujúca. Na to, aby bol časový rad slabo stacionárny, stačí, aby bol invariantný voči posunom v čase iba v rámci momentov do druhého rádu. Pre každý časový posun h a všetky časové okamihy s a t potom musí platiť

$$E(y_t) = \mu = \textit{konšt.},$$

$$\text{var}(y_t) = \sigma_y^2 = \textit{konšt.},$$

$$\text{cov}(y_s, y_t) = E(y_s - \mu)(y_t - \mu) = \text{cov}(y_{s+h}, y_{t+h}).$$

4.2 Autokorelačná funkcia a jej odhad

Ako sa uvádza v [14], autokorelačná funkcia *ACF* sa pre $k = \dots, -1, 0, 1, \dots$ definuje ako

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\text{cov}(y_t, y_t)} = \frac{\text{cov}(y_t, y_{t-k})}{\sigma_y^2},$$

Autokorelačná funkcia je párnou funkciou, preto sa pri jej popise stačí obmedziť na prípad $k \geq 0$, pričom platí $\rho_0 = 1$ a $|\rho_k| \leq 1$.

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Grafické znázornenie ρ_k pre jednotlivé k sa nazýva korelogram. Korelogram pomocou niekoľkých hodnôt popisuje krátkodobú dynamiku stacionárneho časového radu. Ako sa uvádza v [15], model stacionárneho časového radu vysvetľuje, na rozdiel od modelu lineárnej regresie, súčasnú hodnotu časového radu práve na základe korelácií ρ_k niekoľkých minulých hodnôt y_{t-k} so súčasnou hodnotou y_t .

Ak označíme odhad strednej hodnoty

$$\bar{y} = \frac{1}{T} \cdot \sum_{t=1}^T y_t,$$

potom môžeme podľa [16] zapísať odhad autokorelačnej funkcie pre $k = 0, 1, \dots, T-1$ v tvare

$$r_k = \frac{\frac{1}{T} \cdot \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\frac{1}{T} \cdot \sum_{t=1}^T (y_t - \bar{y})^2}.$$

Tvar autokorelačnej funkcie je v Boxovej – Jenkinsovej metodológii veľmi dôležitý, pretože naznačuje, aký typ modelu by bolo vhodné pre daný časový rad použiť. Pre identifikáciu modelu je dôležité predovšetkým určiť hodnotu $k = k_0$, za ktorou začína byť autokorelačná funkcia nulová. Ak hodnota k_0 existuje, podľa [15] sa označuje ako bod useknutia.

Pre konkrétny časový rad však v praxi teoretickú autokorelačnú funkciu ρ_k nepoznáme. Najpodstatnejšou otázkou preto je, ako blízko nuly musí byť r_k , aby sme mohli tvrdiť, že $\rho_k = 0$. Pre tento účel sa používa Barlettova aproximácia, ktorá podľa [16] hovorí o tom, že ak je $\rho_k = 0$ pre $k > k_0$, potom za predpokladu asymptotickej normality pre $k > k_0$ platí

$$r_k \sim N\left(0, \sqrt{\frac{1}{T} \cdot \left(1 + 2 \sum_{j=1}^{k_0} r_j^2\right)}\right).$$

4.3 Parciálna autokorelačná funkcia a jej odhad

Ako sa uvádza v [14], okrem autokorelačnej funkcie ρ_k sa pri analýze časového radu využíva aj parciálna autokorelačná funkcia *PACF*, ktorú označujeme ako ρ_{kk} .

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Hodnota ρ_{kk} je definovaná ako parciálny korelačný koeficient medzi y_t a y_{t-k} pri pevných hodnotách $y_{t-k-1}, \dots, y_{t-1}$ a platí pre ňu $\rho_{00} = 1$ a $\rho_{11} = \rho_1$. V praxi sa podľa [15] zvyčajne používa rekurentný spôsob odhadu r_{kk} parciálnej autokorelačnej funkcie ρ_{kk} podľa vzorcov

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j},$$

$$r_{11} = r_1,$$

kde $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$, pre $k > 1$ a $j = 1, \dots, k-1$.

Podobne ako autokorelačná funkcia, aj parciálna autokorelačná funkcia môže mať bod useknutia k_0 , ktorý je dôležitým identifikačným nástrojom modelov Boxovej – Jenkinsovej metodológie. V tomto prípade sa využíva Quenouilleova aproximácia, ktorá podľa [16] hovorí o tom, že ak je $\rho_{kk} = 0$ pre $k > k_0$, potom za predpokladu asymptotickej normality pre $k > k_0$ platí

$$r_{kk} \sim N\left(0, \sqrt{\frac{1}{T}}\right).$$

4.4 Proces kĺzavých súčtov MA

Ako sa uvádza v [41], proces kĺzavých súčtov rádu q , ktorý sa označuje ako $MA(q)$, má tvar

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \theta(B) \cdot \varepsilon_t,$$

kde $\theta_1, \dots, \theta_q$ sú parametre a $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ je operátor kĺzavých súčtov.

Proces $MA(q)$ je podľa [15] vždy stacionárny s nulovou strednou hodnotou, rozptylom

$$\sigma_y^2 = (1 + \theta_1^2 + \dots + \theta_q^2) \cdot \sigma^2$$

a autokorelačnou funkciou

$$\rho_k = \frac{\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}; k = 1, \dots, q,$$

$$\rho_k = 0; k > q.$$

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Autokorelačná funkcia $MA(q)$ procesu má bod useknutia k_0 rovný rádu modelu q . Ako sa uvádza v [16], parciálna autokorelačná funkcia ρ_{kk} procesu $MA(q)$ nemá bod useknutia, ale je obmedzená lineárnymi kombináciami geometrických postupností a sínusoid s geometricky klesajúcimi amplitúdami.

4.5 Autoregresný proces AR

Autoregresný proces rádu p , označovaný ako $AR(p)$, má podľa [41] tvar

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t,$$

pričom podľa [15] platí

$$y_t - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} = \varphi(B) \cdot y_t = \varepsilon_t,$$

kde $\varphi_1, \dots, \varphi_p$ sú parametre a $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ je autoregresný operátor.

Autokorelačná funkcia ρ_k procesu $AR(p)$ je podľa [16] lineárnou kombináciou klesajúcich geometrických postupností a sínusoid rôznych frekvencií s geometricky klesajúcimi amplitúdami. Parciálna autokorelačná funkcia ρ_{kk} procesu $AR(p)$ má bod useknutia k_0 rovný rádu modelu p .

4.6 Zmiešaný ARMA proces

Ako sa uvádza v [45], zmiešaný proces rádu p a q , ktorý sa označuje ako $ARMA(p, q)$, má tvar

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

pričom podľa [15] platí

$$\varphi(B) \cdot y_t = \theta(B) \cdot \varepsilon_t,$$

kde $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ je autoregresný operátor a $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ je operátor klzavých súčtov.

Ako sa uvádza v [16], autokorelačná funkcia ρ_k procesu $ARMA(p, q)$ nemá bod useknutia, ale je lineárnou kombináciou klesajúcich geometrických postupností a sínusoid rôznych frekvencií s geometricky klesajúcimi amplitúdami, s výnimkou počiatočných hodnôt $\rho_0, \rho_1, \dots, \rho_{q-p}$, ak $q \geq p$.

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Ani parciálna autokorelačná funkcia procesu $ARMA(p, q)$ nemá bod useknutia. Podľa [16] je obmedzená lineárnou kombináciou klesajúcich geometrických postupností a sínusoid rôznych frekvencií s geometricky klesajúcimi amplitúdami, s výnimkou počiatočných hodnôt $\rho_{00}, \dots, \rho_{p-q, p-q}$, ak $p \geq q$.

4.7 Testy jednotkového koreňa

Väčšina časových radov v praxi je nestacionárna. Prítomnosť jednotkového koreňa v autoregresnom operátore príslušného modelu podľa [33] hovorí o možnosti ich stacionarizácie pomocou diferencovania. Pri identifikácii prítomnosti jednotkového koreňa môže byť nápomocný napríklad veľmi pomalý pokles korelogramu. Okrem neho však existujú aj testy jednotkového koreňa.

4.7.1 Dickeyov – Fullerov test

Dickeyov – Fullerov test má podľa [20] tri varianty, niekedy označované súhrnne ako τ -testy. Prvým variantom je τ -test, ktorý je jednostranným testom náhodnej prechádzky proti stacionárnemu $AR(1)$ procesu, teda

$$H_0 : y_t = y_{t-1} + \varepsilon_t \text{ proti } H_1 : y_t = \varphi_1 y_{t-1} + \varepsilon_t \text{ pre } \varphi_1 < 1.$$

Druhým variantom je jednostranný test náhodnej prechádzky proti stacionárnemu $AR(1)$ procesu s nenulovou úrovňou, označovaný ako τ_μ -test. Zapísať ho môžeme ako

$$H_0 : y_t = y_{t-1} + \varepsilon_t \text{ proti } H_1 : y_t = \alpha + \varphi_1 y_{t-1} + \varepsilon_t \text{ pre } \varphi_1 < 1.$$

Posledný variant, τ_t -test, je jednostranným testom náhodnej prechádzky proti stacionárnemu $AR(1)$ procesu s lineárnym trendom, teda

$$H_0 : y_t = y_{t-1} + \varepsilon_t \text{ proti } H_1 : y_t = \alpha + \beta t + \varphi_1 y_{t-1} + \varepsilon_t \text{ pre } \varphi_1 < 1.$$

Ako sa uvádza v [15], nulová hypotéza sa dá vo všetkých troch prípadoch jednoducho zapísať ako

$$H_0 : \Delta y_t = \psi y_{t-1} + \varepsilon_t; \psi = 0,$$

kým alternatívna hypotéza je vo všeobecnosti

$$H_1 : \Delta y_t = \alpha + \beta t + \psi y_{t-1} + \varepsilon_t; \psi < 0,$$

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

kde $\psi = \varphi_1 - 1$. Pre τ_μ -test $\beta = 0$ navyše platí, pre τ -test $\alpha = \beta = 0$. Testovacia štatistika vo všetkých troch variantoch Dickeyovho – Fullerovho testu je podľa [16] klasický t -pomer, teda test významnosti parametra ψ . Má tvar

$$DF = \frac{\hat{\psi}}{\hat{\sigma}(\hat{\psi})}$$

a kritický obor

$$DF \leq t_{1-\alpha}^*(T).$$

Ako sa ďalej uvádza v [16], pri platnosti nulovej hypotézy testovacia štatistika nemá ani asymptoticky t -rozdelenie. Z toho dôvodu boli kritické hodnoty napočítané simulačne a zvlášť pre jednotlivé testy. Vo všeobecnosti má dané rozdelenie ťažšie konce ako zodpovedajúce t -rozdelenie, preto sú tieto kritické hodnoty v absolútnej hodnote väčšie ako pri t -rozdelení. Uvedené sú napríklad v [49].

4.7.2 Rozšírený Dickeyov – Fullerov test

Dickeyov – Fullerov test je podľa [61] použiteľný iba v prípade, že reziduálna zložka ε_t predstavuje biely šum. Ak je totiž závislá premenná Δy_t autokorelovaná, potom má test pravdepodobnosť zamietnutia H_0 , ak H_0 platí, väčšiu ako hladina spoľahlivosti α . Pre tento prípad bol navrhnutý rozšírený Dickeyov – Fullerov test, ktorý podľa [16] používa nulovú hypotézu v tvare

$$H_0 : \Delta y_t = \psi y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + \varepsilon_t; \psi = 0,$$

pričom testovacia štatistika a kritické hodnoty pre jednotlivé varianty zostávajú rovnaké ako pred rozšírením.

4.8 Proces ARIMA

Pre nestacionárne časové rady, ktoré sa však dajú stacionarizovať diferencovaním, sú v rámci Boxovej – Jenkinsovej metodológie určené procesy typu *ARIMA*. Integrovaný zmiešaný proces rádu p , d , q , ktorý sa označuje ako *ARIMA*(p, d, q), má podľa [16] tvar

$$\varphi(B) \cdot \Delta^d y_t = \alpha + \theta(B) \cdot \varepsilon_t,$$

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

kde $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ je autoregresný operátor, $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ je operátor kĺzavých súčtov a $\Delta^d y_t$ je d -tá diferencia modelovaného časového radu y_t . Parameter α modeluje prípadnú nenulovú úroveň procesu $\Delta^d y_t$.

V modeli *ARIMA* sa vykoná stacionarizácia pomocou vhodnej diferencie modelovaného časového radu, následkom čoho sa vzniknutý stacionárny časový rad môže modelovať pomocou zmiešaného procesu *ARMA*. Pre $d > 0$ je model *ARIMA* invariantný voči prípadnému posunu o ľubovoľnú konštantu, preto je podľa [15] zbytočné časový rad vopred centrovať odčítaním výberového priemeru.

Konštrukcia *ARIMA*(p, d, q) modelu je teda založená na konštrukcii stacionárneho *ARMA*(p, q) modelu pre príslušne diferencovaný časový rad. Rád diferencovania d pritom v praxi obvykle neprekročí hodnotu 2. Možností, ako ho stanoviť pre konkrétny časový rad, je veľa. Najpoužívanejšími sú podľa [14] testy na prítomnosť jednotkového koreňa, prípadne aplikácia informačných kritérií.

4.9 Predpovedanie

Ako sa uvádza v [45], veľkou výhodou Boxovej – Jenkinsovej metodológie je pomerne jednoduchá konštrukcia predpovedí, ktorá sa dá ilustrovať na príklade stacionárneho *ARMA*(p, q) procesu v tvare

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}.$$

Symbolom $\hat{y}_{t+k}(t)$ označíme predpoveď hodnoty y_{t+k} vytvorenú v čase t . Pre jednoduchosť budeme podľa [15] konštruovať lineárnu predpoveď, ktorá bude lineárnou funkciou hodnôt $\varepsilon_t, \varepsilon_{t-1}, \dots$, keďže predpokladáme stacionaritu procesu.

Predpoveď budeme hľadať v tvare

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

pričom napríklad pri minimalizácii *MSE* je úlohou nájsť koeficienty $\psi_k^*, \psi_{k+1}^*, \dots$, ktoré by minimalizovali výraz

$$\left(1 + \psi_1^2 + \dots + \psi_{k-1}^2 + \sum_{j=k}^{\infty} (\psi_j - \psi_j^*)^2\right) \sigma^2.$$

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

Tento výraz nadobúda minimálne hodnoty pri $\psi_j^* = \psi_j$, kde $j = k, k+1, \dots$. To znamená, že predpoveď má tvar

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

pričom jej chyba definovaná ako $e_{t+k}(t) = y_{t+k} - \hat{y}_{t+k}(t)$ je rovná

$$e_{t+k}(t) = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \dots + \psi_{k-1} \varepsilon_{t+1},$$

má nulovú strednú hodnotu a rozptyl

$$\text{var}(e_{t+k}(t)) = (1 + \psi_1^2 + \dots + \psi_{k-1}^2) \sigma^2.$$

Špeciálne podľa [16] platí

$$e_t(t-1) = y_t - \hat{y}_t(t-1) = \varepsilon_t,$$

teda biely šum predstavuje jednokrokové chyby predpovede.

4.10 Kointegrácia

Pri lineárnej kombinácii nestacionárnych časových radov je výsledný časový rad zvyčajne tiež nestacionárny. V praxi sme však podľa [28] pomerne často schopní skombinovať nestacionárne časové rady aj takým spôsobom, že výsledná lineárna kombinácia je stacionárna. Taký prípad sa označuje ako kointegrácia a môže byť interpretovaný ako vzťah dlhodobej rovnováhy medzi veličinami.

Ako sa uvádza v [16], ak y_1, \dots, y_m sú nestacionárne časové rady, pričom platí $y_1 \sim I(1), \dots, y_m \sim I(1)$, potom ak existuje ich netriviálna lineárna kombinácia, ktorá je stacionárna, y_1, \dots, y_m sú kointegrované časové rady. Všeobecnejšie sa dá kointegrácia rádu (d, b) , pričom $d > 0, b > 0$ a $y_1 \sim I(d), \dots, y_m \sim I(d)$, definovať tak, že existuje lineárna kombinácia daných časových radov, ktorá je integrovaná rádu $I(d-b)$.

4.10.1 Test založený na rezíduách

Na testovanie kointegrácie časových radov môžu byť použité napríklad rezíduá z modelu lineárnej regresie. Tento typ testu je podľa [52] založený na myšlienke, že v prípade kointegračného vzťahu medzi y, x_1, \dots, x_k by MNS -rezíduá $\hat{\varepsilon}_t$ vypočítané z modelu

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t$$

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

mali byť stacionárnym procesom.

Ako sa uvádza v [16], stačí modifikovať Dickeyov – Fullerov test jednotkového koreňa a testovať nulovú hypotézu

$$H_0 : \Delta \hat{\varepsilon}_t = \psi \varepsilon_{t-1} + u_t; \psi = 0.$$

Kritické hodnoty tohto testu sú však odlišné od kritických hodnôt, ktoré majú Dickeyove – Fullerove testy. Uvedené sú napríklad v [56]. Pri zamietnutí nulovej hypotézy o tom, že $\hat{\varepsilon}_t$ obsahujú jednotkový koreň, zamietame aj nulovú hypotézu o tom, že medzi veličinami neexistuje kointegračný vzťah.

4.10.2 Johansenov test

Iným typom testu kointegrácie, ktorý sa v praxi používa najčastejšie, sú Johansenove testy. Podľa [40] sú založené na *MMV*-odhade kanonických korelácií, ktoré merajú parciálnu závislosť medzi m -rozmernými vektormi Δy_t a y_{t-1} pri pevných hodnotách vektorov Δy_{t-1} , Δy_{t-2} , Tieto kanonické korelácie sú druhými odmocninami hodnôt λ_1 , ..., λ_m , pre *MMV*-odhady ktorých platí

$$1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_m \geq 0.$$

Počet nenulových *MMV*-odhadov $\hat{\lambda}_1$, ..., $\hat{\lambda}_m$ je rovný počtu nájdených kointegračných vzťahov. Johansenove testy sú síce založené na pomere vierohodností, ale ich kritické hodnoty sa generujú simulačne.

V praxi sú rozšírené dva typy týchto testov. Prvým je podľa [16] Johansenov test so štatistikou

$$\lambda(r) = -T \cdot \sum_{t=r+1}^m \ln(1 - \hat{\lambda}_t).$$

Je to test nulovej hypotézy, že počet kointegračných vzťahov je najviac r , oproti alternatíve, že je väčší ako r . Test zamietne nulovú hypotézu, ak $\lambda(r)$ má hodnotu väčšiu než príslušná kritická hodnota. Vykonáva sa postupne pre $r = 0, 1, \dots, m-1$.

Ako sa ďalej uvádza v [16], Johansenov test so štatistikou

$$\lambda(r) = -T \cdot \ln(1 - \hat{\lambda}_{r+1})$$

4 AUTOKORELAČNÉ METÓDY ANALÝZY ČASOVÝCH RADOV

je testom nulovej hypotézy, že počet kointegračných vzťahov je práve r , oproti alternatíve, že ich je práve $r+1$. Rovnako ako v predchádzajúcom prípade, test zamietne nulovú hypotézu, ak hodnota $\lambda(r)$ je väčšia než príslušná kritická hodnota. Vykonáva sa tiež postupne pre $r = 0, 1, \dots, m-1$.

4.11 Grangerova kauzalita

Ak sú dve $I(1)$ premenné kointegrované, potom medzi nimi podľa [27] musí existovať kauzálne pôsobenie minimálne v jednom smere. Ak jeden časový rad kauzálne ovplyvňuje iný časový rad, potom by malo jeho zohľadnenie zlepšiť predpoveď daného radu. Kauzalita podľa Grangera znamená silnú koreláciu medzi súčasnou hodnotou jedného časového radu a minulými hodnotami iného časového radu.

V praxi sa testovanie kauzality vykonáva pomocou VAR modelov. Uvažujme podľa [48] napríklad dvojrozmerný VAR model v tvare

$$\begin{aligned}y_{1t} &= \varphi_{11}y_{1,t-1} + \varphi_{12}y_{2,t-1} + \varepsilon_{1t} \\y_{2t} &= \varphi_{21}y_{1,t-1} + \varphi_{22}y_{2,t-1} + \varepsilon_{2t}\end{aligned}$$

Ak $\varphi_{12} \neq 0$, tak y_2 kauzálne pôsobí na y_1 . Naopak, ak $\varphi_{21} \neq 0$, potom y_1 kauzálne pôsobí na y_2 . Ak platí súčasne $\varphi_{12} \neq 0$ a $\varphi_{21} \neq 0$, potom existuje spätná väzba medzi y_1 a y_2 . Ak $\varphi_{12} \neq 0$ a $\varphi_{21} = 0$, potom existuje jednosmerná závislosť y_1 na y_2 . Naopak, ak $\varphi_{12} = 0$ a $\varphi_{21} \neq 0$, potom existuje jednosmerná závislosť y_2 na y_1 . Ak $\varphi_{12} = 0$ a $\varphi_{21} = 0$, potom y_1 a y_2 sú nezávislé.

5 Dáta

V diplomovej práci sme pracovali s reálnymi vnútrobankovými dátami Všeobecnej úverovej banky, a.s.. Tieto dáta mali formu vývoja hodnôt jednotlivých ukazovateľov pre dve odlišné časti retailového portfólia a boli pripravené na mesačnej báze. Vývoj každého ukazovateľa pokrýval 119 mesiacov od januára 2006 po november 2015. Podarilo sa nám teda zohľadniť obdobie takmer desiatich rokov.

Makroekonomické ukazovatele nám, podobne ako vnútrobankové premenné, poskytla Všeobecná úverová banka, a.s.. Pokrývali rovnaké časové obdobie a obsahovali aj predpovede hodnôt pre rok 2016. Ukazovatele boli pripravené na štvrťročnej báze. My sme však potrebovali získať ich mesačný vývoj, preto sme najprv priradili ich štvrťročné hodnoty všetkým trom mesiacom štvrťroka.

Po korelačnej analýze vzťahu makroekonomických ukazovateľov s vysvetľovanou premennou sa však nakoniec ako lepšia voľba vytvorenia mesačného vývoja ukázala metóda, v ktorej sme hodnoty jednotlivých ukazovateľov prislúchajúce danému štvrťroku priradili jeho prvému mesiacu a hodnoty pre zvyšné dva mesiace sme dopočítali interpoláciou.

Hodnotu každého ukazovateľa pre druhý mesiac štvrťroka sme vypočítali sčítaním dvojnásobku hodnoty pre daný štvrťrok a hodnoty pre nasledujúci štvrťrok a vydelením tohto súčtu tromi. Hodnotu tretieho mesiaca v štvrťroku sme vypočítali tak, že sme súčet hodnoty pre daný štvrťrok a dvojnásobku hodnoty pre nasledujúci štvrťrok vydělili tromi.

5.1 Vnútrobankové vysvetľujúce premenné

Čo sa týka interných ukazovateľov, do modelu sa ponúkalo množstvo údajov, ktoré má banka o klientoch a ich úveroch k dispozícii. Tieto údaje sa dajú rozdeliť na dve základné skupiny. Do prvej skupiny patria údaje, ktoré sa viažu k charakteristikám klienta a jeho úveru v čase schválenia žiadosti. Druhou skupinou údajov sú aktuálne údaje o portfóliu a jeho charakteristikách.

5.1.1 Údaje zo žiadostí

K údajom zo žiadosti patrí napríklad schválený objem. Je to objem peňazí, ktorý banka klientovi poskytla formou úveru. Úver môže byť v závislosti od dohodnutých

5 DÁTA

podmienok čerpaný aj postupne, nemôže však prekročiť tento objem. Zo samotnej podstaty výpočtu miery omeškania je zrejmé, že výška schváleného objemu za posledné mesiace môže mať na jej aktuálnu hodnotu veľký vplyv.

Vek klienta sa uvádza v rokoch a vzťahuje sa na obdobie schválenia úveru. Domnievame sa, že aj tento ukazovateľ by mohol mať určitý vplyv na správanie sa klienta pri splácaní úveru. Mladší klienti sa totiž vo všeobecnosti javia ako rizikovejší, kým splácanie úverov starších klientov so stálym zamestnaním a vytvoreným zázemím zvykne byť bezproblémovjšie.

5.1.2 Údaje o portfóliu

Miera omeškania od 90 do 360 dní, ktorá v modeli slúžila ako vysvetľovaná premenná, môže byť významne ovplyvnená aj zostatkami s menším alebo väčším omeškaním z minulých mesiacov. Časť úverov v omeškaní menšom ako 90 dní sa totiž postupom času môže presunúť do vyššej kategórie delikvencie. Menej častým javom je presun úveru v omeškaní viac ako 360 dní do nižšej kategórie.

Podiel bezúčelových úverov by mohol byť takisto podstatným faktorom vplyvujúcim na mieru omeškania. Zo svojej podstaty sa totiž bezúčelové úvery zdajú byť v porovnaní s ostatnými rizikovejšie. Sú totiž rýchlejšie a jednoduchšie dostupné, no úroková miera je pri nich zvyčajne podstatne vyššia. Preto môžu mať neskôr klienti s ich splácaním väčšie problémy.

5.2 Makroekonomické vysvetľujúce premenné

Veľmi dôležitou súčasťou modelu na predpovedanie miery omeškania retailového portfólia sú nepochybne externé vysvetľujúce premenné. Aj bez hlbšej analýzy totiž vieme odhadnúť, že podiel zostatkov v omeškaní môže závisieť aj od aktuálnej finančnej situácie domácností. Táto môže byť do veľkej ovplyvnená hrubým domácim produktom krajiny, nezamestnanosťou či infláciou.

5.2.1 Hrubý domáci produkt

Hrubý domáci produkt je podľa [65] hlavným meradlom ekonomickej výkonnosti štátu. Definuje sa ako hodnota všetkých vyprodukovaných tovarov a služieb, s výnimkou tých, ktoré boli použité pri ich výrobe. Výpočet percentuálnej miery rastu objemu hrubého domáceho produktu umožňuje porovnať dynamiku hospodárskeho rastu štátu v čase.

5 DÁTA

Čo sa týka tohto ukazovateľa, mali sme k dispozícii vývoj percentuálnej miery zmien objemu hrubého domáceho produktu Slovenskej republiky pri konštantných cenách. Konštantné ceny sú v tomto prípade dôležité, pretože nie je želaným efektom, aby pohyb cien ovplyvňoval mieru rastu, a tým skresľoval obraz o skutočnom hospodárskom raste štátu.

5.2.2 Nezamestnanosť

Údaje o nezamestnanosti v Slovenskej republike pochádzali z prieskumu Labour Force Survey. Nezamestnanosť je v tomto prípade štatistický odhad na základe reprezentatívneho prieskumu domácností. Respondenti sa v prieskume kategorizujú na základe otázok v dotazníku. Podľa [24] sa údaje získané na limitovanej vzorke respondentov následne prepočítajú na celú populáciu krajiny.

Za nezamestnaných sú podľa [65] považovaní všetci obyvatelia Slovenskej republiky vo veku od 15 rokov, s výnimkou cudzincov s prechodným pobytom a osôb žijúcich v zahraničí, ktorí v referenčnom týždni nemali platenú prácu, v posledných štyroch týždňoch si ju aktívne hľadali, prípadne do nej nastúpia v priebehu troch mesiacov. Títo obyvatelia musia byť schopní nastúpiť do práce do dvoch týždňov.

5.2.3 Inflácia

Inflácia je podľa [65] stály rast cenovej hladiny, ktorý sa prejavuje poklesom kúpnej sily peňažnej jednotky. Jej zrejme najrozšírenejším meradlom je index spotrebiteľských cien (*CPI*). Počíta sa na základe cien potravín, šatstva, bývania a iných tovarov a služieb každodennej potreby. Ceny jednotlivých tovarov a služieb sú vážené ich ekonomickým významom.

Index spotrebiteľských cien nie je jediným meradlom inflácie. Táto sa dá merať aj indexom cien výrobcov (*PPI*), ktorý meria cenovú hladinu na úrovni veľkoobchodu a výroby, alebo deflátorom *HDP*. V tejto práci sme sa však rozhodli pracovať s indexom spotrebiteľských cien, keďže spomedzi všetkých meradiel najlepšie charakterizuje dopad inflácie na domácnosti.

5.2.4 Finančná kríza

Dáta, s ktorými sme pracovali, v sebe zahŕňali aj obdobie finančnej krízy. Rozhodli sme sa preto vytvoriť umelú vysvetľujúcu premennú, ktorá mala za účel simulovať v modeli jej trvanie.

6 Predikčné modely pre prvú časť portfólia

6.1 Model lineárnej regresie

Pri vytváraní modelu lineárnej regresie sme začali výberom ukazovateľov, ktoré by na základe skúseností z dlhodobého sledovania kvality aktív retailového portfólia mohli mieru omeškania ovplyvňovať. Všetky ukazovatele sme následne pomocou štatistického softvéru *R* analyzovali z pohľadu ich rozdelenia a vzťahu k miere omeškania.

Na základe toho sme si ako vysvetľujúce premenné zvolili schválený objem za posledných 12 mesiacov (*APPAMT*), podiel sumy zostatkov klientov, ktorí mali v čase schválenia viac ako 40 rokov (*AGEA40*), o dva mesiace oneskorenú sumu zostatkov v delikvencii od 30 do 89 dní (*DLQ3089*) a podiel sumy zostatkov bezúčelových úverov (*WOPURP*), ktorý sme oneskorili o pol roka.

Čo sa týka externých vysvetľujúcich premenných, zvolili sme si finančnú krízu (*CRISIS*) a zmenu hrubého domáceho produktu (*GDP*) s trojmesačným oneskorením. Nezamestnanosť (*UNEMPL*) sme voči miere omeškania oneskorili o jeden mesiac, zmenu indexu spotrebiteľských cien (*CPI*) nebolo potrebné voči miere omeškania posúvať.

Prvý vytvorený model lineárnej regresie bol reprezentovaný rovnicou 6.1. Významy jednotlivých premenných sme si už načrtli, t v rovnici znázorňuje jednotlivé mesiace od júla 2006 po november 2015 a $\hat{\varepsilon}_t$ je rezíduum v čase t . Takmer všetky odhady parametrov mali očakávané znamienka. Výnimkou bol iba parameter prislúchajúci podielu sumy zostatkov klientov starších ako 40 rokov.

$$DLQ90360_t = -0,0108 - 0,0134 \cdot APPAMT_t + 0,0105 \cdot AGEA40_t + \\ + 0,2260 \cdot DLQ3089_{t-2} + 0,0506 \cdot WOPURP_{t-6} + 0,0555 \cdot CRISIS_t + \\ - 0,0110 \cdot GDP_{t-6} + 0,0304 \cdot UNEMPL_{t-1} + 0,0371 \cdot CPI_t + \hat{\varepsilon}_t$$

Rovnica 6.1: Prvý vytvorený model lineárnej regresie

Na základe skúseností zo sledovania kvality aktív retailového portfólia sme pri premennej *AGEA40* očakávali negatívne znamienko odhadu parametra. Z obrázka 6.1 sme však mohli vidieť, že tento odhad nebol na hladine významnosti päť percent

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

signifikantný. Test multikolinearity pomocou variančného inflačného faktora *VIF* navyše v modeli odhalil vzájomnú lineárnu závislosť vysvetľujúcich premenných.

```
Time series regression with "ts" data:
Start = 2006(7), End = 2015(11)

Call:
dynlm(formula = DLQ90360 ~ APPAMT + AGEA40 + L(DLQ3089, 2) +
      L(WOPURP, 6) + CRISIS + L(GDP, 6) + L(UNEMPL, 1) + CPI)

Residuals:
      Min       1Q   Median       3Q      Max
-0.141072 -0.032793 -0.002363  0.039639  0.160597

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.010774   0.221026  -0.049 0.961215
APPAMT       -0.013413   0.001886  -7.113 1.49e-10 ***
AGEA40        0.010542   0.009356   1.127 0.262407
L(DLQ3089, 2) 0.225957   0.078686   2.872 0.004949 **
L(WOPURP, 6)  0.050649   0.006988   7.248 7.73e-11 ***
CRISIS        0.055503   0.027162   2.043 0.043541 *
L(GDP, 6)     -0.011027   0.002807  -3.928 0.000154 ***
L(UNEMPL, 1)  0.030351   0.008802   3.448 0.000816 ***
CPI           0.037122   0.006186   6.001 2.90e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05927 on 104 degrees of freedom
Multiple R-squared:  0.9701, Adjusted R-squared:  0.9678
F-statistic: 421.2 on 8 and 104 DF, p-value: < 2.2e-16
```

Obrázok 6.1: Výstup R pre prvý vytvorený model lineárnej regresie

Na základe toho sme krokovou regresiou, založenou na minimalizácii *AIC*, z modelu vynechali premenné *AGEA40* a *DLQ3089*. Ako môžeme vidieť na obrázku 6.2, zjednodušenie nemalo vplyv na kvalitu modelu. Všetky odhady parametrov v rovnici 6.2 už mali očakávané znamienka. Zjednodušením sme zároveň vyriešili problém multikolinearity, čo sme opäť skontrolovali prostredníctvom *VIF*.

$$DLQ90360_t = 0,1208 - 0,0110 \cdot APPAMT_t + 0,0587 \cdot WOPURP_{t-6} + 0,0643 \cdot CRISIS_t - 0,0152 \cdot GDP_{t-6} + 0,0463 \cdot UNEMPL_{t-1} + 0,0312 \cdot CPI_t + \hat{\varepsilon}_t$$

Rovnica 6.2: Zjednodušený model lineárnej regresie

```
Time series regression with "ts" data:
Start = 2006(7), End = 2015(11)

Call:
dynlm(formula = DLQ90360 ~ APPAMT + L(WOPURP, 6) + CRISIS + L(GDP,
      6) + L(UNEMPL, 1) + CPI)

Residuals:
      Min       1Q   Median       3Q      Max
-0.183443 -0.038679 -0.003223  0.037911  0.164669

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.120772   0.152030   0.794 0.42874
APPAMT       -0.010963   0.001514  -7.242 7.36e-11 ***
L(WOPURP, 6)  0.058681   0.005628  10.426 < 2e-16 ***
CRISIS        0.064349   0.023972   2.684 0.00844 **
L(GDP, 6)     -0.015191   0.002502  -6.072 2.00e-08 ***
L(UNEMPL, 1)  0.046324   0.007080   6.543 2.19e-09 ***
CPI           0.031195   0.005746   5.429 3.62e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06115 on 106 degrees of freedom
Multiple R-squared:  0.9675, Adjusted R-squared:  0.9657
F-statistic: 526.2 on 6 and 106 DF, p-value: < 2.2e-16
```

Obrázok 6.2: Výstup R pre zjednodušený model lineárnej regresie

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

Test ani grafické zobrazenia nepreukázali v modeli prítomnosť odľahlých ani vplyvných pozorovaní. Pokračovali sme preto testom homoskedasticity, teda konštantnosti rozptylu. Využili sme naň Breuschov – Paganov test, ktorý na hladine významnosti 5 percent zamietol nulovú hypotézu o konštantnom rozptyle, hoci veľmi tesne, pri p -hodnote 4,9 percenta.

Vytvorili sme preto nový model lineárnej regresie s rovnakými vysvetľujúcimi premennými, ktorý sa od pôvodného líšil iba tým, že modeloval druhú odmocninu z miery omeškania. Na novovytvorený model sme opäť aplikovali krokovú regresiu založenú na minimalizácii AIC . Tento postup z modelu vynechal premennú $CRISIS$, ktorá modelovala finančnú krízu.

$$\sqrt{DLQ90360}_t = 0,4678 - 0,0059 \cdot APPAMT_t + 0,0381 \cdot WOPURP_{t-6} + \\ - 0,0085 \cdot GDP_{t-6} + 0,0257 \cdot UNEMPL_{t-1} + 0,0117 \cdot CPI_t + \hat{\varepsilon}_t$$

Rovnica 6.3: Model lineárnej regresie s transformovanou vysvetľovanou premennou

Z rovnice 6.3 novovytvoreného modelu je zrejmé, že odhady jednotlivých parametrov si zachovali svoje pôvodné znamienka, ktoré boli konzistentné s očakávaniami, keď nepriamoúmerne mieru omeškania ovplyvňovali iba schválený objem za posledných 12 mesiacov a zmena HDP . Na obrázku 6.3 môžeme vidieť, že všetky odhady boli signifikantné na hladine významnosti 5 percent.

```
Time series regression with "ts" data:
Start = 2006(7), End = 2015(11)

Call:
dynlm(formula = sqrt(DLQ90360) ~ APPAMT + L(WOPURP, 6) + L(GDP,
6) + L(UNEMPL, 1) + CPI)

Residuals:
    Min       1Q   Median       3Q      Max
-0.095279 -0.019062 -0.001013  0.017820  0.090147

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4678191  0.0634538   7.373 3.72e-11 ***
APPAMT      -0.0059166  0.0006813  -8.685 4.75e-14 ***
L(WOPURP, 6)  0.0380565  0.0020154  18.883 < 2e-16 ***
L(GDP, 6)    -0.0084719  0.0012991  -6.521 2.37e-09 ***
L(UNEMPL, 1)  0.0257046  0.0032282   7.963 1.92e-12 ***
CPI          0.0116954  0.0027124   4.312 3.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

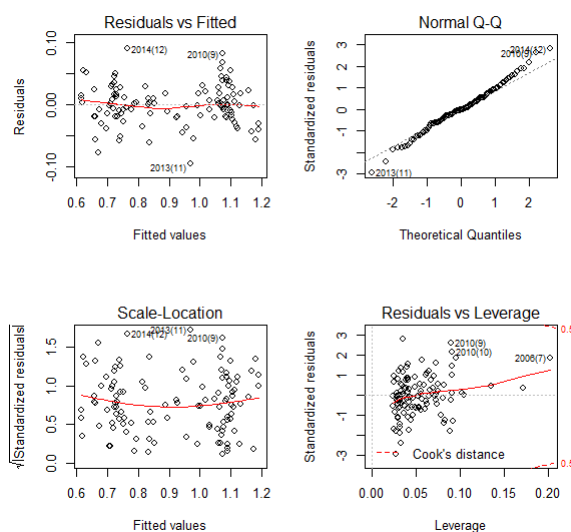
Residual standard error: 0.03277 on 107 degrees of freedom
Multiple R-squared:  0.9692,    Adjusted R-squared:  0.9678
F-statistic: 673.6 on 5 and 107 DF,  p-value: < 2.2e-16
```

Obrázok 6.3: Výstup R pre model lineárnej regresie s transformovanou vysvetľovanou premennou

Z hľadiska splnenia základných predpokladov sa model zdal byť na základe obrázka 6.4 v poriadku. Ako prvý sme pri ňom opäť otestovali predpoklad o konštantnom rozptyle. Breuschov – Paganov test nulovú hypotézu o homoskedasticite

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

nezamietol pri p -hodnote takmer 80 percent. Nový model už teda splňal predpoklad konštantného rozptylu. Mohli sme pristúpiť k jeho ďalšiemu testovaniu.



Obrázok 6.4: Grafy modelu lineárnej regresie s transformovanou vysvetľovanou premennou

Predpoklad normálneho rozdelenia rezíduí modelu sme otestovali Shapirovým – Wilkovým testom. Tento test nulovú hypotézu o tom, že rezíduá pochádzajú z normálneho rozdelenia, nezamietol. Jeho p -hodnota dosiahla takmer 80 percent, čo bol pomerne jednoznačný výsledok, na základe ktorého sme mohli rezíduá modelu považovať za normálne rozdelené.

Posledným testom bol test autokorelácie rezíduí. Durbinov – Watsonov test nulovú hypotézu o tom, že rezíduá nie sú autokorelované, zamietol. Jeho p -hodnota bola veľmi blízka nule, čo znamenalo, že v rezíduách modelu bola prítomná autokorelácia. Podľa [46] sú vo väčšine prípadov v praxi rezíduá procesom typu $AR(1)$. To, či to platilo aj v našom prípade, sme si však overili pomocou testu.

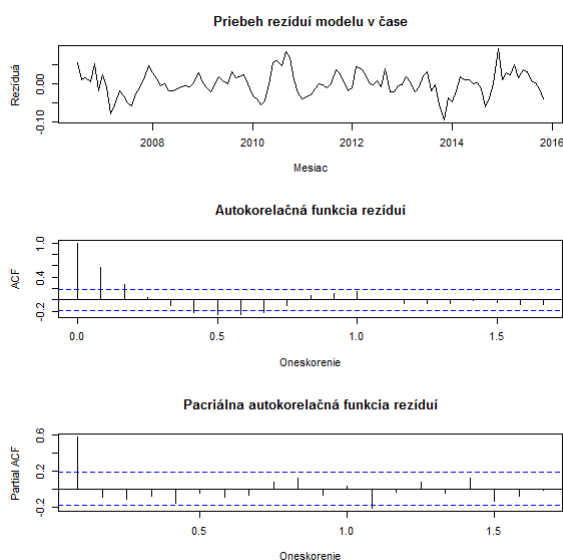
```
Value of test-statistic is: -5.1941
Critical values for test statistics:
1pct 5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Obrázok 6.5: Výstup R pre test jednotkového koreňa rezíduí modelu lineárnej regresie s transformovanou vysvetľovanou premennou

Na obrázku 6.5 môžeme vidieť, že rozšírený Dickeyov – Fullerov test zamietol prítomnosť jednotkového koreňa, keď jeho testovacia štatistika bola výrazne nižšia ako kritická hodnota prislúchajúca hladine významnosti 5 percent. Z výsledku testu

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

v spojení s obrázkom 6.6, popisujúcim charakter rezíduí, vyplývalo, že sme rezíduá modelu lineárnej regresie mohli charakterizovať procesom $AR(1)$.



Obrázok 6.6: Priebeh rezíduí, autokorelačná a parciálna autokorelačná funkcia modelu lineárnej regresie s transformovanou vysvetľovanou premennou

Prvým riešením autokorelácie rezíduí bolo pridanie oneskorenej vysvetľovanej premennej medzi vysvetľujúce premenné. Novovytvorený model však aj napriek splneniu všetkých predpokladov predpovedal horšie ako pôvodný model. Podobná situácia nastala pri modeli lineárnej regresie, ktorý modeloval vzťahy medzi medzimesačnými zmenami miery omeškania a vysvetľujúcich premenných.

S využitím rovnakých premenných ako v pôvodnom modeli sme preto nakoniec vytvorili model lineárnej regresie, v ktorom sme rezíduá modelovali procesom $AR(1)$. Pre účely odhadu jednotlivých parametrov sme si zvolili metódu maximálnej vierohodnosti. Novovzniknutý model je popísaný rovnicou 6.4, v ktorej \hat{u}_t reprezentuje biely šum. Parameter autokorelácie rezíduí ρ nadobudol hodnotu 0,6109.

$$\sqrt{DLQ90360}_t = 0,3854 - 0,0050 \cdot APPAMT_t + 0,0393 \cdot WOPURP_{t-6} + \\ -0,0082 \cdot GDP_{t-6} + 0,0293 \cdot UNEMPL_{t-1} + 0,0101 \cdot CPI_t + \hat{\varepsilon}_t$$

$$\hat{\varepsilon}_t = 0,6109 \cdot \hat{\varepsilon}_{t-1} + \hat{u}_t$$

Rovnica 6.4: Model lineárnej regresie s autokorelovanými rezíduami

Odhady jednotlivých parametrov sa v porovnaní s pôvodným modelom výrazne nezmenili. Všetky boli významné na hladine významnosti 5 percent. Vysvetľujúce

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

premenné neboli navzájom silno korelované, čo môžeme vidieť na obrázku 6.7. Model bol v poriadku aj z pohľadu splnenia základných predpokladov, mohli sme preto pristúpiť ku konštrukcii predpovedí.

Jednokrokové ani viackrokové predpovede tohto modelu na testovacej vzorke nefungovali ideálne. Ich najväčšou nevýhodou bola, popri neschopnosti zachytiť výkyvy vo vývoji miery omeškania, aj absencia intervalovej predpovede, keďže softvér *R* neposkytuje odhady predikčných intervalov pre model lineárnej regresie s autokorelovanými rezíduami.

Predikčné intervaly sme sa pokúsili nasimulovať odhadom dolnej a hornej hranice pomocou intervalov spoľahlivosti pre odhady parametrov. Vytvorené intervaly však boli veľmi široké, keď sa horná hranica 80 -percentného intervalu pre predpovedané hodnoty pohybovala na hranici najvyšších historicky nameraných hodnôt miery omeškania, pochádzajúcich z obdobia finančnej krízy.

```
Generalized least squares fit by maximum likelihood
Model: sqrt(DLQ90360) ~ APPAMT + lag(WOPURP, 6) + lag(GDP, 6)
Data: MO
      AIC      BIC  logLik
-490.8892 -469.0701 253.4446

Correlation Structure: AR(1)
Formula: ~1
Parameter estimate(s):
      Phi
0.6109324

Coefficients:
      Value Std.Error t-value p-value
(Intercept)  0.3854429 0.10641779  3.621978  0.0004
APPAMT      -0.0050290 0.00111190 -4.522870  0.0000
lag(WOPURP, 6) 0.0392962 0.00374091 10.504447  0.0000
lag(GDP, 6)   -0.0082464 0.00216892 -3.802057  0.0002
lag(UNEMPL, 1) 0.0293473 0.00555260  5.285323  0.0000
CPI          0.0101067 0.00489223  2.065876  0.0413

Correlation:
      (Intr) APPAMT 1(WOP6 1(GDP6 1(UNE1
APPAMT      -0.806
lag(WOPURP, 6) -0.543  0.500
lag(GDP, 6)   0.023 -0.408  0.087
lag(UNEMPL, 1) -0.905  0.581  0.202  0.078
CPI          0.313 -0.318 -0.565 -0.358 -0.173

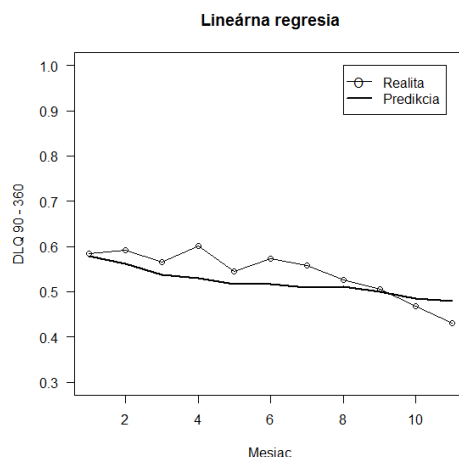
Standardized residuals:
      Min      Q1      Med      Q3      Max
-2.93049623 -0.49191796  0.01631629  0.65241834  2.71189879

Residual standard error: 0.03237775
Degrees of freedom: 113 total; 107 residual
```

Obrázok 6.7: Výstup *R* pre model lineárnej regresie s autokorelovanými rezíduami

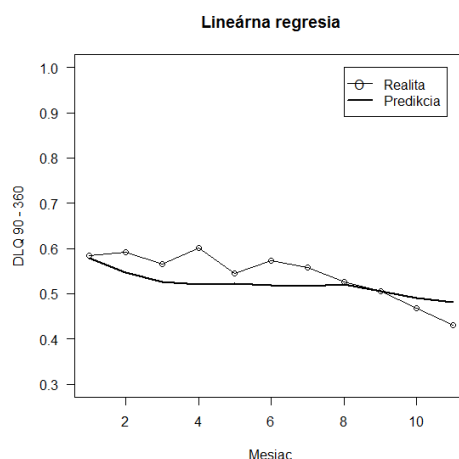
Ako môžeme vidieť na obrázku 6.8, aj keď viackroková bodová predpoveď modelu lineárnej regresie pomerne dobre odhadla klesajúci trend miery omeškania na testovacej vzorke, jej detailnejšie pohyby už zachytiť nedokázala. Odmocninová priemerná štvorcová chyba bola nižšia ako 0,04 percentuálneho bodu, priemerná absolútna percentuálna chyba bola pod hranicou 6 percent.

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA



Obrázok 6.8: Grafické znázornenie viackrokovej bodovej predpovede modelom lineárnej regresie s autokorelovanými rezíduami

Jednokrokové predpovede mali v porovnaní s ako viackrokovou predpoveďou mierne menšie rozpätie intervalov spoľahlivosti, ale aj v tomto prípade bola ich šírka príliš veľká. Z obrázka 6.9 je zrejmé, že ani jednokrokové bodové predpovede modelu lineárnej regresie nedokázali v dostatočnej miere zachytiť medzimesačné výkyvy miery omeškania, a až priveľmi pripomínali predpoveď viackrokovú.



Obrázok 6.9: Grafické znázornenie jednokrokových bodových predpovedí modelom lineárnej regresie s autokorelovanými rezíduami

Odmocninová priemerná štvorcová chyba bola v tomto prípade dokonca mierne vyššia ako pri viackrokovej predpovedi, no stále bola približne na úrovni 0,04 percentuálneho bodu. Priemerná absolútna percentuálna chyba pri jednokrokových predpovediach presiahla 6 percent. Úspešnosť predpovedania medzimesačných zmien vo vývoji miery omeškania bola vyššia ako 60 percent.

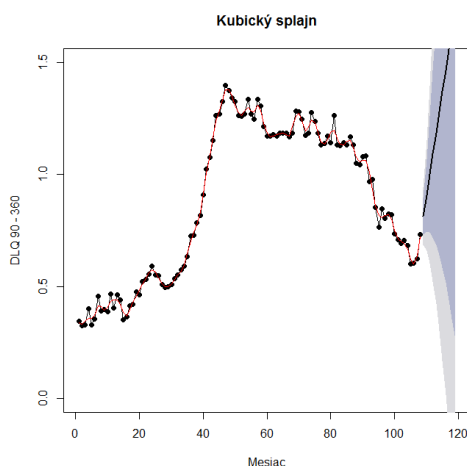
6.2 Modely časových radov

Keďže predpovede modelu lineárnej regresie, aj napriek zahrnutiu početnej skupiny vysvetľujúcich premenných, nefungovali podľa našich predstáv, prešli sme k predpovedaniu pomocou časových radov. Využili sme pritom dva prístupy k ich analýze. Najprv sme vyskúšali modelovanie na základe dekompozície, a následne sme vytvorili predikčný model aj na základe Boxovej – Jenkinsovej metodológie.

6.2.1 Kubický vyhladzovací splajn

Prvá metóda patrí do kategórie dekompozičných metód analýzy časových radov. Ich trendovú zložku sa snaží modelovať kubickým splajnom. Úspešnosť predpovedania tejto metódy sme najprv otestovali viackrokovou predpoveďou, a neskôr aj jednokrokovými predpoveďami na testovacej vzorke. Za ňu sme si aj pri modeloch časových radov zvolili obdobie od januára do novembra 2015.

Pri viackrokovej predpovedi sa ukázalo, že predpovedanie kubickým vyhladzovacím splajnom by pre banku mohlo byť veľmi rizikové. Táto metóda totiž pre jedenásť mesiacov testovacej vzorky predpovedala prudký nárast miery omeškania, ktorý by dokonca prekonal aj jej historické hodnoty z obdobia finančnej krízy. V skutočnosti pritom miera omeškania v danom období zaznamenala mierny pokles.



Obrázok 6.10: Grafické znázornenie viackrokovej predpovede kubickým vyhladzovacím splajnom

Ako môžeme vidieť na obrázku 6.10, skutočné hodnoty miery omeškania síce vo všetkých mesiacoch padli do 95 -percentného predikčného intervalu, tento mal ale veľmi široké rozpätie, ktoré pri posledných mesiacoch testovacej vzorky kleslo až pod

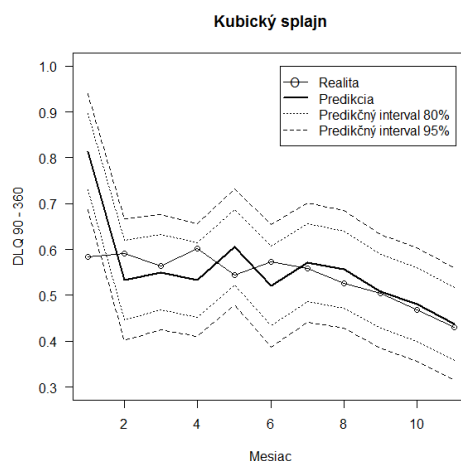
6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

nulu. Vtedy sa síce skutočné hodnoty zmestili aj do užšieho, 80 -percentného, predikčného intervalu, tento bol však stále pomerne široký.

Odmocninová priemerná štvorcová chyba nadobudla hodnotu viac ako 0,8 percentuálneho bodu, priemerná absolútna chyba bola o niečo menšia. Priemerná absolútna percentuálna chyba sa blížila k hranici 150 percent, čo bola skutočne veľká hodnota. Odmocninová priemerná štvorcová chyba na posledných piatich pozorovaniach testovacej vzorky bola takmer 1,1 percentuálneho bodu.

Čo sa týka jedнокrokových predpovedí, z obrázka 6.11 je zrejmé, že hneď prvú z nich táto metóda výrazne nadhodnotila, keď skutočná hodnota miery omeškania nespadla ani do 95 -percentného predikčného intervalu. Predpovede vo všetkých ostatných mesiacoch už spadli aj do 80-percentného predikčného intervalu, tento však bol stále pomerne široký, hoci nie až tak ako pri lineárnej regresii.

Aj v ďalších piatich mesiacoch boli rozdiely predpovedí a skutočných hodnôt pomerne veľké. Naopak, keď vývoj skutočných hodnôt miery omeškania začal stabilne klesať, predpovede sa stabilizovali na uspokojivých hodnotách, s malými odchýlkami od reality. Ako sa neskôr ukázalo, na posledných mesiacoch testovacej vzorky fungoval kubický vyhladzovací splajn najlepšie zo všetkých modelov.



Obrázok 6.11: Grafické znázornenie jedнокrokových predpovedí kubickým vyhladzovacím splajnom

Aby sme jedнокrokové predpovede tejto metódy mohli ohodnotiť aj kvantitatívne, pozreli sme sa na miery presnosti predpovedí. Odmocninová priemerná štvorcová chyba bola vzhľadom k hodnotám miery omeškania v sledovanom období

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

stále pomerne veľká, keď dosiahla takmer 0,08 percentuálneho bodu, podobne ako priemerná absolútna chyba, blížiac sa k hodnote 0,05 percentuálneho bodu.

Priemerná absolútna percentuálna chyba dosiahla takmer 9 percent. Poslednou mierou presnosti predpovedí bola percentuálna úspešnosť odhadu medzimesačných zmien vo vývoji miery omeškania. Predpoveď kubickým vyhladzovacím splajnom zachytila správne iba 6 z 11 medzimesačných pohybov miery omeškania, čo predstavovalo síce nadpolovičnú, ale nie veľmi uspokojivú úspešnosť predpovedania.

6.2.2 Holtova metóda

Pri Holtovej metóde bolo na začiatku potrebné určiť jej vstupné parametre. Po vyskúšaní viacerých kombinácií sme ich voľbu nakoniec nechali na softvér, keďže automaticky zvolené parametre, ktoré môžeme vidieť v rovnici 6.5, fungovali z hľadiska presnosti predpovedí najlepšie. Viackroková predpoveď Holtovou metódou bola podstatne lepšia ako predpoveď kubickým vyhladzovacím splajnom.

$$DLQ90360_{t+1} = L_t + B_t$$

$$L_0 = 0,3311; L_t = 0,8753 \cdot DLQ90360_t + (1 - 0,8753) \cdot (L_{t-1} + B_{t-1})$$

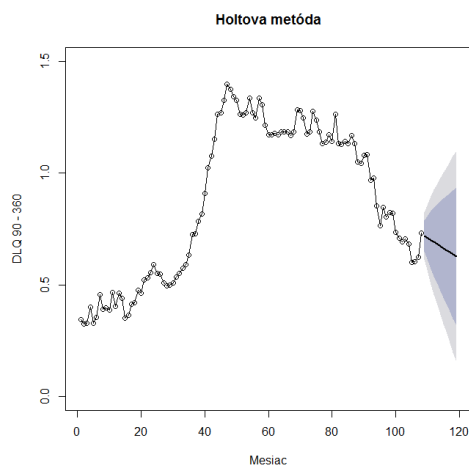
$$B_0 = 0,0087; B_t = 0,0729 \cdot (L_t - L_{t-1}) + (1 - 0,0729) \cdot B_{t-1}$$

Rovnica 6.5: Model Holtovej metódy

Klesajúcosť viackrokovej predpovede, ktorú môžeme vidieť na obrázku 6.12, zodpovedala realite. Miery presnosti predpovedí mali v porovnaní s predchádzajúcou metódou takisto lepšie hodnoty. Všetky skutočné hodnoty miery omeškania v rámci testovacej vzorky dát spadli do 80 -percentného predikčného intervalu, ktorý mal v tomto prípade oveľa menšie rozpätie ako pri kubickom vyhladzovacom splajne.

Odmocninová priemerná štvorcová chyba a priemerná absolútna chyba na testovacej vzorke boli menšie ako 0,14 percentuálneho bodu. Priemerná absolútna percentuálna chyba bola iba mierne nad hranicou 25 percent. Odmocninová priemerná štvorcová chyba na posledných piatich pozorovaniach testovacej vzorky bola vyššia ako 0,15 percentuálneho bodu.

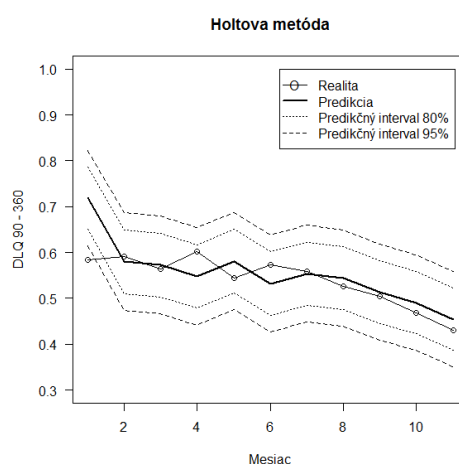
6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA



Obrázok 6.12: Grafické znázornenie viackrokovej predpovede Holtovou metódou

Čo sa týka jednokrokových predpovedí, z obrázka 6.13 môžeme pozorovať, že výkyvy v turbulentnejších mesiacoch, najmä od apríla do júna 2015, boli menšie ako pri kubickom vyhladzovacom splajne. Tomu zodpovedali aj lepšie hodnoty všetkých mier presnosti predpovedí. Na ustálenejších posledných mesiacoch testovacej vzorky mal však v kvalite predpovedí navrch kubický vyhladzovací splajn.

Predikčné intervaly mali v porovnaní s predchádzajúcimi metódami menšie rozpätie. Prvá hodnota testovacej vzorky navyše nebola nadhodnotená až tak ako pri kubickom splajne, hoci skutočná hodnota ani v tomto prípade nespadla ani do 95-percentného predikčného intervalu. Tento problém však nastal u takmer všetkých modelov, zrejme kvôli vyššej poslednej hodnote tréningovej vzorky dát.



Obrázok 6.13: Grafické znázornenie jednokrokových predpovedí Holtovou metódou

Odmocninová priemerná štvorcová chyba sa dostala pod 0,05 percentuálneho bodu. Priemerná absolútna percentuálna chyba bola dokonca nižšia ako 6 percent.

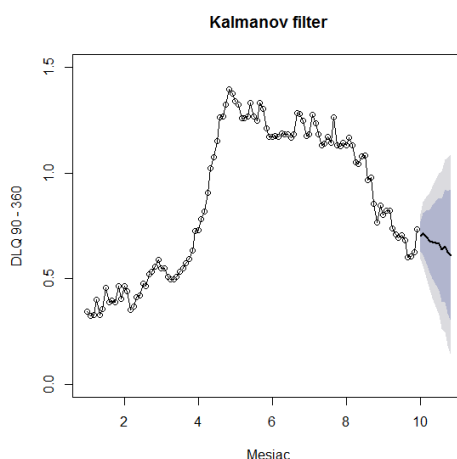
6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

Model bol v porovnaní s kubickým vyhladzovacím splajnom a lineárnou regresiou lepší aj v predpovedaní vývoja miery omeškania, keď správne predpovedal 8 z 11 medzimesačných pohybov. To znamenalo úspešnosť vyššiu ako 70 percent.

6.2.3 Kalmanov filter

Tretím modelom, ktorý sme na prvej časti portfólia vyskúšali, bol Kalmanov filter. Opäť sme pomocou neho vytvorili viackrokovú predpoveď aj jednokrokové predpovede, s rovnakou tréningovou a testovacou vzorkou dát ako v predchádzajúcich prípadoch. Tieto predpovede sme porovnaním so skutočnými hodnotami miery omeškania ohodnotili z hľadiska ich presnosti.

Už z obrázka 6.14 je zrejmé, že táto predpoveď bola lepšia ako pri kubickom vyrovnávacom splajne a Holtovej metóde, čo potvrdzovali aj miery presnosti predpovedí. Tieto boli zatiaľ najlepšie zo všetkých modelov. Na obrázku je tiež vidno, že viackroková predpoveď Kalmanovým filtrom nebola priamočiara ako v predchádzajúcich prípadoch, ale predpovedala aj výkyvy miery omeškania.



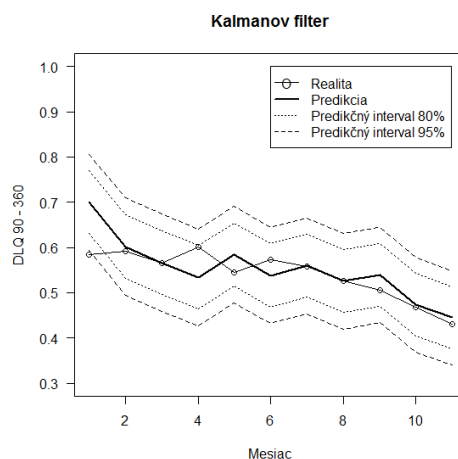
Obrázok 6.14: Grafické znázornenie viackrovej predpovede Kalmanovým filtrom

Odmocninová priemerná štvorcová chyba na testovacej vzorke bola nižšia ako 0,13 percentuálneho bodu, priemerná absolútna chyba bola tiež pod touto hranicou. Priemerná absolútna percentuálna chyba bola nižšia ako 24 percent. Odmocninová priemerná štvorcová chyba na posledných piatich pozorovaniach testovacej vzorky iba tesne presiahla 0,14 percentuálneho bodu.

Aj jednokrokové predpovede Kalmanovým filtrom fungovali lepšie ako predpovede kubickým vyhladzovacím splajnom alebo Holtovou metódou. Kalmanov

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

filter predpovedal správne, rovnako ako Holtova metóda, 8 z 11 medzimesačných zmien vo vývoji miery omeškania, čo opäť znamenalo úspešnosť presahujúcu 70 percent.



Obrázok 6.15: Grafické znázornenie jednokrokových predpovedí Kalmanovým filtrom

Za povšimnutie pri jednokrokových predpovediach Kalmanovým filtrom stojí hneď prvá predpoveď na testovacej vzorke, ktorú tento model nadhodnotil najmenej zo všetkých vyskúšaných modelov časových radov. Ako môžeme vidieť na obrázku 6.15, v tomto prípade padla skutočná hodnota miery omeškania približne na dolnú hranicu 95-percentného predikčného intervalu.

6.2.4 ARIMA model

Na prvej časti portfólia sme nakoniec vyskúšali aj model časových radov, ktorý nepatrí medzi modely založené na dekompozícii, ale využíva autokorelačné vlastnosti časového radu. Po otestovaní prítomnosti jednotkového koreňa rozšíreným Dickeovým – Fullerovým testom sme zistili, že pre získanie stacionárneho časového radu je potrebné vysvetľovanú premennú dvakrát diferencovať.

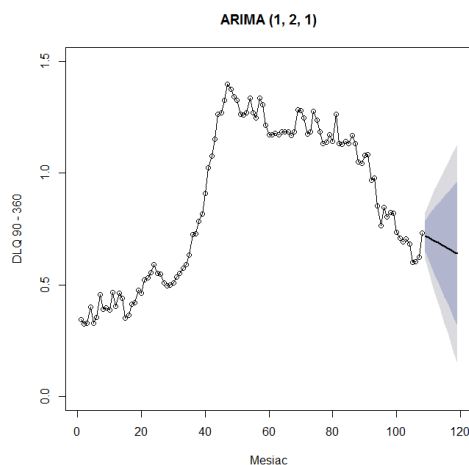
Po dvojitém diferencovaní bolo pomerne ťažké určiť typ modelu na základe autokorelačnej a parciálnej autokorelačnej funkcie. Rozhodli sme sa preto založiť jeho výber na minimalizácii AIC . Dostali sme sa tak k modelu $ARIMA(1,2,1)$, ktorý bol reprezentovaný rovnicou 6.6. Viackroková predpoveď týmto modelom fungovala pomerne uspokojivo, čo môžeme vidieť na obrázku 6.16.

$$\Delta^2 DLQ90360_t = -0,1195 \cdot \Delta^2 DLQ90360_{t-1} + \hat{\varepsilon}_t - 0,9109 \cdot \hat{\varepsilon}_{t-1}$$

Rovnica 6.6: Model ARIMA (1, 2, 1)

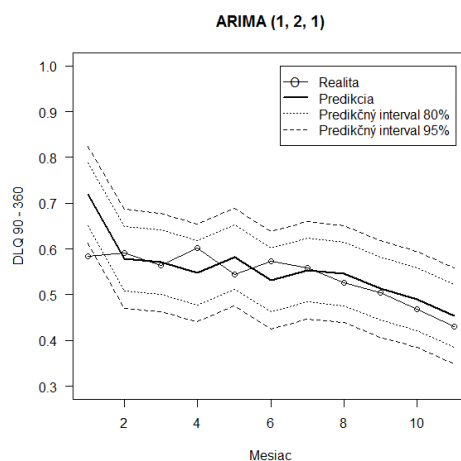
6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

Odmocninová priemerná štvorcová chyba viackrokovej predpovede, ktorá bola podobná ako pri Holtovej metóde, mierne presiahla 0,14 percentuálneho bodu, pričom na posledných piatich pozorovaniach testovacej vzorky bola vyššia ako 0,16 percentuálneho bodu. Priemerná absolútna chyba na testovacej vzorke bola mierne nižšia. Priemerná absolútna percentuálna chyba bola nad úrovňou 26 percent.



Obrázok 6.16: Grafické znázornenie viackrokovej predpovede modelom ARIMA (1, 2, 1)

Čo sa týka jednokrokových predpovedí, aj tieto boli podobné ako v prípade Holtovej metódy. Tomu zodpovedali aj takmer identické hodnoty mier presnosti predpovedí, ktoré však boli mierne horšie ako pri Kalmanovom filtri. *ARIMA(1,2,1)* model predpovedal správne 8 z 11 medzimesačných pohybov miery omeškania, čo znamenalo úspešnosť viac ako 70 percent.



Obrázok 6.17: Grafické znázornenie jednokrokových predpovedí modelom ARIMA (1, 2, 1)

6.3 Vyhodnotenie výsledkov

Na základe porovnania sme si zvolili jeden model na krátkodobú a jeden model na dlhodobú predpoveď miery omeškania prvej časti portfólia. Pomocou nich sme sa pokúsili predpovedať jej vývoj počas nasledujúcich troch mesiacov, ktoré neboli súčasťou vzorky dát, na ktorej sme modely vyvíjali. Výsledky predpovedí pre tieto mesiace sme následne porovnali s realitou.

6.3.1 Jednokrokové predpovede

V tabuľke 6.1 sa nachádza porovnanie jednokrokových predpovedí miery omeškania pre jedenásť mesiacov testovacej vzorky dát. *RMSE* označuje odmocninovú priemernú štvorcovú chybu v percentuálnych bodoch, v ktorých sa uvádza aj priemerná absolútna chyba *MAE*. Priemerná absolútna percentuálna chyba *MAPE* a úspešnosť predpovedania zmien vo vývoji miery omeškania *PSP* sú udávané v percentách.

MODEL	RMSE	MAE	MAPE	PSP
Kalmanov filter	0,0449	0,0294	5,2364	72,73
Holtova metóda	0,0483	0,0329	5,9424	72,73
ARIMA (1, 2, 1)	0,0485	0,0330	5,9447	72,73
Lineárna regresia	0,0406	0,0333	6,1302	63,64
Kubický splajn	0,0789	0,0499	8,7598	54,55

Tabuľka 6.1: Porovnanie predikčných modelov jednokrokovými predpoveďami

Pri jednokrokových predpovediach sme sa rozhodli používať Kalmanov filter, ktorý mal najnižšie hodnoty odmocninovej priemernej štvorcovej chyby, priemernej absolútnej chyby aj priemernej absolútnej percentuálnej chyby. Tento model mal zároveň, spolu s Holtovou metódou a *ARIMA(1,2,1)* modelom, najvyššiu úspešnosť predpovedania medzimesačných zmien vo vývoji miery omeškania.

KALMANOV FILTER	12.15	1.16	2.16
Predpoveď	0,4427	0,3800	0,4006
Skutočnosť	0,4319	0,3928	0,3987
Rozdiel	-0,0108	0,0128	-0,0019

Tabuľka 6.2: Jednokrokové predpovede Kalmanovým filtrom

V tabuľke 6.2 sú porovnané predpovedané hodnoty miery omeškania s jej skutočnými hodnotami v už spomínaných troch mesiacoch. Môžeme z nich vidieť, že Kalmanov filter najlepšie predpovedal mieru omeškania pre posledný z nich, teda

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

február 2016. V tabuľke 6.3 sú uvedené miery presnosti predpovedí počas tohto trojmesačného obdobia.

RMSE	MAE	MAPE	PSP
0,0097	0,0085	2,0723	100,00

Tabuľka 6.3: Miery presnosti jednokrokových predpovedí Kalmanovým filtrom

Hodnoty odmocninovej priemernej štvorcovej chyby aj priemernej absolútnej chyby sa zmestili pod 0,01 percentuálneho bodu, priemerná absolútna percentuálna chyba bola iba mierne vyššia ako 2 percentá. Za povšimnutie však stojí najmä stopercentná úspešnosť predpovedania medzimesačných zmien vo vývoji miery omeškania.

6.3.2 Viackrokové predpovede

Čo sa týka viackrokových predpovedí, *RMSE5* označuje v tabuľke 6.4 odmocninovú priemernú štvorcovú chybu na posledných piatich mesiacoch testovacej vzorky dát. Z hľadiska mier presnosti by sa mohlo zdať, že jednoznačne najlepšie predpovede poskytol model lineárnej regresie, z ich grafického znázornenia sme ale mohli vidieť, že nedokázali detailnejšie zachytiť vývoj miery omeškania.

MODEL	RMSE	MAE	MAPE	RMSE 5
Lineárna regresia	0,0379	0,0319	5,8965	0,03
Kalmanov filter	0,1277	0,1246	23,75	0,1436
Holtova metóda	0,1350	0,1319	25,15	0,1506
ARIMA (1, 2, 1)	0,1423	0,1388	26,51	0,1612
Kubický splajn	0,8115	0,7370	143,65	1,0729

Tabuľka 6.4: Porovnanie predikčných modelov viackrokovou predpoved'ou

Najlepšie miery presnosti predpovedí mal model lineárnej regresie iba z toho dôvodu, že ako jediný nenadhodnotil predpoveď miery omeškania v prvom mesiaci testovacej vzorky. Jeho veľkou nevýhodou boli aj absentujúce predikčné intervaly, ktoré sme sa síce pokúsili nasimulovať intervalmi spoľahlivosti, tieto však mali veľmi široké rozpätie a boli prakticky nepoužiteľné.

Pri viackrokových predpovediach sme sa preto opäť rozhodli používať model Kalmanovho filtra, ktorý mal spomedzi všetkých modelov časových radov najnižšie hodnoty odmocninovej priemernej štvorcovej chyby na celej testovacej vzorke dát, a aj na jej posledných piatich mesiacoch. Podobná situácia nastala pri priemernej absolútnej chybe a priemernej absolútnej percentuálnej chybe.

6 PREDIKČNÉ MODELY PRE PRVÚ ČASŤ PORTFÓLIA

KALMANOV FILTER	12.15	1.16	2.16
Predpoveď	0,4427	0,3897	0,3998
Skutočnosť	0,4319	0,3928	0,3987
Rozdiel	-0,0108	0,0031	-0,0010

Tabuľka 6.5: Viackroková predpoveď Kalmanovým filtrom

V tabuľke 6.5 sú porovnané predpovedané hodnoty miery omeškania s jej skutočnými hodnotami v troch mesiacoch mimo vývojovej vzorky. Môžeme si všimnúť, že Kalmanov filter opäť najpresnejšie predpovedal mieru omeškania pre mesiac február 2016. V tabuľke 6.6 sú zobrazené miery presnosti predpovedí počas tohto trojmesačného obdobia.

RMSE	MAE	MAPE
0,0065	0,0050	1,1790

Tabuľka 6.6: Miery presnosti viackrokovej predpovede Kalmanovým filtrom

Zaujímavým zistením sú lepšie hodnoty všetkých troch mier presnosti predpovedí v porovnaní s jednokrokovými predpoveďami. Pokles, aj keď iba o niekoľko tisícín percentuálneho bodu, nastal u všetkých troch mier. Najvýraznejšie sa pritom zlepšila priemerná absolútna percentuálna chyba, ktorá sa v tomto prípade priblížila bližšie k 1 percentu.

7 Predikčné modely pre druhú časť portfólia

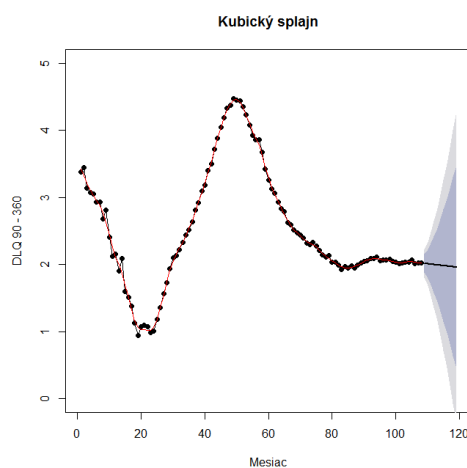
7.1 Model časových radov

Po skúsenosti s modelom lineárnej regresie z prvej časti portfólia sme sa pri druhej časti portfólia rozhodli zvoliť mierne odlišný postup. Pri výstavbe predikčného modelu miery omeškania sme začali modelmi časových radov bez vysvetľujúcich premenných. Až neskôr sme otestovali existenciu kointegračných vzťahov a kauzálneho pôsobenia makroekonomických ukazovateľov na mieru omeškania.

7.1.1 Kubický vyhladzovací splajn

V prvej časti portfólia sme mohli vidieť, že predpoveď kubickým vyhladzovacím splajnom nepatrila medzi najlepšie modely predpovedania budúcich hodnôt miery omeškania. Aj napriek tomu sme sa však pozreli na to, ako fungovala pri druhej časti portfólia. Opäť sme ju otestovali najprv viackrokovou a následne aj jednokrokovými predpoveďami na jedenástich mesiacoch testovacej vzorky dát.

Na obrázku 7.1 môžeme vidieť, že viackroková predpoveď predpovedala na testovacej vzorke, na ktorej miera omeškania v skutočnosti rástla, mierne klesajúci trend vysvetľovanej premennej. To sa odrazilo aj na hodnotách mier presnosti predpovedí.. Predpoveď kubickým vyhladzovacím splajnom však v tomto prípade fungovala o poznanie lepšie ako pri prvej časti portfólia.



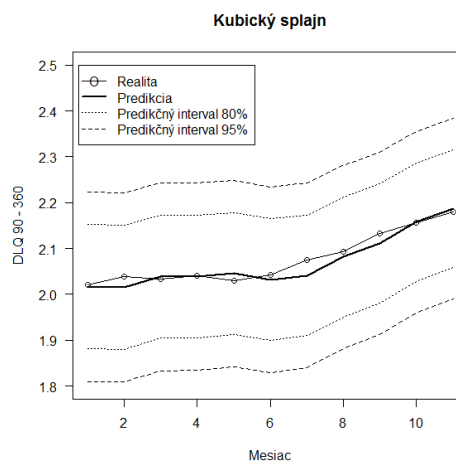
Obrázok 7.1: Grafické znázornenie viackrovej predpovede kubickým vyhladzovacím splajnom

Odmocninová priemerná štvorcová chyba presiahla 0,11 percentuálneho bodu, priemerná absolútna percentuálna chyba bola vyššia ako 4 percentá a priemerná

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

absolútna chyba nižšia ako 0,09 percentuálneho bodu. Odmocninová priemerná štvorcová chyba na posledných piatich mesiacoch testovacej vzorky dát dosiahla hodnotu 0,16 percentuálneho bodu.

Jednokrokové predpovede kubickým vyhladzovacím splajnom, ktorých grafické znázornenie môžeme vidieť na obrázku 7.2, mali odmocninovú priemernú štvorcovú chybu aj priemernú absolútnu chybu menšiu ako 0,02 percentuálneho bodu, pričom priemerná absolútna percentuálna chyba nedosiahla ani 0,6 percenta. Táto metóda správne predpovedala 7 z 11 medzimesačných zmien vo vývoji miery omeškania.



Obrázok 7.2: Grafické znázornenie jednokrokových predpovedí kubickým vyhladzovacím splajnom

7.1.2 Holtova metóda

Viacukroková bodová predpoveď Holtovou metódou, ktorej model je popísaný rovnicou 7.1, mala podobnú presnosť ako predpoveď kubickým vyhladzovacím splajnom. Predpoveď hodnôt miery omeškania na testovacej vzorke dát mala opäť klesajúci trend. Ako však môžeme vidieť na obrázku 7.3, intervalová predpoveď Holtovou metódou bola predsa len o niečo presnejšia, keď jej predikčné intervaly mali menšie rozpätie ako pri kubickom vyhladzovacom splajne.

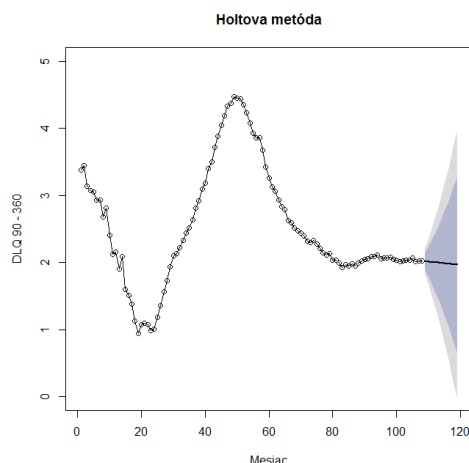
$$DLQ90360_{t+1} = L_t + B_t$$

$$L_0 = 3,5144; L_t = 0,6897 \cdot DLQ90360_t + (1 - 0,6897) \cdot (L_{t-1} + B_{t-1})$$

$$B_0 = -0,0947; B_t = 0,3541 \cdot (L_t - L_{t-1}) + (1 - 0,3541) \cdot B_{t-1}$$

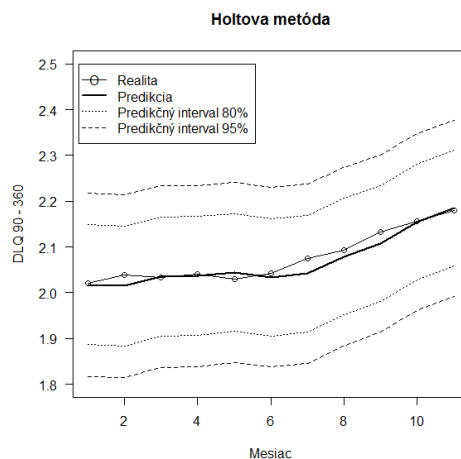
Rovnica 7.1: Model Holtovej metódy

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA



Obrázok 7.3: Grafické znázornenie viackrokovej predpovede Holtovou metódou

Jednokrokové predpovede, ktorých grafické znázornenie sa nachádza na obrázku 7.4, boli opäť porovnateľné s predpoveďami kubickým vyhladzovacím splajnom. Holtova metóda však mala tentokrát mierne horšie hodnoty odmocninovej priemernej štvorcovej chyby, priemernej absolútnej chyby aj priemernej absolútnej percentuálnej chyby. Na druhej strane však dokázala úspešne predpovedať o jednu medzimesačnú zmenu vo vývoji miery omeškania viac.

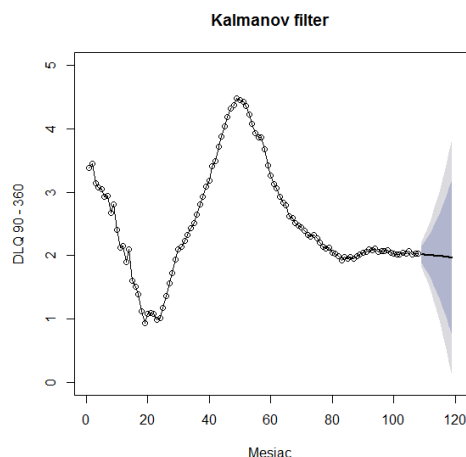


Obrázok 7.4: Grafické znázornenie jednokrokových predpovedí Holtovou metódou

7.1.3 Kalmanov filter

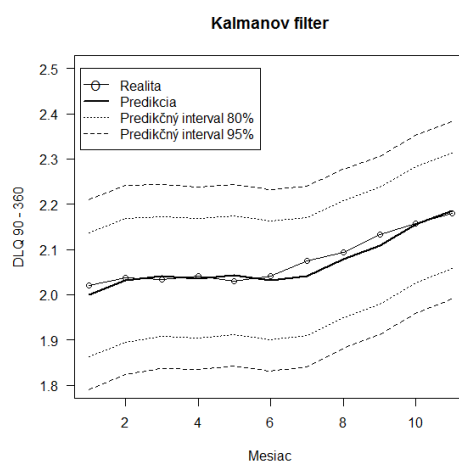
Viacroková predpoveď Kalmanovým filtrom, znázornená na obrázku 7.5, bola v porovnaní s Holtovou metódou mierne horšia, keď mala opäť mierne klesajúci trend. V porovnaní s predchádzajúcimi metódami síce mala menšie rozpätie predikčných intervalov, no miery presnosti predpovedí boli v porovnaní s Holtovou metódou o stotiny percentuálneho bodu horšie.

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA



Obrázok 7.5: Grafické znázornenie viackrokovej predpovede Kalmanovým filtrom

Čo sa týka jednokrokových predpovedí, tieto fungovali opäť o niečo horšie ako pri Holtovej metóde. Znázornené sú na obrázku 7.6. Predpovede Kalmanovým filtrom mali o tisíce percentuálneho bodu vyššie hodnoty priemerných absolútnych chýb. Rovnako ako Holtova metóda, aj Kalmanov filter správne predpovedal 8 z 11 medzimesačných zmien miery omeškania.



Obrázok 7.6: Grafické znázornenie jednokrokových predpovedí Kalmanovým filtrom

7.1.4 ARIMA model

Po troch metódach časových radov, z ktorých ani jedna nepatrila medzi modely Boxovej – Jenkinsovej metodológie, sme na účely predpovedania hodnôt miery omeškania druhej časti portfólia vyskúšali aj *ARIMA* model. Po otestovaní prítomnosti jednotkového koreňa sme zistili, že pre získanie stacionárneho časového radu je potrebné vysvetľovanú premennú diferencovať.

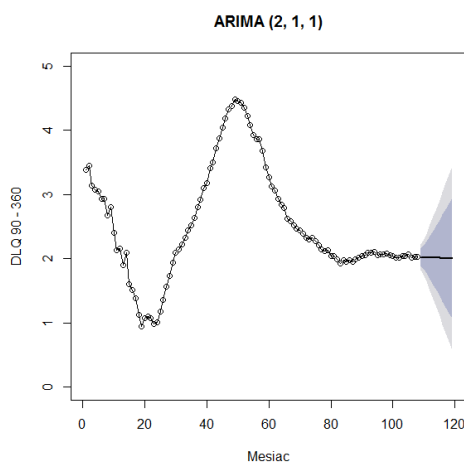
7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

Podobne ako pri prvej časti portfólia, aj v tomto prípade bolo po diferencovaní pomerne ťažké určiť vhodný typ modelu na základe priebehu autokorelačnej a parciálnej autokorelačnej funkcie. Rozhodli sme sa preto založiť jeho výber opäť na minimalizácii AIC . Tak sme sa dostali k modelu $ARIMA(2,1,1)$, ktorý bol reprezentovaný rovnicou 7.2.

$$\Delta DLQ90360_t = 0,4959 \cdot \Delta DLQ90360_{t-1} + 0,3968 \cdot \Delta DLQ90360_{t-2} + \hat{\varepsilon}_t - 0,4860 \cdot \hat{\varepsilon}_{t-1}$$

Rovnica 7.2: Model ARIMA (2, 1, 1)

Model $ARIMA(2,1,1)$ bol prvým modelom, ktorý na obdobie testovacej vzorky dát nepredpovedal výraznejšie klesajúci trend miery omeškania. Jeho predpoveď, ktorú môžeme vidieť na obrázku 7.7, mala preto v porovnaní s predchádzajúcimi modelmi výrazne nižšie hodnoty priemerných chýb. Predikčné intervaly mali navyše menšie rozpätie ako pri kubickom splajne, Holtovej metóde alebo Kalmanovom filtri.



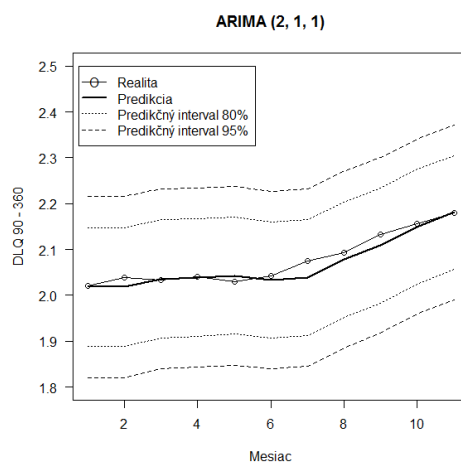
Obrázok 7.7: Grafické znázornenie viackrokovej predpovede modelom ARIMA (2, 1, 1)

Odmocninová priemerná štvorcová chyba viackrokovej predpovede bola menšia ako 0,09 percentuálneho bodu, kým priemerná absolútna percentuálna chyba iba mierne presiahla 3 percentá. Priemerná absolútna chyba bola nižšia ako 0,07 percentuálneho bodu a odmocninová priemerná štvorcová chyba na posledných piatich mesiacoch nepresiahla hodnotu 0,13 percentuálneho bodu.

Jednokrokové predpovede sú znázornené na obrázku 7.8. Fungovali podobne ako pri Holtovej metóde a Kalmanovom filtri, čo potvrdzovali aj takmer zhodné hodnoty odmocninovej priemernej štvorcovej chyby a priemernej absolútnej

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

percentuálnej chyby. Úspešnosť predpovedania medzimesačných zmien vo vývoji miery omeškania presiahla 70 percent.



Obrázok 7.8: Grafické znázornenie jednokrokových predpovedí modelom ARIMA (2, 1, 1)

7.2 Modely lineárnej regresie

V predchádzajúcej podkapitole sme videli, že modely časových radov fungovali pri viackrokových aj jednokrokových predpovediach pomerne dobre. Ich hlavnou nevýhodou však bolo že okrem minulých hodnôt miery omeškania nedokázali pri jej predpovedaní zohľadniť aj iné vysvetľujúce premenné. Tento problém sme sa rozhodli odstrániť vytvorením modelov lineárnej regresie s autokorelovanými rezíduami.

Čo sa týka vysvetľujúcich premenných, zamerali sme sa najmä na makroekonomické ukazovatele. Pri každom z nich sme najprv testom jednotkového koreňa zistili rád integrovanosti. Všetky makroekonomické premenné boli, rovnako ako miera omeškania druhej časti portfólia, integrované rádu 1. Znamenalo to, že sme mohli testovať, či sú s mierou omeškania kointegrované.

Pre účel testovania kointegrácie sme použili Johansenov kointegračný test. Pri každej premennej, u ktorej test existenciu kointegračného vzťahu s mierou omeškania nezamietol, sme ďalej otestovali, či ovplyvňuje mieru omeškania. Kauzálne pôsobenie makroekonomických ukazovateľov na vysvetľovanú premennú sme testovali Grangerovým testom kauzality.

Johansenov kointegračný test na hladine významnosti 5 percent zamietol hypotézu o existencii kointegračného vzťahu medzi indexom spotrebiteľských cien a mierou omeškania. Jeho výsledok môžeme vidieť na obrázku 7.9. Test Grangerovej

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

kauzality podobne zamietol hypotézu o kauzálnom pôsobení tohto indexu, reprezentujúceho infláciu, na mieru omeškania.

```
Values of teststatistic and critical values of test:
      test 10pct  5pct  1pct
r <= 1 |  2.33 10.49 12.25 16.26
r = 0  | 10.29 16.85 18.96 23.65
```

Obrázok 7.9: Výstup R pre test kointegrácie medzi mierou omeškania a zmenou CPI

Pri zvyšných dvoch makroekonomických premenných, nezamestnanosti a zmene hrubého domáceho produktu, testy existenciu kointegračného vzťahu ani kauzálného pôsobenia makroekonomických premenných na mieru omeškania nezamietli. V nasledujúcich častiach podkapitoly si priblížime predikčné modely, ktoré sa nám pomocou týchto premenných podarilo vytvoriť.

7.2.1 Model lineárnej regresie s ARIMA rezíduami a nezamestnanosťou

Ako sme už spomínali, Grangerov test kauzality nezamietol kauzálne pôsobenie nezamestnanosti na mieru omeškania. Ešte predtým Johansenov kointegračný test na hladine významnosti 5 percent nezamietol existenciu kointegračného vzťahu medzi týmito dvoma premennými. Jeho výsledok môžeme vidieť na obrázku 7.10. Na základe toho sme mohli pristúpiť k vytvoreniu modelu lineárnej regresie.

```
Values of teststatistic and critical values of test:
      test 10pct  5pct  1pct
r <= 1 |  4.07 10.49 12.25 16.26
r = 0  | 23.13 16.85 18.96 23.65
```

Obrázok 7.10: Výstup R pre test kointegrácie medzi mierou omeškania a nezamestnanosťou

Konkrétne sa jednalo o model lineárnej regresie vytvorený z diferencií premenných, ktorého rezíduá boli procesom $ARMA(3,1)$. Model v takomto tvare by sa dal previesť na tvar využívajúci pôvodné premenné. Rezíduá takéhoto modelu, ktorý by mohol obsahovať aj konštantný člen, by boli procesom $ARIMA(3,1)$. Spomínaný konštantný člen sa diferencovaním premenných eliminuje.

$$\Delta DLQ90360_t = 0,0104 \cdot \Delta UNEMPL_t + \Delta \hat{\varepsilon}_t$$

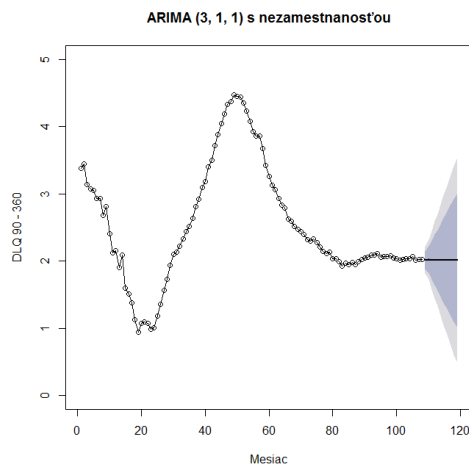
$$\Delta \hat{\varepsilon}_t = -0,5819 \cdot \Delta \hat{\varepsilon}_{t-1} + 0,5745 \cdot \Delta \hat{\varepsilon}_{t-2} + 0,5545 \cdot \Delta \hat{\varepsilon}_{t-3} + \hat{u}_t + 0,7003 \cdot \hat{u}_{t-1}$$

Rovnica 7.3: Model lineárnej regresie s nezamestnanosťou a ARIMA rezíduami

Viackroková predpoveď modelom lineárnej regresie fungovala o niečo lepšie ako pri $ARIMA$ modeli, čo sme okrem grafického zobrazenia na obrázku 7.11 mohli vidieť aj z hodnôt mier presnosti predpovedí. Zahnutie vysvetľujúcej premennej

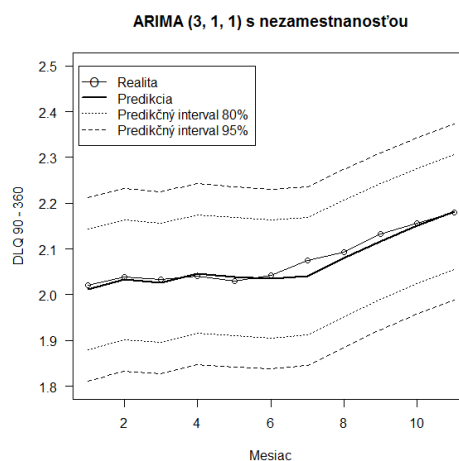
7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

zohralo pri predpovedi úlohu v tom, že predpovedaná miera omeškania na testovacej vzorke prvýkrát neklesala, čo sa prejavilo aj na spomínaných mierach presnosti.



Obrázok 7.11: Grafické znázornenie viackrokovej predpovede modelom lineárnej regresie s nezamestnanosťou a ARIMA rezíduami

Pri viackrokovej predpovedi bola odmocninová priemerná štvorcová chyba menšia ako 0,09 percentuálneho bodu, priemerná absolútna percentuálna chyba iba mierne presiahla hodnotu 3 percent. Priemerná absolútna chyba bola nižšia ako 0,07 percentuálneho bodu a odmocninová priemerná štvorcová chyba na posledných piatich mesiacoch mala menšiu hodnotu ako 0,13 percentuálneho bodu.



Obrázok 7.12: Grafické znázornenie jednokrokových predpovedí modelom lineárnej regresie s nezamestnanosťou a ARIMA rezíduami

Z obrázka 7.12 je zrejmé, že aj pri jednokrokových predpovediach nastalo zlepšenie, čo sa týka ich presnosti. Najvýraznejšie sa zvýšila úspešnosť predpovedania medzimesačných zmien vo vývoji miery omeškania, ktorá v tomto prípade presiahla 90

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

percent. Prvých 6 hodnôt testovacej vzorky navyše tento model predpovedal oveľa lepšie ako všetky predchádzajúce modely.

Odmocninová priemerná štvorcová chyba jednokrokových predpovedí nedosiahla ani 0,02 percentuálneho bodu, pričom priemerná absolútna chyba iba mierne presiahla hranicu 0,01 percentuálneho bodu. Priemerná absolútna percentuálna chyba bola vyššia ako 0,5 percenta. Ako sme už spomínali, model úspešne predpovedal 10 z 11 medzimesačných pohybov miery omeškania.

7.2.2 Model lineárnej regresie s ARIMA rezíduami a zmenou HDP

Ešte lepšie fungoval model lineárnej regresie, ktorý sme vytvorili na základe zmien *HDP*, oneskorených o jeden mesiac. Grangerov test kauzality na hladine významnosti 5 percent nezamietol hypotézu o ich kauzálnom pôsobení na mieru omeškania. Ako môžeme vidieť na obrázku 7.13, Johansenov test kointegrácie nezamietol ani hypotézu o existencii práve jedného kointegračného vzťahu.

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 1		2.38	10.49	12.25 16.26
r = 0		25.69	16.85	18.96 23.65

Obrázok 7.13: Výstup R pre test kointegrácie medzi mierou omeškania a zmenou HDP

Vytvorený model lineárnej regresie bol podobný ako v predchádzajúcom prípade, keď rezíduá modelu boli opäť procesom *ARIMA(3,1)*. Ako je zrejmé z rovnice 7.4, aj odhady parametrov zostali podobné ako v predchádzajúcom modeli. Parameter prislúchajúcej vysvetľujúcej premennej mal očakávané znamienko. V porovnaní s predchádzajúcim modelom mal aj menšiu štandardnú odchýlku.

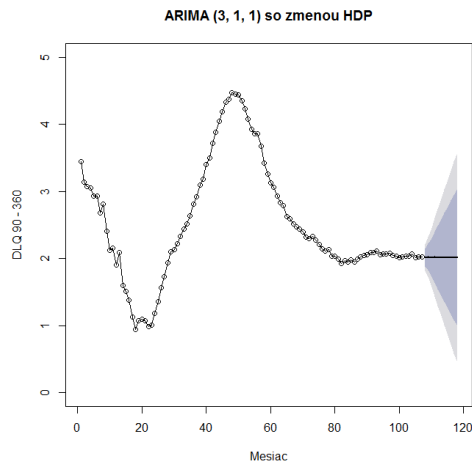
$$\Delta DLQ90360_t = -0,0200 \cdot \Delta GDP_{t-1} + \Delta \hat{\varepsilon}_t$$

$$\Delta \hat{\varepsilon}_t = -0,5877 \cdot \Delta \hat{\varepsilon}_{t-1} + 0,5947 \cdot \Delta \hat{\varepsilon}_{t-2} + 0,5543 \cdot \Delta \hat{\varepsilon}_{t-3} + \hat{u}_t + 0,7271 \cdot \hat{u}_{t-1}$$

Rovnica 7.4: Model lineárnej regresie so zmenou HDP a ARIMA rezíduami

Z obrázka 7.14 sme mohli vidieť, že viackroková predpoveď fungovala podobne ako v predchádzajúcom modeli, ktorý ako vysvetľujúcu premennú využíval nezamestnanosť. Predpovedaná miera omeškania opäť neklesala, až na malé výnimky mala skôr mierne rastúci trend. Predpovede boli v porovnaní s predchádzajúcim modelom o niečo presnejšie.

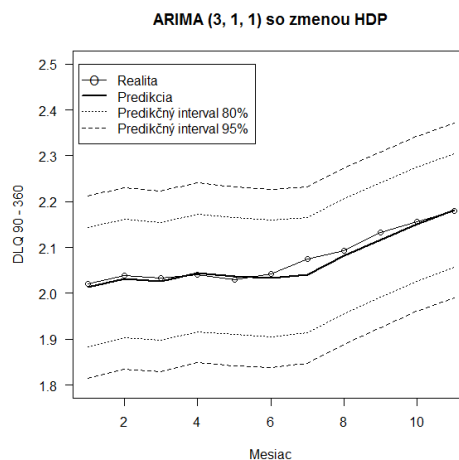
7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA



Obrázok 7.14: Grafické znázornenie viackrokovej predpovede modelom lineárnej regresie so zmenou HDP a ARIMA rezíduami

Odmocninová priemerná štvorcová chyba viackrokovej predpovede bola menšia ako 0,08 percentuálneho bodu, priemerná absolútna percentuálna chyba bola pod úrovňou 3 percent. Priemerná absolútna chyba bola nižšia ako 0,06 percentuálneho bodu a odmocninová priemerná štvorcová chyba na posledných piatich mesiacoch dosiahla hodnotu menšiu ako 0,12 percentuálneho bodu.

Jednokrokové predpovede boli podobné modelu, ktorý predpovedal mieru omeškania pomocou nezamestnanosti. Úspešnosť predpovedania medzimesačných zmien vo vývoji miery omeškania mali tieto modely rovnakú. Z obrázka 7.15 je zrejmé, že prvých 6 predpovedaných hodnôt testovacej vzorky, podobne ako v predchádzajúcom modeli, veľmi dobre kopirovalo vývoj skutočnej miery omeškania.



Obrázok 7.15: Grafické znázornenie jednokrokových predpovedí modelom lineárnej regresie so zmenou HDP a ARIMA rezíduami

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

Odmocninová priemerná štvorcová chyba nedosiahla ani v tomto prípade 0,02 percentuálneho bodu, pričom priemerná absolútna chyba mierne sa dokonca dostala pod hranicu 0,01 percentuálneho bodu. Priemerná absolútna percentuálna chyba bola menšia ako 0,5 percenta. Model úspešne predpovedal 10 z 11 medzimesačných pohybov miery omeškania, čo znamenalo úspešnosť viac ako 90 percent.

7.2.3 Model lineárnej regresie s ARIMA rezíduami, zmenou HDP a nezamestnanosťou

Posledným predikčným modelom, ktorý sme pre druhú časť portfólia vytvorili, bol model lineárnej regresie, predpovedajúci mieru omeškania pomocou dvoch vysvetľujúcich premenných, o mesiac oneskorenej zmeny *HDP* a nezamestnanosti. Keďže už predpovede predchádzajúceho modelu boli pomerne uspokojivé, boli sme zvedaví, či navýšenie počtu vysvetľujúcich premenných ešte zvýši ich presnosť.

Pred vytvorením modelu sme testom Grangerovej kauzality opäť nezamietli kauzálne pôsobenie makroekonomických ukazovateľov na mieru omeškania. Následne sme Johansenovým testom kointegrácie, ktorého výsledok je zobrazený na obrázku 7.16, na hladine významnosti 5 percent nezamietli hypotézu o existencii kointegračného vzťahu. Pokračovali sme preto konštrukciou samotného modelu.

```
Values of teststatistic and critical values of test:
      test 10pct  5pct  1pct
r <= 2 |  3.45 10.49 12.25 16.26
r <= 1 | 12.55 16.85 18.96 23.65
r = 0  | 39.87 23.11 25.54 30.34
```

Obrázok 7.16: Výstup R pre test kointegrácie medzi mierou omeškania, zmenou HDP a nezamestnanosťou

Odhady parametrov z rovnice 7.5 a ich štandardné odchýlky boli veľmi podobné ako v predchádzajúcich modeloch. Viackroková predpoveď, zobrazená na obrázku 7.17, fungovala v porovnaní so všetkými ostatnými modelmi mierne lepšie, keď miery presnosti jej predpovedí boli o tisíce percentuálneho bodu lepšie ako v modeli, ktorý mal ako vysvetľovanú premennú iba zmenu *HDP*.

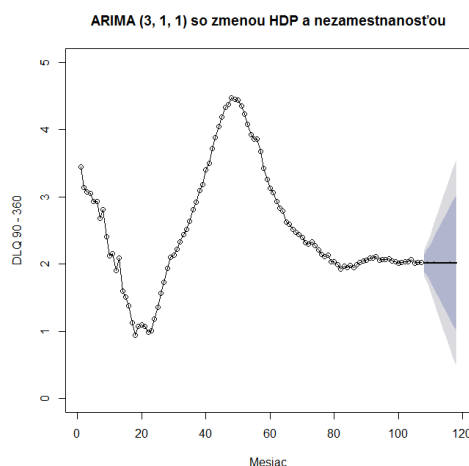
$$\Delta DLQ90360_t = -0,0199 \cdot \Delta GDP_{t-1} + 0,0110 \cdot \Delta UNEMPL_t + \Delta \hat{\varepsilon}_t$$

$$\Delta \hat{\varepsilon}_t = -0,5911 \cdot \Delta \hat{\varepsilon}_{t-1} + 0,5900 \cdot \Delta \hat{\varepsilon}_{t-2} + 0,5509 \cdot \Delta \hat{\varepsilon}_{t-3} + \hat{u}_t + 0,7285 \cdot \hat{u}_{t-1}$$

Rovnica 7.5: Model lineárnej regresie so zmenou HDP, nezamestnanosťou a ARIMA rezíduami

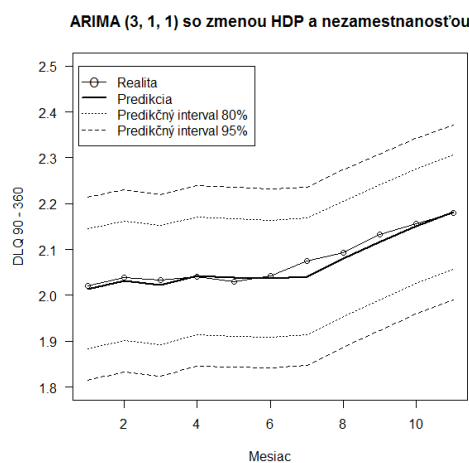
7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

Odmocninová priemerná štvorcová chyba viackrokovej predpovede bola menšia ako 0,08 percentuálneho bodu, priemerná absolútna percentuálna chyba bola pod úrovňou 3 percent. Priemerná absolútna chyba bola nižšia ako 0,06 percentuálneho bodu a odmocninová priemerná štvorcová chyba na posledných piatich mesiacoch dosiahla hodnotu menšiu ako 0,12 percentuálneho bodu.



Obrázok 7.17: Grafické znázornenie viackrokovej predpovede modelom lineárnej regresie so zmenou HDP, nezamestnanosťou a ARIMA rezíduami

Zvýšenie počtu vysvetľujúcich premenných so sebou neprinieslo výrazné zlepšenie presnosti jednokrokových predpovedí. Odmocninová priemerná štvorcová chyba, priemerná absolútna chyba a aj priemerná absolútna percentuálna chyba boli podobné ako v modeli využívajúcom nezamestnanosť. Jednokrokové predpovede sú znázornené na obrázku 7.18.



Obrázok 7.18: Grafické znázornenie jednokrokových predpovedí modelom lineárnej regresie so zmenou HDP, nezamestnanosťou a ARIMA rezíduami

7.3 Vyhodnotenie výsledkov

Aj pri druhej časti portfólia sme si na základe porovnania presnosti predpovedí všetkých vytvorených modelov vybrali jeden model na krátkodobú a jeden model na dlhodobú predpoveď. Tieto modely sme následne aplikovali do praxe, keď sme pomocou nich predpovedali vývoj miery omeškania druhej časti portfólia počas troch mesiacov, ktoré neboli súčasťou vývojovej vzorky dát. Výsledky predpovedí sme následne porovnali s realitou.

7.3.1 Jednokrokové predpovede

MODEL	RMSE	MAE	MAPE	PSP
ARIMA (3, 1, 1)+GDP	0,0133	0,0099	0,4769	90,91
ARIMA (3, 1, 1)+UNEMPL+GDP	0,0133	0,0101	0,4843	90,91
ARIMA (3, 1, 1)+UNEMPL	0,0135	0,0104	0,5020	90,91
ARIMA (2, 1, 1)	0,0158	0,0117	0,5618	72,73
Holtova metóda	0,0161	0,0123	0,5934	72,73
Kalmanov filter	0,0159	0,0128	0,6171	72,73
Kubický splajn	0,0157	0,0123	0,5923	63,64

Tabuľka 7.1: Porovnanie predikčných modelov jednokrokovými predpoveďami

Pri jednokrokových predpovediach sme sa rozhodli používať model lineárnej regresie so zmenou *HDP* a *ARIMA* rezíduami. Z tabuľky 7.1 totiž bolo zrejmé, že tento model mal najnižšie hodnoty odmocniny priemernej štvorcovej chyby, priemernej absolútnej chyby aj priemernej absolútnej percentuálnej chyby. Tento model mal zároveň najvyššiu úspešnosť predpovedania medzimesačných zmien miery omeškania, ktorá presiahla 90 percent.

ARIMA + GDP	12.15	1.16	2.16
Predpoveď	2,2007	2,2397	2,3585
Skutočnosť	2,2228	2,3183	2,3542
Rozdiel	0,0221	0,0786	-0,0043

Tabuľka 7.2: Jednokrokové predpovede modelom lineárnej regresie so zmenou *HDP* a *ARIMA* rezíduami

V tabuľke 7.2 sú porovnané predpovedané hodnoty miery omeškania s jej skutočnými hodnotami v spomínaných troch mesiacoch. Môžeme z nich vidieť, že tento model predpovedal mieru omeškania druhej časti portfólia pre február 2016 presne na stotinu percentuálneho bodu. Takmer presný bol model lineárnej regresie aj

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

v predpovedi pre december 2015. Tabuľka 7.3 vyobrazuje miery presnosti predpovedí počas tohto obdobia.

RMSE	MAE	MAPE	PSP
0,0472	0,0350	1,5219	100,00

Tabuľka 7.3: Miery presnosti jednokrokových predpovedí modelom lineárnej regresie so zmenou HDP a ARIMA rezíduami

Za povšimnutie stojí najmä stopercentná úspešnosť predpovedania medzimesačných zmien miery omeškania. Hodnoty odmocninovej priemernej štvorcovej chyby aj priemernej absolútnej chyby sa, podobne ako pri pôvodnej testovacej vzorke dát, zmestili pod 0,05 percentuálneho bodu. Priemerná absolútna percentuálna chyba iba tesne prekročila 1,5 percenta.

7.3.2 Viackroková predpoveď

Čo sa týka viackrokových predpovedí, jednoznačne najlepšie predpovede z hľadiska priemerných chýb poskytol model lineárnej regresie so zmenou *HDP*, nezamestnanosťou a *ARIMA* rezíduami. Ako môžeme vidieť v tabuľke 7.4, tento model mal najnižšie hodnoty odmocninovej priemernej štvorcovej chyby na celej testovacej vzorke aj na jej posledných piatich mesiacoch. Model tiež vyšiel najlepšie z porovnania priemerných absolútnych chýb a priemerných absolútnych percentuálnych chýb.

MODEL	RMSE	MAE	MAPE	RMSE 5
ARIMA (3, 1, 1)+UNEMPL+GDP	0,0771	0,0559	2,6298	0,1132
ARIMA (3, 1, 1)+GDP	0,0789	0,0576	2,7096	0,1157
ARIMA (3, 1, 1)+UNEMPL	0,0852	0,0645	3,0398	0,1243
ARIMA (2, 1, 1)	0,0877	0,0654	3,0792	0,1281
Holtova metóda	0,1095	0,0849	4,0090	0,1584
Kalmanov filter	0,1108	0,0876	4,1391	0,1600
Kubický splajn	0,1110	0,0866	4,0880	0,1603

Tabuľka 7.4: Porovnanie predikčných modelov viackrokovou predpoveďou

V tabuľke 7.5 sú predpovedané hodnoty miery omeškania troch mesiacov, ktoré neboli súčasťou vývojovej vzorky dát, porovnané s jej skutočnými hodnotami. Model lineárnej regresie nepredpovedal mieru omeškania presne na stotinu percentuálneho bodu ani pre jeden z troch mesiacov. Nezachytenie prudšieho nárastu miery omeškania v prvých dvoch mesiacoch roka 2016 sa mu však vyčítať nedá, keďže tento nárast zrejme nebol spôsobený makroekonomickými faktormi.

7 PREDIKČNÉ MODELY PRE DRUHÚ ČASŤ PORTFÓLIA

ARIMA + UNEMPL + GDP	12.15	1.16	2.16
Predpoveď	2,2012	2,2157	2,2332
Skutočnosť	2,2228	2,3183	2,3542
Rozdiel	0,0215	0,1026	0,1211

Tabuľka 7.5: Viackroková predpoveď modelom lineárnej regresie so zmenou HDP, nezamestnanosťou a ARIMA rezíduami

Čo sa týka mier presnosti predpovedí, odmocninová priemerná štvorcová chyba presiahla 0,09 percentuálneho bodu, priemerná absolútna chyba 0,08 percentuálneho bodu. Priemerná absolútna percentuálna chyba bola iba mierne vyššia ako 3,5 percenta. Tieto výsledky boli, pochopiteľne, o niečo horšie ako pri jednokrokových predpovediach.

RMSE	MAE	MAPE
0,0924	0,0817	3,5118

Tabuľka 7.6: Miery presnosti viackrokovej predpovede modelom lineárnej regresie so zmenou HDP, nezamestnanosťou a ARIMA rezíduami

Celkovo môžeme hodnotiť predikčné schopnosti modelov lineárnej regresie, vytvorených pre druhú časť portfólia, pozitívne. Modely si okrem uspokojujívých výsledkov validácie na testovacej vzorke dajú pomerne dobre poradiť aj s krátkodobým a dlhodobým predpovedaním miery omeškania pre mesiace, ktoré nepatrili do obdobia vývojovej vzorky dát.

Záver

Cieľom tejto diplomovej práce bolo pomocou vhodných štatistických modelov predpovedať hodnoty ukazovateľov kvality aktív pre dve rozdielne časti retailového portfólia Všeobecnej úverovej banky, a.s.. Práca bola rozdelená na dve hlavné časti, teoretickú a praktickú. Teoretickú časť tvorili štyri úvodné kapitoly, praktická časť sa skladala z troch ďalších kapitol.

Prvá kapitola definovala komerčnú banku, jej retailové portfólio a produkty, z ktorých sa toto portfólio skladá. Ďalej sme v tejto kapitole charakterizovali finančné riziko, jeho riadenie a ukazovatele kvality aktív. Prvá kapitola tiež bližšie popísala teóriu predpovedania a zhrnula možnosti predikcie ukazovateľov kvality aktív. Zakončená bola odôvodnením výberu miery omeškania ako vysvetľovanej premennej.

V druhej kapitole bolo vysvetlené teoretické pozadie lineárnej regresie. Okrem dvoch typov odhadu jej parametrov sme sa venovali aj kritériám výberu modelu, konštrukcii predpovedí a častým problémom vyskytujúcim sa v praxi, najmä multikolinearite, heteroskedasticite a autokorelácii rezíduí. Pri každom z nich boli stručne popísané metódy identifikácie, dôsledky ignorovania a možnosti riešenia.

Tretia kapitola popisovala hlavne dekompozičné metódy analýzy časových radov. Charakterizovali sme v nej jednotlivé zložky časového radu. Najviac pozornosti sme pritom venovali adaptívnym prístupom k trendovej zložke, konkrétne metóde kubického vyhladzovacieho splajnu, exponenciálnemu vyhladzovaniu a Holtovej metóde. Na záver tretej kapitoly sme stručne popísali aj model Kalmanovho filtra.

Vo štvrtej kapitole sme sa venovali prevažne Boxovej – Jenkinsovej metodológii. Okrem definície stacionarity a autokorelačných funkcií v nej boli popísané aj základné procesy stacionárnych časových radov. Ďalej sme sa v tejto kapitole zaoberali testovaním jednotkového koreňa, procesom *ARIMA* a konštrukciou predpovedí. Na záver sme sa zmienili o kointegrácii časových radov, Grangerovej kauzalite a o ich testoch.

Piatu kapitolu tvoril popis reálnych vnútrobankových aj makroekonomických dát, s ktorými sme pracovali. V šiestej kapitole sme popísali vývoj modelu lineárnej regresie pre prvú časť portfólia. Po neuspokojivých výsledkoch jeho validácie sme pre

ZÁVER

prvú časť portfólia vyvinuli aj štyri modely časových radov, a to prístupmi kubického vyhladzovacieho splajnu, Holtovej metódy, Kalmanovho filtra a *ARIMA* modelu.

V siedmej kapitole sme mieru omeškania druhej časti portfólia predpovedali najprv modelmi časových radov bez vysvetľujúcich premenných. Následne sme testovali existenciu kointegračných vzťahov a kauzálneho pôsobenia medzi makroekonomickými ukazovateľmi a mierou omeškania. Po potvrdení ich existencie sme vytvorili tri modely lineárnej regresie s autokorelovanými rezíduami.

Pri prvej aj druhej časti portfólia sme vytvorené predikčné modely validovali viackrokovou predpoveďou aj jednokrokovými predpoveďami na testovacej vzorke dát, ktoré sme následne porovnali. Pre každú časť portfólia sme si vybrali jeden model na dlhodobé a jeden model na krátkodobé predpovede. Ich presnosť sme otestovali na troch mesiacoch, ktoré neboli súčasťou vzorky dát, na ktorej boli modely vyvinuté.

Pri jednokrokových aj viackrokových predpovediach hodnôt miery omeškania prvej časti portfólia sme sa rozhodli používať Kalmanov filter. Pre účely jednokrokových predpovedí pri druhej časti portfólia sme si vybrali model lineárnej regresie so zmenou *HDP* a *ARIMA* rezíduami. Čo sa týka viackrokových predpovedí, zvolili sme si model, ktorý ako vysvetľujúcu premennú využíval aj nezamestnanosť.

Prínosom práce bolo spracovanie témy predpovedania hodnôt ukazovateľov kvality aktív na úrovni retailového portfólia komerčnej banky, ktorej doteraz nebola venovaná dostatočná pozornosť, najmä čo sa týka literatúry písanej v slovenskom jazyku. Užitočná bola aj ilustrácia fungovania jednotlivých predikčných modelov na reálnych dátach, s popisom postupu, porovnaním a vyhodnotením výsledkov.

Pre autora spočíval prínos diplomovej práce v podrobnejšom oboznámení sa s osobitosťami jednotlivých modelov a špecifikami ich praktickej aplikácie na účely predpovedania s využitím reálnych dát. Veľmi obohacujúce bolo pre autora aj rozšírenie teoretických poznatkov o využití štatistických metód v oblasti modelovania a predpovedania ukazovateľov kvality aktív, či kreditného rizika vo všeobecnosti.

V diplomovej práci sme sa zaoberali výstavbou predikčného modelu hodnôt ukazovateľov kvality aktív na úrovni portfólia. Rozšírením tejto práce by mohla byť snaha o model, ktorý by ich portfóliové hodnoty predpovedal, aj s využitím rôznych vysvetľujúcich premenných, pomocou čiastkových predpovedí na úrovni jednotlivých účtov, pričom by zohľadňoval aj budúce zmeny v štruktúre portfólia.

Zoznam použitej literatúry

- [1] ANDĚL, J.: *Základy matematické statistiky*. MatfyzPress, Praha, 2005.
- [2] BAESENS, B. a kol.: Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54 (2003), 627–635.
- [3] BANASIK, J., CROOK, J., THOMAS, L.: Not if but when will borrowers default? *Journal of the Operational Research Society*, 50 (1999), 1185–1190.
- [4] BEASLEY, J. E.: *Forecasting*. Elektronické študijné materiály, Imperial College, London, 2016. Dostupné na internete (29.2.2016): <http://people.brunel.ac.uk/~mastjib/jeb/or/forecast.html>
- [5] BELLOTTI, T., CROOK, J.: Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operations Research Society*, 60 (2009), 1699–1707.
- [6] BELLOTTI, T., CROOK, J.: Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29 (2013), 563-574.
- [7] BELLOTTI, T., CROOK, J.: Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36 (2009), 3302–3308.
- [8] BEŇOVÁ, E. a kol.: *Financie a mena*. Iura Edition, Bratislava, 2007.
- [9] BETANCOURT, L.: Using Markov chains to estimate losses from a portfolio of mortgages. *Review of Quantitative Finance and Accounting*, 12 (1999), 303-317.
- [10] BISHOP, G., WELCH, G.: *An Introduction to the Kalman Filter*. Elektronické študijné materiály, University of North Carolina, Chapel Hill, 2006. Dostupné na internete (6.4.2016): http://cs.unc.edu/~welch/media/pdf/kalman_intro.pdf
- [11] BLUHM, C., OVERBECK, L., WAGNER, C.: *An Introduction to Credit Risk Modelling*. Chapman & Hall/CRC, London, 2010.
- [12] BOX, G. E. P., JENKINS, G. M.: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [13] BREEDEN, J. L.: Portfolio forecasting tools: What you need to know. *The RMA Journal*, 1 (2003), 2-11.
- [14] BROKWCWELL, P. J., DAVIS, R. A.: *Introduction to Time Series and Forecasting*. Springer-Verlag, New York, 2002.
- [15] CIPRA, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL, Praha, 1986.
- [16] CIPRA, T.: *Finanční ekonometrie*. Ekopress, Praha, 2008.

ZOZNAM POUŽITEJ LITERATÚRY

- [17] CROOK, J., BANASIK, J.: Forecasting and explaining aggregate consumer credit delinquency behaviour. *International Journal of Forecasting*, 28 (2012), 145–160.
- [18] CROOK, J., EDELMAN, D., THOMAS, L.: Editorial. *International Journal of Forecasting*, 28 (2012), 128–132.
- [19] CROOK, J., EDELMAN, D., THOMAS, L.: Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183 (2007), 1447–1465.
- [20] DICKEY, D. A., FULLER, W. A.: Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74 (1979), 427–431.
- [21] DIMITRIU, M., OPREA, I. A., SCRIECIU, M. A.: Credit risk modeling using multiple regressions. *International Journal of Advances in Management and Economics*, 5 (2012), 125–131.
- [22] DVORŤÁK, P.: *Bankovníctví pro bankéře a klienty*. LINDE, Praha, 2005.
- [23] EKONOMICKÁ ENCYKLOPÉDIA. Dostupné na internete (11.11.2015): <http://euroekonom.sk>
- [24] EUROSTAT. Dostupné na internete (10.11.2015): <http://ec.europa.eu/eurostat>
- [25] FINANCE.SK. Dostupné na internete (16.11.2015): <http://finance.sk>
- [26] GAFFNEY, E., KELLY, R., McCANN, F.: *A Transitions-Based Framework for Estimating Expected Credit Losses*. Research Technical Papers, Central Bank of Ireland, 2014.
- [27] GRANGER, C. W. J.: Some recent development in a concept of causality. *Journal of Econometrics*, 39 (1988), 199–211.
- [28] GRANGER, C. W. J., ENGLE, R. F.: Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55 (1987), 251–276.
- [29] GREENE, W. H.: *Econometric Analysis*. Prentice Hall, New Jersey, 2003.
- [30] GRIMSHAW, S. D., ALEXANDER, W. P.: Markov chain models for delinquency: Transition matrix estimation and forecasting. *Applied Stochastic Models in Business Industry*, 27 (2011), 267–279.
- [31] HABGOOD, A. a kol.: *Evaluating Forecast Performance*. Independent Evaluation Office, Bank of England, 2015.
- [32] HAMERLE, A., LIEBIG, T., SCHEULE, H.: *Forecasting Credit Portfolio Risk*. Banking and Financial Supervision, Deutsche Bundesbank, 2004.
- [33] HAMILTON, J. D.: *Time Series Analysis*. Princeton University Press, Princeton, 1994.

ZOZNAM POUŽITEJ LITERATÚRY

- [34] HAND, D. J.: Evaluating diagnostic tests: the area under the ROC curve and the balance of errors. *Statistics in Medicine*, 29 (2010), 1502–1510.
- [35] HARVEY, A. C.: *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1991.
- [36] HEIJ, C. a kol.: *Econometric Methods with Applications in Business and Economics*. Oxford University Press, Oxford, 2004.
- [37] HORVÁTOVÁ, E.: *Bankovníctvo*. Súvaha, Bratislava, 2000.
- [38] HYNDMAN, R. J., ATHANASOPOULOS, G.: *Forecasting: Principles and Practice*. Otexts, Melbourne, 2013.
- [39] CHATTERJEE, S., HADI, S.A.: *Regression Analysis by Example*. John Wiley and Sons, New York, 2012.
- [40] JOHANSEN, S.: Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59 (1991), 1551-1580.
- [41] JOHNSTON, J., DiNARDO, J.: *Econometric methods*. McGraw Hill, New York, 1997.
- [42] JURČA, P.: *Kvantitatívne metódy v riadení rizík – kapitola 1*. Elektronické študijné materiály, FMFI UK, Bratislava, 2016. Dostupné na internete (11.1.2016): <http://iam.fmph.uniba.sk/institute/jurca/qrm/Chapter1.pdf>
- [43] JURČA, P.: *Kvantitatívne metódy v riadení rizík – kapitola 5*. Elektronické študijné materiály, FMFI UK, Bratislava, 2016. Dostupné na internete (11.1.2016): <http://iam.fmph.uniba.sk/institute/jurca/qrm/Chapter5.pdf>
- [44] KENNEDY, P. A.: *Guide to Econometrics*. MIT Press, Cambridge, 2003.
- [45] KIRCHGÄSSNER, G., WOLTERS, J.: *Introduction to Modern Time Series Analysis*. Springer-Verlag, Berlin, 2007.
- [46] KLOEK, T.: Loss development forecasting models: an econometrician's view. *Insurance Mathematics and Economics*, 3 (1998), 251-261.
- [47] LEOW, M., CROOK, J.: Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, 236 (2014), 685–694.
- [48] LÜTKEPOHL, H., REIMERS, H.: Granger-causality in cointegrated var processes: The case of the term structure. *Economics Letters*, 40 (1992), 263-268.
- [49] MacKINNON, J. G.: Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11 (1996), 601–618.
- [50] MALIK M., THOMAS, L. C.: Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, 28 (2012), 261-272.

ZOZNAM POUŽITEJ LITERATÚRY

- [51] MARČEK, D., MARČEK, M.: *Analýza, modelovanie a prognózovanie časových radov s aplikáciami v ekonomike*. Edis, Žilina, 2001.
- [52] McCABE, B., LEYBOURNE, S., HARRIS, D.: A residual-based test for stochastic cointegration. *Econometric Theory*, 22 (2006), 429-456.
- [53] McCANN, F.: *Modelling default transitions in the UK mortgage market*. Research Technical Papers, Central Bank of Ireland, 2014.
- [54] MOON, T. H., SOHN, S. Y.: Survival analysis for technology credit scoring adjusting total perception. *Journal of the Operational Research Society*, 62 (2011), 1159–1168.
- [55] PAN, Y., JACKSON, R. T.: Ethnic difference in the relationship between acute inflammation and serum ferritin in US adult males. *Epidemiology and Infection*, 136 (2008), 421-431.
- [56] PHILLIPS, P. C. B., OULIARIS, S.: Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58 (1990), 165–193.
- [57] POLOUČEK, S. a kol.: *Peniaze, banky, finančné trhy*. Iura Edition, Bratislava, 2010.
- [58] PRNO, I.: *Bankovníctvo*. IRIS, Bratislava, 2000.
- [59] ROGERSON, P. A.: *Statistical Methods for Geography*. Sage, London, 2001.
- [60] ROSE, P. S.: *Commercial bank management*. McGraw-Hill, New York, 2002.
- [61] SAID, S. E., DICKEY, D. A.: Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71 (1984), 599-607.
- [62] SELAU, L. P. R., RIBEIRO, J. L. D.: A systematic approach to construct credit risk forecast models. *Pesquisa Operacional*, 31 (2011), 41-56.
- [63] SMITH, L.D., LAWRENCE, E.C.: Forecasting losses on a liquidating long-term loan portfolio. *Journal of Banking and Finance*, 19 (1995), 959-985.
- [64] STEPANOVA, M., THOMAS, L.: Survival analysis for personal loan data. *Operations Research*, 50 (2002), 277–289.
- [65] ŠTATISTICKÝ ÚRAD SLOVENSKEJ REPUBLIKY. Dostupné na internete (18.12.2015): <http://statistics.sk>
- [66] THOMAS, L.: *Consumer Credit Models*. Oxford University Press, Oxford, 2009.
- [67] ZHANG, A.: *Statistical Methods in Credit Risk Modeling*. Dizertačná práca, University of Michigan, Michigan, 2009.
- [68] ZVÁRA, K.: *Regrese*. MatfyzPress, Praha, 2008.