

UNIVERZITA KARLOVA V PRAZE  
Matematicko–fyzikální fakulta

## DIPLOMOVÁ PRÁCE



Tomáš Senft

**Predikce poptávky po oběživu v ekonomice  
z hlediska centrální banky**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Michael Koňák, PhD.

Studijní program: Matematika, Finanční a pojistná matematika

Na tomto místě bych rád poděkoval RNDr. Michaelu Koňákovi, PhD. za vedení diplomové práce, za cenné připomínky a za čas strávený na konzultacích.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 9. dubna 2006

Tomáš Senft

# Obsah

<b>1</b>	<b>Úvod</b>	<b>6</b>
<b>2</b>	<b>Modelování objemu oběživa</b>	<b>8</b>
2.1	Funkce peněz . . . . .	8
2.2	Centrální bankovníctví . . . . .	9
2.3	Modelování objemu oběživa . . . . .	10
<b>3</b>	<b>ARIMA model</b>	<b>12</b>
3.1	Základy BJ metodologie . . . . .	12
3.2	ARIMA model poptávky po oběživu . . . . .	13
3.2.1	Stabilizace řady . . . . .	13
3.2.2	Struktura modelu . . . . .	14
3.2.3	Odhad parametrů modelu . . . . .	15
3.3	Verifikace modelu . . . . .	17
3.4	Predikce . . . . .	20
<b>4</b>	<b>GARCH model</b>	<b>23</b>
4.1	Základní principy . . . . .	23
4.2	GARCH model poptávky po oběživu . . . . .	25
4.2.1	Stabilizace řady . . . . .	25
4.2.2	Struktura modelu . . . . .	26
4.2.3	Odhad parametrů modelu . . . . .	27
4.3	Verifikace modelu . . . . .	27
4.4	Predikce . . . . .	29
<b>5</b>	<b>STS model</b>	<b>32</b>
5.1	Teorie stavových modelů časových řad . . . . .	32
5.1.1	Základní model . . . . .	32
5.1.2	Kalmanovy rekurze . . . . .	33
5.1.3	Inicializace Kalmanova filtru . . . . .	36
5.1.4	Odhad parametrů modelu . . . . .	39
5.2	STS model poptávky po oběživu . . . . .	41
5.2.1	Modelované sezónnosti . . . . .	42
5.2.2	Konstrukce kubických splinů . . . . .	44

5.2.3	Alternativní modelování měsíční sezónnosti . . . . .	47
5.3	Verifikace modelu . . . . .	48
5.4	Predikce . . . . .	49
<b>6</b>	<b>Kombinace modelů</b>	<b>54</b>
<b>7</b>	<b>Porovnání modelů</b>	<b>57</b>
<b>8</b>	<b>Závěr</b>	<b>60</b>
	<b>Literatura</b>	<b>61</b>

**Název práce:** Predikce poptávky po oběživu v ekonomice z hlediska centrální banky

**Autor:** Tomáš Senft

**Katedra:** Katedra pravděpodobnosti a matematické statistiky

**Vedoucí diplomové práce:** RNDr. Michael Koňák, PhD.

**e-mail vedoucího:** Michael.Konak@cnb.cz

**Abstrakt:** Tato diplomová práce se zabývá modelováním a predikcí objemu oběživa, které patří mezi hlavní autonomní veličiny ovlivňující likviditu trhu. V práci jsou objasněny potřeby jeho modelování a prezentovány tři zkonstruované stochastické modely. Jsou jimi ARIMA a GARCH model vycházející z Box-Jenkinsovy metodologie a STS model. STS model je strukturovaný model časové řady využívající Kalmanovy rekurze. Předpovědi jednotlivých modelů jsou dále kombinovány a statisticky porovnány. Výsledky ukazují, že nejvhodnějším modelem pro předpovídání objemu oběživa je kombinace STS a ARIMA modelu, který dosáhl stejné kvality předpovědí jako expertní odhad používaný v ČNB. Lze jej tedy použít přinejmenším jako podpůrný prostředek pro řízení likvidity v ČNB.

**Klíčová slova:** oběživo, ARIMA, GARCH, STS model, Kalmanův filtr

**Title:** Forecasting of the Demand of Currency in Circulation from the View of Central Bank

**Author:** Tomáš Senft

**Department:** Department of Probability and Mathematical Statistics

**Supervisor:** RNDr. Michael Koňák, PhD.

**Supervisor's e-mail address:** Michael.Konak@cnb.cz

**Abstract:** This diploma thesis deals with modeling and forecasting of the daily series of currency in circulation, which is one of the main autonomous factors influencing the liquidity of financial markets. Reasons for its modeling are explained and three constructed stochastic models are presented. There are ARIMA and GARCH models based on Box-Jenkins methodology and STS model. STS model is structured time series model using Kalman equations. Forecasts of models are combined together and statistically compared. The results show that the combination of STS and ARIMA models is the best model for forecasting of the daily series of currency in circulation and it has the same forecasting performance as the current model-judgement practice in the Czech National Bank. The model might be also applied at least as a supportive tool for the liquidity management.

**Key words:** currency in circulation, ARIMA, GARCH, STS model, Kalman filter

# Kapitola 1

## Úvod

Jedním z hlavních úkolů centrální banky v ekonomickém prostředí je dosažení a udržení cenové stability, tj. vytváření nízkoinflačního prostředí v ekonomice. Při plnění svého cíle využívá Česká národní banka (ČNB) několik měnověpolitických nástrojů. Efektivita využití nástrojů měnové politiky závisí na kvalitě odhadu vývoje poptávky a nabídky po penězích (tzv. likvidita trhu). Celková likvidita trhu se skládá z několika komponent, které lze rozdělit na dvě skupiny podle možností jejich kontroly. První skupinu tvoří veličiny, které jsou přesně předpověditelné, respektive předem známé, například vyplývají z již sjednaných obchodů. Druhou skupinu tvoří takzvané autonomní veličiny, které nejsou pod přímou kontrolou centrální banky. Jednou z nejhůře predikovatelných autonomních veličin je právě objem oběživa, tedy celková hodnota bankovek a mincí v držení komerčních bank a nebankovních subjektů.

Poptávka po oběživu se mění v závislosti na celé řadě faktorů. Zpravidla stoupá před výplatami a klesá koncem měsíce. Naopak v průběhu týdne poptávka vzrůstá a největší bývá před víkendem. Tento efekt bývá umocněn pokud je v pátek nebo v pondělí svátek. Faktorů ovlivňujících poptávku po oběživu je však mnohem více, například turistický ruch, podmínky pro výběr z bankomatů atd. V současné době jsou denní predikce objemů oběživa v ČNB realizovány pomocí tzv. expertních odhadů. Jedná se o expertní sledování předchozích hodnot a změn v měnové bázi a následné vyhodnocení. Cílem práce je vyřešení tohoto úkolu aplikací ekonometrických a statistických teorií. Jedná se především o podrobnou analýzu časové řady oběživa, nalezení relevantních exogenních proměnných a následně vytvoření vhodného stochastického modelu vývoje objemu oběživa především s důrazem na jeho předpověďací schopnost.

V kapitole 2 práce je stručně shrnuta problematika peněz, centrálního bankovníctví a důvody modelování objemu oběživa.

V kapitolách 3, 4 a 5 jsou prezentovány sestavené stochastické modely objemu

oběživa. U každého modelu jsou vysvětleny teoretické základy a je popsáno jeho sestavení, včetně odhadu parametrů. Následně je provedena verifikace a testování modelů. Na konci každé kapitoly je část věnovaná predikci pomocí vytvořeného modelu a jsou shrnuty obdržené výsledky.

Kapitola 6 je věnována kombinovaným modelům, které díky tomu, že využívají informace obsažené v dílčích modelech, dávají lepší výsledky předpovědí. Porovnání modelů, včetně k tomu použitých kritérií, je popsáno v kapitole 7, kde je zároveň vybrán finální model.

Pro přípravu dat byl použit program Microsoft Office Excel 2002. Model ARIMA byl zpracován v programu Matlab 6.5, především s použitím aplikačního balíčku System Identification Toolbox. Pro GARCH model byl použit balíček GARCH Toolbox ze stejného programu. Největší část práce – STS model byl naprogramován v systému Mathematica 5.0 s použitím aplikačního balíčku Time Series. Kombinování modelů a jejich porovnání je zpracováno taktéž v systému Mathematica 5.0. Všechna data, výsledky a použité skripty jsou uloženy v CD příloze.

# Kapitola 2

## Modelování objemu oběživa

Obsah této kapitoly je věnován vysvětlení problematiky peněz, centrálního bankovníctví a modelování poptávky po oběživu. Podrobné a aktuální informace lze nalézt na internetových stránkách ČNB.

### 2.1 Funkce peněz

Peníze představují všeobecně přijímaný *prostředek směny*. Za peníze můžeme tedy považovat vše, co je v určitém čase a místě využíváno ekonomickými subjekty při směně statků. Dále mají funkci *zúčtovací jednotky*, tj. lze jimi vyjádřit cenu ostatních statků, a slouží jako *uchovatel hodnoty*.

Nejstarší peníze se objevily zhruba před jedenácti tisíci lety, kdy funkci peněz zastával dobytek a zemědělské plodiny. Forma peněz se během historie postupně vyvíjela. Dnešní peníze jsou penězi nekrytými, neboli penězi s nuceným oběhem, které jsou přijímány proto, že jsou zákonem předepsaným platidlem. Mají dvě základní formy - hotovostní a převládající bezhotovostní. Při tomto dělení je nutné si uvědomit jejich velmi úzkou vzájemnou vazbu. V podstatě lze hotovostní peníze uložením do banky přeměnit na bezhotovostní a naopak. Obě formy jsou také emitovány v zásadě stejně, především úvěrem. Jediným zásadním rozdílem je, že hotovostní peníze, až na drobné výjimky v některých zemích, může emitovat pouze centrální banka. Bezhotovostní peníze oproti tomu mohou emitovat i jiné instituce, například obchodní banky.



## 2.2 Centrální bankovníctví

Pod pojmem centrální bankovníctví rozumíme dvoustupňový bankovní systém, kde prvním stupněm je centrální banka, tzv. banka bank, a druhým stupněm obchodní a další banky, které slouží ostatním subjektům a podnikají na rozdíl od centrální banky za účelem zisku. V České republice je podle Ústavy centrální bankou ČNB, v Evropské měnové unii plní tuto úlohu Evropská centrální banka (ECB). Hlavním cílem ČNB je *cenová stabilita*, která je důležitá pro správný chod ekonomiky. Podobně jako ostatní centrální banky se i ČNB zaměřuje především na stabilitu spotřebitelských cen. V praxi se stabilitou nerozumí neměnnost cen, ale jejich mírný růst. Při plnění své základní měnověpolitické úlohy, tedy cenové stability, volí centrální banka jeden z několika možných měnověpolitických režimů. ČNB používá počínaje rokem 1998 režim cílování inflace, proto jsou také v našem modelu použita data až od tohoto roku. Hlavními rysy cílování inflace jsou střednědobost této strategie, využívání prognózy inflace a veřejné explicitní vyhlášení inflačního cíle, či posloupnosti cílů. Pro dosažení cílů má ČNB k dispozici několik měnověpolitických nástrojů, kterými jsou *operace na volném trhu*, *automatické facility* a *repo sazba a povinné minimální rezervy*. Pro téma této diplomové práce je důležitý první z nich.

O operaci na volném trhu hovoříme, je-li jednou ze dvou obchodujících stran centrální banka a druhou domácí bankovní subjekt. Operace na volném trhu jsou většinou prováděny ve formě tzv. repo operací, při kterých ČNB přijímá od bank přebytečnou likviditu<sup>1</sup> a bankám předává dohodnuté cenné papíry. Obě strany se zároveň zavazují, že po uplynutí doby splatnosti proběhne zpětná transakce, v níž ČNB jako dlužník vrátí věřitelské bance zapůjčenou jistinu zvýšenou o dohodnutý úrok a věřitelská banka vrátí ČNB poskytnuté cenné papíry. Tento hlavní měnový nástroj se provádí ve formě tendrů<sup>2</sup>, kde ČNB přijímá nabídky až do výše predikovaného přebytku likvidity. Horizont predikce je jeden den, což souvisí s tím, že se vyhlášení repo tendru provádí každý pracovní den.

---

<sup>1</sup>Obecně mohou centrální banky v režimu cílování inflace peníze z oběhu nejen stahovat, ale také do oběhu dodávat. Stav, kdy pro udržení úrokových sazeb potřebuje ČNB likviditu stahovat, je dán vývojem počátku 90. let, kdy ČNB pustila do ekonomiky množství peněz za účelem udržení stability devizového kurzu

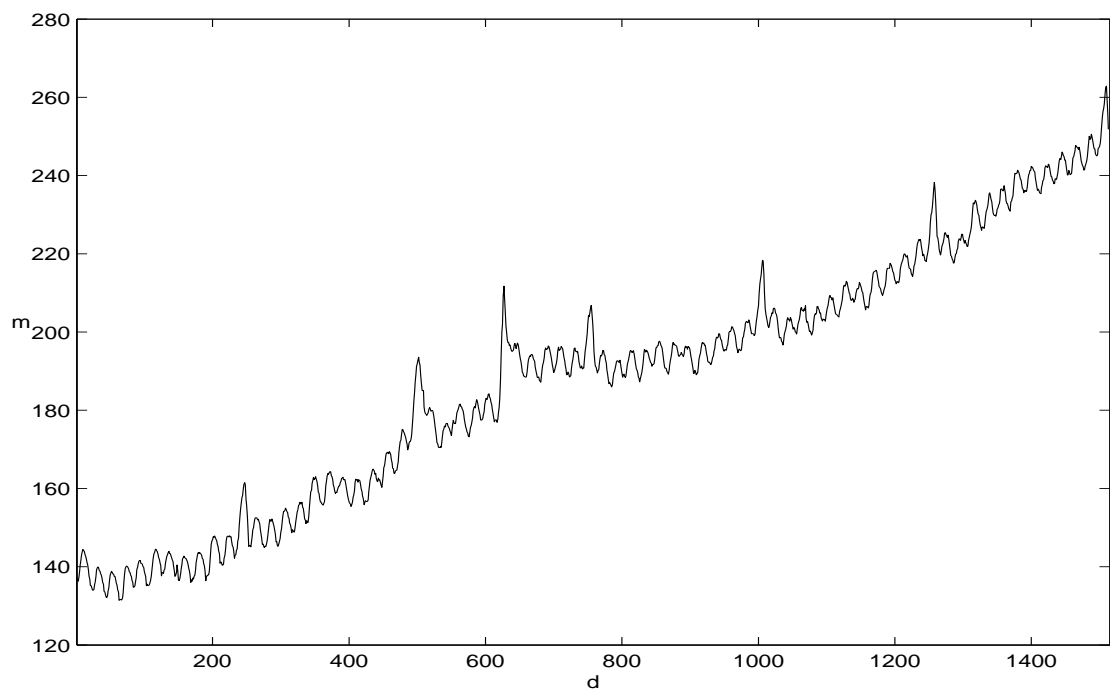
<sup>2</sup>Tendr je způsob operace, kdy banky neznačí podmínky (objem, popř. ani cenu) nabízených, resp. poptávaných cenných papírů dalších bank a nemohou tak vlastní nabídky upravovat. Jedná se o tzv. obálkovou metodu. Nabídky bank vypořádá ČNB podle americké aukční procedury, tj. přijme přednostně nabídky požadující nejnižší úrokovou sazbu a to až do výše predikovaného přebytku likvidity na daný den. V případě, že objem objednaný bankami přesáhne predikovaný přebytek likvidity, ČNB nabídky za nejvyšší sazby buď zcela odmítne, nebo proporcionálně zkrátí.

## 2.3 Modelování objemu oběživa

V souvislosti s řízením likvidity se setkáváme s pojmem měnová báze. Měnová báze je tvořena hotovostními penězi a bezhotovostními penězi. Její zdroje se dělí na autonomní faktory a měnověpolitické faktory. Mezi autonomní faktory patří například vnější sektor, vládní sektor a pohledávky za nebankovními subjekty. Mezi měnověpolitické faktory řadíme závazky z repo operací, automatické facility a krátkodobé úvěry na zachování likvidity. Jak je řízena likvidita? Komerční banky mají u ČNB tzv. clearingové účty (CU), jejichž jednu část tvoří pevně určená výše povinných minimálních rezerv. Množství peněz přesahující tuto hodnotu, které není nijak úročeno, nazýváme přebytek likvidity. Přebytek likvidity se do ekonomiky může dostat např. zvýšením devizových rezerv ČNB nebo státní emisí. V ideálním případě by byl přebytek likvidity nulový, což ale v praxi neplatí. Proto je z důvodu udržení krátkodobých úrokových sazeb třeba tuto přebytečnou likviditu stáhnout, což se řeší pomocí repo tendrů zmíněných v předchozí kapitole.

Objem oběživa je největší autonomní veličinou měnové báze a ovlivňuje množství peněz na clearingových účtech. Konkrétně tak, že o kolik se zvýší objem oběživa, o tolik se sníží množství bezhotovostních peněz na CU a naopak. Predikce objemu oběživa, jež budou poptávat komerční banky, je tedy nutná k určení výše repo tendru na stažení přebytečné likvidity. Tato poptávka vykazuje sezónní charakter (Obr. 2.1) související například s větší poptávkou po oběživu v období Vánoc. Další vlivy jsou popsány v následujících kapitolách.

V současné době jsou denní predikce objemů oběživa v ČNB realizovány pomocí expertních odhadů, tj. expertního sledování předchozích hodnot a změn v měnové bázi a následného vyhodnocení. V této práci je popsáno řešení tohoto úkolu aplikací ekonometrických a statistických teorií, přičemž denní objemy oběživa jsou chápány jako prvky časové řady. Jak již bylo naznačeno, základem jsou data z let 1998 až 2003, přičemž data do roku 2002 jsou použita pro konstrukci modelů a data za rok 2003 jsou ponechána zvláště na testování jejich předpovědácí schopnosti.



Obrázek 2.1: Poptávka po oběživu v letech 1998 - 2003 • m - mld. Kč • d - denní index

# Kapitola 3

## ARIMA model

Prvním modelem, který je použit pro modelování poptávky po oběživu, je ARIMA model. Jeho základem je kombinace stochastické části vycházející z Box-Jenkinsovy (BJ) metodologie s deterministickou regresní komponentou. Tento model dává základní představu o chování časové řady oběživa a poskytuje výchozí bod pro další modely.

V první části jsou nejdříve stručně nastíněny principy BJ metodologie a následně popsán model. V třetí části je model ověřen a jsou popsány k tomu použité statistiky. Nakonec je provedena predikce pomocí sestaveného modelu.

### 3.1 Základy BJ metodologie

Box-Jenkinsova metodologie pracuje se stacionárními náhodnými procesy. Podle *Woldovy věty* lze totiž každý nedeterministický stacionární proces  $\{y_t\}$  rozložit na čistě náhodnou složku a deterministickou část

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + D_t,$$

kde  $\psi_0 = 1$ ,  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ ,  $\{\varepsilon_t\}$  je proces nezávislých stejně rozdělených náhodných veličin s rozdělením  $N(0, \sigma^2)$  a pro všechna  $s$  a  $t$  platí  $\text{Cov}(\varepsilon_s, D_t) = 0$ ,  $\varepsilon_t = P_t \varepsilon_t$ ,  $D_t = P_s D_t$  a  $D_t$  jsou deterministické ( $D_t - P_{t-1} D_t = 0$  pro všechny  $t$ ).  $P_n$  značí operátor nejlepší lineární predikce při znalosti pozorování  $y_t$ ,  $t \leq n$ . Náhodnou složku Woldovy dekompozice označovanou jako  $\text{MA}(\infty)$  proces, lze při dostatečně

velkých  $p$  a  $q$  dobře aproximovat procesem

$$\tilde{y}_t = \sum_{i=1}^p \phi_i \tilde{y}_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t,$$

ekvivalentně zapsáno

$$\phi(B)\tilde{y}_t = \theta(B)\varepsilon_t,$$

kde  $B$  je operátor zpětného posunutí.  $\{\tilde{y}_t\}$  se nazývá ARMA( $p,q$ ) proces a je základním kamenem BJ metodologie. ARMA proces má nulovou střední hodnotu. Podmínka stacionarity je, že všechny kořeny polynomu  $\phi(B)$  leží vně jednotkového kruhu a podmínka invertibility je totožná pro polynom  $\theta(B)$ . Podrobnosti lze nalézt v [1].

## 3.2 ARIMA model poptávky po oběživu

Při konstrukci modelu je třeba nejdříve řadu stabilizovat, pak nalézt vhodnou strukturu stochastické části a relevantní vysvětlující proměnné a následně správně odhadnout jejich vliv na poptávku po oběživu. Jednotlivé kroky jsou popsány podrobněji v podkapitolách. Model vychází z modelů popsaných v [2] a [12].

### 3.2.1 Stabilizace řady

Při pohledu na graf časové řady oběživa (Obr. 2.1) je zřejmé, že se jedná o řadu silně nestacionární, která obsahuje nejen trend, ale i řadu sezónních vlivů. Nestacionaritu potvrzují rovněž korelogramy (Obr. 3.1 a 4.1). Protože chceme při modelování využít BJ metodologie, která pracuje se stacionárními řadami, je třeba řadu vhodným způsobem stabilizovat. Jako nejvhodnější transformaci jsme vyhodnotili první diferenci

$$y_t^d = \Delta y_t = y_t - y_{t-1} = (1 - B)y_t, \quad t = 2, \dots, n.$$

Řadu  $y_t^d$  (Obr. 3.2) dále již neupravujeme, stále přítomnou nestacionaritu odstraníme přidáním relevantních vysvětlujících proměnných. Po odečtení jejich vlivu je časová řada již téměř stacionární.

Další možnou transformací by bylo dodatečné použití difference se zpožděním jednoho roku, tj. 252 dní, protože časová řada oběživa obsahuje určitou meziroční závislost. Pak by ale bylo nutné použít sezónní SARIMA model, což software *Matlab* neumožňuje. Tato varianta byla zpracována v [11].

Při testování stacionarity jsme také použili *Goldfeld-Quandtův test* pro testování konstantnosti rozptylu. Výsledné p-hodnoty ( $p_{13} = 0.5834$  a  $p_{31} = 0.4166$ ) hypotézu homoskedacity nezamítají. Přesto lze pozorovat mírný nárůst variance řady v průběhu času, který by se dal odstranit logaritmováním řady, případně *Box-Coxovou transformací*. Žádná z uvedených variant ale nepřinesla zlepšení předpověďací schopnosti modelu.

### 3.2.2 Struktura modelu

Zvolený model poptávky po oběživu je kombinací ARMA procesu diferencované časové řady a regresních proměnných a je možno jej zapsat ve tvaru

$$\phi(B)\delta(B)y_t = d_t + \theta(B)\varepsilon_t,$$

kde regresní komponenta je sumou jednotlivých kalendářních vlivů  $d_t = \sum_{i=1}^k d_{t,i}$  a  $d_{t,i}$  jsou funkcí pevného vektoru nezávislých „časově závislých“ proměnných. Polynomy  $\phi(B)$  a  $\theta(B)$  určují strukturu a řád stochastické ARMA komponenty,  $\delta(B)$  je diferenční operátor a  $\{\varepsilon_t\}$  je proces nezávislých stejně rozdělených náhodných veličin s normálním rozdělením s nulovou střední hodnotou a rozptylem  $\sigma^2$ . ARMA( $p, q$ ) proces  $d$ -krát diferencované časové řady nazýváme ARIMA( $p, d, q$ ) proces. V našem případě je  $p = 11$ ,  $d = 1$ ,  $q = 1$  a

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_3 B^3 - \phi_4 B^4 - \phi_{10} B^{10} - \phi_{11} B^{11}, \\ \theta(B) &= 1 - \theta_1 B^1, \\ \delta(B) &= 1 - B.\end{aligned}$$

Číselné hodnoty koeficientů polynomů jsou uvedeny v CD příloze. Určení řádů  $p$  a  $q$  ARMA modelu jsme provedli na základě průběhů autokorelační funkce (ACF) a parciální autokorelační funkce (PACF) diferencované časové řady a následným porovnáním výsledků v úvahu přicházejících modelů.

Do modelu jsme zahrnuli pět různých kalendářních vlivů a konstantní člen pro zachycení nenulové střední hodnoty řady prvních diferencí. Číselné hodnoty regresních koeficientů  $w$  jsou opět uvedeny v CD příloze. Výběr a formulace modelovaných kalendářních vlivů byla provedena na základě zkušeností s modelováním poptávky po oběživu v ČNB a ECB a dále analýzou signifikantnosti parametrů a struktury reziduí.

**Konstatní člen** Konstatní člen udávající rychlost růstu časové řady oběživa je modelován regresní proměnnou, která je vektorem jedniček, tedy  $d_{t,1} = w_1$  pro všechna  $t$ .

**Vliv dnů v týdnu** Časová řada poptávky po obězivu vykazuje určitý týdenní cyklus, například ve čtvrtek je nejvyšší poptávka po obězivu z důvodu plnění bankomatů na víkend. Pro zachycení tohoto vlivu jsme použili čtyři v podstatě indikátorové proměnné  $h_t^{po}$ ,  $h_t^{ut}$ ,  $h_t^{st}$ ,  $h_t^{ct}$ . První z nich nabývá hodnoty 1, je-li pondělí, -1 pro pátek a 0 jinak. Další proměnné mají analogické hodnoty pro úterý, středu a čtvrtek. Takto definované proměnné zajistí čistou týdenní sezonalitu, tj. aby součet vlivů všech dnů v týdnu byl nulový.

$$d_{t,2} = w^{po}h_t^{po} + w^{ut}h_t^{ut} + w^{st}h_t^{st} + w^{ct}h_t^{ct}.$$

**Vliv pořadí dne v měsíci** Závislost hodnot řady na pořadí dne v měsíci je ovlivněna například termíny vyplácení mezd. Při modelování měsíčního vlivu jsme stejně jako v ECB (viz [2]) použili trigonometrické funkce ve tvaru

$$d_{t,3} = \sum_{j=1}^p \left( w_j^{sin} \sin \frac{2\pi j m_t}{M_t} + w_j^{cos} \cos \frac{2\pi j m_t}{M_t} \right),$$

kde  $m_k$  udává pořadí dne v měsíci a  $M_t$  je počet dní v daném měsíci. Hodnota  $p$  musí být dostatečně velká, aby dobře zachytila modelovanou sezónnost. Bylo ukázáno v [13], že  $p = 8$  je dostačující.

**Vliv státních svátků a dnů volna** Průběh poptávky po obězivu ovlivňují také státní svátky a dny volna. Jsou to Nový rok, Velikonoce, Svátek práce, Den vítězství nad fašismem, Den slovanských věrozvěstů Cyrila a Metoděje, Den upálení mistra Jana Husa, Den české státnosti, Den vzniku samostatného československého státu, Den boje za svobodu a demokracii a Vánoce. Komponenty svátků mají tvar

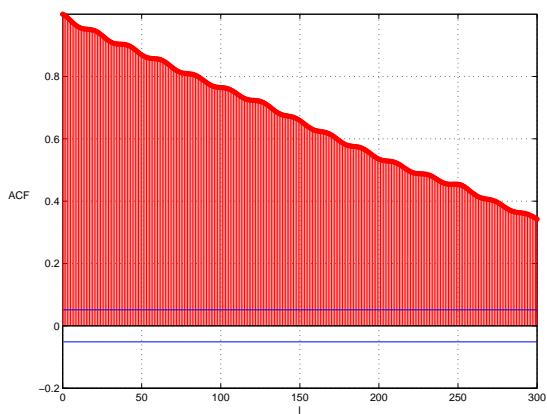
$$d_t = w(B)h_t,$$

kde  $w(B)$  je polynom udávající zpoždění, tj. o kolik časových indexů posuneme proměnnou vpřed a vzad, a  $h_t$  je indikátorová proměnná konkrétního svátku. Tvary polynomů  $w(B)$  jsou uvedeny v tabulce 3.1. Celkový vliv všech svátků a dnů volna dostaneme sečtením dílčích komponent.

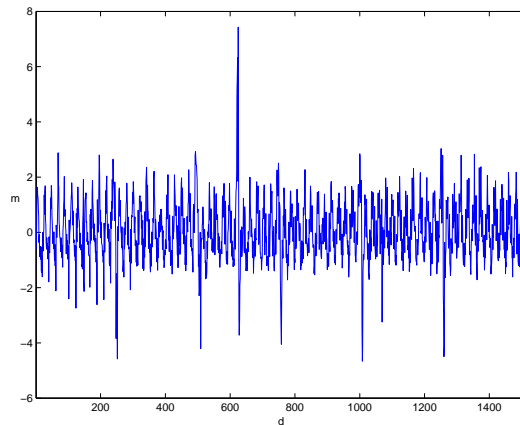
**Vliv mimořádných událostí** V historii významně ovlivnili časovou řadu poptávky po obězivu dvě mimořádné události. Jsou to krach IPB a problém Y2K (problém roku 2000, jednalo se o obavu z nefunkčnosti některých počítačových programů při přechodu z 31.12.1999 na 1.1.2000). Jejich formulace v modelu je stejná jako u státních svátků a dnů volna, tvary polynomů  $w(B)$  jsou taktéž v tabulce 3.1.

### 3.2.3 Odhad parametrů modelu

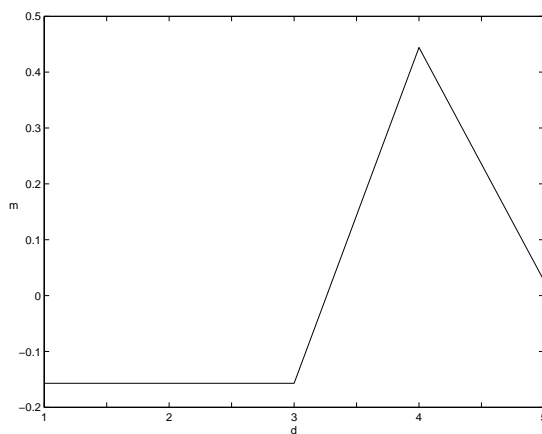
Protože parametry modelu jsou neznámé, je třeba je odhadnout. Nejvhodnější je použít metodu maximální věrohodnosti. Parametry ARMA složky a regresní



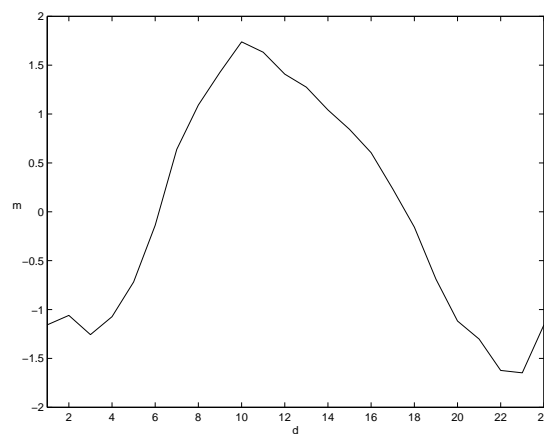
Obrázek 3.1: ACF časové řady oběživa (červeně), hranice významnosti (modře) • l - zpoždění



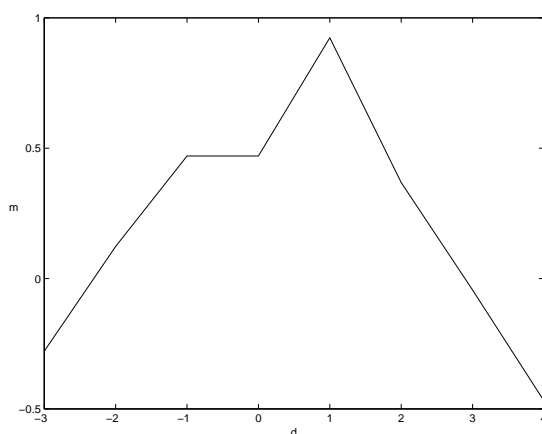
Obrázek 3.2: Řada prvních diferencí • m - mld. Kč • d - denní index



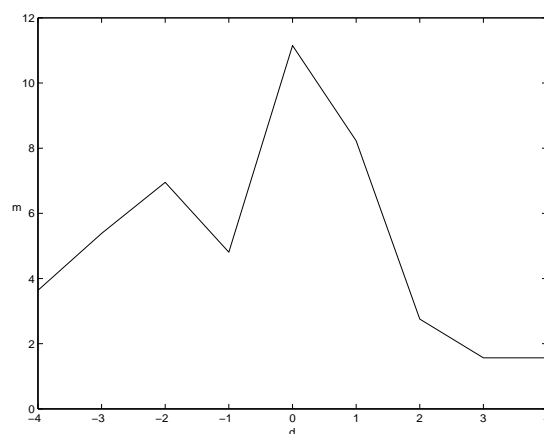
(a) Vliv dnů v týdnu • d - index dne v týdnu



(b) Vliv pořadí dne v měsíci • d - index dne v měsíci



(c) Vliv Dne vítězství nad fašismem • d - index dne od svátku



(d) Vliv krachu IPB • d - index dne od krachu IPB

Obrázek 3.3: Průběhy sezónních vlivů • m - mld. Kč



	Dnů před	Dnů po	Polynom posunutí
státní svátky a dny volna	4	4	$(w_0 + w_1B + \dots + w_8B^8)B^{-4}$
krach IPB	4	4	$(w_0 + w_1B + \dots + w_8B^8)B^{-4}$
problém Y2K	10	4	$(w_0 + \dots + w_{14}B^{14})B^{-10}$

B je operátor zpětného posunutí a  $w_i$  jsou různá pro jednotlivé svátky

Tabulka 3.1: Počty dnů modelovaných před a po svátcích a událostech

koeficienty se odhadují současně. Vzhledem k předpokladu normálního rozdělení lze věrohodnostní funkci vyjádřit jako

$$L(w, \phi, \theta, \sigma^2) = (2\pi)^{-n/2} |\Gamma_n^{-1/2}| \exp \left\{ -\frac{1}{2} (Y - \hat{Y})' \Gamma_n^{-1} (Y - \hat{Y}) \right\},$$

kde  $\Gamma_n(\phi, \theta, \sigma_2)$  je kovarianční matice procesu  $W = Y - \hat{Y}$ ,  $Y$  značí pozorování časové řady a  $\hat{Y}$  odhad řady modelem. Vzhledem k její složitosti je nutné věrohodnostní funkci maximalizovat numericky. K tomu je použita funkce `armax` softwaru *Matlab*. Po odhadu parametrů modelu je ověřováno, zda není některý odhadnutý koeficient nesignifikantní. K testování je použita statistika

$$t_i = \frac{\hat{b}_i}{s_i},$$

kde  $\hat{b}_i$  je odhad  $i$ -tého parametru a  $s_i$  jeho směrodatná odchylka,  $i = 1, 2, \dots, k$ . Nulová hypotéza je, že skutečný parametr je nulový, a při její platnosti má statistika  $t_i$  t-rozdělení o  $n - k$  stupních volnosti. Vybrán je vždy ten nejvíce nesignifikantní parametr, fixován na 0 a znovu je proveden odhad ostatních parametrů. Postupuje se stejně, až není žádný koeficient nesignifikantní. Zvolená hladina významnosti je  $\alpha = 10\%$ .

### 3.3 Verifikace modelu

V této kapitole je ověřována oprávněnost použití našeho modelu a testována kvalita modelu. Použita je řada statistik a testů, blíže jsou popsány jen ty, které nejsou standardně známé. Hladinu významnosti testů volíme  $\alpha = 5\%$ .

Předpokladem ARIMA modelu je, že  $\{\varepsilon_t\}$  je posloupnost nezávislých stejně rozdělených náhodných veličin s normálním rozdělením. Pro testování normality reziduí jsme použili *Kolmogorov-Smirnovův test*, *Lillieforsův test* a *Jarque-Berův test*. Současně jsme analyzovali histogram (Obr. 3.7) a QQ-graf (Obr. 3.6). Pro připomenutí uveďme, že první dva testy jsou založeny na empirické distribuční

funkci, třetí na šikmosti a špičatosti normálního rozložení. Všechny testy hypotézu normálního rozložení zamítají, rezidua mají především výraznou špičatost. Konstantnost rozptylu jsme ověřovali *Goldfeld-Quandtovým testem*, který nulovou hypotézu konstantnosti rozptylu zamítá. Za nejdůležitější ukazatel považujeme vzájemnou nezávislost reziduí. Analýza korelogramů (Obr. 3.4 a 3.5) dává uspokojivé výsledky, drobnou korelovanost lze pozorovat pouze pro roční frekvence, což potvrzuje i *Ljung-Boxův test*. Pro úplnost uvádíme i výsledky Ljung-Boxova testu pro kvadráty reziduí a ověření stacionarity a invertibility procesu. Všechny hodnoty statistik jsou uvedeny v tabulce 4.1.

Principem Ljung-Boxova testu je, že při platnosti hypotézy nekorelovanosti jsou hodnoty autokorelační funkce normálně rozložené. Statistika Ljung-Boxova testu má pak tvar

$$Q = n(n+2) \sum_{j=1}^L \hat{\rho}^2(j)/(n-j),$$

kde  $n$  je délka časové řady,  $L$  počet frekvencí autokorelační funkce zahrnutých do statistiky a  $\hat{\rho}^2(j)$  výběrová autokorelační funkce na frekvenci  $j$ . Při platnosti  $H_0$  má  $Q$  statistika chí-kvadrát rozložení o  $L$  stupních volnosti.

Pro srovnávání kvality alternativních modelů jsme použili věrohodnost modelu, AIC a BIC kritérium. Zároveň jsme prováděli analýzu nenormovaných reziduí (Obr. 3.9), zvláště odchylek přesahujících 1 mld. Kč. Důležitá vlastnost je předpověď schopnost, měřená střední kvadratickou chybou předpovědi. Při jedнокrokové předpovědi za rok 2003 (Obr. 3.8) jsme zaznamenali pět odchylek výrazně větších než 1 mld. Kč. Jedná se o 14.3.2003, 15.5.2003, 5.6.2003, 18.11.2003 a 12.12.2003. Nejzajímavější z nich je 14.3.2003, protože velká chyba predikce v tento den nastává i u ostatních modelů.

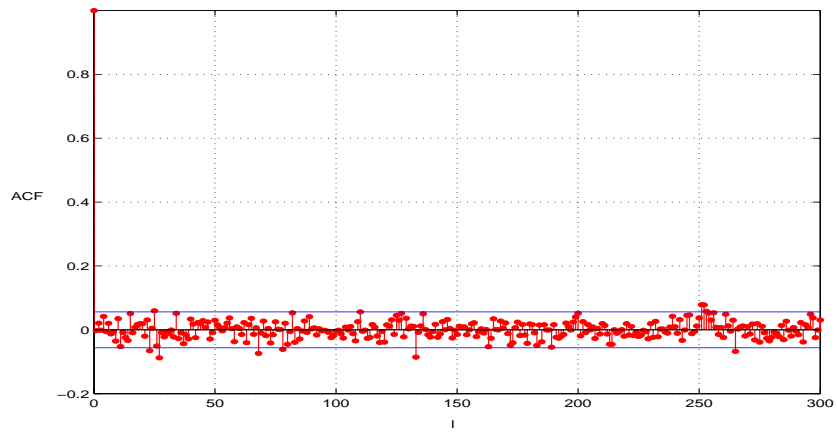
AIC (Akaike's information criterion) a BIC (Bayes information criterion) jsou kritéria určená pro správnou identifikaci modelu, hlavně pro určení počtu parametrů modelu, například řádů  $p$  a  $q$  u ARMA modelu. Vycházejí z věrohodnosti modelu, ale penalizují vzrůstající počet parametrů. BIC je historicky pozdější úpravou kritéria AIC, u kterého dochází k tendenci přeúčtení modelu oproti skutečnosti. Spočítají se jako

$$AIC = -2 \log L + 2 ppar$$

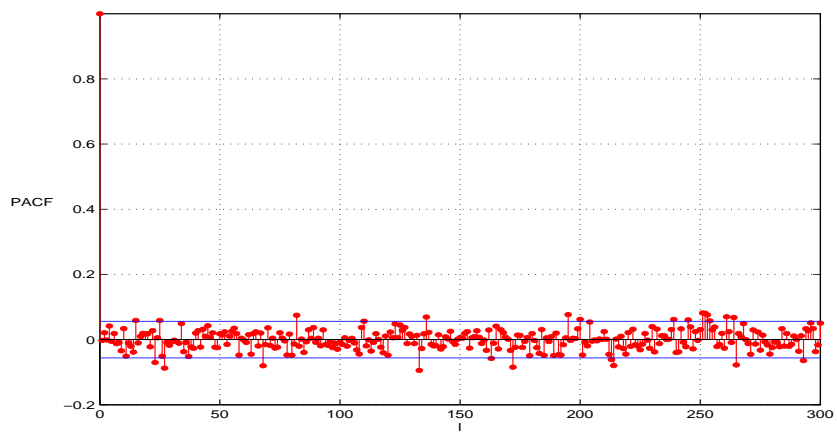
a

$$BIC = -2 \log L + \log(n) ppar,$$

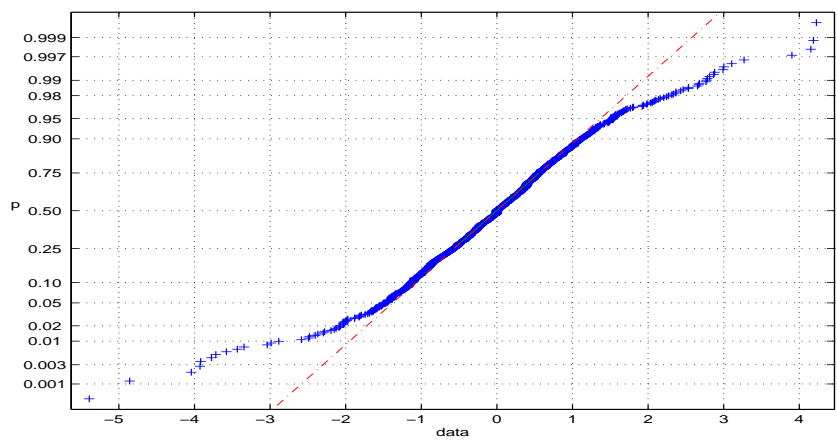
kde  $\log L$  je logaritmická věrohodnost a  $ppar$  počet parametrů. Cílem je minimální hodnota těchto kritérií.



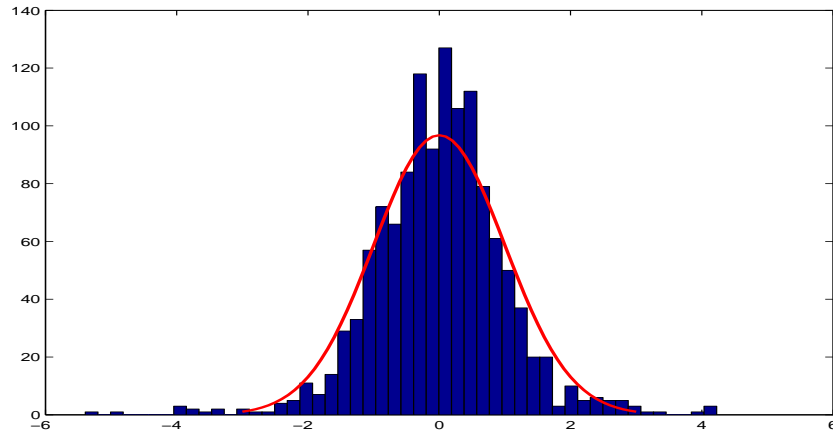
Obrázek 3.4: ACF reziduí (červeně), hranice významnosti (modře) • l - zpoždění



Obrázek 3.5: PACF reziduí (červeně), hranice významnosti (modře) • l - zpoždění



Obrázek 3.6: QQ-graf reziduí



Obrázek 3.7: Histogram reziduí

### 3.4 Predikce

Vzhledem k lineární struktuře modelu lze ARIMA modely snadno předpovídat. Mějme ARIMA proces  $\{y_t\}$ , který lze diferencováním převést na ARMA proces  $\{y_t^d\}$

$$(1 - B)^d y_t = y_t^d, \quad t = 1, \dots, n.$$

Chceme-li vyjádřit z předchozí rovnice  $y_t$  v závislosti na  $y_t^d$  a minulosti  $y_s, s < t$  dostáváme

$$y_t = y_t^d - \sum_{j=1}^d \binom{d}{j} (-1)^j y_{t-j}, \quad t = 1, \dots, n,$$

z čehož můžeme napsat nejlepší lineární předpověď  $y_{t+h}$  při znalosti pozorování časové řady do času  $t$ . Značíme ji  $P_t y_{t+h}$  a máme

$$P_t y_{t+h} = P_t y_{t+h}^d - \sum_{j=1}^d \binom{d}{j} (-1)^j P_t y_{t+h-j}, \quad t = 1, \dots, n, \quad h > 0.$$

V našem případě  $d = 1$  a

$$P_t y_{t+h} = P_t y_{t+h}^d + P_t y_{t+h-1}.$$

Následně vzhledem k tomu, že  $\{y_t^d\}$  je ARMA proces s regresní komponentou platí

$$P_t y_{t+h}^d = d_{t+h} + \sum_{i=1}^p \phi_i P_t y_{t+h-i}^d + \sum_{j=1}^q \theta_j P_t \varepsilon_{t+h-j}.$$

Nyní stačí rekurzivně pokračovat pro časové okamžiky  $t+h, t+h-1, \dots, t+1$ , přičemž využijeme  $P_t y_s^d = y_s^d, s \leq t$  a zároveň  $P_t \varepsilon_s = 0, s > t$  a  $P_t \varepsilon_s = \hat{\varepsilon}_s, s \leq t$ ,

kde

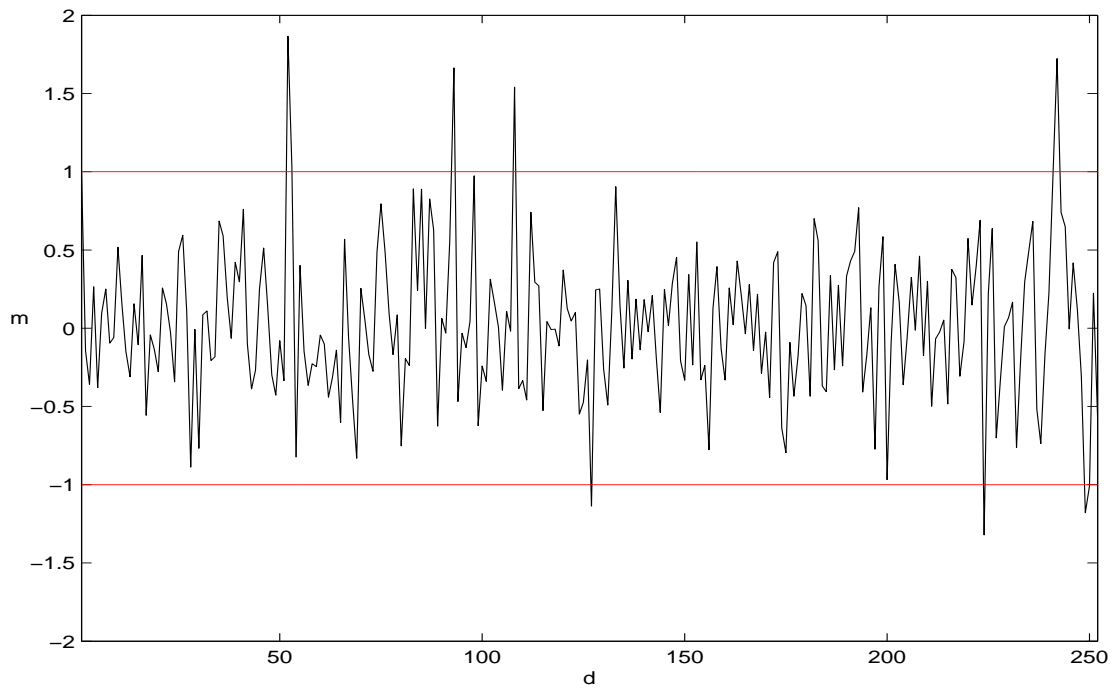
$$\hat{\varepsilon}_s = y_s^d - d_s - \sum_{i=1}^p \phi_i y_{s-i}^d - \sum_{j=1}^q \theta_j \hat{\varepsilon}_{s-j}, \quad s = 1, \dots, t$$

a  $\hat{\varepsilon}_s = 0, y_s^d = 0$  pro  $s \leq 0$ .

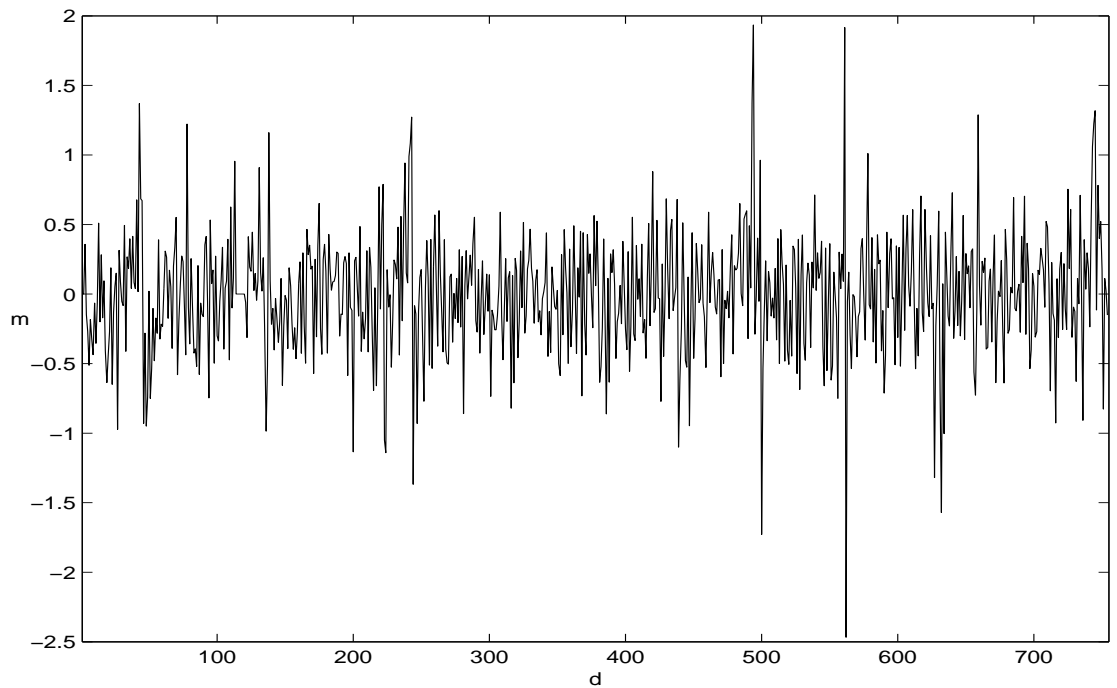
Poptávka po oběživu se v současné době v ČNB předpovídá vždy na následující den, tedy o jeden krok dopředu. V souvislosti s přechodem na jednotnou evropskou měnu euro, a tedy užšímu zapojení České republiky do Evropské měnové unie, ale bude nutno provádět předpověď vždy o jeden a dva týdny dopředu, tedy o 5, respektive 10 dní vpřed. Proto je provedena předpověď s krokem  $h = 1, 5, 10$ . Kvalita předpovědi je měřena kritériem používaným v ČNB, tj. odmocninou střední kvadratické chyby předpovědi (root mean squared error), za testovací rok 2003,

$$\text{RMSE} = \sqrt{\frac{\sum \hat{\varepsilon}_t^2}{m}},$$

kde  $\hat{\varepsilon}_t$  jsou chyby predikce a  $m = 252$  je počet pozorování časové řady (pracovních dní) v testovaném roce. Střední kvadratická chyba se s velikostí kroku zvyšuje, číselné hodnoty jsou uvedeny v tabulce 4.1.



Obrázek 3.8: Rezidua jednokrokové predikce v roce 2003 (černě), hranice  $\pm 1$  mld. Kč (červeně) ● m - mld. Kč ● d - denní index



Obrázek 3.9: Rezidua transformované řady v letech 2000-2002 ● m - mld. Kč ● d - denní index

# Kapitola 4

## GARCH model

Často používanými modely pro modelování finančních časových řad, například cen akcií, jsou modely typu GARCH. Ačkoliv časová řada poptávky po oběživu není klasickou finanční řadou, obsahuje některé její typické znaky, a proto jsme se rozhodli tento typ modelu použít.

Náš model se skládá z ARMA složky, která je z velké části totožná s prvním modelem poptávky po oběživu, a GARCH složky, jejíž principy jsou vysvětleny v první části této kapitoly. Na první pohled se může zdát, že se jedná pouze o rozšíření ARIMA modelu. Protože je však možné GARCH složku použít samostatně a časová řada se i jinak transformuje, jak je ukázáno v druhé části této kapitoly, považujeme tento model za samostatný model. V třetí části kapitoly je model verifikován a ve čtvrté části je zmíněn postup při predikci.

### 4.1 Základní principy

Jak již bylo naznačeno, finanční časové řady mají několik typických vlastností. Definujme nejdříve *výnos* (return) od času  $t - 1$  do času  $t$  jako

$$r_t = \frac{s_t}{s_{t-1}} - 1,$$

kde  $s_t$  je hodnota časové řady v čase  $t$ , například cena akcie, u které se nevyplácí dividenda. Potom typické znaky pro finanční časové řady jsou:

- Časová řada není stacionární, ale vykazuje lokální trendy.

- Hladina výnosů je přibližně stabilní. Extrémní hodnoty se vyskytují ve slucích nerovnoměrně rozložených v čase.
- Výnosy  $r_t$  se zdají být nekorelované, ale  $|r_t|$  a  $r_t^2$  vykazují významnou korelovanost.
- Rozdělení  $r_t$  je leptokurtické, tj. má šikmost  $K > 3$ .

Jestliže předpokládáme hypotézu efektivního trhu (viz [7]), dostaneme, že výnosy nelze předpovídat. Například pokud bychom chtěli aplikovat  $AR(p)$  model na modelování ceny akcií dostali bychom všechny koeficienty  $\phi_i = 0, i = 1, \dots, p$ . V praxi je však například kromě odhadu hodnoty třeba modelovat volatilitu ceny akcií. K tomu se dají dobře využít modely typu ARCH a GARCH.

Náhodný proces  $\{y_t\}$  nazýváme ARCH (autoregressive conditional heteroscedastic process) proces řádu  $p$ , jestliže

$$\begin{aligned} E(y_t | \mathcal{F}_{t-1}) &= \mu, \\ \text{Var}(y_t | \mathcal{F}_{t-1}) &= \sigma_t^2 = \omega + \sum_{k=1}^p \alpha_k y_{t-k}^2, \end{aligned}$$

$\varepsilon_t = \frac{y_t}{\sigma_t}$  je proces nezávislých stejně rozdělených náhodných veličin,  $\omega, \alpha_1, \dots, \alpha_p > 0$  a  $\mathcal{F}_{t-1}$  reprezentuje minulost,  $\mathcal{F}_{t-1} = \sigma\{y_k, k \leq t-1\}$ . Bez omezení obecnosti můžeme uvažovat  $\mu = 0$ . Jestliže druhou rovnici v předchozí definici nahradíme rovnicí

$$\text{Var}(y_t | \mathcal{F}_{t-1}) = \sigma_t^2 = \omega + \sum_{k=1}^p \alpha_k y_{t-k}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2,$$

dostaneme GARCH (generalized ARCH) proces řádu  $p, q$ . Značíme krátce  $ARCH(p)$ , respektive  $GARCH(p, q)$ . V následujících dvou tvrzeních uvedme několik základních vlastností těchto procesů, které potvrzují správnost jejich aplikace na finanční časové řady.

**Tvrzení 1.** *Nechť  $\{y_t\}$  je stacionární  $GARCH(p, q)$  proces s  $\sigma^2 = \text{Var}(y_t) < \infty$ , potom*

$$\sigma^2 = \frac{\omega}{1 - \sum_{k=1}^p \alpha_k - \sum_{k=1}^q \beta_k}$$

*a je-li navíc  $p = q = 1$ ,  $\mathcal{L}(\varepsilon_t) = N(0, 1)$ ,  $E(y_t^4) < \infty$  a  $K$  šikmost  $\{y_t\}$ , platí*

$$K = 3 + \frac{6\alpha^2}{1 - 3\alpha^2 - 2\alpha\beta - \beta^2} > 3, \quad \text{pro } \alpha > 0.$$



**Tvrzení 2.** Necht'  $\{y_t\}$  je stacionární GARCH( $p, q$ ) proces s  $E(y_t^4) < \infty$  a mějme dále  $m = \max(p, q)$ ,  $\alpha_j = 0, j > p$  a  $\beta_j = 0, j > q$ , potom je

$$w_t = y_t^2 - \sigma_t^2$$

proces nezávislých stejně  $N(0, \text{konst.})$  rozdělených náhodných veličin a

$$y_t^2 = \omega + \sum_{k=1}^m \gamma_k y_{t-k}^2 - \sum_{k=1}^q \beta_k w_{t-k} + w_t$$

ARMA( $m, q$ ) proces s  $\gamma_k = \alpha_k + \beta_k$ .

Důkazy obou tvrzení lze nalézt v [7]. Z prvního tvrzení je možno odvodit nutnou a zároveň postačující podmínku pro stacionaritu GARCH procesu, která zní

$$\sum_{k=1}^p \alpha_k + \sum_{k=1}^q \beta_k < 1 \quad \text{a} \quad \omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q > 0.$$

Protože však v praxi předpoklad nezávislosti hodnoty časové řady na minulosti často neplatí a také se objevují různé vnější vlivy, je třeba model vhodně rozšířit, například na ARMA( $r, m$ )-GARCH( $p, q$ ) proces  $\{y_t\}$  s nenulovou střední hodnotou  $\mu$  a vysvětlující exogenní proměnnou  $\{x_t\}$ , který má tvar

$$\begin{aligned} y_t &= \mu + \sum_{k=0}^l \gamma_k x_{t-k} + \sum_{k=1}^r \phi_k y_{t-k} + \sum_{k=1}^m \theta_k \eta_{t-k} + \eta_t, \\ \eta_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega + \sum_{k=1}^p \alpha_k y_{t-k}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2, \end{aligned} \tag{4.1}$$

kde  $\{\varepsilon_t\}$  je proces nezávislých stejně rozdělených náhodných veličin.

## 4.2 GARCH model poptávky po oběživu

Podobně jako u ARIMA modelu je třeba časovou řadu poptávky po oběživu  $\{y_t\}$  nejdříve stabilizovat, poté nalézt vhodnou strukturu modelu a následně odhadnout jeho parametry, čemuž obsahově odpovídají jednotlivé podkapitoly.

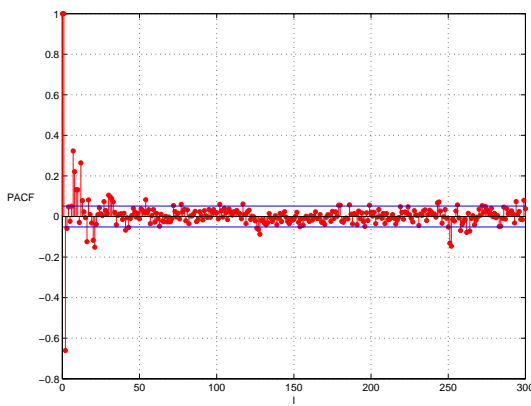
### 4.2.1 Stabilizace řady

Jak již bylo naznačeno, u modelů typu GARCH se používá pro stabilizaci nestacionární řady její první relativní diference, kterou jsme na straně 23 u finančních

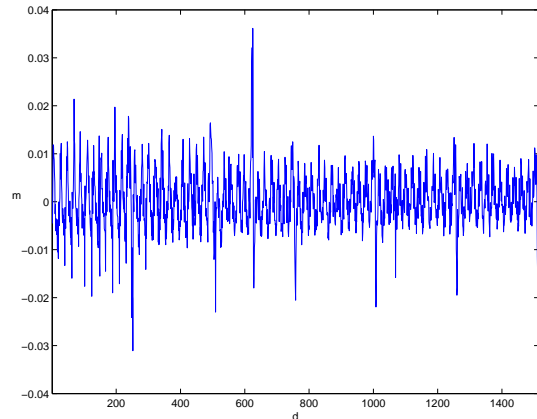
časových řad označili jako výnos. Alternativním a námi použitým způsobem je ale transformace

$$\tilde{y}_t = \log \frac{y_t}{y_{t-1}}.$$

Důvodem použití této metody u finančních časových řad oproti klasickému diferencování je složené úročení, a tedy hypotetický součinnový tvar časové řady, který se při použití logaritmu převede do součtového tvaru. Podrobnosti lze nalézt v [7].



Obrázek 4.1: PACF časové řady oběživa (červeně), hranice významnosti (modře) • 1 - zpoždění



Obrázek 4.2: Transformovaná řada  
• m - mld. Kč • d - denní index

## 4.2.2 Struktura modelu

Námi zkonstruovaný model je model ARMA(11,1)-GARCH(1,1) s několika vysvětlujícími exogenními proměnnými, jehož matematický zápis odpovídá rovnicím (4.1). Navíc jsme předpokládali, že veličiny  $\{\varepsilon_t\}$  mají normované normální rozdělení. Přesné hodnoty všech koeficientů jsou uvedeny v CD příloze, zde pouze uvedeme, že v našem modelu je

$$\mu \neq 0, \quad \omega \neq 0$$

a AR polynom má tvar

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_4 B^4 - \phi_7 B^7 - \phi_{10} B^{10} - \phi_{11} B^{11}.$$

Určení řádů ARMA složky jsme provedli na základě stejných kritérií jako u ARIMA modelu. Řád GARCH komponenty odpovídá běžně používanému řádu GARCH modelů časových řad.

Deterministickou část tvoří exogenní vysvětlující proměnné, které jsme stejně jako u ARIMA modelu využili k zachycení různých sezónních vlivů, a také jejich tvar zůstal v podstatě stejný. Jsou to Vliv dnů v týdnu, Vliv pořadí dne v měsíci, Vliv státních svátků a dnů volna a Vliv mimořádných událostí. Jedinou změnou je sloučení Nového roku, Svátku práce, Dne vítězství nad fašismem, Dne slovanských věrozvěstů Cyrila a Metoděje, Dne upálení mistra Jana Husa, Dne české státnosti, Dne vzniku samostatného československého státu a Dne boje za svobodu a demokracii a jejich implementace jako vícenásobné opakování jednoho pevného svátku. Důvodem pro sloučení byla velká výpočetní náročnost modelu.

### 4.2.3 Odhad parametrů modelu

Parametry modelu jsme odhadli metodou maximální věrohodnosti s postupným vyřazováním nesignifikantních parametrů stejným způsobem jako u ARIMA modelu. Pro úplnost uvedme věrohodnostní funkci modelu GARCH(1,1), která se pak zkombinuje s věrohodnostní funkcí ARMA modelu, čímž dostaneme věrohodnostní funkci našeho modelu poptávky po oběživu. Tuto funkci je třeba numericky maximalizovat, my jsme použili funkci `garchfit` programu *Matlab*. Při předpokladu normovaného normálního rozložení veličin  $\{\varepsilon_t\}$  má podmíněná věrohodnostní funkce procesu GARCH(1,1)  $\{y_t\}$  tvar

$$L(\omega, \alpha, \beta | \varepsilon_1) = \prod_{t=2}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right),$$

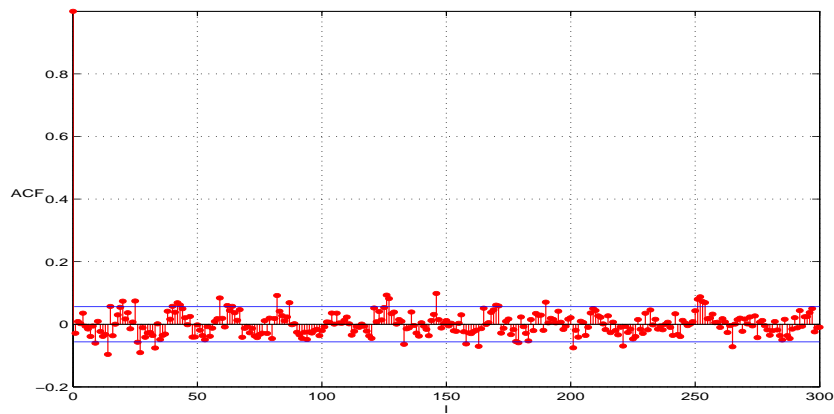
přičemž podmíněná a nepodmíněná věrohodnostní funkce se liší při dostatečně dlouhé časové řadě zanedbatelně. Více o věrohodnostních funkcích ARIMA a GARCH procesů lze nalézt například v [7].

## 4.3 Verifikace modelu

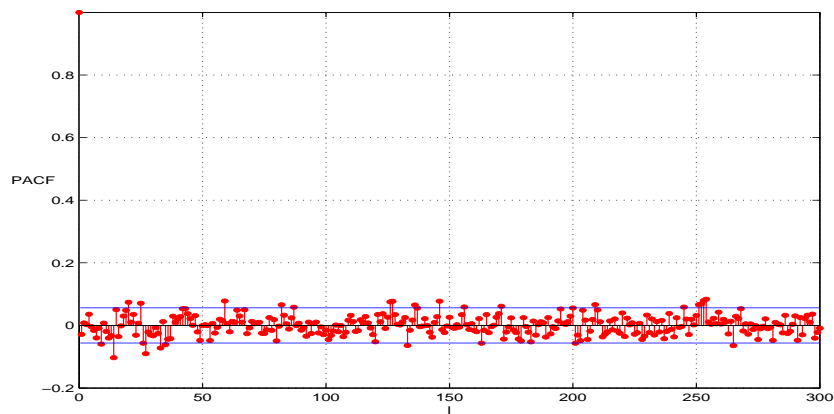
V této kapitole je uvedeno ověření platnosti přijatých předpokladů. Základním předpokladem je, že veličiny

$$\varepsilon_t = \frac{\eta_t}{\sigma_t}$$

jsou nezávislé stejně rozložené s normovaným normálním rozdělením. Nazvěme  $\{\varepsilon_t\}$  normovanými reziduii. Aplikujeme testy na normalitu *Kolmogorov-Smirnovův* test, *Lilleforsův* test a *Jarque-Berův* test a analyzujeme histogram (Obr. 4.5) a QQ - graf. I když dva testy hypotézu normálního rozložení stále zamítají, je zde zřejmá větší evidence pro oprávněnost předpokladu normality oproti ARIMA modelu. Konstantnost rozptylu *Goldfeld-Quandtův* test zamítá, což je způsobeno



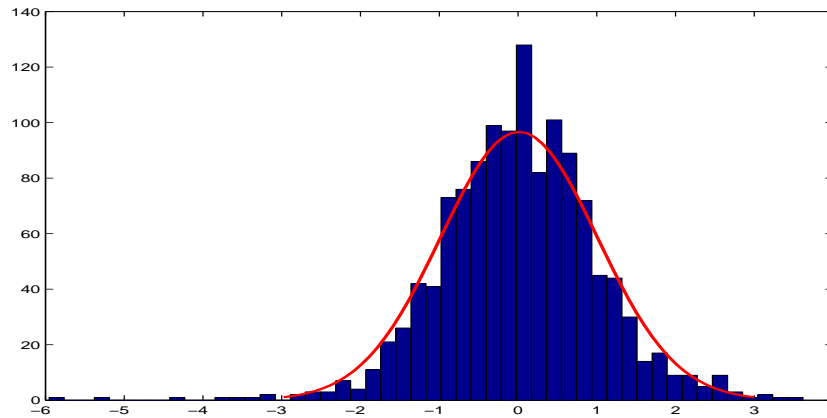
Obrázek 4.3: ACF normovaných reziduí (červeně), hranice významnosti (modře)  
 • 1 - zpoždění



Obrázek 4.4: PACF normovaných reziduí (červeně), hranice významnosti (modře)  
 • 1 - zpoždění

především větším rozptylem samotné transformované řady v roce 1998. Možným řešením tohoto problému by bylo zkrácení časové řady právě o tento rok. Předpoklad nezávislosti je přibližně splněn, což ukazují výsledky *Ljung-Boxova* testu a korelogramy (Obr. 4.3 a 4.4). *Ljung-Boxův* test pro kvadráty reziduí poukazuje na typické vlastnosti finančních časových řad. Samozřejmě jsme také ověřili stacionaritu modelu. Číselné výsledky jsou uvedeny v tabulce 4.1.

Kvalitu modelů jsme porovnávali věrohodností modelu, AIC a BIC kritériem. Zároveň jsme analyzovali graf nenormovaných reziduí (Obr. 4.7). Důležitá vlastnost je samozřejmě předpovědácí schopnost, měřená střední kvadratickou chybou předpovědi. Při jednokrokové předpovědi za rok 2003 (Obr. 4.6) jsme zaznamenali čtyři odchylky výrazně větší než 1 mld. Kč. Jedná se o 14.3.2003, 5.6.2003, 18.11.2003 a 12.12.2003.



Obrázek 4.5: Histogram reziduí

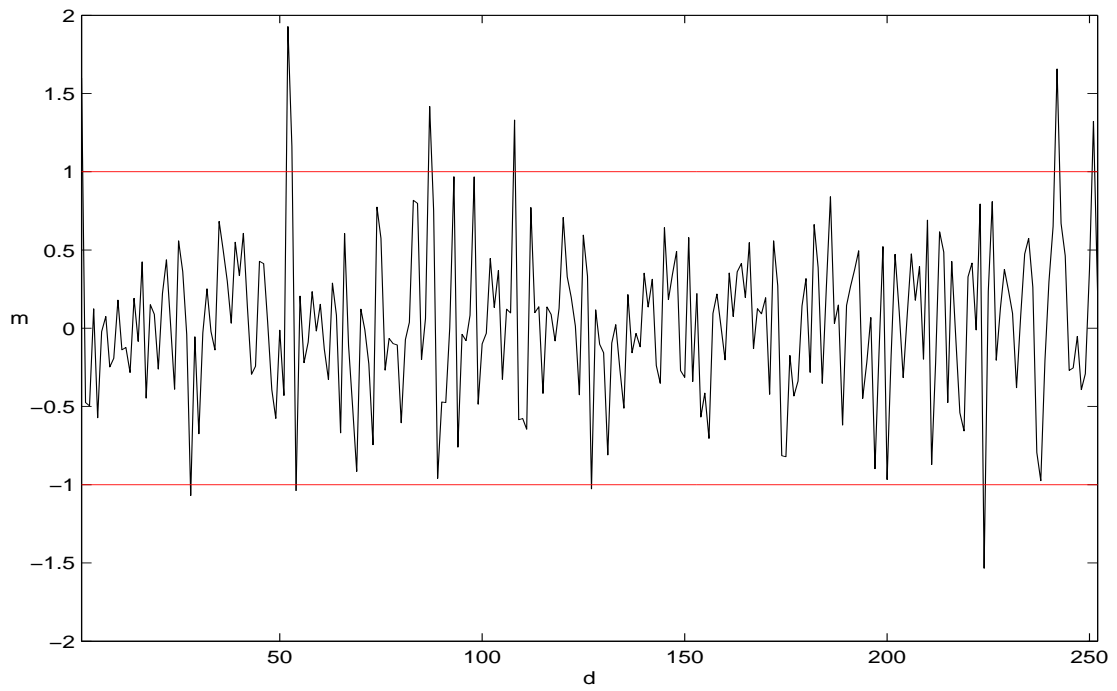
## 4.4 Predikce

Předpověď budoucí hodnoty časové řady pomocí GARCH modelu závisí pouze na jeho ARMA složce, neboť podle předpokladu je  $\{\varepsilon_t\}$  proces nezávislých stejně rozdělených náhodných veličin s normovaným normálním rozložením. Proto označujeme čisté GARCH modely za nepředpověditelné, neboť u nich platí

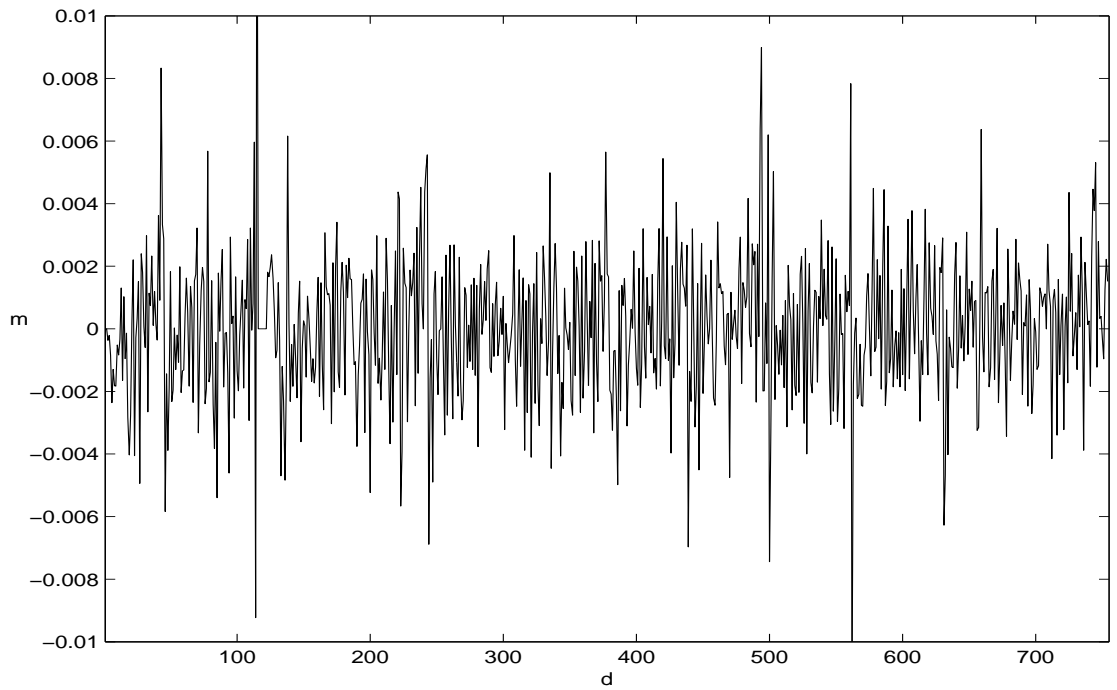
$$P_{t-1}y_t = E(y_t|\mathcal{F}_{t-1}) = Ey_t = \mu \quad (= 0).$$

Naopak možnost využití GARCH komponenty by byla při odhadu volatility. Postup předpovídání ARMA procesu byl již popsán na straně 20.

Opět jsme provedli předpověď s krokem  $h = 1, 5, 10$ . Velikost střední kvadratické chyby se s rostoucím krokem značně zvyšuje, číselné hodnoty jsou uvedeny v tabulce 4.1. Řešením by mohlo být zřejmě zvýšení řádu AR komponenty, což platí i u ARIMA modelu, změna by ale zároveň vedla ke zhoršení chyby jednokrokové předpovědi, kterou jsme zvolili za hlavní kritérium.



Obrázek 4.6: Rezidua jednokrokové predikce v roce 2003 (černě), hranice  $\pm 1$  mld. Kč (červeně) ● m - mld. Kč ● d - denní index



Obrázek 4.7: Rezidua transformované řady v letech 2000-2002 ● m - mld. Kč ● d - denní index

model	ARIMA	GARCH
Nenormovaná rezidua		
Střední hodnota	0.00038733	$9.6617 \times 10^{-5}$
Rozptyl	0.20951	$9.2609 \times 10^{-6}$
Normovaná rezidua		
Střední hodnota	$-6.185 \times 10^{-17}$	0.012291
Rozptyl	1	0.99186
Šikmost	-0.14759	-0.28308
Špičatost	5.7854	5.2802
Kolmogorov-Smirnov test	Zamítá $H_0$	Nezamítá $H_0$
p-hodnota	0.014344	0.14907
Lilliefors test	Zamítá $H_0$	Zamítá $H_0$
p-hodnota	NaN <sup>1</sup>	NaN <sup>1</sup>
Jarque-Bera test	Zamítá $H_0$	Zamítá $H_0$
p-hodnota	0	0
Goldfeld-Quandt test	Zamítá $H_0$	Zamítá $H_0$
p13-hodnota	0.025101	$3.2898 \times 10^{-10}$
p31-hodnota	0.9749	1
Ljung-Box test	rezidua	
Q(5) p-hodnota	0.72279	0.09926
Q(10) p-hodnota	0.74724	0.13377
Q(22) p-hodnota	0.63773	0.001668
Q(261) p-hodnota	0.27431	$5.0804 \times 10^8$
Ljung-Box test	kvadrát reziduí	
Q(5) p-hodnota	0	0
Q(10) p-hodnota	0	0
Q(22) p-hodnota	0	0
Q(261) p-hodnota	$2.7641 \times 10^{10}$	0
Log. věrohodnost	75.7097	5645.7958
AIC	-1.4194	-11163.5917
BIC	384.0551	-10834.6534
Chyby předpovědí	[mld. Kč]	
RMSE krok 1	0.48066	0.50111
RMSE krok 1 v ČNB	0.43787	0.43787
RMSE krok 5	1.2867	1.5057
RMSE krok 10	1.8952	2.2858
Stacionarita	Ano	Ano
Invertibilita	Ano	Ano

<sup>1</sup>Procedura `lillietest` v *Matlab* v některých případech nevrací p-hodnotu, ale pouze udává výsledek testu.

Tabulka 4.1: Tabulka výsledků modelů

# Kapitola 5

## STS model

Obsahem této kapitoly je třetí model použitý pro modelování časové řady oběživa. Jedná se o strukturovaný model časové řady (STS - structural time series model), tedy o její rozklad na trend, sezónní a nepravidelnou složku. Tento model se vyjádří jako stavový model (state space model). Pomocí Kalmanových rekurzí a metody maximální věrohodnosti se odhadnou jeho parametry a provede předpověď.

V první části této kapitoly je popsána obecná teorie stavových modelů časových řad, která vychází z [6]. V druhé je prezentován konkrétní použitý model pro modelování časové řady oběživa. Třetí a čtvrtá část je věnována jeho verifikaci a předpovědi řady pomocí modelu.

### 5.1 Teorie stavových modelů časových řad

#### 5.1.1 Základní model

Obecný lineární Gaussovský stavový model lze zapsat jako

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \quad t = 1, \dots, n, \end{aligned} \quad (5.1)$$

kde  $y_t$  je  $p \times 1$  vektor pozorování a  $\alpha_t$  je nepozorovaný  $m \times 1$  vektor nazývaný *stavový vektor*. Principem modelu je, že vývoj modelu v čase je určen vektorem  $\alpha_t$  podle druhé rovnice (5.1). Protože ale  $\alpha_t$  nelze pozorovat přímo, musíme založit analýzu na pozorovaném  $y_t$ . První rovnice (5.1) se nazývá *rovnice pozorování*, druhá se nazývá *stavová rovnice*. Matice  $Z_t$ ,  $T_t$ ,  $R_t$ ,  $H_t$  a  $Q_t$  jsou považovány za známé a chybové členy  $\varepsilon_t$  a  $\eta_t$  za seriálně nezávislé a nezávislé na sobě pro všechna  $t$ .



Matice  $Z_t$  a  $T_{t-1}$  mohou záviset na  $y_1, \dots, y_{t-1}$ . Dále předpokládáme, že počáteční stavový vektor  $\alpha_1$  je  $N(a_1, P_1)$  rozdělený, nezávislý na  $\varepsilon_1, \dots, \varepsilon_n$  a  $\eta_1, \dots, \eta_n$ . Pro začátek považujeme  $a_1$  a  $P_1$  za známé.

Jeden z nejjednodušších stavových modelů je *model lokálního lineárního trendu* (local linear trend model)

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_{t+1} &= \mu_t + \nu_t + \xi_t, & \xi_t &\sim N(0, \sigma_\xi^2), \\ \nu_{t+1} &= \nu_t + \zeta_t, & \zeta_t &\sim N(0, \sigma_\zeta^2), \end{aligned} \quad (5.2)$$

Když  $\sigma_\xi^2 = \sigma_\zeta^2 = 0$ , potom  $\nu_{t+1} = \nu_t = \nu_0$  a  $\mu_{t+1} = \mu_t + \nu_0$ . Trend je tedy přesně lineární a model se redukuje do modelu deterministického lineárního trendu s náhodným šumem. Tvar (5.2) se  $\sigma_\xi^2 > 0$  a  $\sigma_\zeta^2 > 0$  oproti tomu umožňuje změny hladiny a trendu v čase. Rovnice (5.2) lze přepsat do tvaru

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} + \varepsilon_t, \\ \begin{pmatrix} \mu_{t+1} \\ \nu_{t+1} \end{pmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} + \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix}, \end{aligned}$$

což je speciální případ (5.1).

Výhodou stavových modelů je možnost jejich spojování v komplexnější modely. Jako stavový model lze kromě lokálního lineárního trendu totiž zapsat například měsíční sezónnost modelovanou pomocí kubických splinů (kapitola 5.2.2). Tyto jednotlivé stavové modely lze za předpokladu vzájemné nezávislosti  $\alpha_1^i$ ,  $\{\varepsilon_t\}$  a  $\{\eta_t^i\}$ ,  $i = 1, 2$  spojit do jednoho stavového modelu

$$\alpha_t = \begin{pmatrix} \alpha_t^1 & \alpha_t^2 \end{pmatrix}' \quad (5.3)$$

a

$$\begin{aligned} Z_t &= \begin{pmatrix} Z_t^1 & Z_t^2 \end{pmatrix}, \\ T_t &= \text{diag} \left( T_t^1, T_t^2 \right), \\ R_t &= \text{diag} \left( R_t^1, R_t^2 \right), \\ Q_t &= \text{diag} \left( Q_t^1, Q_t^2 \right), \end{aligned} \quad (5.4)$$

kde  $\alpha_t^i$  je stavový vektor,  $\{\eta_t^i\}$  náhodný člen a  $Z_t^i$ ,  $T_t^i$ ,  $R_t^i$  a  $Q_t^i$  jsou matice stavového modelu lokálního lineárního trendu ( $i = 1$ ) a měsíční sezónnosti ( $i = 2$ ). Značení odpovídá definici (5.1). Postup lze analogicky rozšířit na libovolný počet stavových modelů, což použijeme při konstrukci STS modelu poptávky po oběživu.

## 5.1.2 Kalmanovy rekurze

Základní tři úlohy analýzy časové řady stavovými modely jsou *filtrování*, *vyhlazování* a *predikce*. Jde o problém nalezení nejlepšího (ve smyslu nejmenší střední kvadratické chyby) lineárního odhadu stavového vektoru  $\alpha_t$  při znalosti pozorování

- $y_1, \dots, y_{t-1}$  při predikci,
- $y_1, \dots, y_t$  při filtrování,
- $y_1, \dots, y_n, n > t$  při vyhlazování.

Ve stavových modelech se tyto úlohy řeší různými skupinami Kalmanových rekurzí. Ty využívají markovské vlastnosti stavových modelů a jsou výpočetně efektivní. Základem je *Kalmanův filtr*. Následuje jeho odvození ([6]), značení je stejné jako v základním tvaru (5.1).

Cílem bude odvodit podmíněné rozdělení  $\alpha_{t+1}$  za podmínky  $Y_t$  pro  $t = 1, \dots, n$ , kde  $Y_t = \{y_1, \dots, y_t\}$ , což je jednokroková predikce. Na závěr drobným přeformulováním dostaneme zároveň i výsledek úlohy filtrování. Jelikož rozdělení všech náhodných veličin je normální, podmíněné rozdělení podmnožiny veličin podmíněné jinou podmnožinou těchto veličin je také normální. Požadované rozdělení je tedy určeno znalostí  $a_{t+1} = E(\alpha_{t+1}|Y_t)$  a  $P_{t+1} = \text{Var}(\alpha_{t+1}|Y_t)$ . Předpokládejme, že  $\alpha_t$  za podmínky  $Y_{t-1}$  má rozdělení  $N(a_t, P_t)$ . Jelikož  $\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$ , dostáváme

$$\begin{aligned} a_{t+1} &= E(T_t \alpha_t + R_t \eta_t | Y_t) \\ &= T_t E(\alpha_t | Y_t), \end{aligned} \quad (5.5)$$

$$\begin{aligned} P_{t+1} &= \text{Var}(T_t \alpha_t + R_t \eta_t | Y_t) \\ &= T_t \text{Var}(\alpha_t | Y_t) T_t' + R_t Q_t R_t', \end{aligned} \quad (5.6)$$

pro  $t = 1, \dots, n$ . Označme

$$v_t = y_t - E(y_t | Y_{t-1}) = y_t - E(Z_t \alpha_t + \varepsilon_t | Y_{t-1}) = y_t - Z_t a_t.$$

Potom je  $v_t$  chyba jednokrokové předpovědi  $y_t$  za podmínky  $Y_{t-1}$ . Když  $Y_{t-1}$  a  $v_t$  jsou známé, pak také  $Y_t$  je známé. Tedy  $E(\alpha_t | Y_t) = E(\alpha_t | Y_{t-1}, v_t)$ . Následně  $E(v_t) = 0$  a  $\text{Cov}(y_j, v_t) = E[y_j E(v_t | Y_{t-1})'] = 0$ , pro  $j = 1, \dots, t-1$ . Z vícerozměrné normální regrese dostáváme

$$\begin{aligned} E(\alpha_t | Y_t) &= E(\alpha_t | Y_{t-1}, v_t) \\ &= E(\alpha_t | Y_{t-1}) + \text{Cov}(\alpha_t, v_t) [\text{Var}(v_t)]^{-1} v_t \\ &= a_t + M_t F_t^{-1} v_t, \end{aligned} \quad (5.7)$$

kde  $M_t = \text{Cov}(\alpha_t, v_t)$  a  $F_t = \text{Var}(v_t)$ . Dále

$$\begin{aligned} M_t &= \text{Cov}(\alpha_t, v_t) = E[E\{\alpha_t(Z_t \alpha_t + \varepsilon_t - Z_t a_t)' | Y_{t-1}\}] \\ &= E[E\{\alpha_t(\alpha_t - a_t)' Z_t' | Y_{t-1}\}] = P_t Z_t' \end{aligned}$$

a

$$F_t = \text{Var}(Z_t \alpha_t + \varepsilon_t - Z_t a_t) = Z_t P_t Z_t' + H_t.$$

Předpokládejme, že  $F_t$  je nesingulární. Tato podmínka je běžně splněna pro všechny dobře formulované modely, více v [6]. Substitucí v (5.5) a (5.7) vyjde

$$\begin{aligned} a_{t+1} &= T_t a_t + T_t M_t F_t^{-1} v_t \\ &= T_t a_t + K_t v_t, \quad t = 1, \dots, n, \end{aligned} \quad (5.8)$$

kde

$$K_t = T_t M_t F_t^{-1} = T_t P_t Z_t' F_t^{-1}.$$

Tedy  $a_{t+1}$  je lineární funkce  $a_t$  a  $v_t$ . Podobně máme

$$\begin{aligned} \text{Var}(\alpha_t | Y_t) &= \text{Var}(\alpha_t | Y_{t-1}, v_t) \\ &= \text{Var}(\alpha_t | Y_{t-1}) - \text{Cov}(\alpha_t, v_t) [\text{Var}(v_t)]^{-1} \text{Cov}(\alpha_t, v_t)' \\ &= P_t - M_t F_t^{-1} M_t' \\ &= P_t - P_t Z_t' F_t^{-1} Z_t P_t. \end{aligned}$$

Substitucí do (5.6) dostaneme

$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t', \quad t = 1, \dots, n, \quad (5.9)$$

kde

$$L_t = T_t - K_t Z_t.$$

Rekurze (5.8) a (5.9) jsou základem Kalmanova filtru pro model (5.1).

Pro přehlednost všechny rovnice jsou

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ K_t &= T_t P_t Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t, & t &= 1, \dots, n, \\ a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t', \end{aligned} \quad (5.10)$$

kde  $a_1$  je vektor středních hodnot a  $P_1$  je varianční matice počátečního stavového vektoru  $\alpha_1$ . Rekurze (5.10) nazýváme (standardní) Kalmanův filtr. Přeformulováním Kalmanova filtru lze nyní lehce odvodit rekurentní rovnice pro samotné filtrování. Označme odhad stavového vektoru  $a_{t|t} = E(\alpha_t | Y_t)$  a jemu příslušné varianční matice  $P_{t|t} = \text{Var}(\alpha_t | Y_t)$ . Dostaneme *souběžný Kalmanův filtr*

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ & & M_t &= P_t Z_t' \\ a_{t|t} &= a_t + M_t F_t^{-1} v_t, & P_{t|t} &= P_t - M_t F_t^{-1} M_t', & t &= 1, \dots, n, \\ a_{t+1} &= T_t a_{t|t}, & P_{t+1} &= T_t P_{t|t} T_t' + R_t Q_t R_t'. \end{aligned}$$

Další skupinou rekurentních rovnic je *Kalmanův filtr pro vyhlazování*. Jejich výsledkem je odhad  $\alpha_t$  při znalosti celé časové řady. Označme  $y = (y_1', \dots, y_n')'$ . Nechť  $\hat{\alpha}_t = E(\alpha_t | y)$  a  $V_t = \text{Var}(\alpha_t - \hat{\alpha}_t) = \text{Var}(\alpha_t | y)$  pro  $t = 1, \dots, n$ . Za podmínky známého rozdělení počátečního stavového vektoru  $\alpha_1 \sim N(a_1, P_1)$  má Kalmanův filtr pro vyhlazování tvar

$$\begin{aligned} r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \\ \hat{\alpha}_t &= a_t + P_t r_{t-1}, & V_t &= P_t - P_t N_{t-1} P_t, \end{aligned} \quad (5.11)$$

pro  $t = n, \dots, 1$ , s inicializací  $r_n = 0$  a  $N_n = 0$ . Při samotném výpočtu postupujeme následovně. Nejdříve projdeme řadu dopředně standardním Kalmanovým filtrem, následně zpětně Kalmanovým filtrem pro vyhlazování a tím získáme odhady  $\hat{\alpha}_t$  a  $V_t$ . Při dopředném průchodu musíme uložit veličiny  $v_t$ ,  $F_t$ ,  $K_t$ ,  $a_t$  a  $P_t$  pro  $t = 1, \dots, n$ . Alternativně lze uložit pouze  $a_t$  a  $P_t$  a ostatní veličiny dopočítat při zpětném průchodu, což je použitý způsob v našem modelu časové řady oběživa.

Třetí úloha, a to predikce pomocí Kalmanova filtru je popsána v kapitole 5.4.

Značnou výhodou tohoto přístupu je možnost analýzy časové řady v případě, že není kompletní a její pozorování v některých časových okamžicích chybí. Předpokládejme, že máme obecný stavový model, kde pozorování  $y_j$ ,  $j = \tau, \dots, \tau^* - 1$ , chybí, pro  $1 < \tau < \tau^* \leq n$ . Pro filtrování v časech  $t = \tau, \dots, \tau^* - 1$  stačí použít v Kalmanově filtru vektor  $v_t = 0$  a matici  $K_t = 0$  pro tyto časové okamžiky. Dostaneme

$$a_{t+1} = T_t a_t, \quad P_{t+1} = T_t P_t T_t' + R_t Q_t R_t', \quad t = \tau, \dots, \tau^* - 1.$$

Podobně pro zpětné vyhlazovací rekurze máme

$$r_{t-1} = T_t' r_t, \quad N_{t-1} = T_t' N_t T_t, \quad t = \tau^* - 1, \dots, \tau.$$

Ostatní rekurzivní rovnice zůstanou nezměněny. Ověření těchto výsledků lze nalézt v [6]. V našem modelu byl vzhledem k již hotové implementaci Kalmanova filtru použit alternativní přístup, který je s popsáním ekvivalentní. Byl použit přeformulovaný model

$$y_t^* = \begin{cases} 0 & t \in \{\tau, \dots, \tau^* - 1\}, \\ y_t & \text{jinak,} \end{cases}$$

$$Z_t^* = \begin{cases} 0 & t \in \{\tau, \dots, \tau^* - 1\}, \\ Z_t & \text{jinak,} \end{cases} \quad \varepsilon_t^* = \begin{cases} \delta_t & t \in \{\tau, \dots, \tau^* - 1\}, \\ \varepsilon_t & \text{jinak,} \end{cases}$$

kde  $\{\delta_t\}$  je proces nezávislých stejně rozdělených náhodných veličin s  $\delta_t \sim N(0, I)$ , nezávislých na  $\alpha_1, \varepsilon_s, \eta_s$  pro  $t, s = 1, \dots, n$ . Ostatní části modelu zůstávají nezměněny. V případě více chybějících částí pozorování časové řady se tyto přístupy analogicky rozšíří.

### 5.1.3 Inicializace Kalmanova filtru

V předcházejících dvou kapitolách jsme považovali rozdělení počátečního stavu za známé,  $\alpha_1 \sim N(a_1, P_1)$ ,  $a_1$  a  $P_1$  známé. To však ve většině praktických aplikací, včetně modelu poptávky po oběživu, neplatí. Proces řešící tento problém se nazývá *inicializace*. V obecném případě má počáteční stav modelu tvar

$$\alpha_1 = a + A \delta + R_0 \eta_0, \quad \eta_0 \sim N(0, Q_0), \quad (5.12)$$

kde  $m \times 1$  vektor  $a$  je známý,  $\delta$  je  $q \times 1$  vektor neznámých veličin,  $m \times q$  matice  $A$  a  $m \times (m - q)$  matice  $R_0$  jsou selekční matice. Selekční matice jsou matice, jejichž sloupce jsou sloupce jednotkové matice  $I_m$  a dané dohromady dají množinu  $g$  sloupců matice  $I_m$  tak, že  $g \leq m$  a  $A' R_0 = 0$ . Matice  $Q_0$  je pozitivně semidefinitní a známá. Reprezentace konstantní části je  $a$ ,  $A \delta$  představuje nestacionární část a  $R_0 \eta_0$  stacionární část. V modelu poptávky po oběživu je celý počáteční stav neznámý a veličiny stavového vektoru nestacionární, tedy  $A = I_m$  a  $\alpha_1 = \delta$ . Vektor  $\delta$  může být považován za pevný vektor neznámých parametrů, tedy

$$\delta \sim N(\text{konst.}, 0), \quad (5.13)$$

nebo jako vektor náhodných veličin s normálním rozdělením s nulovými středními hodnotami a nekonečnými rozptyly. První způsob odvodil Rosenberg a je více diskutován v [6]. Druhým způsobem se mimo jiné zabýval De Jong v [4] a nazývá se *difusní inicializace*. Vychází z předpokladu

$$\delta \sim N(0, \kappa I_q), \quad (5.14)$$

kde  $\kappa \rightarrow \infty$ . Při modelování poptávky po oběživu byly použity obě varianty. Ve finálním modelu byla zvolena difusní inicializace, ačkoliv výsledky obou variant byly v podstatě rovnocenné.

Jednou možností řešení je *rozšířený Kalmanův filtr* (augmented Kalman filter), jehož základem je rozšíření pozorovaného vektoru časové řady následujícím způsobem. Pro dané  $\delta$  aplikujeme Kalmanův filtr s  $a_1 = E(\alpha_1) = a + A \delta$ ,  $P_1 = \text{Var}(\alpha_1) = P_* = R_0 Q_0 R_0'$  a označme výslednou hodnotu  $a_t$  jako  $a_{\delta,t}$ . Jelikož  $a_{\delta,t}$  je lineární funkce pozorování a  $a_1 = a + A \delta$  můžeme psát

$$a_{\delta,t} = a_{a,t} + A_{A,t} \delta,$$

kde  $a_{a,t}$  je hodnota  $a_t$  výstupu Kalmanova filtru, u kterého je  $a_1 = a$ ,  $P_1 = P_*$  a kde  $j$ -tý sloupec  $A_{A,t}$  je hodnota  $a_t$  výstupu filtru, u kterého je vektor pozorování nulový a  $a_1 = A_j$ ,  $P_1 = P_*$ , kde  $A_j$  je  $j$ -tý sloupec matice  $A$ . Označme hodnotu  $v_t$  výstupu Kalmanova filtru, u kterého je  $a_1 = a + A \delta$ ,  $P_1 = P_*$  jako  $v_{\delta,t}$ . Podobně jako pro  $a_{\delta,t}$  můžeme psát

$$v_{\delta,t} = v_{a,t} + V_{A,t} \delta,$$

kde  $v_{a,t}$  a  $V_{A,t}$  jsou dané výsledkem stejných filtrů jako  $a_{a,t}$  a  $A_{A,t}$ . Matice  $(a_{a,t}, A_{A,t})$  a  $(v_{a,t}, V_{A,t})$  mohou být spočteny jediným průchodem Kalmanova filtru, do něhož vstupuje vektor pozorování  $y_t$  rozšířený nulovými hodnotami. Toto je možné, protože v každém filtru produkujícím jednotlivý sloupec  $(a_{a,t}, A_{A,t})$  je použita stejná inicializační varianční matice  $P_1 = P_*$ , a tedy výstup týkající se rozptylu, který označíme jako  $F_{\delta,t}$ ,  $K_{\delta,t}$  a  $P_{\delta,t+1}$  je stejný. Nahrazením rovnic Kalmanových filtrů pro  $v_t$  a  $a_t$  odpovídajícími rovnicemi pro matice  $(a_{a,t}, A_{A,t})$  a  $(v_{a,t}, V_{A,t})$  dostaneme

$$\begin{aligned} (v_{a,t}, V_{A,t}) &= (y_t, 0) - Z_t (a_{a,t}, A_{A,t}), \\ (a_{a,t+1}, A_{A,t+1}) &= T_t (a_{a,t}, A_{A,t}) + K_{\delta,t} (v_{a,t}, V_{A,t}), \end{aligned} \quad (5.15)$$

kde  $(a_{a,1}, A_{A,1}) = (a, A)$ . Rekurze pro  $F_t$ ,  $K_t$  a  $P_{t+1}$  zůstanou stejné jako v klasickém Kalmanově filtru, tedy

$$\begin{aligned} F_{\delta,t} &= Z_t P_{\delta,t} Z_t' + H_t, \\ K_{\delta,t} &= T_t P_{\delta,t} Z_t' F_{\delta,t}^{-1}, & L_{\delta,t} &= T_t - K_{\delta,t} Z_t, \\ P_{\delta,t+1} &= T_t P_{\delta,t} L_{\delta,t}' + R_t Q_t R_t', \end{aligned} \quad (5.16)$$

pro  $t = 1, \dots, n$  a  $P_{\delta,1} = P_*$ . Index  $\delta$  je přidán proto, že daná veličina je vypočítána za předpokladu daného  $\delta$  a nikoliv, že na  $\delta$  přímo matematicky závisí. Jako rozšířený Kalmanův filtr budeme označovat rovnice (5.15) a (5.16).

Nechť nyní platí (5.14). Pro dané  $\kappa$  máme

$$a_{t+1} = E(\alpha_{t+1} | Y_t) = a_{a,t+1} + A_{A,t+1} \bar{\delta}_t,$$

kde  $\bar{\delta}_t = E(\delta | Y_t)$ . Dále

$$\begin{aligned} \log p(\delta | Y_t) &= \log p(\delta) + \log p(Y_t | \delta) - \log p(Y_t) \\ &= \log p(\delta) + \sum_{j=1}^t \log p(v_{\delta,j}) - \log p(Y_t) \\ &= -\frac{1}{2\kappa} \delta' \delta - b_t' \delta - \frac{1}{2} \delta' S_{A,t} \delta + \text{členy nezávislé na } \delta, \end{aligned}$$

kde

$$b_t = \sum_{j=1}^t V_{A,j}' F_{\delta,j}^{-1} v_{a,j}, \quad S_{A,t} = \sum_{j=1}^t V_{A,j}' F_{\delta,j}^{-1} V_{A,j}. \quad (5.17)$$

Diferencováním vzhledem k  $\delta$  a položením rovno 0 dostaneme maximálně věrohodný odhad a při normálním rozložení veličin máme

$$\bar{\delta}_t = - \left( S_{A,t} + \frac{1}{\kappa} I_q \right)^{-1} b_t. \quad (5.18)$$

Tedy

$$\begin{aligned} P_{t+1} &= E[(a_{t+1} - \alpha_{t+1})(a_{t+1} - \alpha_{t+1})'] \\ &= E[\{a_{a,t+1} - \alpha_{t+1} + A_{A,t+1}(\delta - \bar{\delta}_t)\} \{a_{a,t+1} - \alpha_{t+1} + A_{A,t+1}(\delta - \bar{\delta}_t)\}'] \\ &= P_{\delta,t+1} + A_{A,t+1} \text{Var}(\delta | Y_t) A_{A,t+1}' \\ &= P_{\delta,t+1} + A_{A,t+1} \left( S_{A,t} + \frac{1}{\kappa} I_q \right)^{-1} A_{A,t+1}', \end{aligned}$$

jelikož  $\text{Var}(\delta | Y_t) = (S_{A,t} + \frac{1}{\kappa} I_q)^{-1}$ . Pro  $\kappa \rightarrow \infty$  dostáváme

$$\begin{aligned} \bar{\delta}_t &= -S_{A,t}^{-1} b_t, \\ \text{Var}(\delta | Y_t) &= S_{A,t}^{-1}, \end{aligned}$$

v případě, že  $S_{A,t}$  je nesingulární. Výpočet  $b_t$  a  $S_{A,t}$  lze snadno přidat k rozšířenému Kalmanovu filtru. Potom

$$\begin{aligned} a_{t+1} &= a_{a,t+1} - A_{A,t+1} S_{A,t}^{-1} b_t, \\ P_{t+1} &= P_{\delta,t+1} + A_{A,t+1} S_{A,t}^{-1} A_{A,t+1}', \end{aligned}$$

při  $\kappa \rightarrow \infty$ . Pro nedegenerativní modely existuje  $d$  takové, že pro  $t < d$  je  $S_{A,t}$  singulární, takže hodnoty  $a_{t+1}$  a  $P_{t+1}$  dané předchozími rovnicemi neexistují. Naopak pro  $t \geq d$  je  $S_{A,t}$  nesingulární, hodnoty  $a_{d+1}$  a  $P_{d+1}$  existují a následně hodnoty  $a_{t+1}$  a  $P_{t+1}$  pro  $t > d$  mohou být vypočítány standardním Kalmanovým filtrem.

Obvykle je  $d$  menší než délka sezóny modelované časovou řadou, a tedy mnohem menší než délka pozorované časové řady. V modelu poptávky po oběživu je však potřeba  $d = 828$ , což jsou tři sezóny (roky) a polovina pozorované časové řady. Důvodem je zřejmě souběh tří modelovaných vlivů ve stejném období, a to na konci kalendářního roku. Jedná se o roční sezónnost, Nový rok a problém Y2K. Takto dlouhá inicializace znemožňuje použití jinak výpočetně efektivnějšího *přesného inicializačního Kalmanova filtru* (exact initial Kalman filter), protože u něj dochází k velkým numerickým chybám. Eventuální možnost postupu by byla použití jeho verze s rozkladem variančních matic na mocninný součin  $P_t = \tilde{P}_t \tilde{P}_t'$  (square root filter), a zaručit tak jejich pozitivní semidefinitnost, podrobnosti lze nalézt v [6].

#### 5.1.4 Odhad parametrů modelu

Model poptávky po oběživu, stejně jako všechny praktické modely, závisí na neznámých parametrech. V této kapitole ukážeme, jak lze tyto parametry odhadnout metodou maximální věrohodnosti. Pro zřejmost uveďme, že například u lokálního lineárního modelu (5.2) jsou neznámé parametry velikosti rozptylů  $\sigma_\varepsilon^2$ ,  $\sigma_\xi^2$  a  $\sigma_\zeta^2$ .

Nejdříve odvodíme věrohodnostní funkci v případě, že rozdělení počátečního stavového vektoru je známé,  $\alpha_1 \sim E(a_1, P_1)$ ,  $a_1$  a  $P_1$  jsou známé. Pak věrohodnost modelu je

$$L(y) = p(y_1, \dots, y_n) = p(y_1) \prod_{t=2}^n p(y_t | Y_{t-1}),$$

kde  $Y_t = \{y_1, \dots, y_t\}$ . Pro praktický výpočet pak použijeme logaritmicou věrohodnostní funkci

$$\log L(y) = \sum_{t=1}^n \log p(y_t | Y_{t-1}),$$

kde  $p(y_1 | Y_0) = p(y_1)$ . Pro model (5.1) máme  $E(y_t | Y_{t-1}) = Z_t a_t$ ,  $v_t = y_t - Z_t a_t$  a  $F_t = \text{Var}(y_t | Y_{t-1})$ . Substituováním  $N(Z_t a_t, F_t)$  za  $p(y_t | Y_{t-1})$  dostaneme

$$\log L(y) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + v_t' F_t^{-1} v_t). \quad (5.19)$$

Hodnoty  $v_t$ ,  $F_t$ , a tedy celá  $\log L(y)$ , se vypočítají pomocí Kalmanova filtru. Předpokládáme jen, že  $F_t$  je nesingulární pro  $t = 1, \dots, n$ . Jestliže tato podmínka

není splněna automaticky, je obvykle možné předefinovat model tak, aby splněna byla. V našem modelu poptávky po oběživu se ale tento problém nevyskytl.

V difusním případě řešeném rozšířeným Kalmanovým filtrem je sdružená hustota  $\delta$  a  $y$  pro dané  $\kappa$

$$\begin{aligned} p(\delta, y) &= p(\delta)p(y|\delta) \\ &= p(\delta) \sum_{t=1}^n p(v_{\delta,t}) \\ &= (2\pi)^{(np+q)/2} \kappa^{-q/2} \prod_{t=1}^n |F_{\delta,t}|^{-1/2} \times \\ &\quad \times \exp \left[ -\frac{1}{2} \left( \frac{\delta' \delta}{\kappa} + S_{a,n} + 2b'_n \delta + \delta' S_{A,n} \delta \right) \right], \end{aligned}$$

kde  $v_{\delta,t}$ ,  $b_n$ ,  $S_{A,n}$  jsou definovány v předcházející kapitole a

$$S_{a,n} = \sum_{t=1}^n v'_{a,t} F_{\delta,t}^{-1} v_{a,t}.$$

Z (5.18) známe  $\bar{\delta}_n = -(S_{A,n} + \kappa^{-1}I_q)^{-1} b_n$  a exponent předcházejícího výrazu může být přepsán jako

$$-\frac{1}{2} [S_{a,n} + (\delta + \bar{\delta}_n)' (S_{A,n} + \kappa^{-1}I_q) (\delta - \bar{\delta}_n) - \bar{\delta}'_n (S_{A,n} + \kappa^{-1}I_q) \bar{\delta}_n].$$

Odintegrovaním  $\delta$  dostaneme marginální hustotu  $y$ . Logaritmická věrohodnostní funkce má tvar

$$\begin{aligned} \log L(y) &= -\frac{np}{2} \log 2\pi - \frac{q}{2} \log \kappa - \frac{1}{2} \log |S_{A,n} + \kappa^{-1}I_q| - \\ &\quad - \frac{1}{2} \sum_{t=1}^n \log |F_{\delta,t}| - \frac{1}{2} [S_{a,n} - \bar{\delta}'_n (S_{A,n} + \kappa^{-1}I_q) \bar{\delta}_n]. \end{aligned}$$

Přičtením  $\frac{q}{2} \log \kappa$  a pro  $\kappa \rightarrow \infty$  dostaneme difusní logaritmickou věrohodnostní funkci

$$\begin{aligned} \log L_d(y) &= -\frac{np}{2} \log 2\pi - \frac{1}{2} \log |S_{A,n}| - \frac{1}{2} \sum_{t=1}^n \log |F_{\delta,t}| \\ &= -\frac{1}{2} (S_{a,n} - b'_n S_{A,n}^{-1} b_n). \end{aligned} \tag{5.20}$$

V předchozí kapitole jsme ukázali, že rozšířený Kalmanův filtr může být přerušen v časovém okamžiku  $t = d$ . Můžeme tedy použít parciální věrohodnost založenou na  $Y_d$  a následně přidat příspěvky inovací  $v_{d+1}, \dots, v_n$  ze standardního Kalmanova filtru. Odhad parametrů pak získáme numerickou maximalizací této věrohodnostní funkce. V našem modelu poptávky po oběživu jsme použili funkci `NMinimize` programu *Mathematica*. Numerická maximalizace je značně výpočetně a časově náročná, možným zrychlením by bylo například použití *EM algoritmu*, nebo převod problému do *spektrální domény*.



## 5.2 STS model poptávky po oběživu

Pro modelování poptávky po oběživu jsme použili STS model

$$y_t = \mu_t + \gamma_t + \varepsilon_t,$$

kde  $\mu_t$  je model lokálního lineárního trendu (5.2),  $\gamma_t$  je stochastická sezónní komponenta a nepravidelná složka  $\varepsilon_t$  je proces nezávislých náhodných veličin s normálním rozdělením s nulovou střední hodnotou a rozptylem  $\sigma_\varepsilon^2$ . Sezónní komponenta je sumou pěti subkomponent  $\gamma_t = \sum_{i=1}^5 \gamma_t^i$ , které modelují rozdílné sezónní vlivy. Jedná se o *sezónnost dnů v týdnu, měsíční sezónnost, roční sezónnost, státní svátky a dny volna a mimořádné události*. Každá subkomponenta má tvar

$$\gamma_t^i = z_t^i \delta_t^i,$$

kde  $z_t^i$  je pevný  $1 \times g_i$  vektor nezávislých „časově závislých“ proměnných. Hodnoty  $z_t^i$  jsou definovány vždy přes celé období které modelují, například 5 dnů pro sezónnost dne v týdnu, a jejich hodnoty se periodicky opakují pro celou časovou řadu. Aby se zajistilo, že  $\gamma_t^i$  je čistě sezónní komponenta, musí být součet  $z_t^i$  přes celé sezónní období nulový vektor, tedy u sezónnosti dne v týdnu  $\sum_{k=t}^{t+4} z_k^i = 0$ .  $\delta_t^i$  je časově proměnlivý parametr, který sleduje náhodnou procházku

$$\delta_t^i = \delta_{t-1}^i + \chi_t^i,$$

kde  $\chi_t^i$  je  $g_i \times 1$  vektor sériálně nezávislých normálně rozdělených náhodných procesů s nulovou střední hodnotou a kovarianční maticí  $E(\chi_t^i \chi_t^{i'}) = \sigma_i^2 I$ . Ne všechny sezónní komponenty v modelu jsou stochastické, některé jsou uvažovány jako deterministické, a potom je  $\sigma_i^2 = 0$ . Jednotlivé komponenty se spojí podle (5.3) a (5.4) a matice stavového zápisu modelu pak jsou

$$Z_t = (1 \quad 0 \quad z_t^1 \quad z_t^2 \quad z_t^3 \quad z_t^4 \quad z_t^5) \quad H_t = \sigma_\varepsilon^2$$

$$T_t = \begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \quad R_t = I$$

$$Q_t = \text{diag}(\sigma_\xi^2, \sigma_\zeta^2, \sigma_1^2, \dots, \sigma_1^2, \sigma_2^2, \dots, \sigma_2^2, \sigma_3^2, \dots, \sigma_3^2, 0, \dots, 0).$$

Aby bylo možné modelovat všechny měsíce a roky, je nutné uvážit měsíc s 23 (pracovními) dny a tedy rok s 276 dny. Jestliže má některý měsíc méně než 23 dnů, považujeme tyto dny za chybějící pozorování. Pro ilustraci uveďme 31. duben, nebo 29. únor v nepřestupných letech, pokud se samozřejmě jedná o pracovní dny.

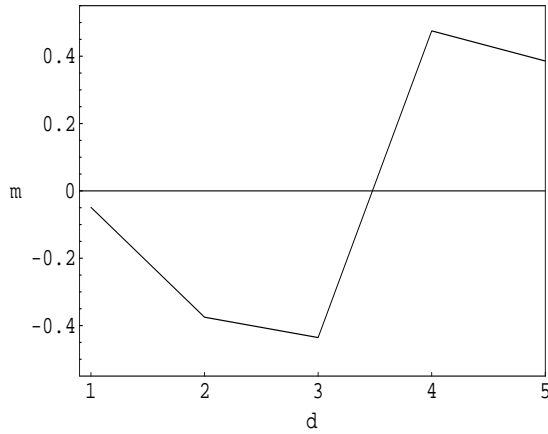
## 5.2.1 Modelované sezónnosti

**Sezónnost dnů v týdnu** K zachycení této závislosti jsme použili model  $\gamma_t^1 = z_t^1 \delta_t^1$ , kde  $z_t^1$  je vektor  $1 \times 4$  a jeho první prvek nabývá 1, je-li den  $t$  pondělí,  $-1$  pátek a 0 pro ostatní dny. Prvky 2, 3 a 4 nabývají stejných hodnot pro úterý, středu a čtvrtek. Z definice je zřejmé, že  $\sum_{k=t}^{t+4} z_k^1$  je nulový vektor. Sezónnost dnů v týdnu je uvažována jako stochastická, a tedy  $\delta_t^1$  je vektor sledující náhodnou procházku s rozptylem  $\sigma_1^2$ .

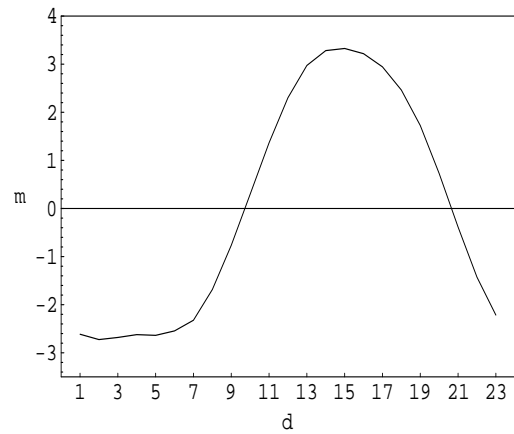
**Měsíční sezónnost** Pro modelování jsme zvolili kubické spliny proměnlivé v čase, jejichž přesná konstrukce je popsána v další kapitole. Nejdůležitějším faktorem je zde určení počtu uzlů a jejich správné umístění v rámci modelovaného období. Na základě korelogramů, kvality proložení a předpověďací schopnosti modelu jsme vybrali počet šesti uzlů a rozmístění (0), 3, 9, 11, 15, 19, 23. Nula je zde uvedena pouze z důvodu objasnění navázání dvou období, jedná se vlastně o uzel na pozici 23 předcházející sezóny. Vzhledem k periodicitě kubických splinů musí mít spliny v deterministickém případě v prvním (0) a posledním (23) uzlu stejnou hodnotu. Z tohoto důvodu je důležité, aby časová řada na začátku modelovaného období měla přibližně stejné hodnoty a tedy žádný prudký trend. Proto první den v modelovaném měsíci neodpovídá prvnímu dni v kalendářním měsíci, ale je posunut až na 5. den kalendářního měsíce. Umístění chybějících dnů do celkového počtu 23 dnů za měsíc je založeno na stejném principu. V modelu poptávky po oběživu jsme umístili tyto chybějící pozorování mezi 4. a 5. den v měsíci. Měsíční komponenta je rovněž stochastická s parametrem  $\sigma_2^2$ .

**Roční sezónnost** Jako příklad závislosti poptávky po oběživu na ročním období uveďme letní prázdniny, kde dochází k zvýšení poptávky z důvodu turistického ruchu. Zde jsme stejně jako u měsíční závislosti využili kubické spliny. Rozmístění i počet uzlů jsme použili stejné jako v modelu ECB, tj. 16 uzlů s umístěním (0), 1, 23, 99, 121, 140, 166, 202, 207, 213, 220, 226, 231, 234, 239, 246 a 276. S počátkem modelovaného roku v posledním únorovém dnu. Důležitá je zde velká koncentrace uzlů okolo Vánoc a konce roku, aby se zachytil prudký nárůst poptávky po oběživu v tomto období. Oproti tomu je jen několik uzlů ve zbývajících částech roku, kde kolísání řady je menší. Zvětšením počtu uzlů se dá sice dosáhnout lepší kvality proložení, ale zároveň se zhorší předpověďací schopnost modelu. Parametr rozptylu stochastické části označíme jako  $\sigma_3^2$ .

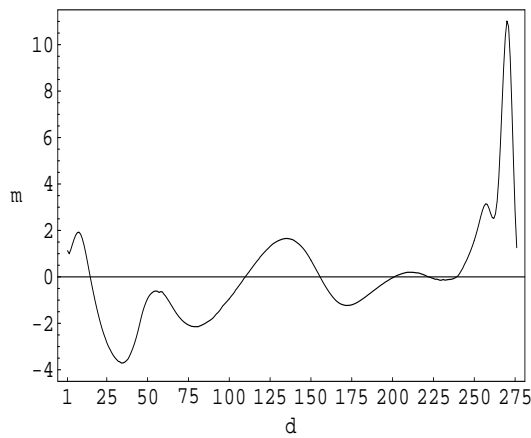
**Státní svátky a dny volna** Vliv státních svátků a dnů volna je modelován deterministickými vysvětlujícími proměnnými. Tato komponenta má tvar  $\gamma_t = z_t^4 \delta_t^4$ , kde  $z_t^4$  je v podstatě indikátor daného svátku a  $\delta_t^4$  je konstantní vektor určující samotnou velikost vlivu svátku na poptávku po oběživu. Rozměr vektoru  $z_t^4$  závisí na počtu modelovaných dnů volna a modelovaných



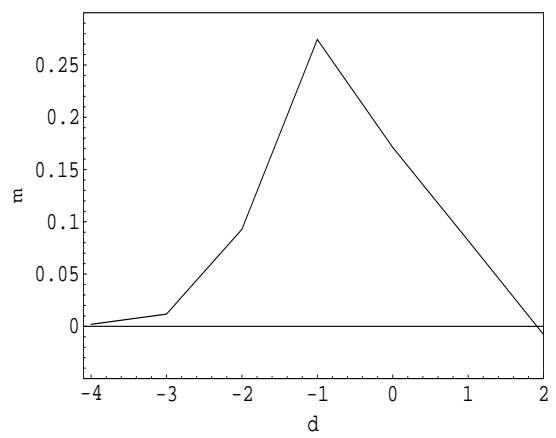
(a) Sezónnost dnů v týdnu • d - index dne v týdnu



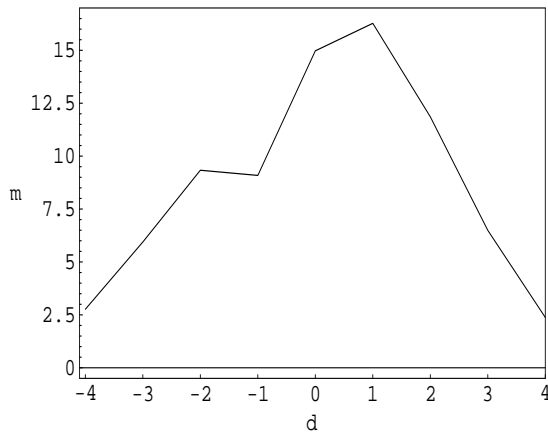
(b) Měsíční sezónnost • d - index dne v měsíci



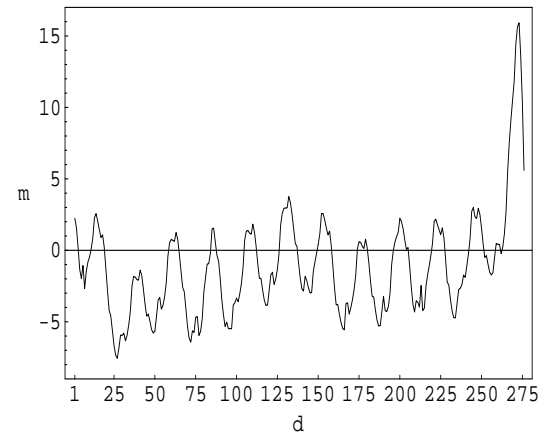
(c) Roční sezónnost • d - index dne v roce



(d) Vliv pevného svátku • d - index dne od svátku



(e) Vliv krachu IPB • d - index dne od krachu IPB



(f) Celkový vliv sezónností • d - index dne v roce

Obrázek 5.1: Průběhy sezónních komponent • m - mld. Kč

dnů před, respektive po jednotlivém svátku, v našem případě  $1 \times 34$ . Hodnota na všech pozicích vektoru  $z_t^4$  nabývá buď 1, jedná-li se o modelovaný den, nebo

$$\frac{\text{počet výskytů modelovaného dne v roce}}{\text{počet zbývajících dnů v roce}},$$

aby bylo zajištěno  $\sum_{k=t}^{t+275} z_k^4 = 0$ . Interpretace je, že zvýšená poptávka při svátku se rovnoměrně rozloží úbytkem poptávky v průběhu zbytku roku. Vzhledem k výpočetní náročnosti modelu jsou Svátek práce, Den osvobození, Den slovanských věrozvěstů Cyrila a Metoděje a Den upálení mistra Jana Husa, Den české státnosti, Den vzniku samostatného československého státu a Den boje za svobodu a demokracii uvažovány jako vícenásobné opakování jednoho pevného svátku.

Svátek	Dnů před	Dnů po	Polynom posunutí
Nový rok	4	4	$(w_0 + w_1B + \dots + w_8B^8)B^{-4}$
Velikonoce	4	4	$(w_0 + w_1B + \dots + w_8B^8)B^{-4}$
Vánoce	4	4	$(w_0 + w_1B + \dots + w_8B^8)B^{-4}$
pevný svátek	4	2	$(w_0 + w_1B + \dots + w_6B^6)B^{-4}$

$B$  je operátor zpětného posunutí a  $w_i$  jsou různá pro jednotlivé svátky

Tabulka 5.1: Počty dnů modelovaných před a po jednotlivých svátcích

**Mimořádné události** Stejně jako v ARIMA a GARCH modelu jsme zahrnuli dvě mimořádné události, krach IPB a problém Y2K, které měly vliv na výši poptávky po oběživu. Zde ale vzhledem k rekurzivní povaze modelu mají tyto události roku 2000 pouze minimální vliv na současnost. Jejich forma v modelu je analogická státním svátkům a dnům volna, pouze s rozdílem, že záporný výraz složek vektoru  $z_t^4$  určený k vynulování vlivu během roku se zde nahradí 0, protože nemá žádný reálný smysl. Označíme  $z_t^5$  a  $\delta_t^5$ .

Událost	Dnů před	Dnů po	Polynom posunutí
IPB	4	4	$(w_0 + w_1B + \dots + w_8B^8)B^{-4}$
Y2K	10	4	$(w_0 + w_1B + \dots + w_{14}B^{14})B^{-10}$

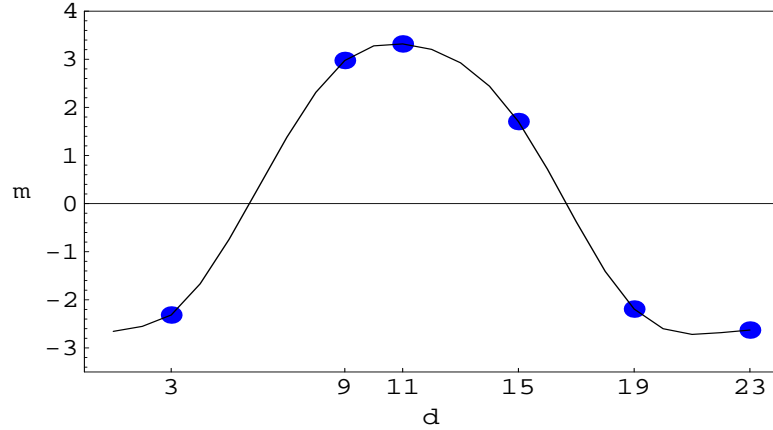
$B$  je operátor zpětného posunutí a  $w_i$  jsou různá pro jednotlivé svátky

Tabulka 5.2: Počty dnů modelovaných před a po mimořádných událostech

## 5.2.2 Konstrukce kubických splinů

Cílem kubických splinů je sestavit funkci  $g(x)$ , která aproximuje neznámou funkci  $f(x)$  tak, že je složením polynomiálních funkcí diferencovatelných do řádu 3 a

má první dvě derivace spojité. Funkce  $f(x)$  je zadána souřadnicemi uzlů  $(x_i^\dagger, y_i^\dagger)$ ,  $i = 0, 1, \dots, k$ . Množinu  $(x_0^\dagger, \dots, x_k^\dagger)$  nazveme *rozmístěním uzlů*.



Obrázek 5.2: Použité rozmístění uzlů měsíční komponenty • m - mld. Kč  
• d - denní index

Nejdříve předpokládejme, že souřadnice uzlů jsou známé, a tedy pevně dané a rozmístění je v rostoucím pořadí

$$x_0^\dagger < x_1^\dagger < \dots < x_k^\dagger.$$

Označme  $h_j = x_j^\dagger - x_{j-1}^\dagger$ ,  $d_j(x)$  první a  $d_j^2(x)$  druhou derivaci funkce splinů  $g(x)$  na intervalu  $[x_{j-1}^\dagger, x_j^\dagger]$  a  $a_j = d_j^2(x_j^\dagger)$ ,  $j = 1, \dots, k$ . Kubické spliny jsou směsí polynomů třetího řádu, a tedy jejich druhá derivace na intervalu  $[x_{j-1}^\dagger, x_j^\dagger]$  je lineární funkce

$$d_j^2(x) = \frac{x_j^\dagger - x}{h_j} a_{j-1} + \frac{x - x_{j-1}^\dagger}{h_j} a_j, \quad j = 1, \dots, k.$$

Výraz pro první derivaci a samotnou funkci splinů dostaneme integrací s podmínkami  $g(x_j^\dagger) = y_j^\dagger$  a  $g(x_{j-1}^\dagger) = y_{j-1}^\dagger$ . Platí

$$d_j(x) = - \left[ \frac{1}{2} \frac{(x_j^\dagger - x)^2}{h_j} - \frac{h_j}{6} \right] a_{j-1} + \left[ \frac{1}{2} \frac{(x - x_{j-1}^\dagger)^2}{h_j} - \frac{h_j}{6} \right] a_j$$

a

$$g_j(x) = (x_j^\dagger - x) \frac{[(x_j^\dagger - x)^2 - h_j^2]}{6h_j} a_{j-1} + \frac{x_j^\dagger - x}{h_j} y_{j-1}^\dagger + (x - x_{j-1}^\dagger) \frac{[(x - x_{j-1}^\dagger)^2 - h_j^2]}{6h_j} a_j + \frac{x - x_{j-1}^\dagger}{h_j} y_j^\dagger, \quad (5.21)$$

kde  $x_{j-1}^\dagger \leq x \leq x_j^\dagger$ ,  $j = 1, \dots, k$ . Poslední rovnici lze zapsat ve vektorovém zápisu

$$g_j(x) = r_j(x)' y^\dagger + s_j(x)' a,$$

kde  $a = (a_0, \dots, a_k)'$ ,  $y^\dagger = (y_0^\dagger, \dots, y_k^\dagger)$  a vektory  $r_j(x)$  a  $s_j(x)$  mají na všech pozicích 0, kromě  $j$ -té a  $(j+1)$ -vé pozice, které odpovídají příslušným vahám v rovnici (5.21). Využijeme podmínku spojitosti první derivace  $d_j(x_j^\dagger) = d_{j+1}(x_j^\dagger)$  a dostaneme soustavu  $k-1$  rovnic

$$\frac{h_j}{h_j + h_{j+1}} a_{j-1} + 2a_j + \frac{h_{j+1}}{h_j + h_{j+1}} a_{j+1} = \frac{6y_{j-1}^\dagger}{h_j(h_j + h_{j+1})} - \frac{6y_j^\dagger}{h_j h_{j+1}} + \frac{6y_{j+1}^\dagger}{h_{j+1}(h_j + h_{j+1})}, \quad j = 1, 2, \dots, k-1.$$

K soustavě  $k-1$  rovnic o  $k+1$  neznámých je třeba přidat další dvě podmínky, aby byla řešitelná. Jelikož chceme periodickou funkci splinů, musí být

$$y_0^\dagger = y_k^\dagger, \quad d_1(x_0^\dagger) = d_k(x_k^\dagger), \quad d_1^2(x_0^\dagger) = d_k^2(x_k^\dagger).$$

Z druhé podmínky plyne

$$\frac{h_k}{h_1 + h_k} a_{k-1} + 2a_k + \frac{h_1}{h_1 + h_k} a_1 = \frac{6y_{k-1}^\dagger}{h_k(h_1 + h_k)} - \frac{6y_k^\dagger}{h_1 h_k} + \frac{6y_1^\dagger}{h_1(h_1 + h_k)}.$$

Poslední podmínku přepíšeme jako  $a_0 = a_k$ , a tím máme soustavu  $k$  rovnic o  $k$  neznámých, jejíž řešení v maticovém tvaru je

$$a = P^{-1} Q y^\dagger,$$

kde

$$P = \begin{pmatrix} 2 & \frac{h_2}{h_1+h_2} & 0 & \dots & 0 & \frac{h_1}{h_1+h_2} \\ \frac{h_2}{h_2+h_3} & 2 & \frac{h_3}{h_2+h_3} & \dots & 0 & 0 \\ 0 & \frac{h_3}{h_3+h_4} & 2 & \dots & 0 & 0 \\ 0 & 0 & \frac{h_4}{h_4+h_5} & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & \frac{h_k}{h_{k-1}+h_k} \\ \frac{h_1}{h_1+h_k} & 0 & 0 & \dots & \frac{h_k}{h_1+h_k} & 2 \end{pmatrix}$$

$$Q = \begin{pmatrix} -\frac{6}{h_1 h_2} & \frac{6}{h_2(h_1+h_2)} & 0 & \dots & 0 & \frac{6}{h_1(h_1+h_2)} \\ \frac{6}{h_2(h_2+h_3)} & -\frac{6}{h_2 h_3} & \frac{6}{h_3(h_2+h_3)} & \dots & 0 & 0 \\ 0 & \frac{6}{h_3(h_3+h_4)} & -\frac{6}{h_3 h_4} & \dots & 0 & 0 \\ 0 & 0 & \frac{6}{h_4(h_4+h_5)} & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\frac{6}{h_{k-1} h_k} & \frac{6}{h_k(h_{k-1}+h_k)} \\ \frac{6}{h_1(h_1+h_k)} & 0 & 0 & \dots & \frac{6}{h_k(h_1+h_k)} & -\frac{6}{h_1 h_k} \end{pmatrix}$$

Nyní můžeme vyjádřit funkci splinů jako lineární funkci  $y^\dagger$

$$g_j(x) = w_j(x)' \tilde{y}^\dagger, \quad (5.22)$$

kde

$$w_j(x)' = \tilde{r}_j(x)' + \tilde{s}_j(x)'P^{-1}Q, \quad x_{j-1}^\dagger \leq x \leq x_j^\dagger, \quad j = 1, \dots, k$$

a  $k \times 1$  vektory  $\tilde{y}^\dagger$ ,  $\tilde{r}_j$  a  $\tilde{s}_j$  jsou upravené vektory  $y^\dagger$ ,  $r_j$  a  $s_j$  vzhledem k  $y_0^\dagger = y_k^\dagger$ . Parametr  $x$  je bod, ve kterém chceme funkci  $g(x)$  vypočítat, a  $j$  značí, mezi kterými dvěma uzly tento bod leží. Rovnici(5.22) přepíšeme do tvaru sezónní komponenty stavového modelu

$$\gamma_t = w_t' \gamma^\dagger,$$

kde  $t$  je časový okamžik, ve kterém se model nachází. Parametr  $t$  určuje, ve které části období (např. který den v roce) se časová řada nachází, a  $j$  dále nepíšeme, protože je zřejmé, mezi kterými uzly  $t$  leží. Dále potřebujeme zajistit, že celkový vliv sezónnosti za období bude nula. Proto vypustíme poslední prvek vektoru  $\gamma^\dagger$  a výsledný vektor označíme  $\delta$ . Následně  $i$ -tý prvek  $1 \times (k-1)$  vektoru  $z_t$  vypočteme jako

$$z_{ti} = w_{ti} - w_{tk}w_{*i}/w_{*k}, \quad i = 1, \dots, k-1,$$

pro všechny dny  $t$  v modelovaném období, a kde  $w_{*i}$  je  $i$ -tý prvek vektoru

$$w_* = \sum_{\text{období}} w_l, \quad \text{např. pro rok} \quad w_* = \sum_{l=1}^{276} w_l.$$

Nakonec přestaneme uvažovat souřadnice uzlů za pevné, ale umožníme jejich proměnlivost v čase. Předpokládáme, že sledují náhodnou procházku

$$\gamma_t^\dagger = \gamma_{t-1}^\dagger + \chi_t,$$

kde  $\chi_t$  je vektor seriálně nekorelovaných náhodných veličin s nulovou střední hodnotou a kovarianční maticí

$$\text{Var}(\chi_t) = \sigma_\chi^2(I - (1/w_*'w_*)w_*w_*').$$

Parametr  $\sigma_\chi^2$  udává rychlost s jakou se spline může měnit a celá kovarianční matice zaručuje, že  $w_*'\chi_t = 0$ . Stejně jako u splinu neměnného v čase označíme vektor  $\gamma_t^\dagger$  zkrácený o poslední prvek  $\delta_t$  a dostaneme definitivní tvar periodického kubického splinu proměnného v čase

$$\gamma_t = z_t \delta_t.$$

### 5.2.3 Alternativní modelování měsíční sezónnosti

Jako alternativní způsob zachycení vlivu dne v měsíci jsme místo proložení kubických splinů použili trigonometrické funkce. Sezónní komponenta při délce sezóny  $s$  má pak tvar

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{jt},$$

kde

$$\begin{aligned}\gamma_{j,t+1} &= \gamma_{jt} \cos \lambda_j + \gamma_{jt}^* \sin \lambda_j + \omega_{jt}, \\ \gamma_{j,t+1}^* &= -\gamma_{jt} \sin \lambda_j + \gamma_{jt}^* \cos \lambda_j + \omega_{jt}^*, \quad j = 1 \dots, \lfloor s/2 \rfloor,\end{aligned}$$

$\lambda_j = \frac{2\pi j}{s}$  a  $\omega_{jt}, \omega_{jt}^*$  jsou nezávislé náhodné veličiny s rozdělením  $N(0, \sigma_\omega^2)$ . V modelu poptávky po oběživu je  $s = 23$  a bylo by tedy třeba  $k = 11$  subkomponent  $\gamma_{jt}$ , ale lze ukázat (viz [13]), že  $k = 8$  plně postačuje. Předchozí rovnice zapíšeme do stavového modelu se stavovým vektorem

$$\alpha_t = (\gamma_{1t}, \gamma_{1t}^*, \gamma_{2t}, \dots, \gamma_{8t})$$

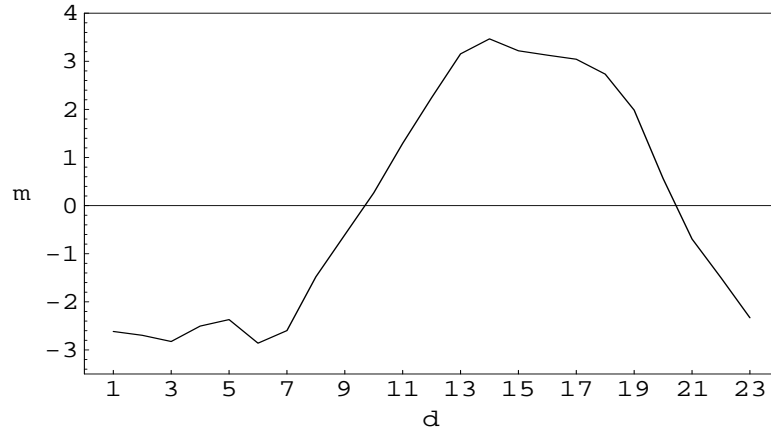
a maticemi

$$\begin{aligned}Z_t &= (1, 0, 1, 0, \dots, 1, 0), & T_t &= \text{diag}(C_1, C_2, \dots, C_8), \\ R_t &= I_{16}, & Q_t &= \sigma_\omega^2 I_{16},\end{aligned}$$

kde

$$C_j = \begin{pmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{pmatrix}, \quad j = 1, \dots, 8.$$

Připojení komponenty do modelu poptávky po oběživu se provede známým způsobem.



Obrázek 5.3: Měsíční sezónnost - trigonometrické funkce • m - mld. Kč  
• d - denní index

## 5.3 Verifikace modelu

Při testování správnosti a kvality modelu jsme použili řadu nástrojů, jejichž číselné hodnoty jsou uvedeny v souhrnné tabulce na konci kapitoly věnované STS modelu.



Protože se ve velké části jedná o již popsaná kritéria u ARIMA a GARCH modelů, je uveden pouze jejich výčet. Hladinu významnosti u testů mějme  $\alpha = 5\%$ .

Předpokladem stavového modelu je, že chybové členy  $\varepsilon_t$  a  $\eta_t$  jsou normálně rozdělené a seriálně nezávislé s konstantními rozptyly. Vzhledem k těmto předpokladům musí být standardizované chyby jedнокrokové předpovědi

$$e_t = \frac{v_t}{\sqrt{F_t}}, \quad t = 1, \dots, n,$$

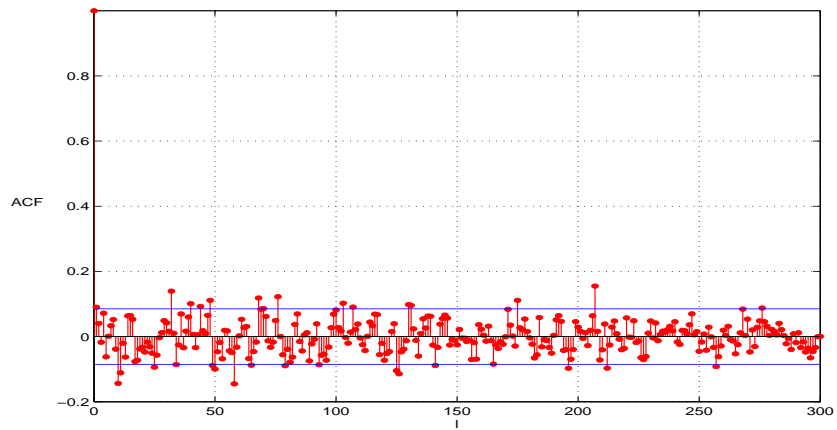
také normálně rozdělené a seriálně nezávislé s jednotkovým rozptylem. Normalitu jsme testovali pomocí *Kolmogorov-Smirnovova testu*, *Lillieforsova testu*, *Jarque-Berova testu* a vykreslením histogramu a QQ - grafu (Obr. 5.7 a 5.6). Všechny testy, podobně jako u ostatních modelů, hypotézu normálního rozložení zamítají. Pro ověření neměnnosti rozptylu jsme použili *Goldfeld-Quandtův test*, který nulovou hypotézu nezamítá. Seriální nekorelovanost jsme analyzovali pomocí korelogramů (Obr. 5.4 a 5.5) a *Ljung-Boxova testu* pro rezidua a také pro jejich druhé mocniny. Výsledky lze charakterizovat jako uspokojivé, významnější korelovanost se objevuje jen u ročních frekvencí.

Při srovnávání kvality jednotlivých modelů jsme použili věrohodnost modelu, AIC a BIC kritérium. Zároveň jsme prováděli analýzu grafu nenormovaných reziduí jedнокrokové predikce (Obr. 5.8 a 5.9), zvláště odchylek přesahujících 1 mld. Kč. Důležitá vlastnost je samozřejmě předpovědicí schopnost, měřená střední kvadratickou chybou předpovědi. Při jedнокrokové předpovědi za rok 2003 jsme zaznamenali tři odchylky výrazně větší než 1 mld. Kč. Jedná se o 14.3.2003, 9.5.2003 a 7.11.2003. Poslední dvě zřejmě souvisí s určitou ztrátou informace způsobenou vložением chybějících dní, první z nich se nepodařilo uspokojivě vysvětlit.

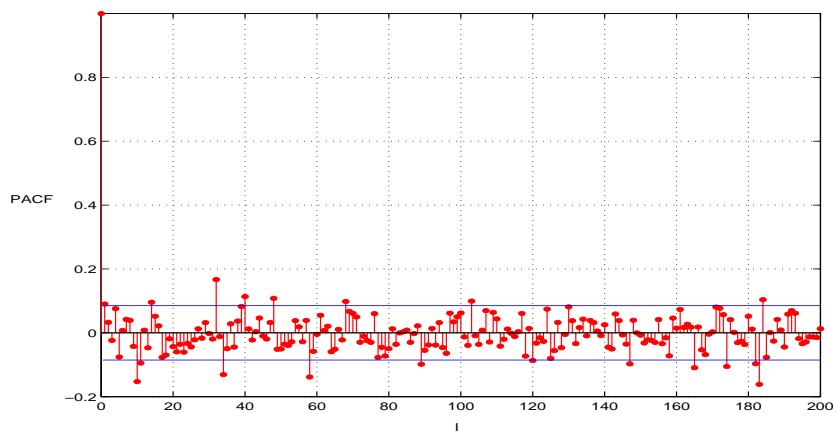
Všechny testy, histogram a QQ-graf byly provedeny na normovaných reziduích  $\{e_t\}$ . Normovaná rezidua uvažujeme bez nulových reziduí v časových okamžicích vložených chybějících dní, protože by došlo k nežádoucímu zkreslení výsledků. Ostatní statistiky provádíme na nenormovaných reziduích  $\{v_t\}$  včetně nul na místech vložených chybějících dní, protože zde jde buď pouze o relativní vztah mezi veličinami, nebo tato varianta nabízí větší přehlednost výstupů.

## 5.4 Predikce

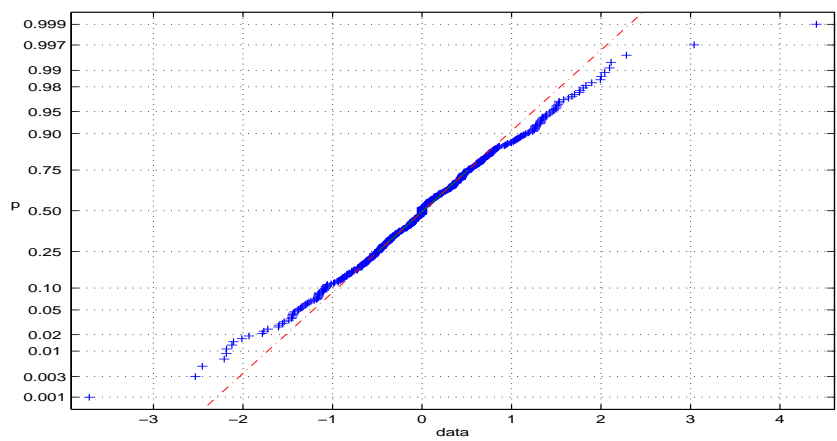
Po zkonstruování a ověření modelu nyní chceme časovou řadu poptávky po oběživu předpovídat. Přesněji formulováno najít nejlepší odhad  $y_{n+j}$ ,  $j > 0$ , za předpokladu znalosti  $y_1, \dots, y_n$ , ve smyslu střední kvadratické chyby. Označme jej  $\bar{y}_{n+j}$  a ze standardních znalostí víme, že  $\bar{y}_{n+j} = E(y_{n+j}|y)$ . Pro  $j = 1$  se jedná o výstup



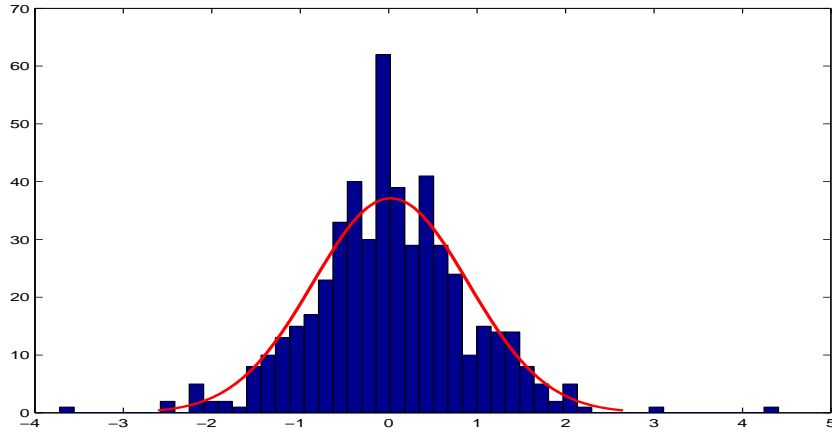
Obrázek 5.4: ACF reziduí (červeně), hranice významnosti (modře) • 1 - zpoždění



Obrázek 5.5: PACF reziduí (červeně), hranice významnosti (modře) • 1 - zpoždění



Obrázek 5.6: QQ-graf reziduí



Obrázek 5.7: Histogram reziduí

Kalmanova filtru, uvažme tedy  $j > 1$ . Jelikož  $y_{n+j} = Z_{n+j} \alpha_{n+j} + \varepsilon_{n+j}$ , dostáváme

$$\bar{y}_{n+j} = Z_{n+j} \mathbb{E}(\alpha_{n+j} | y).$$

Označme  $\bar{\alpha}_{n+j} = \mathbb{E}(\alpha_{n+j} | y)$  a  $\bar{P}_{n+j} = \mathbb{E}[(\bar{\alpha}_{n+j} - \alpha_{n+j})(\bar{\alpha}_{n+j} - \alpha_{n+j})' | y]$ . Současně odvodíme

$$\begin{aligned} \bar{F}_{n+j} &= \mathbb{E}[(\bar{y}_{n+j} - y_{n+j})(\bar{y}_{n+j} - y_{n+j})' | y] \\ &= \mathbb{E}\{Z_{n+j}(\bar{\alpha}_{n+j} - \alpha_{n+j}) - \varepsilon_{n+j}\} \{Z_{n+j}(\bar{\alpha}_{n+j} - \alpha_{n+j}) - \varepsilon_{n+j}\}' \\ &= Z_{n+j} \bar{P}_{n+j} Z_{n+j}' + H_{n+j}. \end{aligned}$$

Dále stačí určit rekurze pro výpočet  $\bar{\alpha}_{n+j}$  a  $\bar{P}_{n+j}$ . Z  $\alpha_{n+j} = T_{n+j-1} \alpha_{n+j-1} + R_{n+j-1} \eta_{n+j-1}$  plyne

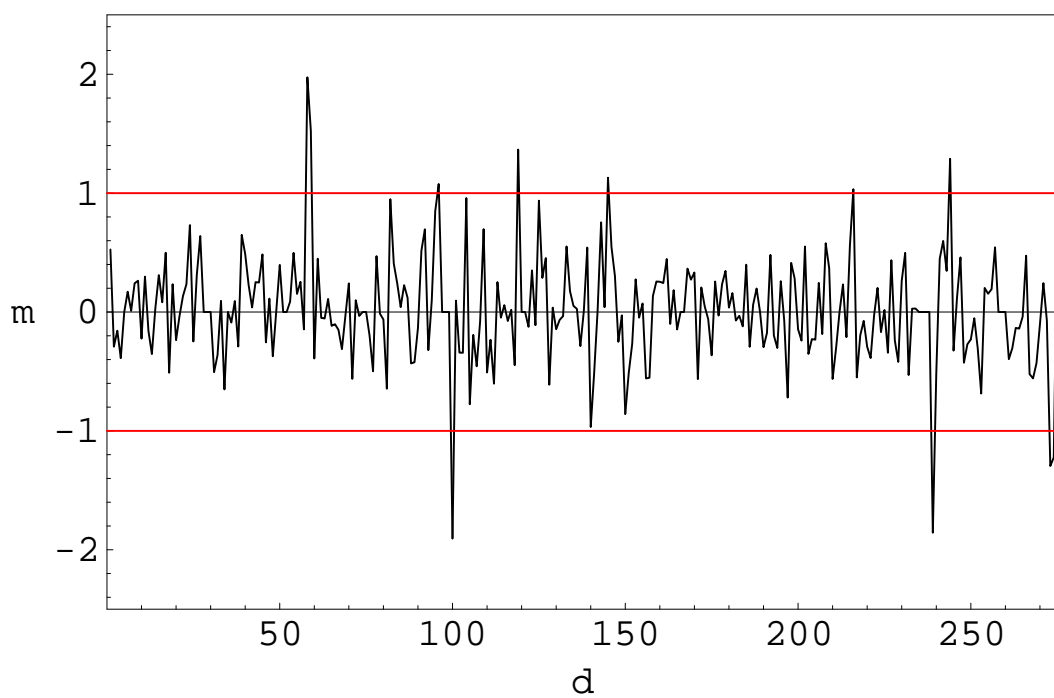
$$\bar{\alpha}_{n+j} = T_{n+j-1} \mathbb{E}(\alpha_{n+j-1} | y) = T_{n+j-1} \bar{\alpha}_{n+j-1}$$

a

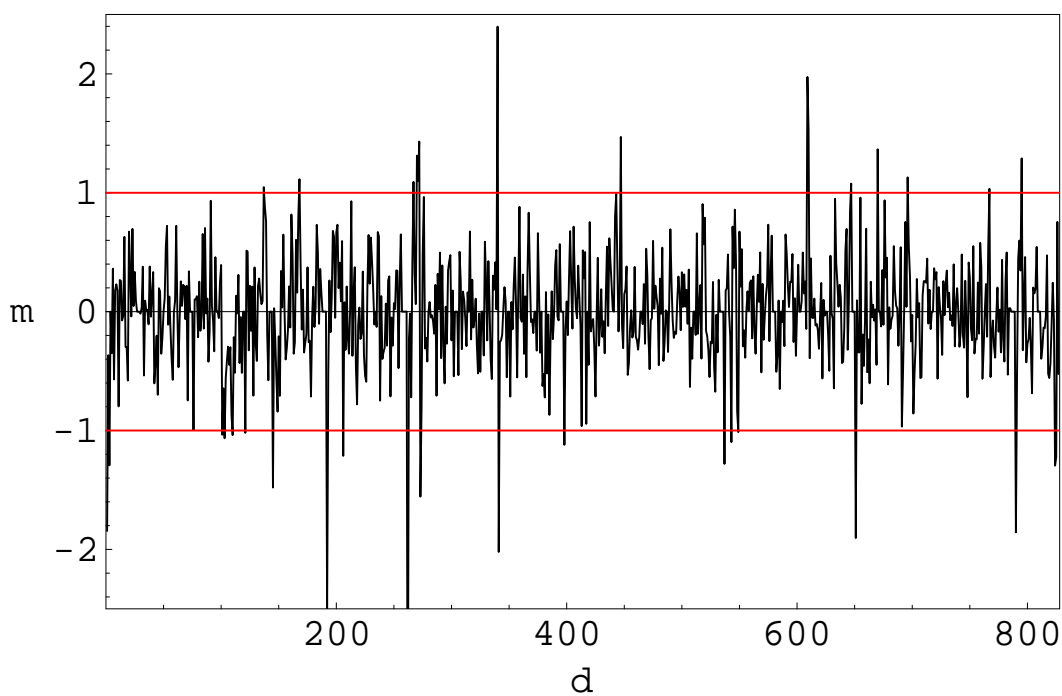
$$\begin{aligned} \bar{P}_{n+j} &= \mathbb{E}[(\bar{\alpha}_{n+j} - \alpha_{n+j})(\bar{\alpha}_{n+j} - \alpha_{n+j})' | y] \\ &= T_{n+j-1} \mathbb{E}[(\bar{\alpha}_{n+j-1} - \alpha_{n+j-1})(\bar{\alpha}_{n+j-1} - \alpha_{n+j-1})' | y] T_{n+j-1}' \\ &\quad + R_{n+j-1} \mathbb{E}(\eta_{n+j-1} \eta_{n+j-1}') R_{n+j-1}' \\ &= T_{n+j-1} P_{n+j-1} T_{n+j-1}' + R_{n+j-1} Q_{n+j-1} R_{n+j-1}'. \end{aligned}$$

Takto můžeme postupovat pro  $n+j, n+j-1, \dots, n+2$ , až  $\bar{\alpha}_{n+1} = \alpha_{n+1}$  a  $\bar{P}_{n+1} = P_{n+1}$ . Všimneme si, že odvozené rovnice jsou rovnice Kalmanova filtru, jestliže  $v_k = 0$  a  $K_k = 0$  pro  $k = n+1, \dots, j-1$ . Pro predikci lze tedy použít standardní Kalmanův filtr, u kterého považujeme pozorování  $y_t$ ,  $t > n$  za chybějící.

V našem modelu poptávky po oběživu jsme provedli předpovědi s krokem  $j = 1, 5, 10$ , tj. stejně jako u ostatních modelů. Střední kvadratická chyba předpovědi se s velikostí kroku zvyšuje, číselné hodnoty jsou uvedeny v tabulce 5.3. Vzhledem k lepším výsledkům uvažujeme v dalších kapitolách jen STS model s měsíční sezónností modelovanou kubickými spliny.



Obrázek 5.8: Rezidua jednokrokové predikce v roce 2003 (černě), hranice  $\pm 1$  mld. Kč (červeně) ● m - mld. Kč ● d - denní index



Obrázek 5.9: Rezidua jednokrokové predikce v letech 2001-2003 (černě), hranice  $\pm 1$  mld. Kč (červeně) ● m - mld. Kč ● d - denní index

Model	s kubickými spliny	s trigon. funkcemi
Nenormovaná rezidua		
Střední hodnota	-0.01422	-0.01317
Rozptyl	0.25276	0.27983
Normovaná rezidua		
Střední hodnota	0.02114	0.00041
Rozptyl	0.76767	0.59438
Šikmost	0.10879	-0.01544
Špičatost	4.73427	5.39046
Kolmogorov-Smirnov test	Zamítá $H_0$	Zamítá $H_0$
p-hodnota	0.00857	0.00053
Lilliefors test	Zamítá $H_0$	Zamítá $H_0$
p-hodnota	0.048	NaN <sup>1</sup>
Jarque-Bera test	Zamítá $H_0$	Zamítá $H_0$
p-hodnota	$2.6312 \times 10^{-14}$	0
Goldfeld-Quandt test	Nezamítá $H_0$	Nezamítá $H_0$
p13-hodnota	0.64623	0.79926
p31-hodnota	0.35377	0.20074
Ljung-Box test	rezidua	
Q(5) p-hodnota	0.08078	0.00024
Q(10) p-hodnota	0.03555	0.00039
Q(22) p-hodnota	0.01281	0.00011
Q(261) p-hodnota	0	0
Ljung-Box test	kvadrát reziduí	
Q(5) p-hodnota	$7.2217 \times 10^{-6}$	0.00032
Q(10) p-hodnota	0.0004	0.00578
Q(22) p-hodnota	0.01707	0.13595
Q(261) p-hodnota	0.66782	0.96007
Log. věrohodnost	$-1.79053 \times 10^6$	730.341
AIC	6487.77	-2.3201
BIC	6488.47	-1.6168
Chyby předpovědí	[mld. Kč]	
RMSE krok 1	0.47018	0.50443
RMSE krok 1 v ČNB	0.43787	0.43787
RMSE krok 5	1.0122	1.22207
RMSE krok 10	1.45305	1.56971

<sup>1</sup>Procedura `lillietest` v *Matlab* v některých případech nevrací p-hodnotu, ale pouze udává výsledek testu.

Tabulka 5.3: Tabulka výsledků modelů

# Kapitola 6

## Kombinace modelů

V předcházejících kapitolách jsme zkonstruovali tři různé modely pro predikci poptávky po oběživu. Nyní bychom mohli podle určitého kritéria, např. velikosti střední kvadratické chyby předpovědi, nebo kombinací více kritérií rozhodnout, který model považujeme za nejlepší a ten dále používat pro predikci řady. Tím bychom ale žádným způsobem nevyužili dodatečnou informaci obsaženou v ostatních modelech, danou jinou strukturou těchto modelů, či použitím jiných vysvětlujících exogenních proměnných. Zvláště pokud by chyby předpovědi dvou modelů byly negativně korelované, mohli bychom jejich kombinací dosáhnout výrazně lepší předpovědi časové řady oproti předpovědi obou jednotlivých modelů. Námi použitá metoda kombinací modelů je založená na regresní metodě navržené v [8]. Pro pozorovanou časovou řadu  $y_t, t = 1, \dots, T$  mějme dvě alternativní předpovědi  $f_t^a$  a  $f_t^b, t = 1, \dots, T$ . Kombinované předpovědi konstruujeme ve tvaru

$$\hat{y}_t = \alpha f_t^a + (1 - \alpha) f_t^b,$$

pro  $t = 1, \dots, T$ . Váhy alternativních modelů nalezneme pomocí metody nejmenších čtverců. Vzhledem ke skutečnosti dlouhé inicializace Kalmanova filtru u STS modelu a jeho následné adaptace po přechodu z rozšířeného na standardní Kalmanův filtr hledáme optimální váhy pouze na části časové řady připadající na rok 2002.

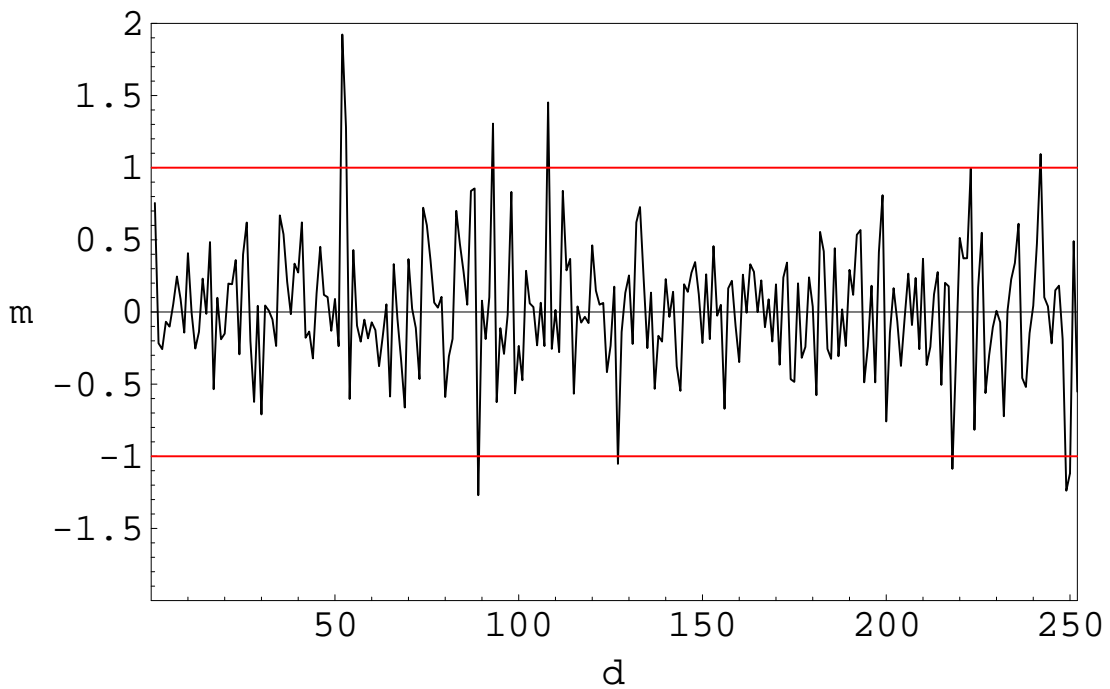
Model/Předpověď	1 krok vpřed	5 kroků vpřed	10 kroků vpřed
sts - arima	0.434918	0.904458	1.23999
sts - garch	0.448697	0.901731	1.29987
arima - garch	0.476364	1.24903	1.84795

Tabulka 6.1: RMSE předpovědi kombinovaných modelů v mld. Kč

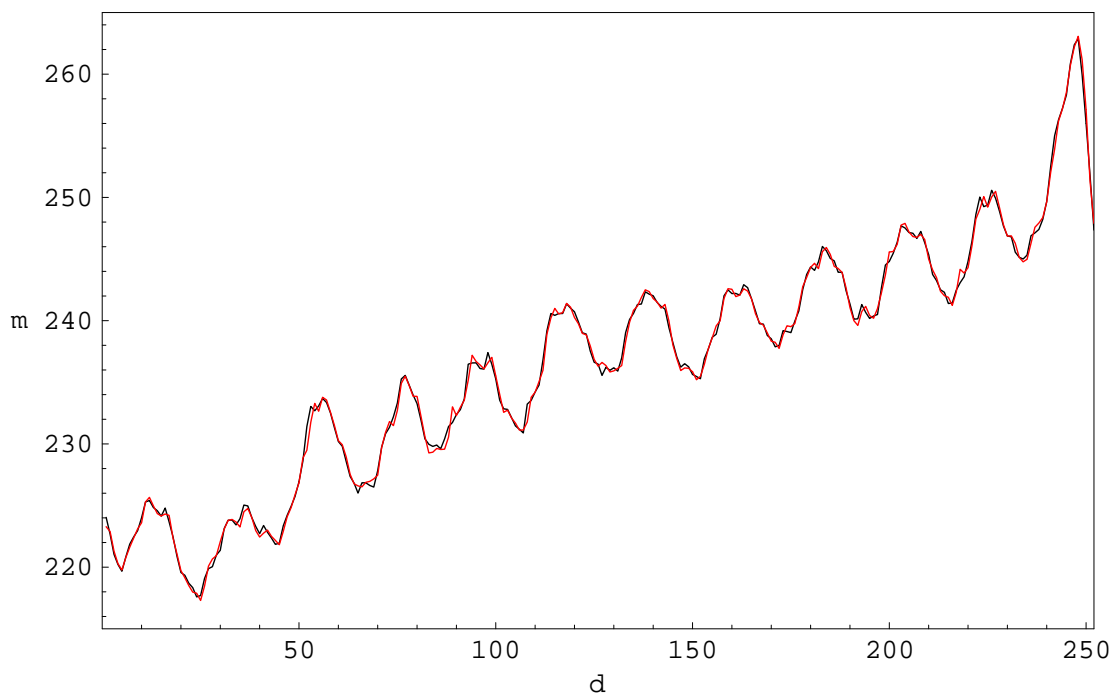
	sts - arima	sts - garch	arima - garch
RMSE krok 1	0.435519	0.449314	0.480947

Tabulka 6.2: RMSE předpovědí kombinovaných modelů po čtvrtletích v mld. Kč

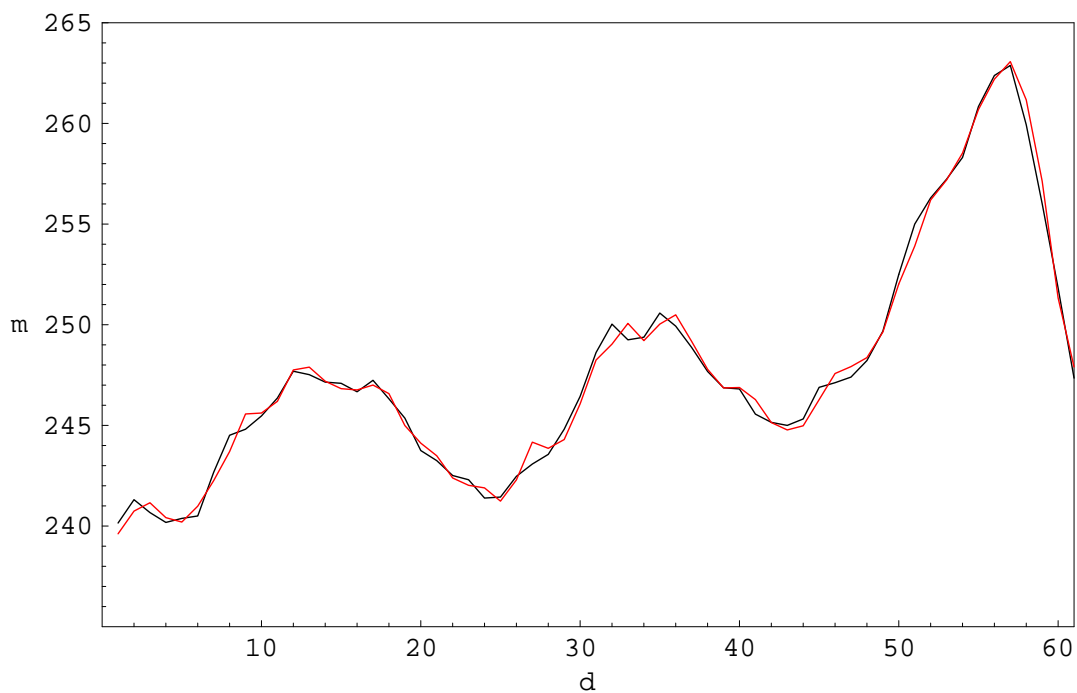
Vzhledem k tomu, že časová řada vykazuje sezonní vlivy, nabízí se možnost nestanovovat váhy předpovědí modelů stejné pro všechna  $t = 1, \dots, T$ , ale proměnlivé po obdobích, a tím využít případné lepší vlastnosti jednotlivých modelů v těchto obdobích. My jsme provedli dělení na čtvrtletí, a tedy určili váhy  $\alpha_{1Q}, \dots, \alpha_{4Q}$ . Kvalita predikce se ovšem nezvýšila. Podrobnější dělení na měsíce jako v ECB jsme neaplikovali, protože pro analýzu používáme pouze hodnoty roku 2002 a v jednotlivých měsících nemáme dostatek pozorování. Kvalita předpovědí měřená odmocninou střední kvadratické chyby se kombinací jednotlivých modelů zlepšila, výsledky uvádíme v tabulce 6.1 a 6.2. Číselné hodnoty koeficientů  $\alpha$  a další statistiky regresní analýzy lze nalézt v CD příloze.



Obrázek 6.1: Rezidua jednokrokové predikce sts-arima modelu v roce 2003 (černě), hranice  $\pm 1$  mld. Kč (červeně) • m - mld. Kč • d - denní index



Obrázek 6.2: Jednokroková předpověď sts-arima modelu (červeně) a skutečnost (černě) v roce 2003 ● m - mld. Kč ● d - denní index



Obrázek 6.3: Jednokroková předpověď sts-arima modelu (červeně) a skutečnost (černě) v 4. čtvrtletí 2003 ● m - mld. Kč ● d - denní index



# Kapitola 7

## Porovnání modelů

Jak již bylo naznačeno v předcházejících částech práce, hlavním kritériem pro porovnání modelů bude jejich předpověď schopnost. Použitými statistikami jsou odmocnina střední kvadratické chyby předpovědi, Diebold-Marianův test pro porovnání přesnosti předpovědí a počet odchylek předpovědi větších než 1 mld. Kč. Nelze navíc říci, že by některý z modelů byl statisticky více správný. Problémy s nenormalitou reziduí nastaly u všech modelů, řešením by mohl být předpoklad jiného rozdělení reziduí, například t-rozdělení. Zamítnutí konstantnosti rozptylu reziduí u ARIMA a GARCH modelu nepovažujeme za velký problém, neboť k němu dochází v důsledku většího rozptylu časové řady oběživa v roce 1998, kdy pravděpodobně docházelo k adaptaci na nový měnověpolitický režim ČNB. Zkrácení časové řady jsme neprovedli z důvodu zachování konzistence podkladů jednotlivých modelů.

Střední kvadratická chyba, včetně výsledných hodnot, již byla uvedena u jednotlivých modelů, zbývá tedy objasnit Diebold-Marianův test. Tento test je určen pro testování nulové hypotézy stejné předpověď schopnosti dvou modelů a vychází z centrální limitní věty pro stacionární náhodné procesy. Diebold a Mariano pracují s obecnou chybovou funkcí (loss function), my jsme použili kvadratickou chybovou funkci. Testová statistika má pak tvar

$$S_{DM} = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}} \xrightarrow{d} N(0, 1),$$

kde  $\bar{d} = \frac{1}{T} \sum_{i=1}^T \hat{d}_i$ ,  $\hat{d}_i = \hat{e}_{A,i}^2 - \hat{e}_{B,i}^2$ ,  $i = 1, \dots, T$ ,  $\hat{e}_{A,i}$  jsou chyby predikce modelu A,  $\hat{e}_{B,i}$  jsou chyby predikce modelu B,  $T$  je délka predikovaného intervalu a  $\hat{f}_d(0)$  je konzistentní odhad spektrální hustoty procesu  $\{d_i\}$  na frekvenci 0. Předpokladem je zde stacionarita procesu  $\{d_i\}$ . Odhad  $2\pi\hat{f}_d(0)$  dostaneme jako váženou sumu

Modely/Předpověď		1 krok vpřed	5 kroků vpřed	10 kroků vpřed
STS vs	ARIMA	0.362	0.063	0.063
	GARCH	0.000	0.001	0.000
	sts-arima	0.977	0.995	0.992
	sts-garch	0.984	0.990	0.979
	arima-garch	0.421	0.094	0.070
ARIMA vs	GARCH	0.000	0.006	0.011
	sts-arima	0.998	0.980	0.974
	sts-garch	0.891	0.980	0.966
	arima-garch	0.938	0.842	0.891
GARCH vs	sts-arima	1.000	1.000	1.000
	sts-garch	1.000	1.000	1.000
	arima-garch	1.000	1.000	0.999
sts-arima vs	sts-garch	0.158	0.532	0.049
	arima-garch	0.006	0.030	0.026
sts-garch vs	arima-garch	0.143	0.028	0.034

Tabulka 7.1: Výsledky Diebold-Marianova testu

dostupných výběrových autokovariancí  $\hat{\gamma}_d(\tau)$ ,

$$2\pi\hat{f}_d(0) = \sum_{\tau=-(T-1)}^{(T-1)} l\left(\frac{\tau}{S(T)}\right) \hat{\gamma}_d(\tau),$$

kde funkce  $l\left(\frac{\tau}{S(T)}\right)$  přiřazuje váhy autokovarianční funkci na jednotlivých zpožděních a je nazývána jádrem (kernel) a  $S_t$  udává zpoždění, po kterém autokovariance zanedbáváme (truncation lag). Stejně jako Diebold a Mariano jsme použili jednotkové (obdélníkové) jádro

$$l\left(\frac{\tau}{S(T)}\right) = \begin{cases} 1 & \text{pro } \left|\frac{\tau}{S(T)}\right| \leq 1 \\ 0 & \text{jinak} \end{cases}$$

a  $S(T) = (k - 1)$  v případě  $k$ -krokové předpovědi, neboť dle teoretických výsledků jsou chyby optimální  $k$ -krokové předpovědi nejvýše  $(k - 1)$ -závislé. V praxi toto sice z mnoha důvodů nemusí platit, přesto použití této hodnoty  $S(T)$  dává dobré výsledky, podrobnosti lze najít v [5]. Vypočtené hodnoty Diebold-Marianova testu, převedené na pravděpodobnosti, jsou uvedeny v tabulce 7.1. Jestliže je hodnota malá, například pro hladinu testu  $\alpha = 5\%$  menší než 0.025, potom zamítáme nulovou hypotézu ve prospěch modelu A, tj. předpověď modelem A je statisticky přesnější než modelem B. Je-li naopak hodnota velká, zamítáme nulovou hypotézu ve prospěch modelu B. V tabulce 7.2 následují četnosti překročení hranice 1 mld. Kč u chyb předpovědí jednotlivými modely.

Z výsledků je patrné, že nejlepším modelem je sts-arima model, a to pro všechny provedené horizonty předpovědí. Při jednokrokové předpovědi je předpověď

Model/Předpověď	1 krok vpřed	5 kroků vpřed	10 kroků vpřed
STS	11	63	98
ARIMA	9	95	130
GARCH	44	122	185
sts-arima	10	59	89
sts-garch	8	62	87
arima-garch	10	96	138

Tabulka 7.2: Počet chyb predikce větších než 1 mld. Kč

sts-arima modelem statisticky významně přesnější než všechny ostatní modely, s výjimkou sts-garch modelu. Nejhorším modelem je samotný model GARCH, který je signifikantně horší než ostatní modely a také počet chyb predikce přesahujících 1 mld. Kč je výrazně větší než u ostatních modelů. Všechny kombinace modelů vykazují lepší výsledky než příslušné samostatné modely. Největší přínos z kombinace modelů lze pozorovat u sts-garch a sts-arima modelu. Naopak u kombinace arima-garch je zlepšení předpovědi minimální, neboť struktura obou modelů je velmi podobná.

Budeme tedy považovat sts-arima za náš finální model. Při analýze predikce (Obr. 6.1) je zajímavé, že k největším odchylkám predikce od skutečnosti nedochází v období Vánoc, ale v květnu a červnu, neboť právě kvalita zachycení vlivu období Vánoc se podařila zvýšit kombinací modelů. Největší odchylka, téměř 2 mld. Kč, nastala 14.3.2003 a je v této části roku zcela ojedinělá. V ostatních měsících je predikce modelem velmi přesná. Srovnáme-li finální model s triviální předpovědí, kdy jako předpověď  $y_{t+1}$  při znalosti  $y_1, \dots, y_t$  bereme  $y_t$ , která má RMSE jedнокrokové předpovědi 1.09739, snížili jsme chybu predikce na necelých 40% oproti této triviální předpovědi.

# Kapitola 8

## Závěr

Cílem této práce bylo vytvoření stochastického modelu, který by mohl být použit pro předpovídání nejvýznamnějšího autonomního faktoru ovlivňujícího likviditu peněžního trhu v ČR. Tímto faktorem je objem oběživa, neboli objem hotovostních peněz držených mimo centrální banku.

Jak již bylo zmíněno, tato časová řada vykazuje velké množství sezónních závislostí, které se navzájem kombinují a způsobují četné komplikace při tvorbě modelu. Při nalezení relevantních vysvětlujících proměnných jsme vycházeli ze zkušeností expertů ČNB, předchozích modelů ([2] a [12]) a vlastní analýzy časové řady. Náročnost správného zachycení sezónností potvrzuje například 75 nenulových parametrů u ARIMA modelu.

Zkonstruovali jsme tři různé modely. Jsou jimi ARIMA model, založený na Box-Jenkinsově metodologii, GARCH model, kterým jsme se snažili postihnout některé rysy reziduí ARIMA modelu a STS model založený na zcela odlišném přístupu k časovým řadám pomocí stavových prostorů a Kalmanových rekurzí. Všechny tři modely dávají dobré výsledky. Nejlepší z nich je STS model, který po zkombinování s ARIMA modelem vytváří finální model. Oproti tomu aplikace GARCH modelu nespĺnila očekávání. Finální sts-arima model dosáhl nepatrně menší chyby jednokrokové predikce než expertní odhad prováděný v ČNB.

V práci byl zkonstruován stochastický model oběživa, který dává výsledky na úrovni dosud používaných odhadů. Přidá-li se k výsledkům modelu expertní zkušenost, zkoriguje se předpověď modelem v kritických obdobích roku a při mimořádných událostech. Vznikne tak lepší odhad poptávky po penězích, který umožní efektivnější aplikaci měnověpolitických nástrojů ČNB. Vytvořený model nabízí zároveň predikci s delším časovým horizontem, která bude nutná v souvislosti se zavedením jednotné evropské měny euro.

# Literatura

- [1] Brockwell, P.J., Davis, R.A. (2001): Introduction to Time Series and Forecasting. Springer-Verlag, New York.
- [2] Cabrero, A., Camba-Mendez, G., Hirsch, A. (2001): Modeling the Daily Series of Banknotes in Circulation in the Context of the Liquidity Management Operations of the ECB. *ECB Working Paper no. 142*.
- [3] Cipra, T. (1986): Analýza časových řad s aplikacemi v ekonomii. SNTL, Praha.
- [4] De Jong, P. (1991): The Diffuse Kalman Filter. *The Annals of Statistics* **19**, 1073-1083.
- [5] Diebold, F.X., Mariano, R.S. (1995): Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* **13**, 253-263.
- [6] Durbin, J., Koopman, S.J. (2001): Time Series Analysis by State Space Methods. Oxford University Press.
- [7] Franke, J., Härdle, W., Hafner, C. (2001): Einführung in die Statistik der Finanzmärkte. Springer-Verlag, Heidelberg.
- [8] Granger, C.V.J., Ramanathan, R. (1984): Improved Methods of Forecasting. *Journal of Forecasting* **3**, 197-204.
- [9] Harvey, A.C. (1989): Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.
- [10] Harvey, A.C., Koopman, S.J., Riani M. (1997): The Modeling and Seasonal Adjustment of Weekly Observations. *Journal of Business and Economic Statistics* **15**, 354-368.
- [11] Hlaváček, M. (2005): The Application of Structured Feedforward Neural Networks to the Modeling of the Daily Series of Currency in Circulation. *Lecture Notes In Computer Science*, Springer-Verlag, New York.

- [12] Hlaváček, M. (2004): Modelování objemu oběživa. Bakalářská práce, FJFI ČVUT, Praha.
- [13] Pierce, D.A., Grupe, M.R., Cleveland, W.P. (1984): Seasonal Adjustment of the Weekly Monetary Aggregates: A Model-Based Approach. *Journal of Business & Economic Statistics* **2**, 260-270.