



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

**DIPLOMOVÁ PRÁCE**

Bc. Matyáš Krahulík

**Modely multiplikatívnych chyb**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Michal Pešta, Ph.D.

Studijní program: Finanční a pojistná matematika

Studijní obor: Finanční a pojistná matematika

Praha 2024

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Rád bych tímto poděkoval panu doc. RNDr. Michalu Peštovi, Ph.D. za cenné rady a trpělivost při vedení práce.

Název práce: Modely multiplikatívních chyb

Autor: Bc. Matyáš Krahulík

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Michal Pešta, Ph.D.

Abstrakt: Tato práce se věnuje tzv. modelům multiplikatívních chyb (MEM), které se využívají k modelování nezáporných časových řad, nejčastěji ve finančním sektoru. Obsahem první kapitoly jsou modely ARCH a GARCH, které sice nepatří do skupiny modelů multiplikatívních chyb, ale úzce s nimi souvisí. Druhá kapitola se již zaměřuje přímo na modely MEM a jejich další rozšíření, jako jsou modely MEM rozšířené v nule (ZA-MEM) nebo semiparametrické modely MEM (SpMEM). Tyto modely jsou nejprve definovány a poté jsou představeny metody pro odhady parametrů v těchto modelech. Ve třetí kapitole, která obsahuje praktickou část práce, jsou postupy z druhé kapitoly aplikovány na reálná data v podobě časové řady škod z jedné z českých pojišťoven. V závěru jsou navrženy další postupy pro rozšíření aplikací modelů MEM na pojišťovnická či jiná data.

Klíčová slova: multiplikatívni chyba, časová řada, MEM, GARCH

Title: Multiplicative Error Models

Author: Bc. Matyáš Krahulík

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Michal Pešta, Ph.D.

Abstract: This thesis is devoted to the so-called multiplicative error models (MEM), which are used to model non-negative time series, most often in the financial sector. The first chapter focuses on ARCH and GARCH models, which do not belong to the group of multiplicative error models, but are closely related to them. The second chapter focuses directly on the MEM and their further extensions, such as zero-augmented MEM (ZA-MEM) or semiparametric MEM (SpMEM). These models are first defined and then methods for parameter estimation in these models are presented. In the third chapter, which contains the practical part of the thesis, the practices from the second chapter are applied to real data in the form of a time series of claims from one of the Czech insurance companies. In the conclusion, further extensions to the the applications of the MEM to insurance or other data are proposed.

Keywords: multiplicative error, time series, MEM, GARCH

# Obsah

Úvod	3
<b>1 Modely ARCH a GARCH</b>	<b>4</b>
1.1 Modely ARCH	4
1.2 Modely GARCH	5
<b>2 Modely multiplikativních chyb</b>	<b>7</b>
2.1 Základní modely multiplikativních chyb	7
2.2 Modely multiplikativních chyb rozšířené v nule	10
2.3 Dynamické modely multiplikativních chyb rozšířené v nule	12
2.4 Vektorové modely multiplikativních chyb	13
2.4.1 Vícerozměrné gamma rozdělení	13
2.4.2 Specifikace pomocí kopul	14
2.5 Semiparametrické modely multiplikativních chyb	18
2.5.1 Zobecněná momentová metoda	20
2.5.2 Metoda kvazi-maximální věrohodnosti	21
2.6 Vektorové semiparametrické modely multiplikativních chyb	21
2.6.1 Zobecněná momentová metoda	22
2.7 Předpovědi v modelech SpMEM	23
2.8 Předpovědi v modelech SpvMEM	24
<b>3 Praktická část</b>	<b>25</b>
3.1 Popis a grafické zobrazení dat	25
3.2 Základní modely MEM	29
3.3 Modely MEM rozšířené v nule	34
3.4 Semiparametrické modely MEM	40
<b>Závěr</b>	<b>46</b>
<b>Seznam použité literatury</b>	<b>49</b>
<b>Seznam obrázků</b>	<b>50</b>
<b>Seznam tabulek</b>	<b>51</b>
<b>A Přílohy</b>	<b>52</b>
A.1 Důležité části kódu	52
A.1.1 Odhad parametrů základního modelu metodou maximální věrohodnosti	52
A.1.2 Odhad parametrů v modelu ZA-MEM pomocí exponenciální kvazi-maximální věrohodnosti	52
A.1.3 Odhad parametrů v modelu SpMEM pomocí zobecněné momentové metody	53

A.1.4	Odhad parametrů v modelu SpMEM pomocí kvazi-maximální věrohodnosti s rozdělením gamma . . . .	56
-------	--	----

# Úvod

Modely multiplikativních chyb (MEM) zavedl před více než 20 lety významný americký ekonom Robert F. Engle ve článku (Engle, 2002). Jsou určeny k modelování nezáporných časových řad a jejich nejčastější aplikace jsou ve financích (volatilita, cena aktiv nebo objem obchodu). Dají se však využít téměř kdekoli, kde je z podstaty věci potřeba pracovat právě s pozitivní či nezápornou časovou řadou. Jde o téma, které se dodnes vyvíjí a matematici stále přicházejí s nejrůznějšími úpravami a rozšířeními těchto modelů. Jako příklad lze uvést přehledový článek (Cipollini a Gallo, 2022) od italských ekonomů a matematiků Fabrizio Cipolliniho a Gampiera M. Galla, kteří zavedli tzv. semiparametrické modely MEM (SpMEM), kde model obohatili o nízkofrekvenční složku, standardně odhadovanou jádrovými odhady.

Cílem a motivací této práce je postupně představit vybrané modely z třídy MEM od jejího zavedení až po současnost, ukázat možné přístupy ke statistické inferenci v těchto modelech a teoretické postupy následně aplikovat na reálná data.

# 1. Modely ARCH a GARCH

Základy modelům, které budeme v této práci primárně zkoumat, položil již v roce 1982 americký ekonom, statistik a držitel Nobelovy ceny Robert F. Engle, který ve svém článku (Engle, 1982) zavedl třídu modelů ARCH. Zkratka vychází z anglického „autoregressive conditional heteroskedasticity“, což lze přeložit jako „autoregresivní podmíněná heteroskedasticita“. To znamená, že podmíněný rozptyl se modeluje jako funkce předchozích (zpožděných) hodnot pozorované proměnné. Na Engleovu práci později navázal dánský ekonom Tim Bollerslev, který ve článku (Bollerslev, 1986) zobecnil ARCH modely tak, aby mohl být současný podmíněný rozptyl časové řady modelován nejen za pomoci zpožděných pozorování, ale i s využitím zpožděných podmíněných rozptylů. Modely typu ARCH a GARCH si nyní definujeme a uvedeme i některé jejich důležité vlastnosti.

V následujících částech textu budeme vždy uvažovat diskrétní časovou řadu  $\{Y_t, t \in \mathbb{Z}\}$  s hodnotami v množině reálných čísel. Zavedeme značení  $\Psi_t$  pro sigma algebru generovanou náhodnými veličinami pozorovanými do času  $t$ , tedy  $\Psi_t = \sigma(Y_t, Y_{t-1}, \dots)$ ,  $t \in \mathbb{Z}$ .

## 1.1 Modely ARCH

Nejprve si definujeme proces typu ARCH tak, jak byl zaveden ve článku (Engle, 1982).

**Definice 1.** *Nechť pro náhodné veličiny  $\{Y_t, t \in \mathbb{Z}\}$  platí:*

$$\begin{aligned} Y_t | \Psi_{t-1} &\sim N(\mu_t, h_t), \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2, \\ \varepsilon_t &= Y_t - \mu_t, \\ \boldsymbol{\alpha} &= (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_p)^\top, t \in \mathbb{Z}, \end{aligned}$$

kde  $\boldsymbol{\alpha}$  je vektor neznámých parametrů a platí  $\alpha_0 > 0$  a  $\alpha_i \geq 0$ ,  $i = 1, \dots, p$ . Potom řekneme, že časová řada  $\{Y_t, t \in \mathbb{Z}\}$  je generována **ARCH procesem** řádu  $p$ .

*Poznámka.* Je-li

$$\mu_t = \mathbf{X}_t^\top \boldsymbol{\gamma}, t \in \mathbb{Z},$$

kde  $\mathbf{X}_t = (X_{t1}, \dots, X_{tm})^\top$ ,  $t \in \mathbb{Z}$  jsou náhodné vektory a  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^\top$  je vektor neznámých parametrů, řekneme, že se jedná o **lineární ARCH proces** řádu  $p$ .

Vidíme, že podmíněná střední hodnota je modelována přístupem klasické lineární regrese, zatímco podmíněný rozptyl závisí na předchozích pozorovaných odchylkách, tedy i na zpožděných pozorováních. Jde proto o model autoregresivní podmíněné heteroskedasticity. To s sebou přináší několik výhod, z nichž se nejčastěji uvádí zachycení efektu tzv. shlukování volatility (anglicky „volatility clustering“), jenž se často vyskytuje v časových řadách finančního charakteru. Tento efekt lze zjednodušeně popsat tak, že vysoké výkyvy např. v cenách aktiv jsou obvykle následovány dalšími vysokými výkyvy a naopak nízké výkyvy jsou



následovány dalšími nízkými výkyvy.

Engle ve článku dále odvodil věrohodnostní funkci modelu a ukázal některé jeho důležité vlastnosti. Takovými vlastnostmi jsou například existence momentů a stacionarita procesu.

**Věta 1.** Pro  $r \in \mathbb{Z}$ ,  $2r$ -tý moment ARCH procesu prvního řádu s  $\alpha_0 > 0$  a  $\alpha_1 \geq 0$  existuje právě tehdy, když

$$\alpha_1^r \prod_{j=1}^r (2j - 1) < 1.$$

*Důkaz.* Důkaz je podrobně popsán ve článku (Engle, 1982). □

**Věta 2.** ARCH proces řádu  $p$  s  $\alpha_0 > 0$  a  $\alpha_1 \geq 0, \alpha_2 \geq 0, \dots, \alpha_p \geq 0$  je kovariančně stacionární právě tehdy, když polynom  $\alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \dots + \alpha_p\lambda^p$  má všechny kořeny mimo jednotkový kruh v komplexní rovině. Dále platí pro všechna  $t \in \mathbb{Z}$ :

$$\text{var}(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{j=1}^p \alpha_j}.$$

*Důkaz.* Jednotlivé kroky důkazu jsou podrobně popsány ve článku (Engle, 1982). □

## 1.2 Modely GARCH

Procesy typu ARCH se ihned po jejich zavedení začaly hojně používat pro nejrůznější aplikace, což vedlo k odhalení některých nedostatků. K adekvátnímu zachycení vývoje volatility je často potřeba vysoký řád  $p$ , což vede k nutnosti odhadovat velký počet parametrů. Některý z odhadů pak může porušit podmínku nezápornosti. Tento a další nedostatky ARCH procesů vyřešil ekonom Tim Bollerslev, který ve svém článku (Bollerslev, 1986) zavedl jejich zobecněnou verzi nazvanou GARCH.

**Definice 2.** Necht pro náhodné veličiny  $\{Y_t, t \in \mathbb{Z}\}$  platí:

$$\begin{aligned} Y_t | \Psi_{t-1} &\sim N(\mu_t, h_t), \\ h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}, \\ \varepsilon_t &= Y_t - \mu_t, t \in \mathbb{Z}, \end{aligned}$$

kde  $\alpha_i, i = 0, 1, \dots, q$  a  $\beta_i, i = 1, \dots, p$  jsou neznámé parametry a platí:

$$\begin{aligned} p &\geq 0, \quad q \geq 0, \\ \alpha_0 &> 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, q, \\ \beta_i &\geq 0, \quad i = 1, \dots, p. \end{aligned}$$

Potom řekneme, že časová řada  $\{Y_t, t \in \mathbb{Z}\}$  je generována **GARCH procesem** řádů  $p$  a  $q$ .

*Poznámka.* Stejně jako v případě ARCH procesu lze volit

$$\mu_t = \mathbf{X}_t^\top \boldsymbol{\gamma}, t \in \mathbb{Z},$$

kde  $\mathbf{X}_t = (X_{t1}, \dots, X_{tm})^\top$ ,  $t \in \mathbb{Z}$  jsou náhodné vektory a  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^\top$  je vektor neznámých parametrů. Potom se jedná o **lineární GARCH proces** řádů  $p$  a  $q$ .

*Poznámka.* Alternativně lze v předchozí definici psát

$$h_t = \alpha_0 + A(L) \varepsilon_t^2 + B(L) h_t,$$

kde  $A(z) = \alpha_1 z + \alpha_2 z^2 + \dots + \alpha_q z^q$  a  $B(z) = \beta_1 z + \beta_2 z^2 + \dots + \beta_p z^p$  jsou polynomy a  $L$  značí operátor zpoždění, tedy  $L^k y_t = y_{t-k}$  pro  $k \in \mathbb{Z}_0$  a  $t \geq k$ .

S využitím zápisu v předchozí poznámce lze zapsat následující větu.

**Věta 3.** *Proces GARCH řádů  $p$  a  $q$  je kovariančně stacionární právě tehdy, když  $A(1) + B(1) < 1$ . Potom platí pro  $t \in \mathbb{Z}$ ,  $s \in \mathbb{Z}$  a  $t \neq s$ :  $\text{var}(\varepsilon_t) = \alpha_0 (1 - A(1) - B(1))^{-1}$  a  $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ .*

*Důkaz.* Jednotlivé kroky důkazu jsou podrobně popsány ve článku (Bollerslev, 1986). □

Dalším nedostatkem ARCH modelů, který se modelům typu GARCH vyřešit nepodařilo, je ignorování tzv. pákového efektu, jenž lze také často pozorovat na finančních trzích. Zjednodušeně to znamená, že kladné a záporné odchylky, v modelu označené jako  $\{\varepsilon_t, t \in \mathbb{Z}\}$ , mohou mít na volatilitu odlišný vliv (volatilita má obvykle tendenci více narůstat po výrazných cenových poklesech než po nárůstech). Potřebovali bychom tedy pracovat s asymetrickým modelem, jenž by takový fenomén byl schopen zachytit. Podobných modelů bylo v literatuře navrženo nespočet, avšak jedním z nejpoužívanějších se stal model typu EGARCH představený ve článku (Nelson, 1991).

**Definice 3.** *Nechť pro náhodné veličiny  $\{Y_t, t \in \mathbb{Z}\}$  platí:*

$$Y_t | \Psi_{t-1} \sim N(\mu_t, h_t),$$

$$\log(h_t) = \alpha_0 + \sum_{i=1}^q \alpha_i \left| \frac{\varepsilon_{t-i}}{h_{t-i}} \right| + \sum_{i=1}^p \beta_i \log(h_{t-i}) + \sum_{i=1}^m \gamma_i \frac{\varepsilon_{t-i}}{h_{t-i}},$$

$$\varepsilon_t = Y_t - \mu_t, t \in \mathbb{Z},$$

kde  $\alpha_i, i = 0, \dots, q$ ,  $\beta_i, i = 1, \dots, p$  a  $\gamma_i, i = 1, \dots, m$  jsou neznámé parametry. Potom řekneme, že časová řada  $\{Y_t, t \in \mathbb{Z}\}$  je generována **EGARCH procesem** řádů  $p, q$  a  $m$ .

Kromě zachycení pákového efektu patří mezi výhody modelu i fakt, že díky modelování rozptylu exponenciální funkcí není nutné specifikovat žádné požadavky na odhadované parametry. V předchozích modelech musely být parametry (a tudíž i jejich odhady) nezáporné.

## 2. Modely multiplikativních chyb

Modely ARCH a GARCH se od svého zavedení dočkaly mnoha různých úprav, rozšíření a aplikací. Dalšího velkého pokroku dosáhl opět Robert F. Engle, který je ve svém článku (Engle, 2002) ještě více zobecnil a zavedl tzv. modely multiplikativních chyb (anglicky „Multiplicative Error Models“, často zkracováno jako MEM).

### 2.1 Základní modely multiplikativních chyb

Engle se ve článku zaměřil na modelování nezáporných časových řad. Uvažujme časovou řadu  $\{Y_t, t \in \mathbb{Z}\}$  takovou, že  $Y_t \geq 0$  s.j.  $\forall t \in \mathbb{Z}$ . Předpokládejme, že

$$\forall t \in \mathbb{Z} : \mathbb{P}(Y_t < \xi | \Psi_{t-1}) > 0 \quad \forall \xi > 0, \quad (1)$$

což znamená, že pravděpodobnost výskytu nul a hodnot blízkých nule je nenulová. Dále označíme:

$$\mu_t = \mathbb{E}(Y_t | \Psi_{t-1}), \quad \sigma_t^2 = \text{var}(Y_t | \Psi_{t-1}), \quad t \in \mathbb{Z}.$$

Nejběžnějším typem modelu, který by se pro takovou řadu dal využít, je aditivní model tvaru

$$Y_t = \mu_t + \varepsilon_t, \quad t \in \mathbb{Z},$$

kde

$$\varepsilon_t | \Psi_{t-1} \sim N(0, \sigma_t^2).$$

Mezi takové modely patří i ARCH a GARCH představené v předchozí kapitole (zde ovšem časová řada nemusí být nezáporná). Chceme-li takto zachytit nezápornou časovou řadu, musíme si dávat pozor na to, aby odchylky  $\{\varepsilon_t, t \in \mathbb{Z}\}$  nebyly záporné a v absolutní hodnotě větší než podmíněná střední hodnota, což znamená, že množina přípustných hodnot bude odlišná pro každou odchylku a odhadování metodou maximální věrohodnosti bude velmi složité. Pravděpodobnost výskytu hodnoty blízké nule je:

$$\mathbb{P}(Y_t < \xi | \Psi_{t-1}) = \mathbb{P}(\varepsilon_t < \xi - \mu_t | \Psi_{t-1}).$$

Z toho vyplývá, že aby mohla být splněna podmínka (1), musí být rozdělení odchylky  $\varepsilon_t$  nespojitě v bodě  $-\mu_t$ . Jedním z možných řešení těchto problémů je logaritmičtý model ve tvaru

$$\log(Y_t) = m_t + u_t,$$

kde

$$\mu_t = e^{m_t} \mathbb{E}(e^{u_t}), \quad \sigma_t^2 = e^{2m_t} \text{var}(e^{u_t}).$$

Toto řešení však nebude fungovat, budou-li se v řadě vyskytovat pozorování přímo rovna nule. To se nám může stát především v případě vysokofrekvenčních dat, např. pokud bychom zkoumali časovou řadu ukazující objem zobchodovaných akcií konkrétní společnosti za velmi krátký časový úsek. Model, který tyto problémy řeší, si zavedeme v následující definici.

**Definice 4.** Necht pro náhodné veličiny  $\{Y_t, t \in \mathbb{Z}\}$  platí:

$$Y_t = \mu_t \varepsilon_t,$$

$$\varepsilon_t | \Psi_{t-1} \sim (1, \phi_t^2), t \in \mathbb{Z},$$

kde  $\phi_t^2 \geq 0$ ,  $t \in \mathbb{Z}$  jsou nezáporné rozptyly odchylek. Necht jsou dále  $\{\varepsilon_t, t \in \mathbb{Z}\}$  podmíněně nezávislé. Potom řekneme, že časová řada  $\{Y_t, t \in \mathbb{Z}\}$  je generována **modelem multiplikativních chyb**, zkráceně **MEM**.

Přirozený obor hodnot pro  $\{\varepsilon_t, t \in \mathbb{Z}\}$  je interval  $[0, \infty)$ , tedy bude podmínka (1) automaticky splněna. Velmi podobně lze model definovat i ve vícerozměrném případě.

**Definice 5.** Necht pro náhodné vektory  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  s hodnotami v  $\mathbb{R}^K$  platí:

$$\mathbf{Y}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\varepsilon}_t | \boldsymbol{\Psi}_{t-1} \sim (\mathbf{1}, \mathbf{V}_t), t \in \mathbb{Z},$$

kde  $\mathbf{1}$  je jednotkový vektor,  $\mathbf{V}_t$ ,  $t \in \mathbb{Z}$  jsou symetrické a pozitivně semidefinitní kovarianční matice vektorových odchylek a  $\odot$  značí Hadamardův součin vektorů (součin po složkách). Dále necht jsou  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  podmíněně nezávislé. Potom řekneme, že  $m$ -rozměrná časová řada  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  je generována **vektorovým modelem multiplikativních chyb**, zkráceně **vMEM**.

*Poznámka.* Z definice modelů vMEM vyplývá:

$$\text{var}(\mathbf{Y}_t | \boldsymbol{\Psi}_{t-1}) = \text{diag}(\boldsymbol{\mu}_t) \mathbf{V}_t \text{diag}(\boldsymbol{\mu}_t), t \in \mathbb{Z},$$

a tedy podmíněná pozitivní definitnost kovarianční matice je automaticky zaručena strukturou modelu.

Odhady v modelech typu MEM jsou obvykle prováděny metodou maximální věrohodnosti, k čemuž je potřeba nejdříve zvolit vhodné podmíněné rozdělení odchylek  $\{\varepsilon_t, t \in \mathbb{Z}\}$ . Přirozeně se nabízí volba exponenciálního rozdělení  $\varepsilon_t | \Psi_{t-1} \sim \text{Exp}(1)$ . Hustota takového rozdělení je:

$$f(x) = e^{-x}, x \geq 0.$$

Podmíněná hustota  $Y_t, t \in \mathbb{Z}$  je potom:

$$g(y) = \frac{e^{-y/\mu_t}}{\mu_t}, y \geq 0,$$

a tedy pro vektor odhadovaných parametrů  $\boldsymbol{\theta}$  a pozorování  $Y_1 = y_1, \dots, Y_n = y_n$  máme podmíněnou logaritmickou věrohodnost ve tvaru

$$\ell_n(\boldsymbol{\theta} | \Psi_{n-1}) = \sum_{t=1}^n \left[ -\log(\mu_t(\boldsymbol{\theta})) - \frac{y_t}{\mu_t(\boldsymbol{\theta})} \right]. \quad (2)$$

Položíme-li první derivaci podle  $\boldsymbol{\theta}$  rovnu nule (tentokrát již pro jednoduchost vynecháme  $\boldsymbol{\theta}$  v argumentu  $\mu_t$ ), získáme podmínku:

$$\frac{\partial \ell_n(\boldsymbol{\theta} | \Psi_{n-1})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^n \left[ \left( -\frac{1}{\mu_t} + \frac{y_t}{\mu_t^2} \right) \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \right] = \sum_{t=1}^n \left[ \left( \frac{y_t - \mu_t}{\mu_t^2} \right) \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \right] = \mathbf{0}.$$

Střední hodnota posledního výrazu bude v bodě skutečného parametru  $\boldsymbol{\theta}$  vždy nulová bez ohledu na to, jestli je hustota  $\{Y_t, t \in \mathbb{Z}\}$  skutečně exponenciální, protože platí:

$$\mathbb{E}[Y_t - \mu_t] = \mathbb{E}[Y_t - \mathbb{E}(Y_t | \Psi_{t-1})] = 0.$$

Z toho vyplývá, že funkci (2) lze interpretovat jako tzv. kvazi-věrohodnost a argument jejího maxima bude kvazi-maximálně věrohodný odhad  $\boldsymbol{\theta}$ . Nyní předpokládejme, že  $\{\mu_t, t \in \mathbb{Z}\}$  je dáno rovnicí:

$$\mu_t = \omega_0 + \alpha_0 y_{t-1} + \beta_0 \mu_{t-1},$$

což odpovídá modelu GARCH řádů  $p = 1$  a  $q = 1$ . Tato situace je blíže zkoumána ve článku (Lee a Hansen, 1994), z jehož poznatků je možné formulovat následující větu.

**Věta 4.** *Nechť pro časovou řadu  $\{Y_t, t \in \mathbb{Z}\}$  platí:*

- $\mathbb{E}(Y_t | \Psi_{t-1}) \equiv \mu_{0,t} = \omega_0 + \alpha_0 y_{t-1} + \beta_0 \mu_{0,t-1}, t \in \mathbb{Z},$
- $\varepsilon_t \equiv \frac{Y_t}{\mu_{0,t}}, t \in \mathbb{Z}$  je:
  - *striktně stacionární a ergodická,*
  - *nedegenerovaná (má nenulový rozptyl),*
  - *existuje  $K < \infty$  takové, že pro všechna  $t \in \mathbb{Z} : \mathbb{E}(\varepsilon_t^2 | \Psi_{t-1}) \leq K,$*
  - *$\sup_t \mathbb{E}[\log(\beta_0 + \alpha_0 \varepsilon_t) | \Psi_{t-1}] < 0$  s.j.,*
- $\boldsymbol{\theta}_0 \equiv (\omega_0, \alpha_0, \beta_0)$  je uvnitř  $\Theta \subseteq \mathbb{R}^3,$
- $\ell_n(\boldsymbol{\theta} | \Psi_{n-1}) = -\sum_{t=1}^n \left( \log(\mu_t) + \frac{y_t}{\mu_t} \right),$  kde
  - $\mu_t = \omega + \alpha y_t + \beta \mu_{t-1}$  pro  $t > 1,$
  - $\mu_1 = \frac{\omega}{1-\beta}.$

Potom odhad  $\hat{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell_n(\boldsymbol{\theta} | \Psi_{n-1})$  je konzistentní a asymptoticky normální. Platí:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, \mathbf{V}_0),$$

kde  $\mathbf{V}_0 = \mathbf{B}_0^{-1} \mathbf{A}_0 \mathbf{B}_0^{-1},$   $\mathbf{B}_0 = -\mathbb{E} \left[ \nabla^2 \left( \log(\mu_t(\boldsymbol{\theta})) + \frac{y_t}{\mu_t} \right) \right]$  a

$$\mathbf{A}_0 = \mathbb{E} \left[ \nabla \left( \log(\mu_t(\boldsymbol{\theta})) + \frac{y_t}{\mu_t} \right) \nabla \left( \log(\mu_t(\boldsymbol{\theta})) + \frac{y_t}{\mu_t} \right)^\top \right].$$

*Důkaz.* Věta je důsledkem tvrzení prezentovaných ve článku (Lee a Hansen, 1994). Všechna tato tvrzení jsou v příloze článku podrobně dokázána. □

*Poznámka.* Vidíme, že na rozdíl od běžné specifikace modelu typu GARCH nevyžadujeme, aby  $\alpha_0 + \beta_0 < 1$ . To nám umožňuje konzistentně odhadovat parametry i v případě, že toto omezení nespĺňují.

Ve článku (Engle a Gallo, 2006) je potom bližšie popsána situace, kdy mají odchylky rozdělení gamma a model je aplikován na časovou řadu denních absolutních logaritmických výnosů z držení nějakého aktiva (např. akcií). Modely typu MEM byly v literatuře dále zobecňovány a aplikovány v mnoha dalších případech. Jedním z důležitých článků v této oblasti je (Hautsch a kol., 2014), který podrobněji řeší situaci, kdy se v datech nachází velké množství nulových pozorování (např. v případě již zmiňovaných vysokofrekvenčních dat). Německý ekonom a matematik Nikolaus Hautsch a jeho kolegové v něm navrhli přístup pomocí speciálního typu rozdělení, které má určitou pravděpodobnostní masu soustředěnu právě v nule, zatímco rozdělení na intervalu  $(0, \infty)$  je spojitě. Modely multiplikativních chyb, jejichž odchylky mají takovýto kombinovaný typ rozdělení, se nazývají **ZA-MEM**, což vychází z anglického výrazu „zero-augmented MEM“, neboli model typu MEM rozšířený v nule. Těm se budeme věnovat v následující části.

## 2.2 Modely multiplikativních chyb rozšířené v nule

Uvažujme nejprve nezápornou náhodnou veličinu  $X$  a její nezávislá pozorování  $\{X_t\}_{t=1}^n$ . Tato veličina může odpovídat například odchylkám v modelu. Abychom zachytili pravděpodobnost nulových pozorování, označíme:

$$p_0 := \mathbb{P}(X = 0), \quad 1 - p_0 := \mathbb{P}(X > 0).$$

Předpokládejme, že za podmínky  $X > 0$  má tato náhodná veličina spojitě rozdělení s hustotou

$$g_X(x) = f_X(x|X > 0), \quad x \in (0, \infty).$$

Odtud plyne, že nepodmíněné rozdělení  $X$  je nespojitě v nule a spojitě na intervalu  $(0, \infty)$  a má nepodmíněnou hustotu ve tvaru:

$$f_X(x) = p_0 \delta(x) + (1 - p_0) g_X(x) \mathbb{1}(x > 0), \quad x \geq 0,$$

kde  $0 < p_0 < 1$  a  $\delta(\cdot)$  značí Diracovu delta funkci. K pravděpodobnosti  $p_0$  přistupujeme jako k parametru rozdělení náhodné veličiny  $X$ , který určuje, jaká je pravděpodobnost nulových pozorování, zatímco  $1 - p_0$  nám říká, jak velká pravděpodobnost se přisuzuje striktně pozitivní části nosiče  $X$ . Logaritmická věrohodnost implikovaná touto nepodmíněnou hustotou lze potom zapsat jako:

$$\ell_n(\boldsymbol{\nu}) = n_0 \log(p_0) + n_+ \log(1 - p_0) + \sum_{t \in I_{n_+}} \log\{g_X(x_t, \boldsymbol{\nu}^g)\},$$

kde  $n_0$  je počet nulových pozorování,  $n_+$  je počet nenulových (pozitivních) pozorování,  $n$  je celkový počet pozorování,  $I_{n_+} \subseteq \{1, \dots, n\}$  je množina všech indexů nenulových pozorování a konečně  $\boldsymbol{\nu} = (p_0, \boldsymbol{\nu}^g)^\top$  je vektor všech odhadovaných parametrů, kde  $\boldsymbol{\nu}^g$  je vektor parametrů určujících podmíněnou hustotu  $g_X$ .

Ve článku (Hautsch a kol., 2014), ze kterého v této části čerpáme, se dále věnuje pozornost situaci, kdy podmíněná hustota  $g_X$  odpovídá tzv. zobecněnému F-rozdělení (často označováno jako GF-rozdělení). V takovém případě lze psát:

$$g_X(x) = \frac{ax^{am-1} [\eta + (x/\lambda)^a]^{-\eta-m} \eta^\eta}{\lambda^{am} \mathcal{B}(m, \eta)}, \quad x > 0,$$

kde  $a > 0$ ,  $m > 0$ ,  $\eta > 0$  a  $\lambda > 0$  jsou neznámé parametry a  $\mathcal{B}(m, \eta) = \frac{\Gamma(m)\Gamma(\eta)}{\Gamma(m+\eta)}$  je beta funkce. Z tohoto rozdělení lze odvodit například tvar podmíněných obecných momentů uvažované náhodné veličiny  $X$ :

$$\mathbb{E}[X^s | X > 0] = \frac{\lambda^s \eta^{s/a} \Gamma(m + s/a) \Gamma(\eta - s/a)}{\Gamma(m) \Gamma(\eta)}, \quad a\eta > s.$$

Náhodná veličina  $X$  má potom nepodmíněné rozdělení s hustotou:

$$f_X(x) = p_0 \delta(x) + (1 - p_0) \frac{ax^{am-1} [\eta + (x/\lambda)^a]^{-\eta-m} \eta^\eta}{\lambda^{am} \mathcal{B}(m, \eta)} \mathbb{1}(x > 0), \quad x \geq 0, \quad (3)$$

jehož obecné momenty lze snadno odvodit z tvaru hustoty a předchozího vzorce pro podmíněné momenty jako:

$$\begin{aligned} \mathbb{E}[X^s] &= p_0 \mathbb{E}[X^s | X = 0] + (1 - p_0) \mathbb{E}[X^s | X > 0] \\ &= (1 - p_0) \frac{\lambda^s \eta^{s/a} \Gamma(m + s/a) \Gamma(\eta - s/a)}{\Gamma(m) \Gamma(\eta)}, \quad a\eta > s. \end{aligned}$$

Logaritmická věrohodnost (se zachováním předchozího značení) je potom ve tvaru:

$$\begin{aligned} \ell_n(\boldsymbol{\nu}) &= n_0 \log(p_0) + n_+ \log(1 - p_0) + \sum_{t \in I_{n_+}} [\log(a) + (am - 1) \log(x_t) \\ &\quad + \eta \log(\eta) - (\eta + m) \log\{\eta + (x_t \lambda^{-1})^a\} - \log\{\mathcal{B}(m, \eta)\} - am \log(\lambda)], \end{aligned}$$

kde  $\boldsymbol{\nu} = (p_0, a, m, \eta, \lambda)^\top$ . Na základě výše popsaných principů je nyní možné definovat modely typu ZA-MEM.

**Definice 6.** *Nechť pro náhodné veličiny  $\{Y_t, t \in \mathbb{Z}\}$  platí:*

$$Y_t = \mu_t \varepsilon_t, \quad t \in \mathbb{Z},$$

kde odchylky  $\{\varepsilon_t, t \in \mathbb{Z}\}$  mají podmíněné rozdělení s hustotou  $f_x$  definovanou v (3) s  $\lambda = \frac{1}{\xi(1-p_0)}$  a  $\xi = \eta^{1/a} [\Gamma(m + 1/a) \Gamma(\eta - 1/a)] [\Gamma(m) \Gamma(\eta)]^{-1}$ , kde  $\eta > 1/a$  (takovou hustotu označíme  $f_\varepsilon$ ), píšeme

$$\mathcal{L}(\varepsilon_t | \Psi_{t-1}) = f_\varepsilon.$$

Nechť jsou dále  $\{\varepsilon_t, t \in \mathbb{Z}\}$  podmíněně nezávislé. Potom řekneme, že časová řada  $\{Y_t, t \in \mathbb{Z}\}$  je generována **modelem multiplikativních chyb rozšířeným v nule**, zkráceně **ZA-MEM**.

*Poznámka.* Podmínku  $\lambda = \frac{1}{\xi(1-p_0)}$ , kde

$$\xi = \eta^{1/a}[\Gamma(m + 1/a)\Gamma(\eta - 1/a)][\Gamma(m)\Gamma(\eta)]^{-1},$$

klademe proto, abychom zachovali vlastnost jednotkové podmíněné střední hodnoty odchylek  $\{\varepsilon_t, t \in \mathbb{Z}\}$ .

Pravděpodobnost nulových hodnot značenou jako  $p_0$  jsme dosud považovali za konstantní. Autoři však ve článku prozkoumali i situaci, kdy se tato pravděpodobnost s časem mění. Takové modely nesou zkrácený název DZA-MEM, tedy „dynamic zero-augmented MEM“ (v překladu dynamický model typu MEM rozšířený v nule).

## 2.3 Dynamické modely multiplikativních chyb rozšířené v nule

Předpokládejme, že pravděpodobnost výskytu nul se s časem mění. Označme:

$$p_{0,t} := \mathbb{P}(\varepsilon_t = 0 | \Psi_{t-1}) = \mathbb{P}(\varepsilon_t = 0 | \mathcal{F}_{t-1}) = p_0(\mathcal{F}_{t-1}; \boldsymbol{\theta}), t \in \mathbb{Z},$$

kde  $\forall t \in \mathbb{Z} : \mathcal{F}_{t-1} \subset \Psi_{t-1}$  a  $\boldsymbol{\theta}$  je vektor neznámých parametrů. Hustota  $f_x$  definovaná v (3) bude tímto předpokladem mírně pozměněna na tvar:

$$f_{x,t}(x) = p_{0,t}\delta(x) + (1 - p_{0,t}) \frac{ax^{am-1} [\eta + (x/\lambda)^a]^{-\eta-m} \eta^\eta}{\lambda_t^{am} \mathcal{B}(m, \eta)} \mathbb{1}(x > 0), x \geq 0, \quad (4)$$

Nyní si můžeme dynamické modely multiplikativních chyb rozšířené v nule formálně definovat.

**Definice 7.** *Nechť pro náhodné veličiny  $\{Y_t, t \in \mathbb{Z}\}$  platí:*

$$Y_t = \mu_t \varepsilon_t, t \in \mathbb{Z},$$

kde odchylky  $\{\varepsilon_t, t \in \mathbb{Z}\}$  mají podmíněné rozdělení s hustotou  $f_{x,t}$  definovanou v (4) s  $\lambda_t = \frac{1}{\xi(1-p_{0,t})}$  a  $\xi = \eta^{1/a}[\Gamma(m + 1/a)\Gamma(\eta - 1/a)][\Gamma(m)\Gamma(\eta)]^{-1}$ , kde  $\eta > 1/a$  (takovou hustotu označíme  $f_{\varepsilon,t}$ ), píšeme

$$\mathcal{L}(\varepsilon_t | \mathcal{F}_{t-1}) = f_{\varepsilon,t}.$$

*Nechť jsou dále  $\{\varepsilon_t, t \in \mathbb{Z}\}$  podmíněně nezávislé. Potom řekneme, že časová řada  $\{Y_t, t \in \mathbb{Z}\}$  je generována **dynamickým modelem multiplikativních chyb rozšířeným v nule**, zkráceně **DZA-MEM**.*

*Poznámka.* Podmínku  $\lambda_t = \frac{1}{\xi(1-p_{0,t})}$ , kde

$$\xi = \eta^{1/a}[\Gamma(m + 1/a)\Gamma(\eta - 1/a)][\Gamma(m)\Gamma(\eta)]^{-1},$$

stejně jako v případě modelů ZA-MEM klademe proto, abychom zachovali vlastnost jednotkové podmíněné střední hodnoty odchylek  $\{\varepsilon_t, t \in \mathbb{Z}\}$ .



Zavedeme značení  $\mathcal{I}_t = \mathbb{1}(Y_t > 0)$ ,  $t \in \mathbb{Z}$ , s jehož pomocí již můžeme zapsat logaritmickou věrohodnost ve tvaru:

$$\begin{aligned} \ell_n(\boldsymbol{\nu}) &= \sum_{t=1}^n \{\mathcal{I}_t \log(1 - p_{0,t}) + (1 - \mathcal{I}_t) \log(p_{0,t})\} + \sum_{t \in I_{n+}} [\log(a) + (am - 1) \log(y_t)] \\ &\quad + \eta \log(\eta) - (\eta + m) \log\left\{\eta + \left(\frac{y_t}{\mu_t} \lambda_t^{-1}\right)^a\right\} - \log\{\mathcal{B}(m, \eta)\} - am \log(\lambda_t), \end{aligned}$$

kde  $\boldsymbol{\nu} = (\boldsymbol{\theta}^\top, a, m, \eta, \boldsymbol{\nu}_\mu^\top)^\top$  je vektor neznámých parametrů a  $\boldsymbol{\nu}_\mu$  značí vektor neznámých parametrů v rovnici pro střední hodnotu  $\mu_t$ .

## 2.4 Vektorové modely multiplikativních chyb

Nyní se ještě vrátíme k modelům multiplikativních chyb ve vícerozměrném případě. Tyto vektorové modely zkráceně označované jako vMEM jsme si zatím pouze definovali v definici 5, avšak nyní se blíže podíváme na možné způsoby jejich specifikace. Čerpat budeme především ze článku (Cipollini a kol., 2006).

### 2.4.1 Vícerozměrné gamma rozdělení

Abychom mohli s modelem vMEM pracovat, musíme si nejprve zvolit podmíněné rozdělení vektorů  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ . První možnou volbou je zobecnění rozdělení gamma na vícerozměrný případ, jež se nazývá GammaCR podle příjmení autorů Cheriyaana a Ramabhadrana. Předpokládáme tedy:

$$\boldsymbol{\varepsilon}_t | \Psi_{t-1} \sim \text{GammaCR}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\phi}), t \in \mathbb{Z},$$

kde  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)^\top$  a  $0 < \phi_0 < \min(\phi_1, \dots, \phi_K)$ . Hustota takového rozdělení má poměrně složitý tvar obsahující integrál, avšak klíčové jsou jeho vlastnosti. Pro  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{K,t})^\top$  totiž máme:

$$\varepsilon_{i,t} | \Psi_{t-1} \sim \text{Gamma}(\phi_i, \phi_i), i = 1, \dots, K, t \in \mathbb{Z},$$

a pro prvky kovarianční a korelační matice vektorů  $\boldsymbol{\varepsilon}_t$  platí:

$$\text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t} | \Psi_{t-1}) = \frac{\phi_0}{\phi_i \phi_j}$$

a

$$\rho(\varepsilon_{i,t}, \varepsilon_{j,t} | \Psi_{t-1}) = \frac{\phi_0}{\sqrt{\phi_i \phi_j}}, i, j = 1, \dots, K, t \in \mathbb{Z}.$$

Z toho vyplývá, že korelace mezi komponenty rozdělení GammaCR jsou vždy pozitivní. Druhý možný přístup ke specifikaci rozdělení odchylek je použití tzv. kopul.

## 2.4.2 Specifikace pomocí kopul

Kopuly jsou často používaným nástrojem k popisu závislostní struktury ve vícerozměrném rozdělení. My si kopuly definujeme a uvedeme si některé jejich základní vlastnosti. Čerpat budeme z publikace (Nelsen, 2006).

**Definice 8.** *Distribuční funkce  $C$   $K$ -rozměrného náhodného vektoru, jehož jednorozměrná marginální rozdělení jsou rovnoměrná na intervalu  $(0,1)$ , se nazývá  **$K$ -rozměrná kopula**. Píšeme:*

$$C(u_1, \dots, u_K) = \mathbb{P}(U_1 \leq u_1, \dots, U_K \leq u_K), \quad 0 \leq u_i \leq 1, \quad i = 1, \dots, K,$$

kde

$$U_i \sim R(0,1), \quad i = 1, \dots, K.$$

Důležitou vlastnost kopul popisuje tzv. Sklarova věta.

**Věta 5.** (Sklarova) *Nechť  $F$  je  $K$ -rozměrná distribuční funkce se spojitými jednorozměrnými marginálními distribučními funkcemi  $F_1, \dots, F_K$ . Potom existuje jednoznačně určená  $K$ -rozměrná kopula  $C$  taková, že pro všechna  $\mathbf{x} \in \mathbb{R}^K$ :*

$$F(x_1, \dots, x_K) = C(F_1(x_1), \dots, F_K(x_K)). \quad (5)$$

Naopak, jestliže  $C$  je  $K$ -rozměrná kopula a  $F_1, \dots, F_K$  jsou jednorozměrné distribuční funkce, pak funkce  $F$  definovaná rovnicí (5) je  $K$ -rozměrná distribuční funkce s marginálami  $F_1, \dots, F_K$ .

*Důkaz.* Věta je dokázána v knize (Nelsen, 2006) na straně 21. □

*Poznámka.* Má-li náhodný vektor  $\mathbf{X} = (X_1, \dots, X_K)^\top$  sdruženou distribuční funkci  $F$  se spojitými jednorozměrnými marginálami  $F_1, \dots, F_K$ , lze psát:

$$\begin{aligned} F(x_1, \dots, x_K) &= \mathbb{P}(X_1 \leq x_1, \dots, X_K \leq x_K) \\ &= \mathbb{P}(F_1(X_1) \leq F_1(x_1), \dots, F_K(X_K) \leq F_K(x_K)). \end{aligned}$$

Protože náhodné veličiny  $F_i(X_i)$ ,  $i = 1, \dots, K$  mají za předpokladu spojitosti jednorozměrných marginálů rozdělení  $R(0,1)$ , vidíme, že hledanou kopulou je distribuční funkce vektoru  $(F_1(X_1), \dots, F_K(X_K))^\top$ . Na základě tohoto poznatku lze formulovat následující definici.

**Definice 9.** *Nechť má náhodný vektor  $\mathbf{X} = (X_1, \dots, X_K)^\top$  sdruženou distribuční funkci  $F$  se spojitými jednorozměrnými marginálními distribučními funkcemi  $F_1, \dots, F_K$ . Potom **kopulou vektoru  $\mathbf{X}$**  (nebo **kopulou distribuční funkce  $F$** ) nazveme sdruženou distribuční funkci náhodného vektoru  $(F_1(X_1), \dots, F_K(X_K))^\top$ .*

*Poznámka.* Lze snadno dokázat, že je-li  $C$  kopulou náhodného vektoru  $\mathbf{X} = (X_1, \dots, X_K)^\top$  a  $G_1, \dots, G_K$  jsou rostoucí funkce, potom je  $C$  rovněž kopulou náhodného vektoru  $(G_1(X_1), \dots, G_K(X_K))^\top$ .

Kopul existuje nepřeberné množství, avšak nás bude zajímat tzv. normální (Gaussovská) kopula.

**Definice 10.** *Kopulu náhodného vektoru s vícerozměrným normálním rozdělením nazýváme **normální (Gaussovská) kopula**.*

*Poznámka.* Nechť má náhodný vektor  $\mathbf{X} = (X_1, \dots, X_K)^\top$   $K$ -rozměrné normální rozdělení  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Potom náhodný vektor  $\mathbf{Y} = (Y_1, \dots, Y_K)^\top$ , kde

$$Y_i = \frac{X_i - \mathbb{E}X_i}{\sqrt{\text{var}(X_i)}}, \quad i = 1, \dots, K \quad (6)$$

má vícerozměrné normované normální rozdělení se střední hodnotou v podobě nulového vektoru a korelační maticí  $\mathbf{R}$ , jež má na místě  $(i, j)$  prvek  $\rho(Y_i, Y_j)$ ,  $i, j = 1, \dots, K$ . Protože transformace (6) je rostoucí, má dle poznámky pod definicí 9 náhodný vektor  $\mathbf{Y}$  stejnou kopulu jako náhodný vektor  $\mathbf{X}$ . Z toho plyne, že kopula náhodného vektoru s vícerozměrným normálním rozdělením je jednoznačně určena pouze korelační maticí  $\mathbf{R}$ .

Nyní se vrátíme k modelům vMEM a využijeme kopuly k definici podmíněného rozdělení vektorů  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ . Budeme pracovat s následujícím předpokladem:

$$\boldsymbol{\varepsilon}_t | \Psi_{t-1} \sim N(\mathbf{R}) - \prod_{i=1}^K \text{Gamma}(\phi_i, \phi_i), \quad t \in \mathbb{Z}. \quad (7)$$

To znamená, že závislostní struktura je určena Gaussovskou kopulou s korelační maticí  $\mathbf{R}$ , zatímco pro marginální rozdělení platí  $\varepsilon_{i,t} | \Psi_{t-1} \sim \text{Gamma}(\phi_i, \phi_i)$ ,  $i = 1, \dots, K$ ,  $t \in \mathbb{Z}$ . Nejen vlastnosti tohoto rozdělení byly důkladně prozkoumány ve článku (Xue-Kun Song, 2000), z jehož poznatků vyplývají následující aproximace:

$$\text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t} | \Psi_{t-1}) \simeq \frac{R_{i,j}}{\sqrt{\phi_i \phi_j}}$$

a

$$\rho(\varepsilon_{i,t}, \varepsilon_{j,t} | \Psi_{t-1}) \simeq R_{i,j}, \quad i, j = 1, \dots, K, \quad t \in \mathbb{Z}.$$

Kromě rozdělení odchylek je potřeba zvolit také rovnici pro vektor střední hodnoty  $\boldsymbol{\mu}_t$ . K tomu se nejprve vrátíme k případu jednorozměrného modelu typu MEM, který poté rozšíříme na více rozměrů. Jednoduchá rovnice pro střední hodnotu by mohla mít následující tvar:

$$\mu_t = \omega + \alpha y_{t-1} + \beta \mu_{t-1}, \quad t \in \mathbb{Z}.$$

Takový model však nereflektuje dříve zmiňovaný pákový efekt, a proto budeme volit raději přístup s asymetrickým vlivem kladných a záporných výnosů. Rovnice se přepíše do tvaru:

$$\mu_t = \omega^* + \alpha \left( y_{t-1}^{1/2} \text{sign}(r_{t-1}) + \delta^* \right)^2 + \gamma y_{t-1} \mathbb{1}(r_{t-1} < 0) + \beta \mu_{t-1}, \quad t \in \mathbb{Z},$$

kde parametry  $\delta^*$  a  $\gamma$  zachycují pákový efekt a  $r_t$ ,  $t \in \mathbb{Z}$  chápeme jako výnosy (mohou to být např. denní logaritmické výnosy z držení nějakého aktiva). Spočteme-li druhou mocninu závorky v přechodí rovnici, dostaneme:

$$\mu_t = \omega + \alpha y_{t-1} + \gamma y_{t-1}^{(-)} + \delta y_{t-1}^{(s)} + \beta \mu_{t-1}, \quad t \in \mathbb{Z},$$

kde  $y_{t-1}^{(s)} = y_{t-1}^{1/2} \text{sign}(r_{t-1})$ ,  $y_{t-1}^{(-)} = y_{t-1} \mathbb{1}(r_{t-1} < 0)$ ,  $\omega = \omega^* + \alpha \delta^{*2}$  a  $\delta = 2\alpha \delta^*$ . Po rozšíření na vícerozměrný případ získáme pro vektor střední hodnoty rovnici:

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha} \mathbf{y}_{t-1} + \boldsymbol{\gamma} \mathbf{y}_{t-1}^{(-)} + \boldsymbol{\delta} \mathbf{y}_{t-1}^{(s)} + \boldsymbol{\beta} \boldsymbol{\mu}_{t-1}, \quad t \in \mathbb{Z}, \quad (8)$$

kde  $\mathbf{y}_{t-1}^{(s)} = \mathbf{y}_{t-1}^{1/2} \text{sign}(r_{t-1})$ ,  $\mathbf{y}_{t-1}^{1/2} = (y_{1,t-1}^{1/2}, \dots, y_{K,t-1}^{1/2})^\top$ ,  $\mathbf{y}_{t-1}^{(-)} = \mathbf{y}_{t-1} \mathbb{1}(r_{t-1} < 0)$ , dimenze parametru  $\boldsymbol{\omega}$  je  $K \times 1$  a dimenze parametrů  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\delta}$ ,  $\boldsymbol{\gamma}$  a  $\boldsymbol{\beta}$  je  $K \times K$ . V následující větě jsou popsány podmínky pro nezápornost všech komponentů  $\boldsymbol{\mu}_t$  v každém čase.

**Věta 6.** *Nechť pro parametry v rovnici (8) platí pro všechna  $i, j = 1, \dots, K$  následující podmínky:*

1.  $\beta_{i,j} \geq 0$ ,  $\alpha_{i,j} \geq 0$  a  $\alpha_{i,j} + \gamma_{i,j} \geq 0$ .
2. Je-li  $\alpha_{i,j} = 0$ , pak  $\delta_{i,j} \geq 0$ .
3. Je-li  $\alpha_{i,j} + \gamma_{i,j} = 0$ , pak  $\delta_{i,j} \leq 0$ .
4.  $\omega_i - \frac{1}{4} \sum_{j=1}^K \delta_{i,j}^2 \left[ \frac{\mathbb{1}(\delta_{i,j} < 0) \mathbb{1}(\alpha_{i,j} > 0)}{\alpha_{i,j}} + \frac{\mathbb{1}(\delta_{i,j} > 0) \mathbb{1}(\alpha_{i,j} + \gamma_{i,j} > 0)}{\alpha_{i,j} + \gamma_{i,j}} \right] \geq 0$ .

Potom pro všechna  $t \in \mathbb{Z}$  a  $i = 1, \dots, K$ :  $\mu_{i,t} \geq 0$ .

*Důkaz.* Věta je dokázána v příloze E článku (Cipollini a kol., 2006). □

Specifikovali jsme si tedy podmíněné rozdělení odchylek i tvar rovnice střední hodnoty a můžeme přejít k inferenci metodou maximální věrohodnosti. Podle předpokladu (7) má  $\boldsymbol{\varepsilon}_t$  podmíněnou hustotu:

$$f(\boldsymbol{\varepsilon}_t | \Psi_{t-1}) = c(F_1(\varepsilon_{1,t}), \dots, F_K(\varepsilon_{K,t})) \prod_{i=1}^K f_i(\varepsilon_{i,t}),$$

kde  $f_i$  a  $F_i$  jsou podmíněná hustota a distribuční funkce  $i$ -té složky vektoru  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{K,t})^\top$  pro  $i = 1, \dots, K$  a  $c$  je hustota odvozená ze zvolené kopuly  $C$ . Tedy v případě normální kopuly a marginálních rozdělení *Gamma*  $(\phi_i, \phi_i)$  dostáváme:

$$f_i(\varepsilon_{i,t}) = \frac{\phi_i^{\phi_i}}{\Gamma(\phi_i)} \varepsilon_{i,t}^{\phi_i-1} \exp(-\phi_i \varepsilon_{i,t}),$$

$$F_i(\varepsilon_{i,t}) = \Gamma(\phi_i, \phi_i \varepsilon_{i,t}),$$

a

$$c(\boldsymbol{\varepsilon}_t) = |\mathbf{R}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{q}_t^\top (\mathbf{R}^{-1} - \mathbf{I}) \mathbf{q}_t\right], \quad t \in \mathbb{Z}, \quad i = 1, \dots, K,$$

kde  $\Gamma$  značí tzv. neúplnou gamma funkci,

$$\mathbf{q}_t = \left( \Phi^{-1}(F_1(\varepsilon_{1,t})), \dots, \Phi^{-1}(F_K(\varepsilon_{K,t})) \right), \quad t \in \mathbb{Z}$$

a  $\Phi^{-1}$  značí kvantilovou funkci normovaného normálního rozdělení. Podmíněné rozdělení  $\mathbf{y}_t$  má proto hustotu:

$$f(\mathbf{y}_t | \Psi_{t-1}) = c(F_1(y_{1,t}/\mu_{1,t}), \dots, F_K(y_{K,t}/\mu_{K,t})) \prod_{i=1}^K \frac{f_i(y_{i,t}/\mu_{i,t})}{\mu_{i,t}},$$

kde

$$\frac{f_i(y_{i,t}/\mu_{i,t})}{\mu_{i,t}}$$

je hustota rozdělení *Gamma*  $(\phi_i, \phi_i/\mu_{i,t})$ ,  $i = 1, \dots, K$ . Nyní již není problém odvodit tvar podmíněné logaritmické věrohodnosti. Máme:

$$\begin{aligned} \ell_n = \sum_{t=1}^n & [\log c(F_1(y_{1,t}/\mu_{1,t}), \dots, F_K(y_{K,t}/\mu_{K,t})) \\ & + \sum_{i=1}^K (\log(f_i(y_{i,t}/\mu_{i,t})) - \log(\mu_{i,t}))]. \end{aligned}$$

Vzhledem k tomu, že tvar funkcí  $c$  a  $f_i$ ,  $i = 1, \dots, K$  známe, lze podmíněnou logaritmickou věrohodnost rozepsat jako:

$$\begin{aligned} \ell_n = \sum_{t=1}^n & \left[ \frac{1}{2} \log |\mathbf{R}^{-1}| - \frac{1}{2} \mathbf{q}_t^\top \mathbf{R}^{-1} \mathbf{q}_t + \frac{1}{2} \mathbf{q}_t^\top \mathbf{q}_t \right. \\ & \left. + \sum_{i=1}^K \left( \phi_i \log \phi_i - \ln \Gamma(\phi_i) - \ln y_{i,t} + \phi_i \left( \ln y_{i,t} - \log \mu_{i,t} - \frac{y_{i,t}}{\mu_{i,t}} \right) \right) \right]. \end{aligned}$$

Můžeme si všimnout, že každý sčítanec v podmíněné logaritmické věrohodnosti lze rozdělit na příspěvek kopuly

$$\frac{1}{2} \log |\mathbf{R}^{-1}| - \frac{1}{2} \mathbf{q}_t^\top \mathbf{R}^{-1} \mathbf{q}_t + \frac{1}{2} \mathbf{q}_t^\top \mathbf{q}_t$$

a příspěvek marginálních hustot

$$\sum_{i=1}^K \left( \phi_i \log \phi_i - \ln \Gamma(\phi_i) - \ln y_{i,t} + \phi_i \left( \ln y_{i,t} - \log \mu_{i,t} - \frac{y_{i,t}}{\mu_{i,t}} \right) \right).$$

Zatímco příspěvek marginálních hustot závisí pouze na parametrech rovnice střední hodnoty a na parametru  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)^\top$ , příspěvek kopuly závisí také na korelační matici  $\mathbf{R}$ . Pro derivaci dle korelační matice proto platí:

$$\frac{\partial \ell_n}{\partial \mathbf{R}^{-1}} = \frac{1}{2} (n\mathbf{R} - \mathbf{q}^\top \mathbf{q}),$$

kde jsme označili  $\mathbf{q} = (\mathbf{q}_1^\top, \dots, \mathbf{q}_n^\top)$  matici dimenzí  $n \times K$ . Odtud lze odvodit explicitní formu maximálně věrohodného odhadu korelační matice  $\mathbf{R}$  jako:

$$\widehat{\mathbf{R}} = \frac{1}{n} \mathbf{q}^\top \mathbf{q}.$$

Díky tomuto vyjádření můžeme v původní podmíněné logaritmické věrohodnosti nahradit korelační matici  $\mathbf{R}$  jejím odhadem  $\widehat{\mathbf{R}}$  a získáme tzv. koncentrovanou logaritmickou věrohodnost, kterou označíme jako  $\ell_n^c$ :

$$\begin{aligned} \ell_n^c = -\frac{n}{2} \log |\widehat{\mathbf{R}}| - \frac{1}{2} \sum_{t=1}^n \mathbf{q}_t^\top (\widehat{\mathbf{R}}^{-1} - \mathbf{I}) \mathbf{q}_t \\ + \sum_{t=1}^n \sum_{i=1}^K \left( \phi_i \log \phi_i - \log \Gamma(\phi_i) - \log y_{i,t} + \phi_i \left( \ln y_{i,t} - \log \mu_{i,t} - \frac{y_{i,t}}{\mu_{i,t}} \right) \right). \end{aligned}$$

Dosud jsme však na odhad matice  $\mathbf{R}$  neaplikovali žádná omezení související s faktem, že se jedná o korelační matici. Potřebujeme tedy, aby byla pozitivně definitní a měla jednotkovou diagonálu. Místo původně uvažovaného odhadu  $\widehat{\mathbf{R}}$  proto zavedeme nový odhad

$$\tilde{\mathbf{R}} = \mathbf{D}_Q^{-\frac{1}{2}} \mathbf{Q} \mathbf{D}_Q^{-\frac{1}{2}},$$

kde

$$\mathbf{Q} = \frac{\mathbf{q}^\top \mathbf{q}}{n} \quad \mathbf{D}_Q = \text{diag}(Q_{1,1}, \dots, Q_{K,K}).$$

Za těchto okolností se tvar koncentrované logaritmické věrohodnosti zjednoduší na:

$$\begin{aligned} \ell_n^c &= -\frac{n}{2} \log |\tilde{\mathbf{R}}| \\ &+ \sum_{t=1}^n \sum_{i=1}^K \left( \phi_i \log \phi_i - \log \Gamma(\phi_i) - \log y_{i,t} + \phi_i \left( \ln y_{i,t} - \log \mu_{i,t} - \frac{y_{i,t}}{\mu_{i,t}} \right) \right). \end{aligned}$$

## 2.5 Semiparametrické modely multiplikativních chyb

Nyní se budeme věnovat dalšímu zajímavému rozšíření modelů MEM, které popsali v přehledovém článku (Cipollini a Gallo, 2022) italsí matematici a ekonomové Fabrizio Cipollini a Giampiero M. Gallo. Rovnice modelu multiplikativních chyb definovaného v definici 4 se rozepíše na:

$$Y_t = \mu_t \varepsilon_t = \mu \tau_t \xi_t \varepsilon_t, \quad t \in \mathbb{Z}, \quad (9)$$

kde

- $\mu = \mathbb{E}(Y_t)$  je (konstantní) nepodmíněná střední hodnota  $\{Y_t, t \in \mathbb{Z}\}$ ,
- $\{\tau_t, t \in \mathbb{Z}\}$  je tzv. nízkofrekvenční složka (anglicky „low-frequency component“) zachycující systematické a pomalu se vyvíjející změny (např. dlouhodobý hospodářský růst v případě ekonomických dat),
- $\{\xi_t, t \in \mathbb{Z}\}$  je krátkodobá složka (anglicky „short-run component“) představující dočasné a nárazové změny, jež nejsou součástí dlouhodobých trendů, obvykle parametrizovaná podobným způsobem jako v procesu typu GARCH a
- $\{\varepsilon_t, t \in \mathbb{Z}\}$  stejně jako v původním modelu značí odchylky či poruchy. Předpokládáme, že jsou podmíněně nezávislé a stejně rozdělené s rozdělením s nezáporným nosičem s jednotkovou střední hodnotou a kladným nenulovým rozptylem, tedy lze psát  $\varepsilon_t | \Psi_{t-1} \stackrel{\text{i.i.d.}}{\sim} d^+(1, \sigma^2)$ .

Díky předpokladu jednotkové podmíněné střední hodnoty odchylek platí v každém  $t \in \mathbb{Z}$ :

$$\mathbb{E}(Y_t | \Psi_{t-1}) = \mu \tau_t \xi_t = \mu_t.$$

Na nepodmíněné střední hodnoty nízkofrekvenční složky a krátkodobé složky dále aplikujeme pro všechna  $t \in \mathbb{Z}$  předpoklad  $\mathbb{E}(\tau_t) = \mathbb{E}(\xi_t) = 1$ , čehož lze vždy docílit vhodnou normalizací. Pokud bychom nízkofrekvenční složku  $\tau_t$  položili ve všech časech rovnu jednotce, dostali bychom původní model typu MEM jako v definici 4. My však zvolíme neparametrickou specifikaci a na  $\tau_t$  budeme klást jediný požadavek, a to aby šlo o hladkou funkci času. Z tohoto důvodu se výsledný model nazývá **semiparametrický model multiplikativních chyb** (anglicky „semi-parametric MEM“) a zkracujeme jej jako **SpMEM**.

K odhadu  $\{\tau_t, t \in \mathbb{Z}\}$  je nyní potřeba využít tzv. jádrové odhady. Volíme například:

$$\hat{\tau}_t = \frac{\sum_{s=1}^n y_s^{(\tau)} K\left(\frac{z_t - z_s}{h}\right)}{\sum_{s=1}^n K\left(\frac{z_t - z_s}{h}\right)}, \quad (10)$$

kde

$$y_t^{(\tau)} = \frac{y_t}{\hat{\mu} \hat{\xi}_t},$$

$\hat{\mu}$  a  $\hat{\xi}_t$  jsou vhodné odhady nepodmíněné střední hodnoty a krátkodobé složky,  $z_t = \frac{t}{n}$ ,  $K$  je jádrová funkce a  $h > 0$  je příslušná šířka pásma. Jádrová funkce je nezáporná, spojitá, symetrická a její integrál přes množinu  $\mathbb{R}$  je roven 1.

Krátkodobý komponent  $\xi_t$  se nejčastěji parametrizuje podobným způsobem jako v modelu typu GARCH, který jsme si zavedli v definici 2. Lze psát:

$$\xi_t = \left[1 - \left(\beta_1 + \alpha_1 + \frac{\gamma_1}{2}\right)\right] + \beta_1 \xi_{t-1} + \alpha_1 Y_{t-1}^{(\xi)} + \gamma_1 Y_{t-1}^{(\xi-)}, \quad t \in \mathbb{Z}, \quad (11)$$

kde

$$Y_t^{(\xi)} = \frac{Y_t}{\mu \tau_t}, \quad Y_t^{(\xi-)} = Y_t^{(\xi)} D_t^-,$$

$D_t^- = \mathbb{I}(R_t < 0)$  a náhodnou veličinu  $R_t$  chápeme jako výnos, který do modelu vnáší dodatečnou informaci. Může jít například o klasický finanční (logaritmický) výnos z držení aktiva. Lze si všimnout, že poslední sčítanec se v rovnici nachází kvůli zachycení pákového efektu, podobně jako tomu bylo v modelu typu EGARCH zavedenému v definici 3.

Nyní se budeme věnovat statistické inferenci ve výše popsaném modelu. Nepodmíněnou střední hodnotu budeme vždy odhadovat výběrovým průměrem dosud naměřených dat, tedy  $\hat{\mu} = \frac{1}{n} \sum_{t=1}^n y_t$ . Pro odhad ostatních komponentů nejprve položíme všechna  $\hat{\xi}_t$  rovna jednotce a poté budeme opakovat následující dva kroky:

1. Odhadneme  $\{\tau_t, t \in \mathbb{Z}\}$  s využitím jádrového odhadu představeného v rovnici (10) za pomoci naměřených dat  $\{y_t^{(\tau)} = \frac{y_t}{\hat{\mu} \hat{\xi}_t}, t = 1, \dots, n\}$ .
2. Odhadneme  $\{\xi_t, t \in \mathbb{Z}\}$  a  $\sigma^2 = \text{var}(\varepsilon_t | \Psi_t)$  za pomoci naměřených dat  $\{y_t^{(\xi)} = \frac{y_t}{\hat{\mu} \tau_t}, t = 1, \dots, n\}$  jedním ze způsobů popsaných v sekcích níže. Předpokládáme přitom, že  $\xi_t$  jsou specifikována rovnicí (11) a označíme  $\theta = (\alpha_1, \beta_1, \gamma_1)^\top$  vektor parametrů, které do rovnice vstupují.

S iteracemi skončíme při dosažení vhodného konvergenčního kritéria. Vidíme, že v případě základního modelu typu MEM z definice 4 není první krok nutné provádět.

### 2.5.1 Zobecněná momentová metoda

Výhodou zobecněné momentové metody je, že ji lze použít i bez explicitní specifikace podmíněného rozdělení odchylek. Máme:

$$\varepsilon_t = \frac{Y_t^{(\xi)}}{\xi_t}, \quad (12)$$

kde  $Y_t^{(\xi)} = \frac{Y_t}{\mu \tau_t}$ . Za platnosti předpokladů modelu je výraz  $\varepsilon_t - 1$  podmíněně homoskedastický s nulovou podmíněnou střední hodnotou a s podmíněným rozptylem  $\sigma^2$ . Dle poznatků ze článku (Brownless a kol., 2012) řeší nejlepší nestranný odhad vektorového parametru  $\boldsymbol{\theta} = (\alpha_1, \beta_1, \gamma_1)^T$  zobecněnou momentovou metodou (označíme jej  $\hat{\boldsymbol{\theta}}_M$ ) soustavu rovnic:

$$\sum_{t=1}^n (\varepsilon_t - 1) \mathbf{a}_t = \mathbf{0},$$

kde

$$\mathbf{a}_t = \frac{1}{\xi_t} \nabla_{\boldsymbol{\theta}} \xi_t \quad (13)$$

a  $\mathbf{0}$  značí nulový vektor. Asymptotická rozptylová matice tohoto odhadu je potom

$$\text{Avar}(\hat{\boldsymbol{\theta}}_M) = \lim_{n \rightarrow \infty} \text{var}(\hat{\boldsymbol{\theta}}_M) = \sigma^2 \mathbf{A}^{-1},$$

kde

$$\mathbf{A} = \lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\mathbf{a}_t \mathbf{a}_t^\top) \right].$$

Konzistentní odhad rozptylové matice odhadu je tedy

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_M) = \hat{\sigma}^2 \widehat{\mathbf{A}}^{-1}, \quad (14)$$

kde

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (\hat{\varepsilon}_t - 1)^2 \quad (15)$$

je momentový odhad  $\sigma^2$ ,

$$\widehat{\mathbf{A}} = \frac{1}{n} \sum_{t=1}^n \hat{\mathbf{a}}_t \hat{\mathbf{a}}_t^\top$$

a  $\hat{\varepsilon}_t$  a  $\hat{\mathbf{a}}_t$  odpovídají rovnicím (12) a (13) vyhodnoceným v bodě  $\hat{\boldsymbol{\theta}}_M$ . Abychom mohli popsanou proceduru provést v praxi, potřebujeme ještě v rovnici pro  $Y_t^{(\xi)}$  nahradit neznámé parametry  $\mu$  a  $\tau_t$  jejich odhady  $\hat{\mu}$  a  $\hat{\tau}_t$ .



## 2.5.2 Metoda kvazi-maximální věrohodnosti

Vrátíme-li se k předpokladu ze článku (Engle a Gallo, 2006), že odchylky  $\{\varepsilon_t, t \in \mathbb{Z}\}$  mají rozdělení gamma, můžeme přistoupit k odhadu  $\boldsymbol{\theta}$  metodou kvazi-maximální věrohodnosti. Konkrétně předpokládáme  $\varepsilon_t | \Psi_{t-1} \sim \text{Gamma}(\phi, \phi)$ ,  $t \in \mathbb{Z}$ , z čehož plyne pro všechna  $t \in \mathbb{Z}$ :  $\mathbb{E}(\varepsilon_t | \Psi_{t-1}) = 1$  a  $\text{var}(\varepsilon_t | \Psi_{t-1}) = \sigma^2 = 1/\phi$ . Podobně jako v sekci 2.4.2 lze nyní odvodit tvar podmíněné logaritmické věrohodnosti jako:

$$\ell_n = \sum_{t=1}^n [\phi \log(\phi) - \log \Gamma(\phi) + \phi \log(\varepsilon_t) - \phi \varepsilon_t - \log(y_t^{(\xi)})].$$

Z toho je vidět, že pro maximalizaci v  $\boldsymbol{\theta}$  stačí maximalizovat výraz

$$\sum_{t=1}^n (\log(\varepsilon_t) - \varepsilon_t),$$

který na  $\phi$  nezávisí. Protože platí (12), lze pro derivaci dle  $\boldsymbol{\theta}$  psát podmínka:

$$\sum_{t=1}^n \frac{\partial \xi_t y_t^{(\xi)} - \xi_t}{\partial \boldsymbol{\theta} \xi_t^2} = \mathbf{0}.$$

Analogicky k sekci 2.1 si lze všimnout, že střední hodnota levé strany rovnice bude vždy nulová, i kdybychom špatně specifikovali rozdělení odchylek.

## 2.6 Vektorové semiparametrické modely multiplikativních chyb

Podobným způsobem jako v jednorozměrném případě nyní rozšíříme vektorové modely multiplikativních chyb. Čerpat budeme stále z přehledového článku (Cipollini a Gallo, 2022) a rovnici modelu vMEM zavedeného v definici 5 přepíšeme do tvaru:

$$\mathbf{Y}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \tau_t \boldsymbol{\mu} \odot \boldsymbol{\xi}_t \odot \boldsymbol{\varepsilon}_t, t \in \mathbb{Z},$$

kde

- $\boldsymbol{\mu} = \mathbb{E}(Y_t)$  je  $K$ -rozměrný vektor nepodmíněných středních hodnot  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ ,
- $\{\tau_t, t \in \mathbb{Z}\}$  jsou skaláry představující nízkofrekvenční složku, podobně jako v sekci 2.5 jde o hladkou funkci času,
- $\{\boldsymbol{\xi}_t, t \in \mathbb{Z}\}$  jsou  $K$ -rozměrné vektory vzájemně závislých krátkodobých složek s podobným významem jako v jednorozměrném případě a
- $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  značí  $K$ -rozměrné vektory poruch či odchylek, u kterých znovu předpokládáme, že jsou podmíněně nezávislé a stejně rozdělené, rozdělení každého  $\varepsilon_{t,i}$ ,  $t \in \mathbb{Z}$ ,  $i = 1, \dots, K$  má nezáporný nosič a jednotkovou střední hodnotu a platí  $\text{var}(\boldsymbol{\varepsilon}_t | \Psi_{t-1}) = \boldsymbol{\Sigma}$ , kde  $\boldsymbol{\Sigma}$  je pozitivně definitní matice. Lze tedy zkráceně zapisovat  $\boldsymbol{\varepsilon}_t | \Psi_{t-1} \stackrel{\text{i.i.d.}}{\sim} d^+(\mathbf{1}, \boldsymbol{\Sigma})$ .

Krátkodobé složky  $\{\boldsymbol{\xi}_t, t \in \mathbb{Z}\}$  splňují v každém  $t \in \mathbb{Z}$  předpoklad  $\mathbb{E}(\boldsymbol{\xi}_t) = \mathbf{1}$  a jsou specifikovány analogicky s rovnicí (11), tedy:

$$\boldsymbol{\xi}_t = \left[ \mathbf{1} - \left( \boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1 + \frac{\boldsymbol{\gamma}_1}{2} \right) \mathbf{1} \right] + \boldsymbol{\beta}_1 \boldsymbol{\xi}_{t-1} + \boldsymbol{\alpha}_1 \mathbf{Y}_{t-1}^{(\xi)} + \boldsymbol{\gamma}_1 \mathbf{Y}_{t-1}^{(\xi^-)}, t \in \mathbb{Z}, \quad (16)$$

kde

$$Y_{t,i}^{(\xi)} = \frac{Y_{t,i}}{\mu_i \tau_t}, \quad Y_{t,i}^{(\xi^-)} = Y_{t,i}^{(\xi)} \mathbb{1}(R_t < 0), \quad i = 1, \dots, K$$

a  $\boldsymbol{\beta}_1$ ,  $\boldsymbol{\alpha}_1$  a  $\boldsymbol{\gamma}_1$  jsou matice rozměrů  $K \times K$ . Analogicky s jednorozměrným případem také platí pro všechna  $t \in \mathbb{Z}$  rovnice

$$\mathbb{E}(\mathbf{Y}_t | \Psi_{t-1}) = \tau_t \boldsymbol{\mu} \odot \boldsymbol{\xi}_t = \boldsymbol{\mu}_t.$$

Výše popsanému modelu říkáme vektorový semiparametrický model multiplikačních chyb a zkracujeme jej jako **SpvMEM**.

Model tedy máme dostatečně specifikovaný a můžeme se přesunout ke statistické inferenci. Budeme postupovat velmi podobně jako v sekci 2.5. Nejprve odhadneme  $\boldsymbol{\mu}$  jako  $\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{t=1}^n \mathbf{y}_t$  a všechna  $\hat{\boldsymbol{\xi}}_t$  položíme rovna jednotkovému vektoru a  $\hat{\boldsymbol{\Sigma}}$  rovnu jednotkové matici. Poté opakujeme následující dva kroky.

1. Odhadneme  $\{\tau_t, t \in \mathbb{Z}\}$  jádrovým odhadem

$$\hat{\tau}_t = \frac{\sum_{s=1}^n \mathbf{y}_s^{(\tau)} K\left(\frac{z_t - z_s}{h}\right)}{\sum_{s=1}^n K\left(\frac{z_t - z_s}{h}\right)},$$

kde

$$\mathbf{y}_t^{(\tau)} = \sum_{j=1}^K \frac{y_{j,t}}{\hat{\mu}_j \hat{\xi}_{j,t}} \frac{\hat{\sigma}_j^{-2}}{\sum_{j=1}^K \hat{\sigma}_j^{-2}}$$

a  $\hat{\boldsymbol{\mu}}$ ,  $\hat{\boldsymbol{\xi}}_t$  a  $\hat{\sigma}_j^2 = \hat{\Sigma}_{j,j}$  značí současné odhady příslušných parametrů.

2. Odhadneme  $\{\boldsymbol{\xi}_t, t \in \mathbb{Z}\}$  a  $\boldsymbol{\Sigma}$  za pomoci naměřených hodnot

$$\{y_{t,i}^{(\xi)} = \frac{y_{t,i}}{(\hat{\mu}_i \hat{\tau}_t)}, t = 1, \dots, n, i = 1, \dots, K\} \text{ níže popsaným způsobem.}$$

I tentokrát je vidět, že je-li pro všechna  $t \in \mathbb{Z}$   $\tau_t = 1$ , jedná se o základní vektorový model multiplikačních chyb vMEM zavedený v definici 5 a první krok není nutné provádět. Iterace znovu provádíme do chvíle dosažení vhodného konvergenčního kritéria.

### 2.6.1 Zobecněná momentová metoda

Přistoupíme k odhadu zobecněnou momentovou metodou. Máme

$$\varepsilon_{t,i} = \frac{Y_{t,i}^{(\xi)}}{\xi_{t,i}}. \quad (17)$$

Platí-li předpoklady modelu, výraz  $\varepsilon_t - \mathbf{1}$  je podmíněně homoskedastický s nulovou podmíněnou střední hodnotou a podmíněnou rozptylovou maticí  $\Sigma$ . Nejlepší nestranný odhad zobecněnou momentovou metodou maticového parametru  $\theta$  (označíme jej jako  $\hat{\theta}_M$ ) splňuje rovnici

$$\sum_{t=1}^T \mathbf{A}_t (\varepsilon_t - \mathbf{1}),$$

kde

$$\mathbf{A}_t = \frac{\partial \xi_t^\top}{\partial \theta} \text{diag}(\xi_t)^{-1} \Sigma^{-1}. \quad (18)$$

Asymptotická rozptylová matice takového odhadu je

$$\text{Avar}(\hat{\theta}_M) = \lim_{n \rightarrow \infty} \text{var}(\hat{\theta}_M) = \left( \lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\mathbf{A}_t \Sigma \mathbf{A}_t^\top) \right] \right)^{-1}.$$

Její konzistentní odhad je proto

$$\widehat{\text{Avar}}(\hat{\theta}_M) = \left( \frac{1}{n} \sum_{t=1}^n \widehat{\mathbf{A}}_t \widehat{\Sigma} \widehat{\mathbf{A}}_t^\top \right)^{-1},$$

kde

$$\widehat{\Sigma} = \frac{1}{n} \sum_{t=1}^n (\widehat{\varepsilon}_t - \mathbf{1})(\widehat{\varepsilon}_t - \mathbf{1})^\top$$

je momentový odhad  $\Sigma$  a  $\widehat{\varepsilon}_t$  a  $\widehat{\mathbf{A}}_t$  odpovídají rovnicím 17 a 18 vyhodnoceným v bodě  $\hat{\theta}_M$ . Podobně jako v jednorozměrném případě je pro praktické provedení procedury ještě nutné nahradit v rovnici pro  $Y_{t,i}^{(\xi)}$  neznámé parametry  $\mu_i$  a  $\tau_t$  jejich odhady  $\hat{\mu}_i$  a  $\hat{\tau}_t$ .

## 2.7 Předpovědi v modelech SpMEM

Nyní se krátce podíváme na způsob předpovídání v modelech SpMEM. Předpokládejme, že chceme na základě dat do času  $t$  získat předpověď v čase  $t+h$ ,  $h \in \mathbb{N}$ . Využijeme obecné vyjádření podmíněné střední hodnoty zkoumané časové řady v čase  $t+h$ :

$$\widehat{Y}_{t+h} = \mathbb{E}(Y_{t+h} | \Psi_t) = \mu_{t+h|t} = \mu \tau_{t+h|t} \xi_{t+h|t}.$$

Předpovědi  $\tau_{t+h|t}$  a  $\xi_{t+h|t}$  můžeme získat každou zvlášť. Vzhledem ke způsobu odhadování  $\{\tau_t, t \in \mathbb{Z}\}$  lze (alespoň pro malá  $h$ ) zvolit  $\tau_{t+h|t} := \hat{\tau}_t$ . Pro předpověď  $\xi_{t+h|t}$  lze využít rovnici 11 a psát:

$$\xi_{t+h|t} = \begin{cases} (1 - (\beta_1 + \alpha_1 + \gamma_1/2)) + \beta_1 \xi_t + \alpha_1 y_t^{(\xi)} + \gamma_1 y_t^{(\xi-)}, & \text{pokud } h = 1 \\ (1 - (\beta_1 - \alpha_1 - \gamma_1/2)) + \beta_1 \xi_{t+h-1|t}, & \text{pokud } h > 1. \end{cases}$$

Pro praktické vyjádření předpovědi již stačí nahradit parametry  $\alpha_1$ ,  $\beta_1$  a  $\gamma_1$  jejich odhady  $\hat{\alpha}_1$ ,  $\hat{\beta}_1$  a  $\hat{\gamma}_1$  jedním ze způsobů představených v sekci 2.5.

## 2.8 Předpovědi v modelech SpvMEM

Ve vícerozměrném případě budeme postupovat analogicky. Předpokládejme, že máme data do času  $t$  a chceme získat předpověď v čase  $t + h$ ,  $h \in \mathbb{Z}$ . Opět využijeme podmíněnou střední hodnotu ve tvaru:

$$\widehat{\mathbf{Y}}_{t+h} = \mathbb{E}(\mathbf{Y}_{t+h} | \Psi_t) = \boldsymbol{\mu}_{t+h|t} = \tau_{t+h|t} \boldsymbol{\mu} \odot \boldsymbol{\xi}_{t+h|t}.$$

Stejně jako v sekci 2.7 zvolíme  $\tau_{t+h|t} := \widehat{\tau}_t$ ,  $h \in \mathbb{N}$  a pro vyjádření  $\boldsymbol{\xi}_{t+h|t}$  využijeme rovnici 16. Získáme:

$$\boldsymbol{\xi}_{t+h|t} = \begin{cases} (\mathbf{1} - (\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1 + \boldsymbol{\gamma}_1/2) \mathbf{1}) + \boldsymbol{\beta}_1 \boldsymbol{\xi}_t + \boldsymbol{\alpha}_1 \mathbf{y}_t^{(\xi)} + \boldsymbol{\gamma}_1 \mathbf{y}_t^{(\xi-)}, & \text{pokud } h = 1 \\ (\mathbf{1} - (\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_1 + \boldsymbol{\gamma}_1/2) \mathbf{1}) + \boldsymbol{\beta}_1 \boldsymbol{\xi}_{t+h-1|t}, & \text{pokud } h > 1. \end{cases}$$

V praxi znovu nahradíme maticové parametry  $\boldsymbol{\alpha}_1$ ,  $\boldsymbol{\beta}_1$  a  $\boldsymbol{\gamma}_1$  jejich odhady  $\widehat{\boldsymbol{\alpha}}_1$ ,  $\widehat{\boldsymbol{\beta}}_1$  a  $\widehat{\boldsymbol{\gamma}}_1$  získanými způsobem představeným v sekci 2.6.

# 3. Praktická část

## 3.1 Popis a grafické zobrazení dat

V následujících částech práce se pokusíme některé představené modely aplikovat na reálná data. Z důvodu výpočetní náročnosti se zaměříme pouze na jedno-rozměrné modely. K dispozici máme denní časové řady nahlášených a vyplacených škod pro několik pojistných odvětví v jedné z českých pojišťoven (na základě data vzniku škody). Jde tedy o nezáporné časové řady s možným výskytem nulových pozorování. Data jsou vhodnými úpravami anonymizována tak, aby byl zachován jejich charakter. Dle přání pojišťovny nebudou data v příloze práce zveřejněna. Co se týče pojistných odvětví, jde o pojištění domů a domácností, pojištění odpovědnosti, cestovní pojištění a pojištění úvěrů.

Pojištění domů a domácností patří k nejběžnějším typům pojištění a obvykle pokrývá škody způsobené událostmi, jako jsou přírodní katastrofy, únik vody, vandalismus, krádeže a mnohé další. Pro pojišťovnu v tomto případě představují největší riziko velké přírodní katastrofy, které způsobí škody mnoha pojištěným za krátkou dobu. Jako příklad lze uvést extrémní bouři s krupobitím a tornádem, která se v roce 2021 přehnal přes Moravu a způsobila obrovské škody na majetku.

Pojištění odpovědnosti je též velmi frekventované. Pojištěný se v tomto případě chrání před škodami, které sám způsobí třetí osobě. Může se jednat o odškodnění za smrt či ublížení na zdraví, ale i škody na majetku.

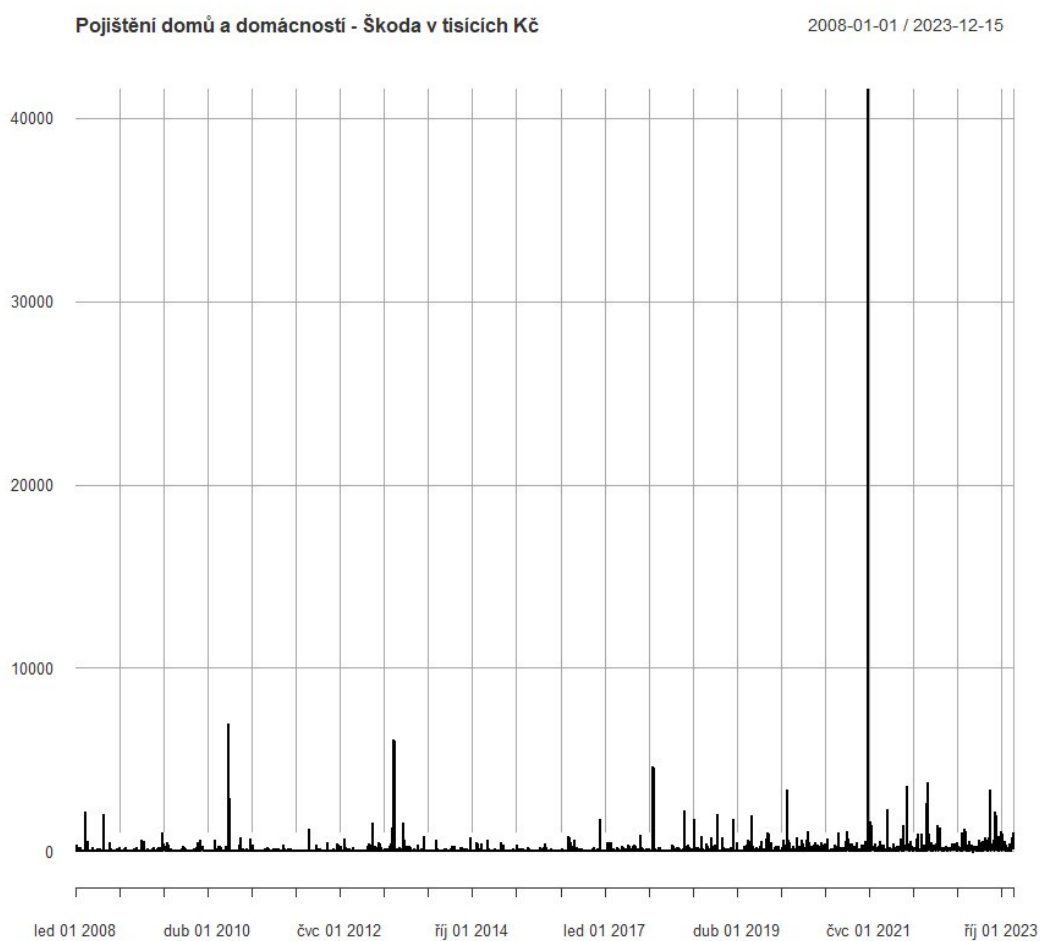
Cestovní pojištění je většinou krátkodobé (v horizontu dnů či týdnů) a chrání pojištěného před nejrůznějšími typy škod spojenými s cestováním. Může se jednat např. o zrušení zájezdu, léčebné výlohy, škody způsobené zpožděním či zrušením spojů, ztracení zavazadel atd. Velké škody mohou nastat např. pokud dojde k přírodní katastrofě v některém z oblíbených cestovních letovisek, což způsobí rušení zájezdů a škody na majetku a zdraví lidí, kteří již do místa odcestovali.

Pojištění úvěrů nejčastěji chrání věřitele před neschopností dlužníka plnit závazky. Pro pojišťovnu tak může představovat riziko např. náhlé zhoršení ekonomické situace, které zapříčiní neschopnost mnoha dlužníků splácet své půjčky. Existují pojišťovny, které se zaměřují výhradně na tento typ pojištění, avšak pro pojišťovnu, ze které data pochází, se jedná o jedno z nejméně významných pojistných odvětví. V denní časové řadě je tedy nutné očekávat mnoho nulových pozorování a mohlo by být vhodnější přistoupit k modelování např. na měsíční či dokonce čtvrtletní bázi.

Předtím, než data začneme hlouběji analyzovat, měli bychom si uvědomit, že se jedná skutečně pouze o nahlášené a vyplacené škody, nikoliv nastalé. To znamená, že některá pozorování se v budoucnu mohou (a s velkou pravděpodobností i budou) měnit v důsledku škod, které již nastaly, ale pojišťovna o nich dosud neví, protože nebyly nahlášené. Pro takovéto účely si pojišťovna tvoří tzv. rezervu na nastalé a dosud nenahlášené škody (anglicky „incurred but not reported claims“,

používá se běžná mezinárodní zkratka „IBNR“). To však znamená, že abychom mohli s daty vhodně pracovat a správně je interpretovat, musíme buď odhadnout tuto rezervu, a nebo využít starší data, kde již není pravděpodobné, že by došlo k nahlášení dalších škod. První varianta bohužel v našem případě není možná, jelikož obdržená data neobsahují datum nahlášení škody, které je k výpočtu IBNR nutné znát. Budeme se proto muset spokojit s využitím starších pozorování.

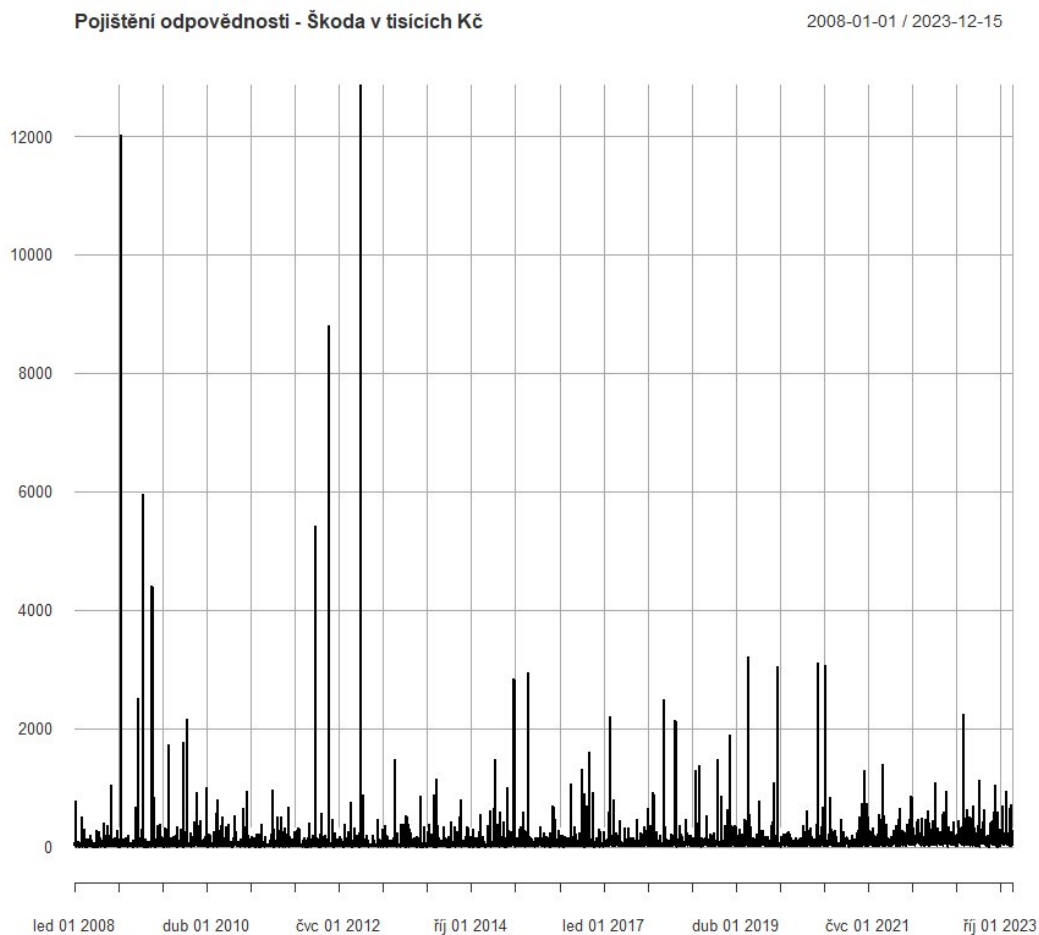
Nyní přejdeme k základnímu grafickému zobrazení dat. Nejprve se podíváme na pojištění domů a domácností v letech 2008 až 2023.



Obrázek 3.1: Nahlášené a vyplacené škody v pojištění domů a domácností od roku 2008.

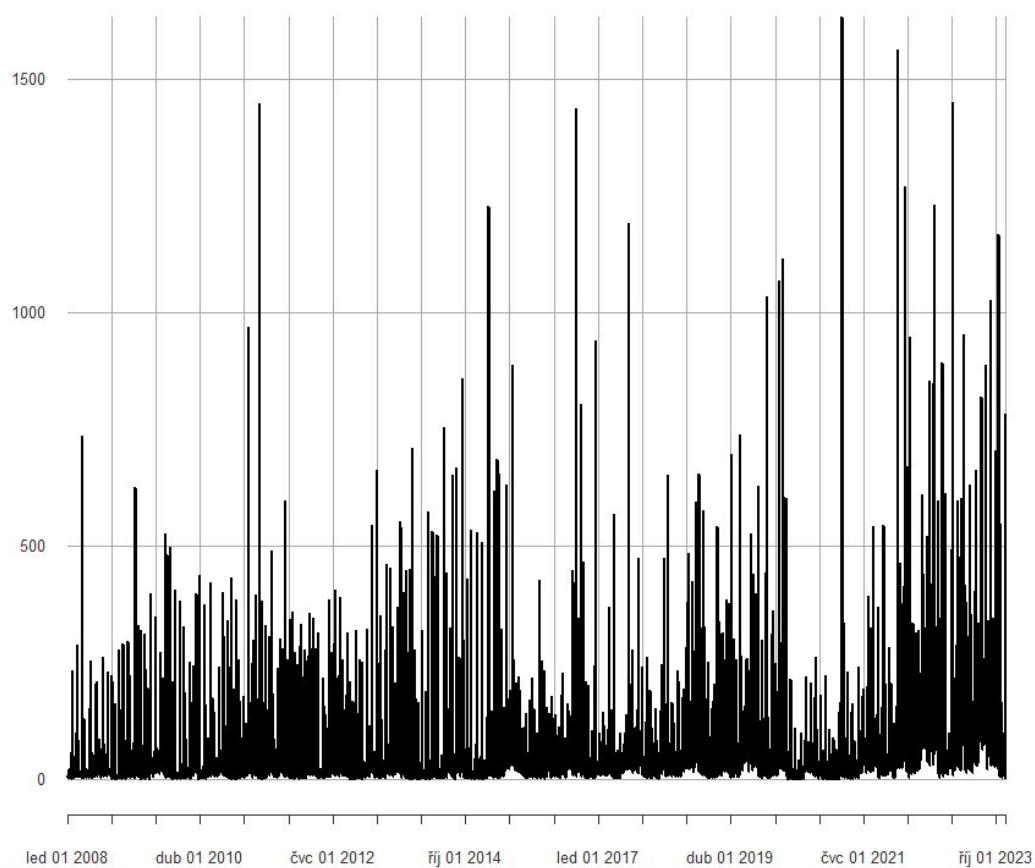
V datech můžeme pozorovat několik mediálně známých přírodních katastrof, jako je např. povodeň na Českokamenicku dne 7. 8. 2010, další povodně v severních, západních a středních Čechách v červnu 2013 nebo orkán, který v říjnu 2017 zasáhl především oblast Krkonoš a Českého středohoří. Ještě mnohonásobně vyšší škody však napáchalo již výše zmiňované tornádo na Moravě v červnu 2021.

Podobně si nyní zobrazíme do grafu škody týkající se pojištění odpovědnosti ve stejném období.



Obrázek 3.2: Nahlášené a vyplacené škody v pojištění odpovědnosti od roku 2008.

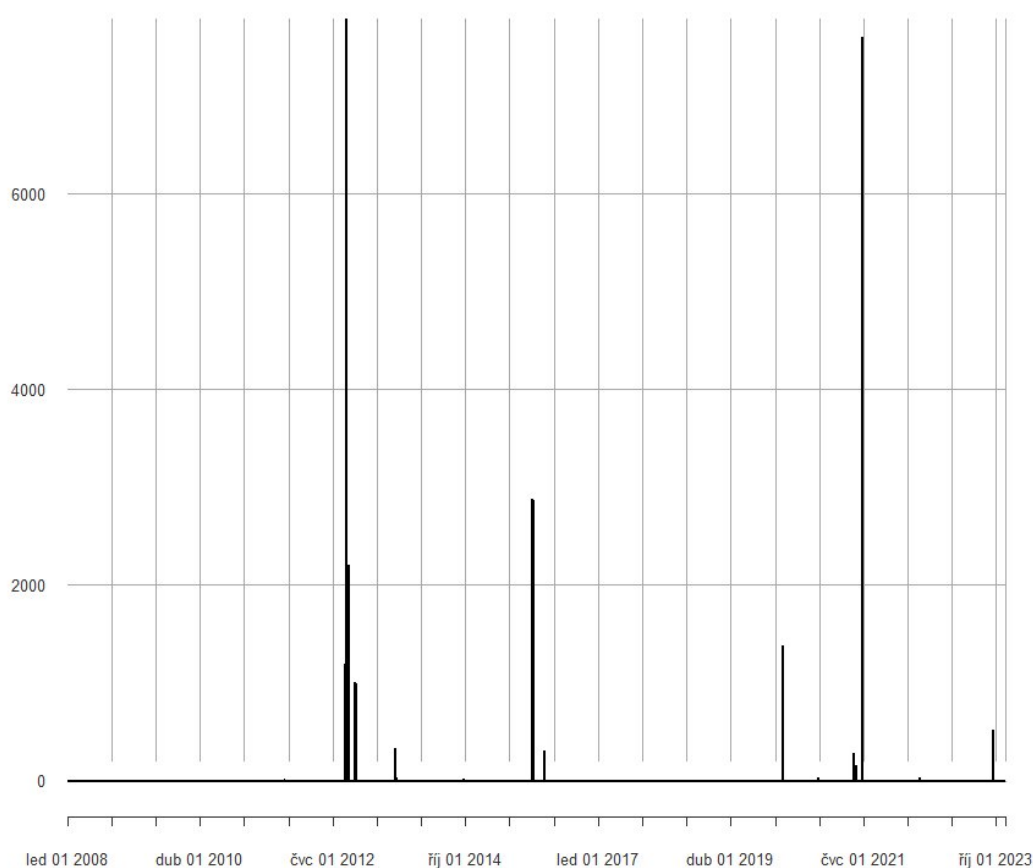
Znovu lze pozorovat dny s výrazně větším objemem škod, avšak odchylky nejsou tak výrazné a již není možné jednoznačně určit, co přesně je způsobilo. Následuje obdobný graf pro cestovní pojištění.



Obrázek 3.3: Nahlášené a vyplacené škody v cestovním pojištění od roku 2008.

Zde již není možné pozorovat mnoho škod, které by výrazně vyčnívaly nad ostatní. Lze si však všimnout významného úpadku škod v době covidových omezení, kdy se i zájem o tento typ pojištění snížil na minimum. Nakonec se ještě podíváme na úvěrové pojištění.





Obrázek 3.4: Nahlášené a vyplacené škody v úvěrovém pojištění od roku 2008.

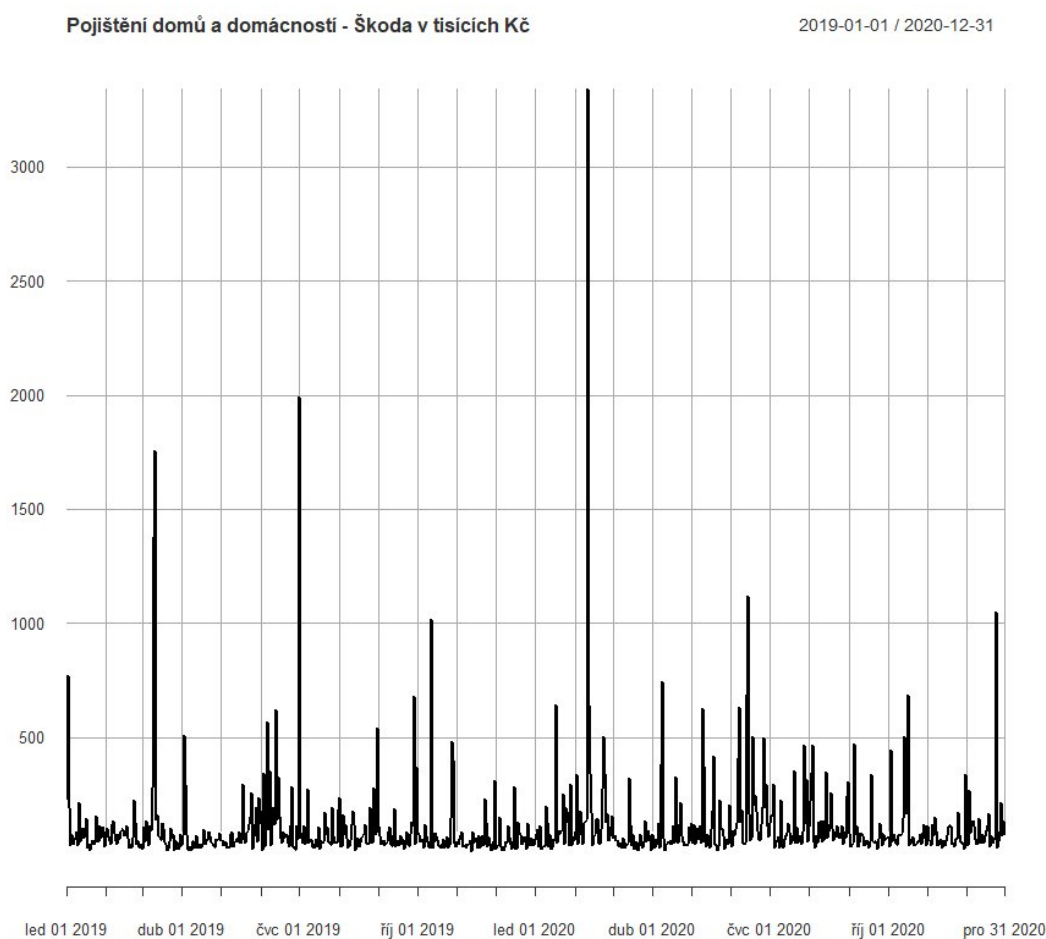
Jak již bylo řečeno, úvěrové pojištění je pro tuto pojišťovnu okrajové, a proto v časové řadě denních škod můžeme pozorovat drtivou většinu nulových pozorování. Konkrétně máme za celou dobu k dispozici pouze 89 nenulových pozorování, tedy modelování škod na denní frekvenci zřejmě nebude vhodné a otázkou zůstává, zdali je vůbec možné na základě takto nízkého množství dat učinit smysluplný závěr. Přesto lze vidět vyšší množství škod např. v roce 2012, kdy česká ekonomika zažívala významnou recesi. Podobně by šlo nárůst škod v roce 2021 připisovat problémům spojeným s covidovou pandemií.

V následujících částech textu se budeme věnovat bližší analýze dat z pojištění domů a domácností a úvěrového pojištění. Podstatné části zdrojového kódu ze softwaru R budou uvedeny v příloze A.1.

## 3.2 Základní modely MEM

V této části práce se pokusíme na zvolená data aplikovat základní model multiplikativních chyb zavedený v definici 4. Tento typ modelu není uzpůsoben na

výskyt nulových pozorování, tedy pokud chceme škody modelovat na denní bázi, je nutné zvolit jeden z významnějších typů pojištění. My budeme konkrétně volit pojištění domů a domácností v letech 2019 a 2020. Důvodem je, že na tyto roky již pojišťovna nedrží téměř žádnou rezervu IBNR, což znamená, že již neočekává dodatečné nahlášení dalších škod. U tohoto typu pojištění navíc není nutné se obávat významného vlivu pandemie. V následujícím grafu vidíme škody právě za toto zvolené období.



Obrázek 3.5: Nahlášené a vyplacené škody v pojištění domů a domácností v letech 2019 a 2020.

V grafu můžeme částečně pozorovat jev podobný dříve popisovanému shlukování volatility na finančních trzích, tedy že po vyšších škodách mají tendenci následovat další vyšší škody a naopak. To může být následkem nepříznivého počasí, které často přetrvává více než jeden den. Do tabulky si nyní shrneme základní statistické charakteristiky dat.

Minimum	1. kvartil	Medián	3. kvartil	Maximum	Průměr	Směrodatná odch.
3.826	33.403	56.642	112.029	3343.574	109.385	204.324

Tabulka 3.1: Základní statistické charakteristiky škod z pojištění domů a domácností v letech 2019 a 2020.

Maximální škoda za jeden den v letech 2019 a 2020 je ve výši přes 3.3 mil. Kč a byla pozorována dne 10. února 2020, kdy se Českou republikou prohnala bouře Sabine. Teď přejdeme k odhadu parametrů modelu MEM pomocí metody maximální věrohodnosti, jak bylo popsáno v části 2.1. Pro střední hodnotu volíme rovnici podobnou modelu GARCH:

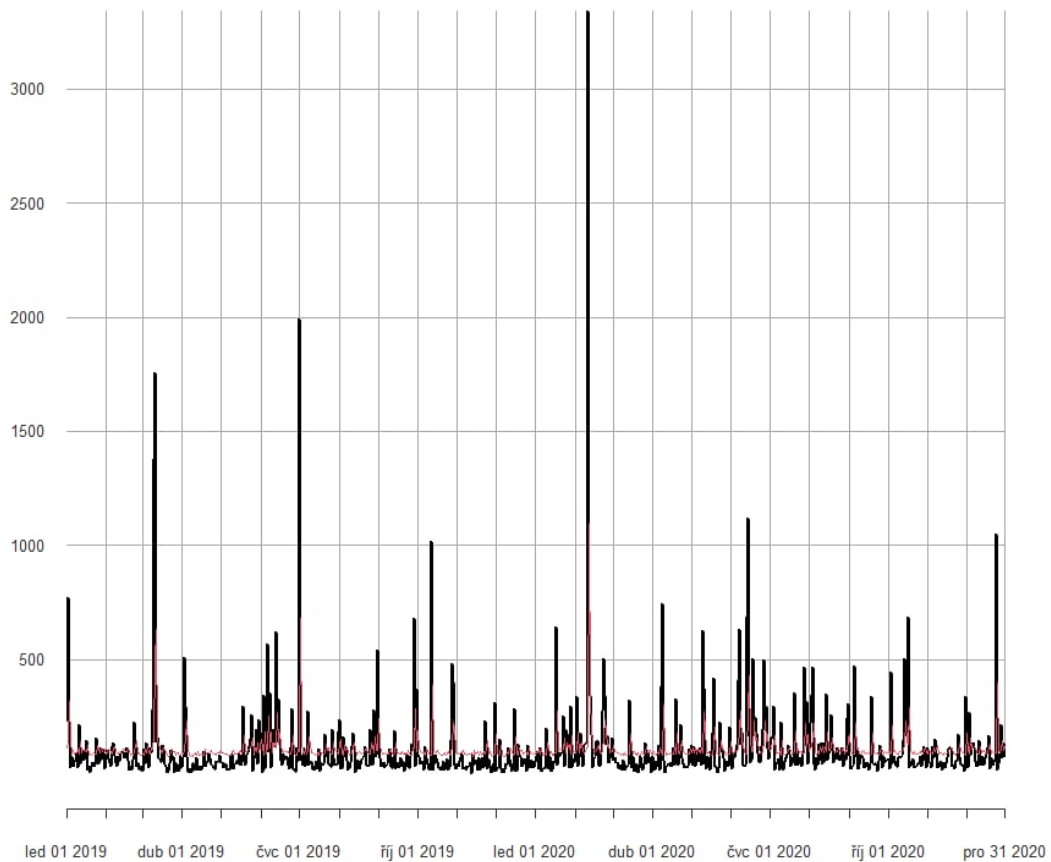
$$\mu_t = \omega + \alpha y_{t-1} + \beta \mu_{t-1}, t \in \mathbb{Z},$$

přičemž  $\mu_1$  nastavíme rovno výběrovému průměru z dat. Jak vyplývá z poznámky k větě 4, není v tomto případě nutné, aby  $\alpha + \beta < 1$ .

$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
71.853	0.303	0.057

Tabulka 3.2: Odhady parametrů v základním modelu MEM metodou maximální věrohodnosti.

Maximalizace logaritmicke věrohodnosti procedurou `optim` v softwaru R v tomto případě proběhla bez problémů a odhad je stabilní i vůči změnám v počátečním odhadu, který je do numerické procedury nutno zadat před jejím spuštěním. V jednom grafu si nyní zobrazíme samotná data společně s tzv. vyrovnanými hodnotami (nebo predikcí „in-sample“) za pomoci základního modelu MEM.



Obrázek 3.6: Proložení dat pomocí základního modelu MEM.

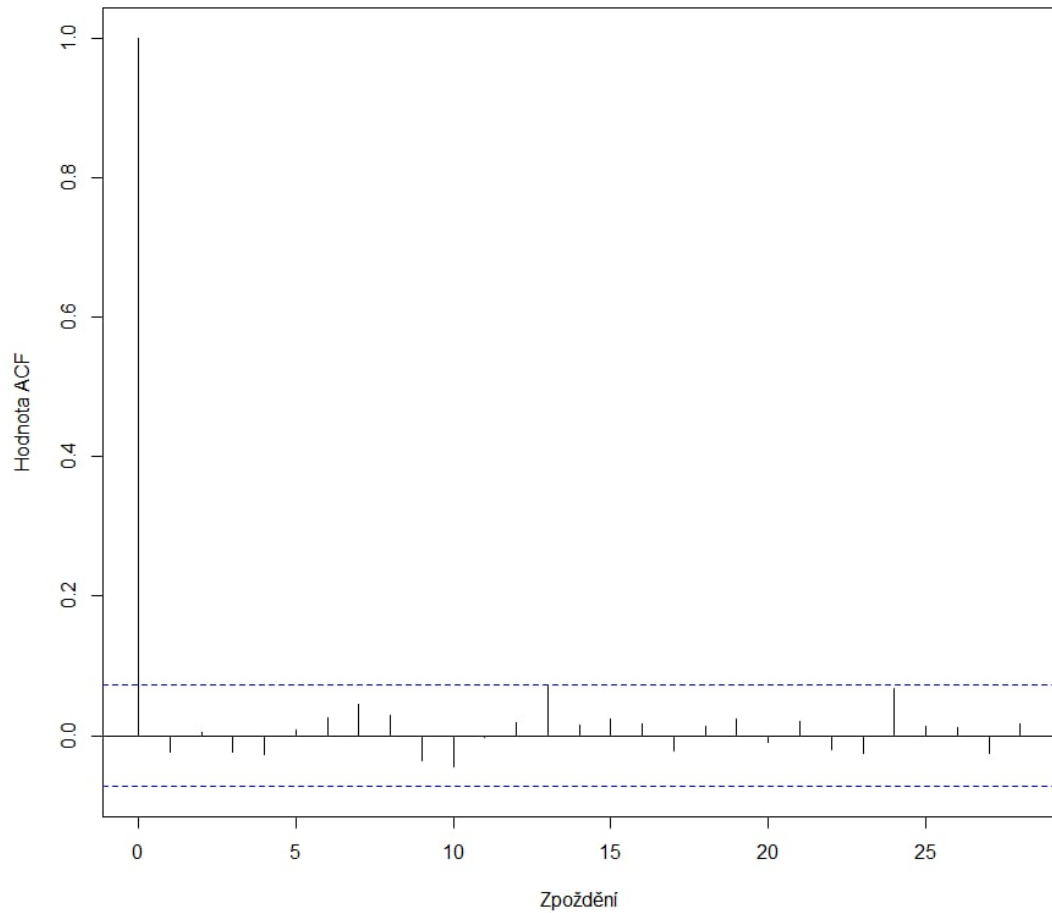
Na první pohled se zdá, že model data zachytil poměrně úspěšně. Abychom mohli kvalitu proložení zhodnotit číselně, spočítáme tzv. střední čtvercovou chybu (anglicky „mean squared error“), což je jedna z nejpoužívanějších metrik pro přesnost předpovědi. Její vzorec je:

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

kde  $\{\hat{y}_t, t = 1, \dots, n\}$  jsou vyrovnané hodnoty. Jde tedy o sumu druhých mocnin aditivních reziduí. Získáme hodnotu 40625.03.

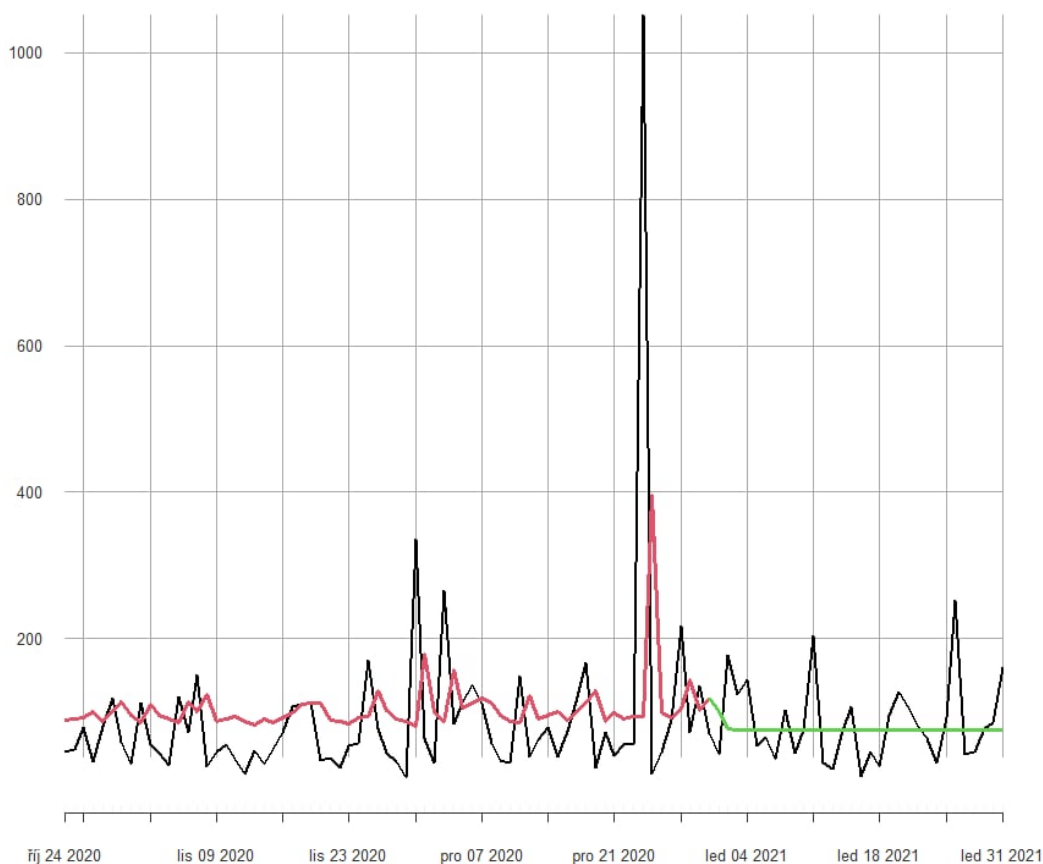
Provedeme též základní diagnostiku modelu. Vydělíme-li data vyrovnanými hodnotami, měly by výsledné multiplikativní odchylky mít jednotkovou střední hodnotu a měly by být nezávislé. Pro test jednotkové střední hodnoty využijeme běžný Studentův t-test. Získáme p-hodnotu 0.978, tedy hypotézu jednotkové střední hodnoty nemůžeme zamítnout. Teď se podíváme na graf odhadnuté autokorelační funkce reziduí.

### Autokorelační funkce reziduí



Obrázek 3.7: Odhadnutá autokorelační funkce reziduí modelu MEM.

Z grafu se nezdá, že by rezidua byla významně korelovaná. Spusíme-li v R tzv. Ljungův-Boxův test se zpožděním 13 (maximum ACF v bodech kromě 0), získáme p-hodnotu 0.649 a opět tedy hypotézu nekorelovanosti nezamítáme. Nakonec se ještě pokusíme o predikci „out-of-sample“ na jeden měsíc dopředu a porovnáme ji s reálnými naměřenými daty v lednu 2021.



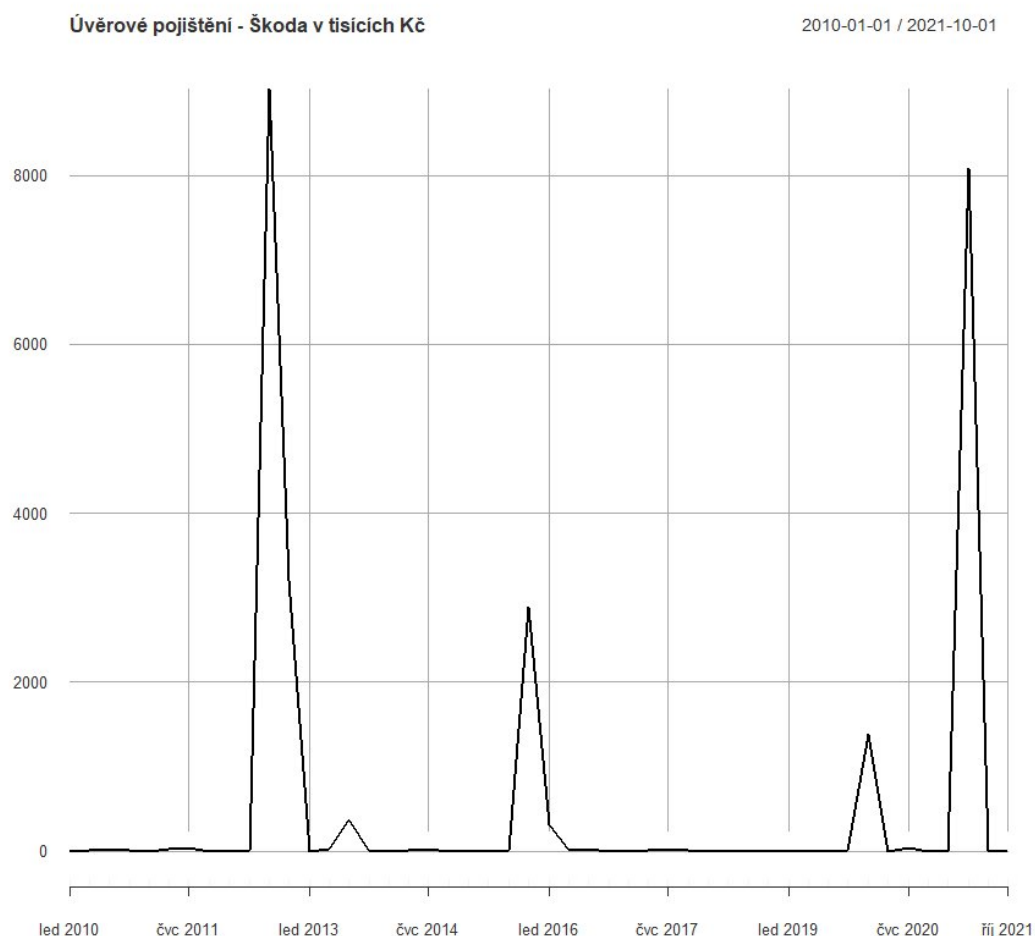
Obrázek 3.8: Predikce škod na jeden měsíc dopředu na základě modelu MEM a porovnání s reálnými daty.

Spočteme-li střední čtvercovou chybu i pro předpověď mimo vzorek, dostaneme hodnotu 3256.571. Vzhledem k jednoduchosti modelu však nelze předpokládat, že by byl schopen kvalitní predikce mimo vzorek, a to především v dlouhodobém horizontu.

### 3.3 Modely MEM rozšířené v nule

V další části zkusíme na data z oblasti pojišťovnictví aplikovat modely MEM rozšířené v nule představené v kapitole 2.2. Vzhledem k tomu, že model je přímo uzpůsoben na výskyt nulových pozorování, je vhodné zvolit časovou řadu, která taková pozorování obsahuje. Podíváme-li se blíže na data, která máme k dispozici, zjistíme, že v pojištění domů a domácností a pojištění odpovědnosti se taková pozorování v celém patnáctiletém období téměř neobjevují a v cestovním pojištění začínají přibývat až v období pandemie koronaviru. Bude tedy potřeba použít data z úvěrového pojištění, kde však zřejmě není vhodné zkoumat škody na denní bázi. Přejdeme proto na čtvrtletní frekvenci a budeme používat data z období od

roku 2010 do roku 2021. Data si opět vykreslíme do grafu.



Obrázek 3.9: Čtvrtletní časová řada škod v úvěrovém pojištění od roku 2010 do roku 2021.

Vidíme, že i v časové řadě se čtvrtletní frekvencí se nulová pozorování objevují často a v roce 2018 pojišťovna dokonce nevyplatila na škodách v tomto odvětví ani jedinou korunu. Následující tabulka obsahuje základní statistické charakteristiky dat.

Minimum	1. kvartil	Medián	3. kvartil	Maximum	Průměr	Směrodatná odch.
0.000	0.000	2.578	18.784	9030.318	531.296	1810.027

Tabulka 3.3: Základní statistické charakteristiky čtvrtletních škod z úvěrového pojištění v letech 2010 až 2021.

Přistoupíme tedy k maximalizaci logaritmické věrohodnosti v softwaru R. Pro střední hodnotu (resp. její přirozený logaritmus) zvolíme opět rovnici podobnou GARCH modelu převzatou ze článku (Hautsch a kol., 2014):

$$\begin{aligned}
\log(\mu_t) &= \omega + \alpha \log(\varepsilon_{t-1}) \mathbb{1}(y_{t-1} > 0) + \alpha_0 \mathbb{1}(y_{t-1} = 0) + \beta \log(\mu_{t-1}) \\
&= \omega + \alpha \log\left(\frac{y_{t-1}}{\mu_{t-1}}\right) \mathbb{1}(y_{t-1} > 0) + \alpha_0 \mathbb{1}(y_{t-1} = 0) + \beta \log(\mu_{t-1}), \quad t \in \mathbb{Z}.
\end{aligned} \tag{19}$$

Pro multiplikatívni odchylky v tomto případě předpokládáme, že mají podmíněnou hustotu  $f_\varepsilon$  zavedenou v definici 6, tedy lze psát  $\mathcal{L}(\varepsilon_t | \Psi_{t-1}) = f_\varepsilon$ ,  $t \in \mathbb{Z}$ . Pro  $Y_t = \mu_t \varepsilon_t$ ,  $t \in \mathbb{Z}$  proto máme podmíněnou hustotu:

$$f_Y(y) = f_\varepsilon\left(\frac{y}{\mu_t}\right) \frac{1}{\mu_t}, \quad y \geq 0. \tag{20}$$

Pracujeme-li s pozorováními  $Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n$  a vektorem neznámých parametrů  $\boldsymbol{\nu}$ , získáme podmíněnou logaritmickou věrohodnost ve tvaru:

$$\ell_n(\boldsymbol{\nu} | \Psi_{n-1}) = \sum_{t=1}^n \log\left(f_\varepsilon\left(\frac{y_t}{\mu_t}\right)\right) - \sum_{t=1}^n \log(\mu_t).$$

Protože tvar hustoty  $f_\varepsilon$  známe, můžeme ji podrobněji rozepsat a se zachováním značení z kapitoly 2.2 máme:

$$\begin{aligned}
\ell_n(\boldsymbol{\nu} | \Psi_{n-1}) &= n_0 \log(p_0) + n_+ \log(1 - p_0) + \sum_{t \in I_{n+}} [\log(a) + (am - 1) \log\left(\frac{y_t}{\mu_t}\right) \\
&\quad + \eta \log(\eta) - (\eta + m) \log\left\{\eta + \left(\frac{y_t}{\mu_t} \lambda^{-1}\right)^a\right\} - \log\{\mathcal{B}(m, \eta)\} \\
&\quad - am \log(\lambda)] - \sum_{t=1}^n \log(\mu_t),
\end{aligned} \tag{21}$$

přičemž v praxi musíme inicializovat  $\mu_1$  např. pomocí výběrového průměru z dat. Naším úkolem by nyní mělo být hledání globálního optima této funkce vzhledem k neznámým parametrům, přičemž je nutné brát do úvahy i omezení, která na parametry klademe. Máme  $0 < p_0 < 1$ ,  $a > 0$ ,  $m > 0$ ,  $\eta > 0$  a též  $\eta > 1/a$ , což vyplývá z definice 6, kde pro parametr  $\lambda$  uvažujeme specifický tvar, aby byl splněn předpoklad jednotkové podmíněné střední hodnoty odchylek.

Dále je nutné si uvědomit, že střední hodnota  $\mu_t$  v sobě obsahuje další neznámé parametry, protože se řídí rovnicí 19. Logaritmickou věrohodnost bychom proto mohli dále rozepsat dosazením

$$\mu_t = \exp\left(\omega + \alpha \log\left(\frac{y_{t-1}}{\mu_{t-1}}\right) \mathbb{1}(y_{t-1} > 0) + \alpha_0 \mathbb{1}(y_{t-1} = 0) + \beta \log(\mu_{t-1})\right).$$

Spustíme-li v R numerickou optimalizaci, ihned narazíme na několik problémů. Článek (Hautsch a kol., 2014), ze kterého v této části práce čerpáme teoretický základ, neuvažuje pro parametry v rovnici 19 žádná omezení. Není však těžké zjistit, že přesáhnou-li některé z parametrů v absolutní hodnotě určitou hranici,



střední hodnota začne velmi rychle konvergovat do nekonečna nebo k nule, což znemožňuje optimalizaci. Parametry rovnice střední hodnoty je tedy určitě nutné omezit, aby bylo možné nalézt stacionární řešení, avšak z teorie není jasné, jaká omezení konkrétně zvolit.

Toto v kombinaci s velkým množstvím dalších parametrů ve zobecněném F-rozdělení činí z optimalizace velmi náročný úkol. Bohužel se nám ani po mnoha neúspěšných pokusech nepodařilo získat smysluplné odhady, které by byly stabilní vůči změnám v počátečním odhadu, který je nutné zadat před spuštěním numerické optimalizace v R. Abychom s odhady mohli vhodně pracovat, bude nutné optimalizační problém zjednodušit.

Přistoupíme proto k využití metody exponenciální kvazi-maximální věrohodnosti, kterou jsme již dříve v práci zmiňovali. Spočívá v tom, že odhad provádíme za předpokladu, že data mají exponenciální rozdělení, i přesto, že tomu tak ve skutečnosti být nemusí. Vlastnosti takových odhadů (konsistence a asymptotická normalita) a podmínky jejich splnění byly diskutovány v kontextu modelů GARCH např. ve článku (Hudecová a Pešta, 2024). Analogické vlastnosti je možné dle článku dokázat i v případě, že pracujeme s modely typu MEM.

Budeme tedy pracovat s exponenciálním rozdělením s hustotou

$$g_X(x) = \exp(-x), x > 0$$

a s podmíněnou hustotou multiplikativních odchylek ve tvaru

$$f_\epsilon(x) = p_0\delta(x) + (1 - p_0)\exp(-x)\mathbb{1}(x > 0), x \geq 0.$$

V této situaci je již velmi jednoduché odvodit podmíněnou logaritmickou věrohodnost:

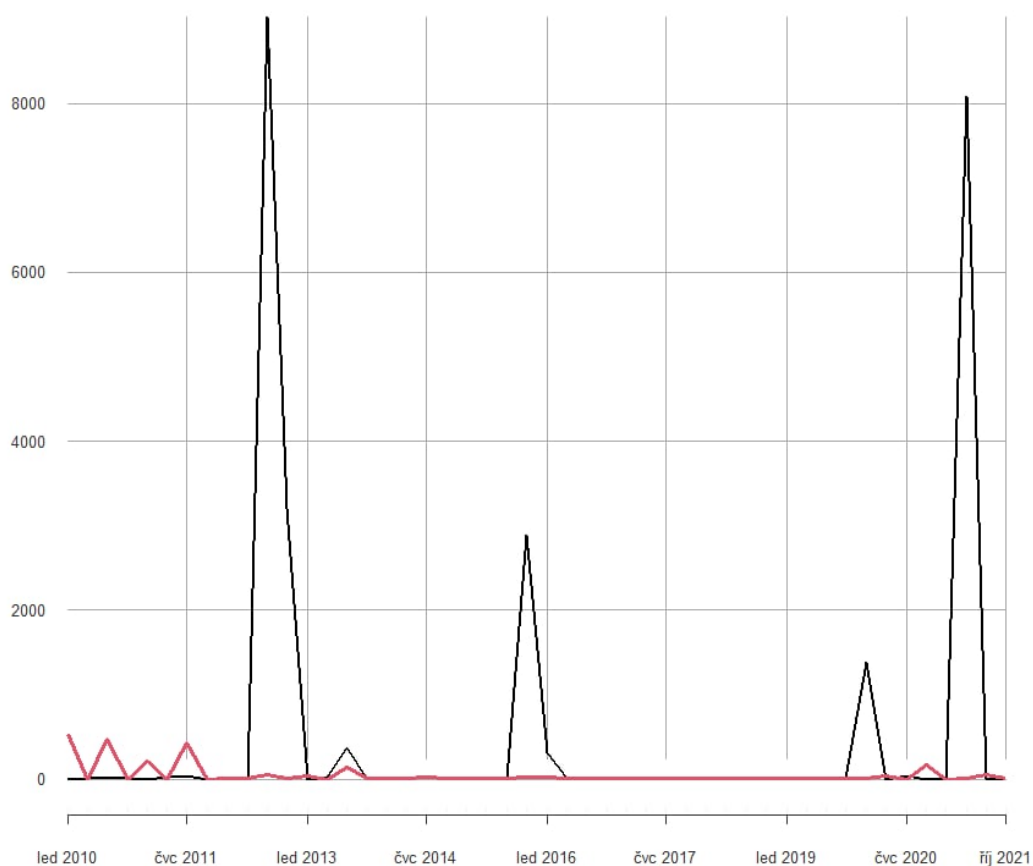
$$\ell_n(\boldsymbol{\nu}|\Psi_{n-1}) = n_0 \log(p_0) + n_+ \log(1 - p_0) - \sum_{t \in I_{n_+}} \frac{y_t}{\mu_t} - \sum_{t=1}^n \log(\mu_t),$$

kterou opět zkusíme v R maximalizovat. Znovu je nutné, aby  $0 < p < 1$  a též je nutné omezit parametry rovnice střední hodnoty z důvodů popsaných výše. Nastavíme tedy  $|\omega| < 10$ ,  $|\alpha| < 1$ ,  $|\alpha_0| < 10$  a  $|\beta| < 1$ . Tato omezení jsou výsledkem empirických pokusů v podobě opakovaného dosazování do rovnice. Nemají tedy exaktní teoretický základ. Přesto se však pokusíme pracovat s následujícími odhady, které jsme optimalizací získali:

$\hat{p}_0$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\alpha}_0$	$\hat{\beta}$
0.479	2.814	0.371	-1.786	-0.602

Tabulka 3.4: Odhady parametrů v modelu ZA-MEM metodou exponenciální kvazi-maximální věrohodnosti (QML).

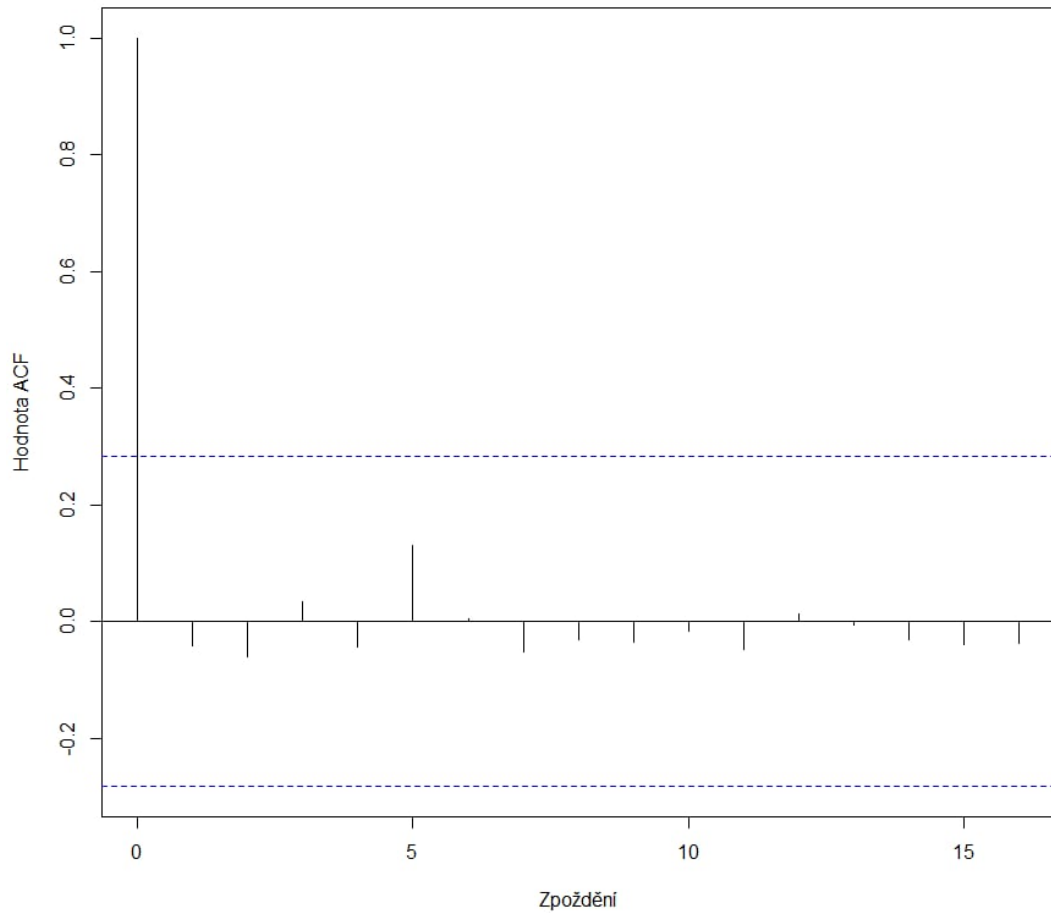
Do grafu si vykreslíme data společně s předpovědí ve vzorku.



Obrázek 3.10: Proložení dat pomocí modelu ZA-MEM.

Proložení dat nevypadá příliš uspokojivě, což potvrzuje i velmi vysoká střední čtvercová odchylka ( $MSE = 3477595$ ). Analogickým způsobem jako u základního modelu provedeme jednoduchou diagnostiku modelu. Nejprve zkusíme, zdali mají multiplikační odchylky (data vydělená vyrovnanými hodnotami) jednotkovou střední hodnotu. Studentův t-test nám dává p-hodnotu 0.03, tedy hypotézu zamítáme a zdá se, že tento předpoklad je porušen. Zobrazíme si nyní graf auto-korelační funkce reziduí.

Autokorelační funkce reziduí



Obrázek 3.11: Odhadnutá autokorelační funkce reziduí modelu ZA-MEM.

Z obrázku se nezdá, že bychom měli mít s autokorelací reziduí problém. To potvrzuje i Ljungův-Boxův test vyhodnocený se zpožděním 5, který dává p-hodnotu 0.925, tedy hypotézu nekorelovanosti nezamítáme. Predikci mimo vzorek tentokrát spočteme na 1 rok dopředu. V následující tabulce je vidět porovnání predikce na rok 2022 se skutečnými pozorováními.

Čtvrtletí	1	2	3	4
<b>Skutečné škody (tis. Kč)</b>	0.000	39.688	10.459	0.000
<b>Predikce modelem ZA-MEM</b>	3.186	8.299	4.662	6.598

Tabulka 3.5: Predikce mimo vzorek na rok 2022 pomocí modelu ZA-MEM a porovnání se skutečnými pozorovanými hodnotami.

Nakonec ještě spočteme MSE této predikce, kde nám vyjde hodnota 268.137.

### 3.4 Semiparametrické modely MEM

Nyní se na krátko vrátíme k datům z pojištění domů a domácností v letech 2019 a 2020, které jsme již používali v části 3.2 a aplikujeme na ně semiparametrické modely multiplikatívních chyb představené v sekci 2.5. K odhadování parametrů v rovnici pro  $\xi$  přitom můžeme využít buď zobecněnou momentovou metodu nebo metodu kvazi-maximální věrohodnosti. Ještě předtím však model mírně zjednodušíme, protože v případě dat z pojišťovnictví nemáme přirozenou interpretaci pro běžný výnos z finančních trhů (v původní rovnici 11 označen jako  $R_t$ ). Proto parametrizaci  $\{\xi_t, t \in \mathbb{Z}\}$  zjednodušíme na tvar:

$$\begin{aligned}\xi_t &= [1 - (\beta + \alpha)] + \beta\xi_{t-1} + \alpha Y_{t-1}^{(\xi)} \\ &= [1 - (\beta + \alpha)] + \beta\xi_{t-1} + \alpha \frac{Y_{t-1}}{\mu\tau_{t-1}} \quad t \in \mathbb{Z}.\end{aligned}\tag{22}$$

Nyní již budeme postupovat přesně tak, jak bylo popsáno v části 2.5. Tedy nejprve si nastavíme

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{t=1}^n y_t = 109.385, \\ \hat{\xi}_t &= 1, \quad t \in \mathbb{Z}\end{aligned}$$

a odhadneme  $\{\tau_t, t = 1, \dots, n\}$  pomocí jádrových odhadů popsaných v rovnici 10, přičemž šířku pásma  $h$  spočítáme pomocí zabudované procedury `dpik` z balíčku `KernSmooth` v R.

Pokud bychom pro odhad parametrů  $\alpha$  a  $\beta$  chtěli použít zobecněnou momentovou metodu, museli bychom pro naše data vyřešit soustavu

$$\sum_{t=1}^n (\varepsilon_t - 1) \mathbf{a}_t = \mathbf{0},$$

kde

$$\mathbf{a}_t = \frac{1}{\xi_t} \nabla_{\boldsymbol{\theta}} \xi_t$$

a  $\boldsymbol{\theta} = (\alpha, \beta)^\top$ . K tomu je nutné spočítat parciální derivace  $\xi_t$  podle  $\alpha$  a  $\beta$ . Inicializujeme  $\xi_1 = 1$ , a tedy  $\nabla_{\boldsymbol{\theta}} \xi_1 = (0, 0)^\top$ . Máme:

$$\frac{\partial \xi_t}{\partial \alpha} = -1 + \beta \frac{\partial \xi_{t-1}}{\partial \alpha} + y_{t-1}^{(\xi)} = -1 + \beta \frac{\partial \xi_{t-1}}{\partial \alpha} + \frac{y_{t-1}}{\mu\tau_t}$$

a

$$\frac{\partial \xi_t}{\partial \beta} = -1 + \xi_{t-1} + \beta \frac{\partial \xi_{t-1}}{\partial \beta}, \quad t \in \mathbb{Z}, \quad t \geq 2.$$

Poté se již můžeme pokusit numericky vyřešit zmiňovanou soustavu rovnic. Tu si nyní ještě rozepíšeme do podrobnějšího tvaru, aby bylo zřejmé, kde do ní vstupují hledané neznámé parametry. Máme:

$$\sum_{t=1}^n \begin{pmatrix} \frac{y_t}{\mu\xi_t\tau_t} - 1 \\ \xi_t \end{pmatrix} \frac{1}{\xi_t} \begin{pmatrix} \frac{\partial \xi_t}{\partial \alpha} \\ \frac{\partial \xi_t}{\partial \beta} \end{pmatrix} = \mathbf{0},\tag{23}$$

kde místo  $\mu$  a  $\tau_1, \tau_2, \dots, \tau_n$  dosadíme jejich výše popsané odhady. Protože  $\xi_t$  je dáno rovnicí 22 a  $\xi_1 = 1$ , je možné dále psát:

$$\begin{aligned}
& \left( \frac{y_1}{\mu\tau_1} - 1 \right) \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \sum_{t=2}^n \left( \frac{y_t}{\mu \left( [1 - (\beta + \alpha)] + \beta\xi_{t-1} + \alpha \frac{y_{t-1}}{\mu\tau_{t-1}} \right) \tau_t} - 1 \right) \\
& \left( [1 - (\beta + \alpha)] + \beta\xi_{t-1} + \alpha \frac{y_{t-1}}{\mu\tau_{t-1}} \right)^{-1} \begin{bmatrix} -1 + \beta \frac{\partial \xi_{t-1}}{\partial \alpha} + \frac{y_{t-1}}{\mu\tau_t} \\ -1 + \xi_{t-1} + \beta \frac{\partial \xi_{t-1}}{\partial \beta} \end{bmatrix} \\
& = \sum_{t=2}^n \left( \frac{y_t}{\mu \left( [1 - (\beta + \alpha)] + \beta\xi_{t-1} + \alpha \frac{y_{t-1}}{\mu\tau_{t-1}} \right) \tau_t} - 1 \right) \\
& \left( [1 - (\beta + \alpha)] + \beta\xi_{t-1} + \alpha \frac{y_{t-1}}{\mu\tau_{t-1}} \right)^{-1} \begin{bmatrix} -1 + \beta \frac{\partial \xi_{t-1}}{\partial \alpha} + \frac{y_{t-1}}{\mu\tau_t} \\ -1 + \xi_{t-1} + \beta \frac{\partial \xi_{t-1}}{\partial \beta} \end{bmatrix} = \mathbf{0}.
\end{aligned} \tag{24}$$

Podobně jako v předchozí části věnované modelům ZA-MEM bychom mohli rovnici dále rozvíjet tak, že bychom dosazovali za  $\xi_{t-1}, \xi_{t-2}, \dots$ . Vidíme tedy, že ačkoliv původní tvar rovnice vypadal jednoduše, ve skutečnosti se jedná o poměrně komplexní problém, kde se nám neznámé parametry  $\alpha$  a  $\beta$  vyskytují na mnoha místech.

Soustavu zkusíme nyní vyřešit v R numericky pomocí procedury `nleqslv` ze stejnojmenného balíčku. Poté, co získáme odhady  $\hat{\alpha}$  a  $\hat{\beta}$  neznámých parametrů  $\alpha$  a  $\beta$ , provedeme další iteraci algoritmu. Tedy znovu odhadneme  $\{\tau_t, t = 1, \dots, n\}$  pomocí jádrových odhadů, přičemž tentokrát již využijeme odhad krátkodobé složky  $\{\xi_t, t = 1, \dots, n\}$  z prvního kroku. Poté opětovným vyřešením soustavy aktualizujeme odhady parametrů  $\alpha$  a  $\beta$ , přičemž jako počáteční odhad pro numerickou proceduru používáme výsledek z předchozího kroku. Algoritmus zastavíme po dosažení vhodného konvergenčního kritéria.

V našem případě můžeme za konvergenční kritérium zvolit např. relativní změnu odhadů. Algoritmus tedy zastavíme ve chvíli, kdy

$$\frac{\hat{\alpha}[i+1] - \hat{\alpha}[i]}{\hat{\alpha}[i]} < k$$

a

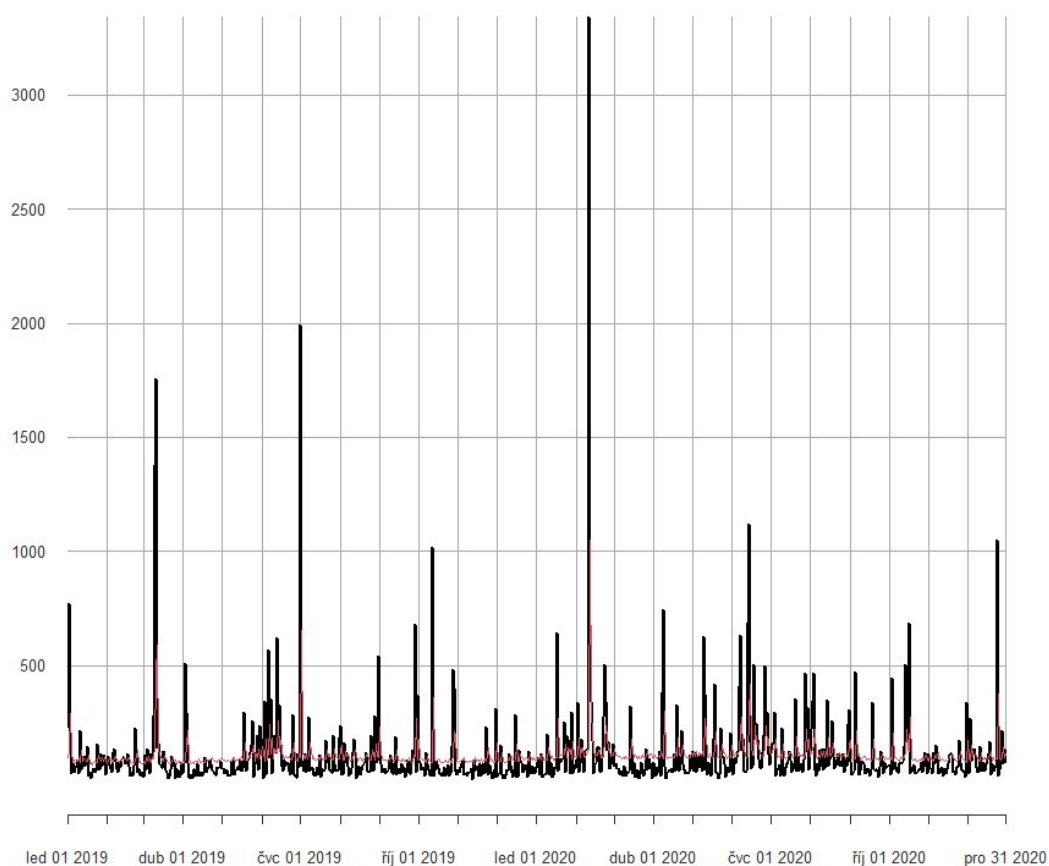
$$\frac{\hat{\beta}[i+1] - \hat{\beta}[i]}{\hat{\beta}[i]} < k,$$

kde  $\hat{\alpha}[i]$  a  $\hat{\beta}[i]$  jsou odhady parametrů z  $i$ -tého kroku a  $k$  je nějaká předem zvolená konstanta. Zvolíme např.  $k = 0.05$  a spustíme algoritmus. Ten se nám zastaví již při třetím kroku a získáme následující odhady:

$$\frac{\hat{\alpha}}{0.287} \quad \frac{\hat{\beta}}{0.043}$$

Tabulka 3.6: Odhady parametrů v modelu SpMEM pomocí zobecněné momentové metody a iteračního algoritmu.

V grafu si nyní vykreslíme data společně s vyrovnanými hodnotami.



Obrázek 3.12: Proložení dat pomocí modelu SpMEM.

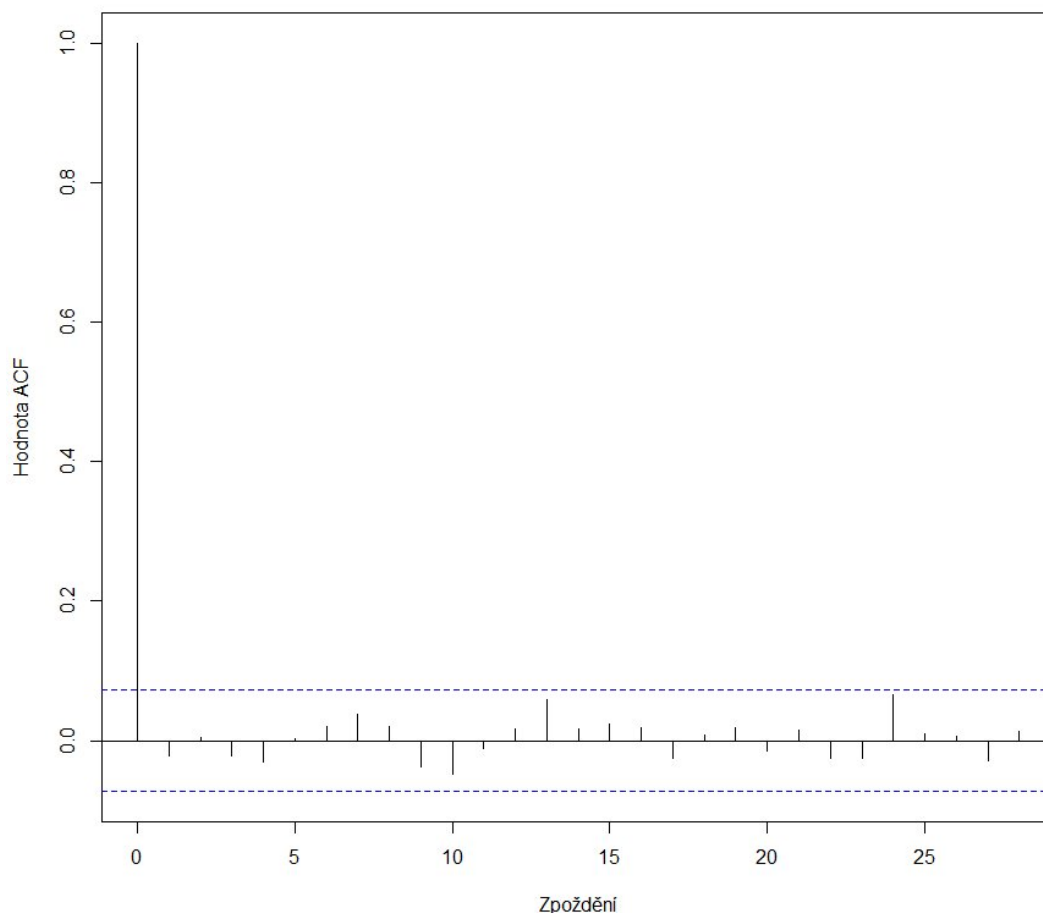
Na tomto místě by bylo vhodné porovnat kvalitu vyrovnání dat (předpovědi „in-sample“) se základním modelem MEM. Využijeme k tomu opět střední čtvercovou chybu.

MEM	SpMEM
40625.03	40356.28

Tabulka 3.7: Porovnání střední čtvercové chyby pro modely MEM a SpMEM a data škod z pojištění domů a domácností.

Jak je vidět, model SpMEM kvalitu proložení dat mírně zlepšil. Nyní provedeme základní diagnostiku. Vydělíme data vyrovnanými hodnotami a na výsledné multiplikační odchylky nejprve aplikujeme t-test, abychom otestovali předpoklad jednotkové střední hodnoty. Získáme p-hodnotu 0.995, tedy hypotézu nemůžeme zamítnout. Opět si vykreslíme i odhadnutou autokorelační funkci reziduí.

**Autokorelační funkce reziduí**



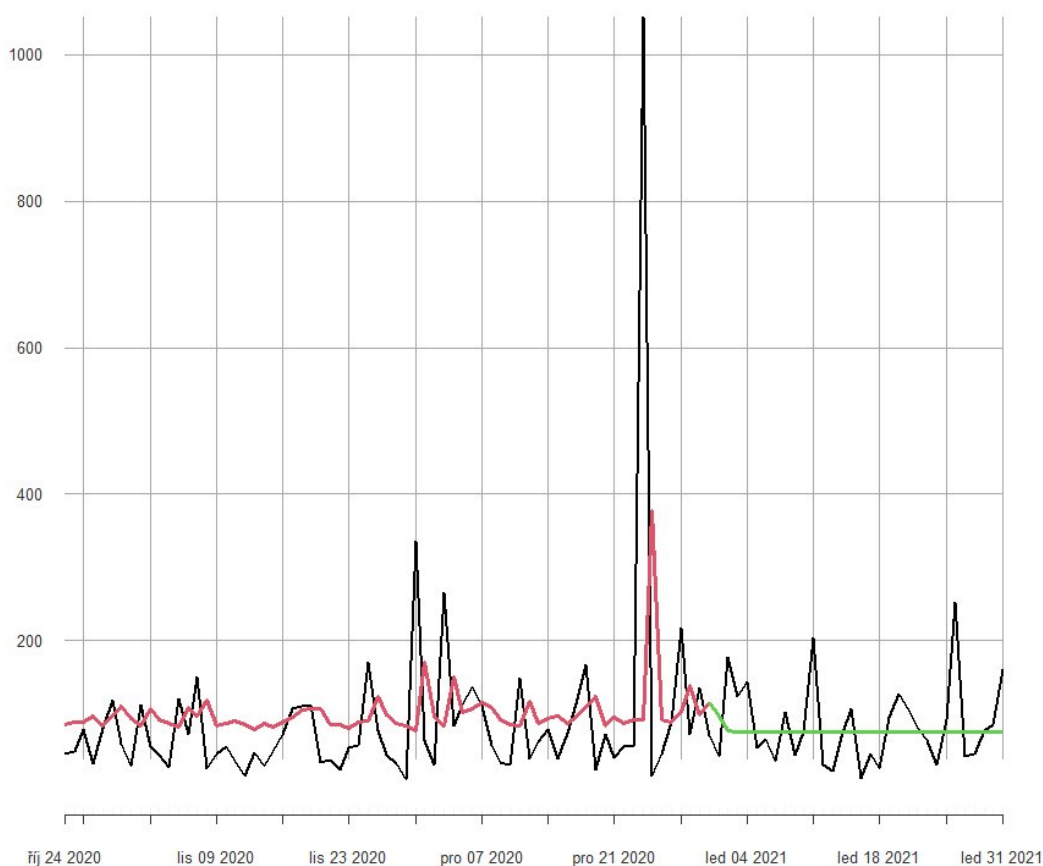
Obrázek 3.13: Odhadnutá autokorelační funkce reziduí modelu SpMEM.

Znovu se nezdá, že by rezidua byla významně korelovaná. Spustíme-li Ljungův-Boxův test (se zpožděním 13), dostaneme p-hodnotu 0.789. I tentokrát hypotézu nekorelovanosti nezamítáme. V následující tabulce si ještě porovnáme výsledky diagnostických testů pro modely MEM a SpMEM.

<b>Model</b>	<b>Studentův t-test</b>	<b>Ljungův-Boxův test</b>
MEM	0.981	0.649
SpMEM	0.995	0.789

Tabulka 3.8: Porovnání p-hodnot diagnostických testů pro modely MEM a SpMEM a data škod z pojištění domů a domácností.

Vidíme, že vyšší p-hodnoty jsme v obou případech získali u modelu SpMEM. Nyní spočteme predikce „out-of-sample“ na měsíc dopředu dle části 2.7 a stejně jako v části 3.2 si je vykreslíme do grafu se skutečnými naměřenými daty.



Obrázek 3.14: Predikce škod na jeden měsíc dopředu na základě modelu SpMEM a porovnání s reálnými daty.

Vzhledem k tvaru rovnice pro  $\xi_t$ , kterou jsme oproti rovnici střední hodnoty v základním modelu MEM téměř nezměnili, nelze očekávat výrazné zvýšení kvality předpovědi „out-of-sample“. Přesto zkusíme kvalitu předpovědi mimo vzorek porovnat se základním modelem MEM a s naivní předpovědí v podobě výběrového průměru. Opět k tomu využijeme střední čtvercovou chybu.

MEM	SpMEM	Průměr
3256.571	3262.091	3695.442

Tabulka 3.9: Porovnání střední čtvercové odchylky předpovědi mimo vzorek modelů MEM a SpMEM s naivní předpovědí za pomoci výběrového průměru.

Vidíme, že střední čtvercová chyba predikce na měsíc dopředu je téměř stejná jako u základního modelu MEM (dokonce se ještě nepatrně zvýšila). Nakonec se ještě pokusíme k odhadu využít metodu kvazi-maximální věrohodnosti tak, jak bylo popsáno v sekci 2.5.2. Do výrazu

$$\sum_{t=1}^n (\log(\varepsilon_t) - \varepsilon_t)$$



za  $\varepsilon_t$  dosadíme  $y_t / (\hat{\tau}_t \xi_t \hat{\mu})$  a opět provedeme numerickou maximalizaci daného výrazu pomocí procedury `optim` v R. Získáme tak odhady neznámých parametrů  $\alpha$  a  $\beta$  a postupujeme úplně stejně jako v případě zobecněné momentové metody. Iterační algoritmus se dle stejného konvergenčního kritéria jako v zobecněné momentové metodě znovu zastaví ve třetím kroku a získáme následující odhady.

$$\frac{\hat{\alpha}}{0.287} \quad \frac{\hat{\beta}}{0.043}$$

Tabulka 3.10: Odhady parametrů v modelu SpMEM pomocí metody kvazi-maximální a iteračního algoritmu.

Jak můžeme vidět, odhady zaokrouhlené na tři desetinná místa se od zobecněné momentové metody vůbec neliší a došli jsme tedy k téměř shodnému výsledku. V závěru si výsledky praktické části shrneme a navrhneme další možné postupy pro rozšíření aplikací modelů třídy MEM.

# Závěr

V této práci jsme se věnovali tzv. modelům multiplikatивních chyb (často se používá anglická zkratka **MEM**). Tyto modely jsou vhodné pro nezáporné časové řady a nejčastěji bývají aplikovány na data z finanční oblasti (např. ceny aktiv, objem obchodovaných aktiv apod.) Vyznačují se tím, že odchylky, které jsou podmíněně nezávislé a mají jednotkovou podmíněnou střední hodnotu, do modelu vstupují multiplikatивně, nikoliv aditivně, jak je běžné u jiných modelů.

V první kapitole jsme se krátce zaměřili na modely ARCH a GARCH, které s modely MEM úzce souvisí a nejčastěji se používají v souvislosti s volatilitou na finančních trzích.

Ve druhé kapitole jsme si již zavedli samotné modely MEM, a to pro jednorozměrný případ i pro vektory. Poté jsme si pro jednorozměrný model odvodili tvar logaritmické věrohodnosti v případě, že pro multiplikatивní odchylky předpokládáme exponenciální rozdělení. Z literatury též vyplývá, že tímto způsobem je možné odhady provádět i v případě, že odchylky ve skutečnosti exponenciální rozdělení nemají. Potom se jedná o tzv. kvazi-maximální věrohodnost. Ve větě 4 jsme si poté shrnuli některé důležité vlastnosti takového odhadu (konzistence a asymptotická normalita).

Následně jsme se přesunuli k modelům MEM, které jsou uzpůsobeny na výskyt nulových pozorování v datech (tzv. modely **ZA-MEM**). Taková pozorování můžeme často vidět v případě vysoké frekvence měření (např. v případě objemu obchodovaných akcií na burze za nějaké velmi krátké časové období). Pro multiplikatивní odchylky je v tomto modelu v literatuře použito tzv. zobecněné F-rozdělení a za tohoto předpokladu jsme si též odvodili tvar logaritmické věrohodnosti. Tento postup jsme poté ještě rozšířili na případ, kdy pravděpodobnost výskytu nulového pozorování v datech není konstantní, ale mění se s časem.

V další části druhé kapitoly jsme si představili možné přístupy k vektorovým modelům MEM. Závislost mezi jednotlivými složkami vektorů jsme si nejprve popsali za pomoci vícerozměrného gamma rozdělení a poté i pomocí kopul. Kopuly jsou velmi často využívány k popisu závislostní struktury ve vícerozměrném rozdělení, a to např. v oblasti pojišťovnictví při modelování škod z vícero pojistných odvětví. Po základním představení kopul jsme si odvodili tvar logaritmické věrohodnosti v případě, že pro vektory odchylek předpokládáme závislostní strukturu danou Gaussovskou kopulou a marginální gamma rozdělení.

V posledních částech druhé kapitoly jsme si představili tzv. semiparametrické modely MEM, znovu pro jednorozměrný i vektorový případ. Tyto modely se od základních modelů MEM liší tím, že v multiplikatивní rovnici pro pozorovanou veličinu se vyskytují další složky, z nichž jedna (v práci označovaná písmenem  $\tau$ ) zachycuje dlouhodobé, systematické a pomalu se vyvíjející změny a k jejímu odhadu využíváme tzv. jádrové odhady, zatímco druhá (označovaná jako  $\xi$ ) obsahuje krátkodobé, dočasné a nárazové změny a bývá parametrizovaná rovnicí

analogickou k modelu GARCH. Opět jsme se zaměřili na statistickou inferenci a postupně jsme si představili možnosti využití zobecněné momentové metody a metody kvazi-maximální věrohodnosti (tentokrát založené na předpokladu gamma rozdělení multiplikativních odchylek) v rámci iteračního algoritmu. Úplný konec teoretické části práce se potom věnuje krátkému pojednání o předpovědích v semiparametrických modelech MEM.

Ve třetí a poslední kapitole práce jsme se již věnovali aplikaci představených modelů na reálná data. Využili jsme přitom data škod z oblasti pojišťovnictví, která představují kladnou či nezápornou časovou řadu a jsou proto k využití modelů vhodná. Z důvodu výpočetní náročnosti jsme se omezili pouze na jedno-rozměrné modely z třídy MEM.

Nejprve jsme data denních škod z pojištění domů a domácností proložili základním modelem MEM s rovnicí střední hodnoty podobající se modelu GARCH. Pomocí procedury „optim“ v softwaru R jsme numericky maximalizovali logaritmickou věrohodnost vycházející z předpokladu exponenciálního rozdělení multiplikativních odchylek a získali jsme odhady neznámých parametrů v rovnici pro střední hodnotu. Optimalizace probíhala rychle a výsledné odhady jsou stabilní i vzhledem ke změnám počátečních odhadů, které je nutné do procedury v R zadat před jejím spuštěním. Proložení dat se též zdá být úspěšné. Model dobře zachycuje povahu dat a základní diagnostika nevykazuje problémy. Jednoduchá rovnice střední hodnoty však není příliš vhodná k predikci „out-of-sample“, a to především v dlouhodobém horizontu.

Následně jsme se zaměřili na modely MEM rozšířené v nule. K tomuto účelu bylo nutné využít data, která v sobě nuly obsahují. Zvolili jsme proto data škod z úvěrového pojištění, které pro pojišťovnu, od níž data pochází, patří k nejméně frekventovaným, a tedy i frekvence škod je velmi nízká. Tato data jsme navíc převedli na čtvrtletní bázi. Ve článku (Hautsch a kol., 2014), ze kterého čerpáme, se pro odchylky používá zobecněné F-rozdělení, jež v sobě obsahuje mnoho dalších parametrů. Numerická optimalizace logaritmické věrohodnosti se tak oproti základnímu modelu MEM značně ztížila. Článek navíc neuvažuje žádná omezení pro parametry v rovnici střední hodnoty, což optimalizaci dále komplikuje. Nakonec se nám nepodařilo získat vhodné odhady, které by byly přijatelně stabilní vůči změnám v počátečních odhadech a bylo by možné s nimi pracovat.

Bylo proto nutné problém zjednodušit, k čemuž jsme využili metodu exponenciální kvazi-maximální věrohodnosti. Po aplikaci vhodných omezení na parametry rovnice střední hodnoty se nám nakonec podařilo odhady získat. Základní diagnostické testy ukázaly, že multiplikativní odchylky vzniklé vydělením dat vyrovnanými hodnotami zřejmě porušují předpoklad jednotkové střední hodnoty, zatímco předpoklad nekorelovanosti se zdá být zachován. Predikci mimo vzorek jsme tentokrát počítali na 1 rok (4 čtvrtletí) dopředu.

Nakonec jsme se posunuli k semiparametrickým modelům MEM, které jsme kvůli možnosti porovnání aplikovali na stejná data z pojištění domů a domácností jako základní modely MEM. V R jsme implementovali iterační algoritmus, který

postupně aktualizuje odhady parametrů z rovnice pro krátkodobou složku  $\xi$ , dokud se dostatečně nestabilizují, přičemž samotné odhadování těchto parametrů probíhá za pomoci zobecněné momentové metody nebo metody kvazi-maximální věrohodnosti založené na předpokladu podmíněného gamma rozdělení multiplikačních odchylek. Zjistili jsme, že obě metody vedou k téměř shodnému výsledku. Model zachycuje data ještě o něco lépe než základní model MEM (střední čtvercová chyba je nižší, zatímco p-hodnoty diagnostických testů jsou vyšší), avšak predikce „out-of-sample“ opět není v dlouhodobém horizontu použitelná, protože pro  $\xi$  se znovu předpokládá parametrizace podobná modelu GARCH.

Co se týče dalšího výzkumu, přirozeným rozšířením by bylo využití vektorových modelů MEM k modelování více vzájemně závislých pojistných odvětví najednou (např. pomocí kopul). Vícerozměrným modelům jsme se v praktické části záměrně vyhnuli kvůli výraznému zvýšení numerické složitosti, avšak v teoretické části práce jsou potřebné postupy popsány a odvozeny.

Dalším logickým krokem by mohlo být uvažování složitějších rovnic pro střední hodnotu nebo krátkodobou složku, které by umožnily lepší využitelnost k předpovědím mimo vzorek. Bylo by též možné v rovnici uvažovat nějakou externí proměnnou, pokud bychom k tomu měli vhodná data.

Práce tedy může sloužit i jako základ pro další využití modelů multiplikačních chyb v pojištění i v jiných oborech, kde se pracuje s nezápornými časovými řadami.

# Seznam použité literatury

- BOLLERSLEV, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**(3), 307–327.
- BROWNLESS, C. T., CIPOLLINI, F. a GALLO, G. M. (2012). Multiplicative Error Models.
- CIPOLLINI, F. a GALLO, G. M. (2022). Multiplicative Error Models: 20 Years on.
- CIPOLLINI, F., ENGLE, R. F. a GALLO, G. M. (2006). Vector Multiplicative Error Models: Representation and Inference. Technical report, National Bureau of Economic Research.
- ENGLE, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, **50**(4), 987–1007.
- ENGLE, R. F. (2002). New Frontiers for ARCH Models. *Journal of Financial Econometrics*, **17**(5), 425–446.
- ENGLE, R. F. a GALLO, G. M. (2006). A Multiple Indicators Model for Volatility Using Intra-daily Data. *Journal of Econometrics*, **131**(1), 3–27.
- HAUTSCH, M., MALEC, P. a SCHIENLE, M. (2014). Capturing the Zero: a New Class of Zero-Augmented Distributions and Multiplicative Error Processes. *Journal of Financial Econometrics*, **12**(1), 89–121.
- HUDECOVÁ, a PEŠTA, M. (2024). Quasi-likelihood Estimation in Volatility Models for Semi-continuous Time Series. *Journal of Time Series Analysis*, **10.1111**.
- LEE, S. W. a HANSEN, B. E. (1994). Asymptotic Theory for the GARCH (1, 1) Quasi-maximum Likelihood Estimator. *Econometric Theory*, **10**(1), 29–52.
- NELSEN, R. B. (2006). *An Introduction to Copulas*. Springer. ISBN 0-387-28659-4.
- NELSON, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica: Journal of the Econometric Society*, **59**(2), 347–370.
- XUE-KUN SONG, P. (2000). Multivariate Dispersion Models Generated from Gaussian Copula. *Scandinavian Journal of Statistics*, **27**(2), 305–320.

# Seznam obrázků

3.1	Nahlášené a vyplacené škody v pojištění domů a domácností od roku 2008. . . . .	26
3.2	Nahlášené a vyplacené škody v pojištění odpovědnosti od roku 2008.	27
3.3	Nahlášené a vyplacené škody v cestovním pojištění od roku 2008.	28
3.4	Nahlášené a vyplacené škody v úvěrovém pojištění od roku 2008. .	29
3.5	Nahlášené a vyplacené škody v pojištění domů a domácností v letech 2019 a 2020. . . . .	30
3.6	Proložení dat pomocí základního modelu MEM. . . . .	32
3.7	Odhadnutá autokorelační funkce reziduí modelu MEM. . . . .	33
3.8	Predikce škod na jeden měsíc dopředu na základě modelu MEM a porovnání s reálnými daty. . . . .	34
3.9	Čtvrtletní časová řada škod v úvěrovém pojištění od roku 2010 do roku 2021. . . . .	35
3.10	Proložení dat pomocí modelu ZA-MEM. . . . .	38
3.11	Odhadnutá autokorelační funkce reziduí modelu ZA-MEM. . . . .	39
3.12	Proložení dat pomocí modelu SpMEM. . . . .	42
3.13	Odhadnutá autokorelační funkce reziduí modelu SpMEM. . . . .	43
3.14	Predikce škod na jeden měsíc dopředu na základě modelu SpMEM a porovnání s reálnými daty. . . . .	44

# Seznam tabulek

3.1	Základní statistické charakteristiky škod z pojištění domů a domácností v letech 2019 a 2020. . . . .	31
3.2	Odhady parametrů v základním modelu MEM metodou maximální věrohodnosti. . . . .	31
3.3	Základní statistické charakteristiky čtvrtletních škod z úvěrového pojištění v letech 2010 až 2021. . . . .	35
3.4	Odhady parametrů v modelu ZA-MEM metodou exponenciální kvazi-maximální věrohodnosti (QML). . . . .	37
3.5	Predikce mimo vzorek na rok 2022 pomocí modelu ZA-MEM a porovnání se skutečnými pozorovanými hodnotami. . . . .	39
3.6	Odhady parametrů v modelu SpMEM pomocí zobecněné momentové metody a iteračního algoritmu. . . . .	41
3.7	Porovnání střední čtvercové chyby pro modely MEM a SpMEM a data škod z pojištění domů a domácností. . . . .	42
3.8	Porovnání p-hodnot diagnostických testů pro modely MEM a SpMEM a data škod z pojištění domů a domácností. . . . .	43
3.9	Porovnání střední čtvercové odchylky předpovědi mimo vzorek modelů MEM a SpMEM s naivní předpovědí za pomoci výběrového průměru. . . . .	44
3.10	Odhady parametrů v modelu SpMEM pomocí metody kvazi-maximální a iteračního algoritmu. . . . .	45

# A. Přílohy

## A.1 Důležité části kódu

### A.1.1 Odhad parametrů základního modelu metodou maximální věrohodnosti

```
x=Household_selected$Claim_ths

likelihood <- function(parameters, data) {
  omega<-parameters[1]
  alpha<-parameters[2]
  bet<-parameters[3]

  n <- length(x)
  mu <- numeric(n)
  mu[1] <- mean(x)

  for (t in 2:n) {
    mu[t]<-omega+alpha*data[t-1]+bet*mu[t-1]
  }

  loglikelihood <-sum(-log(mu[1:n])-data[1:n]/mu[1:n])
  return(loglikelihood)
}

initial_guess<-c(10,0.5,0.5)
result<-optim(par=initial_guess, control = list(fnscale = -1),
fn=likelihood,data=x, method="L-BFGS-B",
lower=c(0,0,0), upper=c(Inf,Inf,Inf))
```

### A.1.2 Odhad parametrů v modelu ZA-MEM pomocí exponenciální kvazi-maximální věrohodnosti

```
x=Credit_quarter$Claim_ths
nearzero<-.Machine$double.eps

likelihoodZAMEM_QML <- function(parameters, data) {
  zeros<- sum(data==0)
  non_zero<-which(data!=0)
  nonzeros<-length(non_zero)

  p<-parameters[1]
  omega<-parameters[2]
  alpha<-parameters[3]
  alphazero<-parameters[4]
```



```

bet<-parameters[5]

n <- length(data)
mu_ZAMEM_QML <- numeric(n)
mu_ZAMEM_QML[1] <- mean(data)

for (t in 2:n) {
  if (data[t-1]>0) {
    mu_ZAMEM_QML[t]=exp(omega
      +alpha*log(data[t-1]/mu_ZAMEM_QML[t-1])
      +bet*log(mu_ZAMEM_QML[t-1]))
  } else {
    mu_ZAMEM_QML[t]=exp(omega+alphazero
      +bet*log(mu_ZAMEM_QML[t-1]))
  }
}
loglikelihood <-zeros*log(p)+(n-zeros)*log(1-p)
-sum(data[non_zero]/mu_ZAMEM_QML[non_zero])
-n*sum(log(mu_ZAMEM_QML[1:n]))

return(loglikelihood)
}

initial_guess<-c(0.5,2,0.3,0.4,0.4)
resultZAMEM_QML<-optim(par=initial_guess,
  control = list(fnscale = -1), fn=likelihoodZAMEM_QML,data=x,
  method="L-BFGS-B", lower=c(nearzero,-10,-1,-10,-1),
  upper=c(1-nearzero,10,1,10,1))

```

### A.1.3 Odhad parametrů v modelu SpMEM pomocí zobecněné momentové metody

```

x=Household_selected$Claim_ths

mu=mean(x)
x_tau= x/mu
n<-length(x_tau)
width<-dpik(x_tau)

kernels=c()
tau_kernel=c() #jádrový odhad tau
for (t in 1:n) {
  for(s in 1:n) {
    kernels[s]=dnorm((t/n-s/n)/width)
  }
  tau_kernel[t]=sum(x_tau*kernels)/sum(kernels)
}

```

```

#výpočet ksi
compute_ksi <- function(parameters, data) {
  alpha<-parameters[1]
  bet<-parameters[2]
  ksi<-numeric(length(data))
  ksi[1]<-1
  n <- length(data)
  for(t in 2:n){
    ksi[t]<-(1-(bet+alpha))+bet*ksi[t-1]+alpha*data[t-1]
  }
  return(ksi)
}

x_ksi=x/(mu*tau_kernel)

#výpočet derivace ksi
compute_ksi_derived <- function(parameters, data) {
  alpha<-parameters[1]
  bet<-parameters[2]
  ksi_derived<-matrix(0,nrow=2, ncol=length(data))
  ksi<-compute_ksi(parameters, data)

  for (i in 2:n) {
    ksi_derived[1,i]<- -1+bet*ksi_derived[1,i-1]+data[i-1]
    ksi_derived[2,i]<- -1+bet*ksi_derived[2,i-1]+ksi[i-1]
  }
  return(ksi_derived)
}

#výpočet epsilon
compute_epsilon<- function(parameters, data) {
  alpha<-parameters[1]
  bet<-parameters[2]
  epsilon=data/compute_ksi(parameters, data)
  return(epsilon)
}

#výpočet matice a
compute_a<- function(parameters, data) {
  alpha<-parameters[1]
  bet<-parameters[2]
  n<-length(data)
  a<-matrix(0,nrow=2, ncol=length(data))
  ksi_derived<-compute_ksi_derived(parameters,data)
  ksi<-compute_ksi(parameters, data)
  for (i in 1:n) {
    a[1,i] <- ksi_derived[1,i]/ksi[i]
    a[2,i]<- ksi_derived[2,i]/ksi[i]
  }
}

```

```

    }
    return(a)
}

#rovnice pro odhad parametrů
equation<-function(parameters,data){
  alpha<-parameters[1]
  bet<-parameters[2]
  clenyl<-matrix(0,nrow=2, ncol=length(data))
  a<-compute_a(parameters,data)
  for (i in 1:n) {
    clenyl[1,i] <-a[1,i]*(compute_epsilon(parameters,data)[i]-1)
    clenyl[2,i]<- a[2,i]*(compute_epsilon(parameters,data)[i]-1)
  }
  eq<-matrix(0,nrow=2, ncol=1)
  for (i in 1:2){
    eq[1,1]<-rowSums(clenyl)[1]
    eq[2,1]<-rowSums(clenyl)[2]
  }
  return(eq)
}

#odhad parametrů - první krok
initial_guess_gmm<-c(0.3,0.7)
result_gmm1<- nleqslv(initial_guess_gmm, equation,data=x_ksi)
ksi_step_1<-compute_ksi(result_gmm1$x,x_ksi)

#opakování postupu - druhý krok algoritmu
x_tau= x/(mu*ksi_step_1)
width<-dpik(x_tau)

kernels=c()
tau_kernel=c()
for (t in 1:n) {
  for(s in 1:n) {
    kernels[s]=dnorm((t/n-s/n)/width)
  }
  tau_kernel[t]=sum(x_tau*kernels)/sum(kernels)
}
x_ksi=x/(mu*tau_kernel)
initial_guess_gmm2<-result_gmm1$x
result_gmm2 <-nleqslv(initial_guess_gmm2, equation,data=x_ksi)
ksi_step_2<-compute_ksi(result_gmm2$x,x_ksi)

#kontrola konvergenčního kritéria
(result_gmm2$x-result_gmm1$x)/result_gmm1$x

#opakování postupu - třetí krok algoritmu

```

```

x_tau= x/(mu*ksi_step_2)
x_tau
width<-dpik(x_tau)

kernels=c()
tau_kernel=c()
for (t in 1:n) {
  for(s in 1:n) {
    kernels[s]=dnorm((t/n-s/n)/width)
  }
  tau_kernel[t]=sum(x_tau*kernels)/sum(kernels)
}

x_ksi=x/(mu*tau_kernel)

initial_guess_gmm3<-result_gmm2$x
result_gmm3<- nleqslv(initial_guess_gmm3, equation,data=x_ksi)
ksi_step_3<-compute_ksi(result_gmm3$x,x_ksi)

#kontrola konvergenčního kritéria
(result_gmm3$x-result_gmm2$x)/(result_gmm2$x)

```

#### A.1.4 Odhad parametrů v modelu SpMEM pomocí kvazi-maximální věrohodnosti s rozdělením gamma

```

x=Household_selected$Claim_ths
mu=mean(x)
x_tau= x/mu
width<-dpik(x_tau)

kernels=c()
tau_kernel=c() #jádrový odhad tau
for (t in 1:n) {
  for(s in 1:n) {
    kernels[s]=dnorm((t/n-s/n)/width)
  }
  tau_kernel[t]=sum(x_tau*kernels)/sum(kernels)
}

x_ksi=x/(mu*tau_kernel)

#definice kvazi-věrohodnosti
quasi_likelihood<-function(parameters, data) {
  alpha<-parameters[1]
  bet<-parameters[2]
  likelihood<-sum(log(compute_epsilon(parameters, data))-
compute_epsilon(parameters, data))
}

```

```

    return(likelihood)
}

#maximalizace - odhad parametrů rovnice ksi
initial_guess<-c(0.2,0.5)
result_quasi_step1<-optim(par=initial_guess,
control = list(fnscale = -1),
fn=quasi_likelihood,data=x_ksi)
ksi_step_1_quasi<-compute_ksi(result_quasi_step1$par,x_ksi)

#opakování postupu - druhý krok algoritmu
x_tau=x/(mu*ksi_step_1_quasi)
width<-dpik(x_tau)

kernels=c()
tau_kernel=c()
for (t in 1:n) {
  for(s in 1:n) {
    kernels[s]=dnorm((t/n-s/n)/width)
  }
  tau_kernel[t]=sum(x_tau*kernels)/sum(kernels)
}

x_ksi=x/(mu*tau_kernel)
initial_guess_quasi2<-result_quasi_step1$par
result_quasi_step2 <-optim(par=initial_guess_quasi2,
control = list(fnscale = -1),
fn=quasi_likelihood,data=x_ksi)
ksi_step_2_quasi<-compute_ksi(result_quasi_step2$par,x_ksi)

#kontrola konvergenčního kritéria
(result_quasi_step2$par-result_quasi_step1$par)
/(result_quasi_step1$par)

#opakování postupu - třetí krok algoritmu
x_tau=x/(mu*ksi_step_2_quasi)
width<-dpik(x_tau)

kernels=c()
tau_kernel=c()
for (t in 1:n) {
  for(s in 1:n) {
    kernels[s]=dnorm((t/n-s/n)/width)
  }
  tau_kernel[t]=sum(x_tau*kernels)/sum(kernels)
}

```

```
x_ksi=x/(mu*tau_kernel)
initial_guess_quasi3<-result_quasi_step2$par
result_quasi_step3 <-optim(par=initial_guess_quasi3,
control = list(fnscale = -1),
fn=quasi_likelihood,data=x_ksi)
result_quasi_step3

#kontrola konvergenčného kritéria
(result_quasi_step3$par-result_quasi_step2$par)
/(result_quasi_step2$par)
```