



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLLOMOVÁ PRÁCE

Luboš Krnáč

**Ověřování předpokladů lineárního
smíšeného modelu**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2021

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

S radosťou by som chcel poďakovať svojmu vedúcemu doc. Mgr. Michalovi Kuličovi, Ph.D. za nezameniteľné rady, ochotnú pomoc pri problémoch a jeho dôraz, ktorý dovedol túto prácu do podoby, ktorá je zrelá na svetlo sveta.

Název práce: Ověřování předpokladů lineárního smíšeného modelu

Autor: Luboš Krnáč

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Diplomová práca sa zaoberá lineárnymi zmiešanými modelmi. V prvej časti sa venujeme metódam odhadov parametrov modelu a testom hypotéz. V druhej časti je kladený dôraz na vplyv porušení jednotlivých predpokladov na vhodné diagnostické grafy pre reziduá a odhady náhodných efektov. Bližšie je opísané, akým spôsobom ktoré porušenie vplyva na diagnostické grafy. V tretej kapitole sa venujeme dôsledkom porušení predpokladov na odhady parametrov. Taktiež sú pre dané porušenia odhadnuté sily a hladiny testov pre fixné efekty modelu.

Klíčová slova: lineárna regresia, lineárny zmiešaný model, diagnostika porušení predpokladov, vplyv porušení predpokladov na odhady parametrov a testy hypotéz

Title: Verification of linear mixed model assumptions

Author: Luboš Krnáč

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Mgr. Michal Kulich, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The diploma thesis deals with linear mixed effects models. In the first chapter, we discuss parameter estimation and hypothesis testing in the linear mixed effects models. The second chapter is dedicated to graphical diagnostics. We look at the suitable diagnostic plots for residuals and random effects estimates. It is closely described, how the violations of assumptions affect the diagnostic plots. In the third chapter we have consequences of the violations of assumptions on the parameter estimates and results of hypothesis testing for fixed effects.

Keywords: linear regression, linear mixed effects model, diagnostics of assumption violations, influence of assumption violations on parameter estimates and hypothesis testing

Obsah

Úvod	3
1 Lineárny zmiešaný model	4
1.1 Motivácia	4
1.1.1 Lineárny model	4
1.1.2 Model analýzy rozptylu (ANOVA model)	5
1.1.3 ANOVA s náhodnými efektami (VARCOMP model)	5
1.2 Definícia zmiešaného lineárneho modelu	6
1.2.1 Jednostupňový lineárny model so zmiešanými efektami	6
1.2.2 Príklady	7
1.3 Odhady parametrov LME modelu	7
1.3.1 Odhady pevných efektov	8
1.3.2 Odhady kovariančnej štruktúry	10
1.4 Testovanie hypotéz v jednostupňovom zmiešanom modeli	18
1.4.1 Pevné efekty	19
1.4.2 Kovariančná štruktúra	19
2 Overovanie predpokladov LME modelu	21
2.1 Predpoklady a diagnostika	21
2.2 Simulácie	22
2.2.1 Model s náhodným počiatkom	23
2.2.2 Model s náhodným sklonom	38
2.2.3 Model s náhodným počiatkom a sklonom	40
3 Vplyv porušení predpokladov na hodnoty odhadov a závery testov hypotéz	52
3.1 Odhady parametrov	53
3.2 Sila a hladina testu	57
Záver	60
Zoznam použitej literatúry	62
Zoznam obrázkov	64
Zoznam tabuliek	68

A	Appendix	70
A.1	Odvozenie aposteriórneho rozdelenia náhodných efektov	70
A.2	Vzorcie na redukciiu dimenzie problému	71
A.3	Vzťahy pre maticové derivácie	71
A.4	Pearsonove rozdelenie typu IV	71
A.5	Grafy modelov s náhodným počiatkom s kombináciou porušení predpokladov	72
A.6	Mnohorozmerné Studentovo rozdelenie	82
A.7	Výsledky simulácii odhadov parametrov	82

Úvod

V tejto práci sa budeme zaoberať nadstavbou lineárnych modelov, lineárnymi modelmi so zmiešanými efektami.

V prvej kapitole sa nachádza samotná definícia lineárneho modelu so zmiešanými efektami spolu s metódami odhadu. Metódy sú rozdelené podľa parametrov, ktoré odhadujú, na metódy pre fixné efekty (GLS, Hendersonove rovnice, Bayesovské empirické odhady) a metódy na odhad kovariančnej štruktúry náhodných efektov (profilové MLE I, profilové MLE II, REML, MINQUE). Na záver kapitoly sú uvedené testy hypotéz o fixných efektoch, o celkovej prítomnosti náhodných efektov v modeli a LR testy pre prítomnosť jednotlivých náhodných efektov.

V druhej kapitole sa pozrieme na diagnostiku LME modelov za prítomnosti určitých porušení predpokladov modelu. Porušenia predpokladov sa týkajú normality chybovej zložky ε_{ij} a nezávislosti chybovej zložky ε_{ij} vrámci subjektov. Pre náhodné efekty \mathbf{b}_i sú to porušenia predpokladov normality. Skúmame množstvo modelov, v ktorých sú porušené predpoklady pre ε_{ij} , alebo pre \mathbf{b}_i , alebo ich kombinácia. Na diagnostiku sú použité histogramy, scatterploty, QQ-grafy a grafy autokorelačnej funkcie, aby bolo možné rozoznať každé porušenie na aspoň jednom z grafov. Uvažované modely, na ktorých boli diagnostické grafy skúmané, sú model s náhodným počiatkom, model s náhodným sklonom a model s náhodným počiatkom aj sklonom. Tiež boli uvažované dve veľkosti týchto modelov (20 subjektov a 200 subjektov), pretože v praxi sú bežné modely s malým počtom subjektov (< 20), ale porušenia predpokladov sa výraznejšie prejavujú na veľkých rozsahoch dát.

V tretej kapitole sa venujeme rovnakým modelom, ako v druhej kapitole s tým rozdielom, že sledujeme odhady jednotlivých parametrov. Pre každé porušenie predpokladov sú odhadnuté parametre modelu - fixné efekty a kovariančná štruktúra. Tieto odhady sú opakované 1000-krát a štatistiky o rozdeleniach odhadov sú zozbierané do tabuliek. V nich sledujeme vychýlenia jednotlivých priemerných odhadov, smerodajné odchýlky, empirické smerodajné odchýlky, prípadne pravdepodobnosť pokrytia skutočnej hodnoty 95% intervalom spoľahlivosti. Navyše pre fixné efekty sledujeme reálnu hladinu a silu testu hypotézy $\mathbf{H}_0 : \beta_j = 0$ proti $\mathbf{H}_A : \beta_j \neq 0$ pre $j = 0, 1$ na hladine $\alpha = 0,05$.

Cieľom tejto práce je poukázať na dôležitosť grafickej diagnostiky pri modelovaní lineárnych modelov so zmiešanými efektami a kvantifikácia vplyvu porušenia predpokladov na hodnoty odhadov jednotlivých parametrov a na závery hypotéz testov.

“Essentially, all models are wrong, but some are useful. The practical question is how wrong do they have to be to not be useful.”

George E. P. Box (1919 – 2013)

Kapitola 1

Lineárny zmiešaný model

1.1 Motivácia

Základným predpokladom lineárnych regresných modelov je nezávislosť jednotlivých pozorovaní. V mnohých situáciách ale tento predpoklad nemôžeme považovať za splnený. Jedným z mnohých prípadov sú skupinové dáta (*clustered data*), kde je v pozorovaniach prítomná určitá príslušnosť do skupín, určitý typ hierarchie.

Uvažujme dve banky B_1 a B_2 . Banka B_1 vyžaduje na založenie účtu vklad, povedzme v rádoch stoviek tisíc CZK, banka B_2 nevyžaduje na založenie účtu žiadny, alebo minimálny vklad. Úlohou je odhadnúť výšku peňažného obnosu na bankovom účte na základe rôznych parametrov (vek, pohlavie, platová skupina, rôzne solventnostné vlastnosti klientov atď.). V klasickej lineárnej regresii predpokladáme, že jednotlivé pozorovania (v tomto prípade peňažné obnosy jednotlivých klientov) sú od seba nezávislé. Lenže v takto postavenej situácii je možné predpokladať, že klienti vrámci jednotlivých bánk nebudú nezávislí. Klienti v banke s počiatočným vkladom budú mať z princípu vyššie obnosy na účte, podobné zhodnotenie peňazí atď.

Iným zdrojom závislosti pozorovaní môžu byť opakované pozorovania (merania) na tom istom subjekte, ktoré sú obzvlášť bežné v medicínskych aplikáciách (napríklad opakované merania tlaku u obvodného doktora atď.).

Pre tento prípad uvažujme banku, ktorá robí opakované prepočty solventnostných pozícií (napr. pre odhady pravdepodobnosti nesplatenia úveru) pre svojich klientov. Je zjavné, že hodnoty solventnosti konkrétneho klienta v čase budú závisieť napríklad na predchádzajúcej solventnostnej pozícii, a preto nemôžeme takéto pozorovania považovať za nezávislé.

V týchto prípadoch je obyčajný lineárny model definovaný nižšie v (1.1) nedostatočným pre pravdivú interpretáciu dát.

1.1.1 Lineárny model

Nech $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ predstavuje vektor pozorovaní, $n > 1$. Nech \mathbb{X} je regresná matica (matica regresorov) $n \times p$, nech $\boldsymbol{\beta} \in \mathbb{R}^p$. Nech $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_n)$, $\sigma^2 > 0$. Potom (*klasickým*) *normálnym lineárnym modelom* nazývame vzťah

$$Y_i = x_i \boldsymbol{\beta} + \varepsilon_i,$$

$i = 1, \dots, n$, x_i značí i -ty riadok matice \mathbb{X} . Sumárne zapísané

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Jedným z pohľadov na lineárny zmiešaný model môže byť nasledujúci.

1.1.2 Model analýzy rozptylu (ANOVA model)

Nech $Y_{ij} \sim \mathcal{N}(\beta_i, \sigma_e^2)$ sú nezávislé náhodné veličiny, kde i označuje i -tý subjekt/skupinu/úroveň kategorickej veličiny, $i = 1, \dots, K$, j počet pozorovaní vrámci skupiny, $j = 1, \dots, n_i$ a $\sum_{i=1}^K n_i = N$ je rozsah výberu. V praxi častým prípadom je práve posledná možnosť zo spomínaných, teda kategorický regresor, kde nás zaujímajú rozdiely medzi jednotlivými kategóriami. V takom prípade naším primárnym záujmom je porovnanie stredných hodnôt kategórií. Model má potom tvar

$$Y_{ij} = \beta_i + \varepsilon_{ij},$$

kde β_i , *pevné*, predstavujú populačné stredné hodnoty pre jednotlivé kategórie $i = 1, \dots, K$, $\mathbb{E}\varepsilon_{ij} = 0$ a $\text{Var } \varepsilon_{ij} = \sigma_e^2$. Klasickým problémom, ktorým sa ANOVA model zaoberá, je test hypotézy

$$H_0 : \beta_1 = \dots = \beta_K,$$

teda, že stredné hodnoty $\mathbb{E}Y_{ij}$ sú rovnaké vo všetkých kategóriách. Podstatným predpokladom klasického problému analýzy rozptylu je nezávislosť pozorovaní. Určitú formu závislosti ale môžeme do modelu vnieť pomocou nasledujúcej úvahy.

1.1.3 ANOVA s náhodnými efektami (VARCOMP model)

Nech Y_{ij} majú zmysel ako vyššie. Teraz ale miesto pevného priemeru kategórie i , β_i , uvažujeme $\beta_i = \beta + b_i$, kde $\beta \in \mathbb{R}$, b_i je náhodná veličina, $\mathbb{E}b_i = 0$ a $\text{Var}b_i = \sigma_b^2$ a vektory $(b_1, \dots, b_K)^\top$ a $(\varepsilon_{11}, \dots, \varepsilon_{Kn_K})^\top$ sú nezávislé. Model je potom tvaru

$$Y_{ij} = \beta + b_i + \varepsilon_{ij},$$

kde $\mathbb{E}\varepsilon_{ij} = 0$ a $\text{Var } \varepsilon_{ij} = \sigma_e^2$. Jednoduchým výpočtom dostaneme podmienené momenty

$$\begin{aligned} \mathbb{E}[Y_{ij}|b_i] &= \beta + b_i, \\ \mathbb{E}Y_{ij} &= \beta, \\ \text{Var}[Y_{ij}|b_i] &= \sigma_e^2, \\ \text{Var } Y_{ij} &= \sigma_b^2 + \sigma_e^2. \end{aligned}$$

Potom pre kovarianciu a koreláciu platí

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{i',j'}) &= 0 && \text{pre } i \neq i', \\ \text{cov}(Y_{ij}, Y_{i,j'}) &= \text{Var } b_i = \sigma_b^2 && \text{pre } j \neq j', \\ \text{cor}(Y_{ij}, Y_{i,j'}) &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2} && \text{pre } j \neq j'. \end{aligned}$$

Je vidieť, že kovariancia (a teda aj korelácia) medzi dvomi pozorovaniami vrámci subjektu je nenulová a rovná $\sigma_b^2 \left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2} \right)$. VARCOMP model preto zachytáva možnú závislostnú štruktúru dát a vie s ňou pracovať tam, kde predpoklad nezávislosti v modeli analýzy rozptylu neplatí.

1.2 Definícia zmiešaného lineárneho modelu

Kombináciou modelu analýzy rozptylu (fixné efekty), VARCOMP modelu (náhodné efekty) sa dostávame k **jednostupňovému lineárnemu modelu so zmiešanými efektami**¹. Tento model zachytáva podstatu modelu analýzy rozptylu (populácia rozdelená do kategórií, resp. na subjekty, vrámci ktorých sú pozorovania zachytávané) s modelom VARCOMP, ktorý uvažuje určitú závislosť vrámci kategórie, resp. subjektu a (viacnásobnej) lineárnej regresie, kedy do modelu vstupuje množstvo ďalších regresorov.

1.2.1 Jednostupňový lineárny model so zmiešanými efektami

Uvažujme nezávislé subjekty $i \in \{1, \dots, K\}$ a označme $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ a $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^\top$. Nech \mathbb{X}_i je $n_i \times p$ regresná matica pre pevné efekty² a \mathbb{Z}_i je $n_i \times q$ regresná matica pre náhodné efekty³.

Definícia 1.1

Vektory odozvy $\mathbf{Y}_1, \dots, \mathbf{Y}_K$ splňujú jednostupňový model so zmiešanými efektami, ak

$$\mathbf{Y}_i = \mathbb{X}_i \boldsymbol{\beta} + \mathbb{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, K, \quad (1.1)$$

kde \mathbf{b}_i (náhodné efekty) sú nezávislé vektory,

$$\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbb{B}),$$

$\boldsymbol{\varepsilon}_i$ sú nezávislé vektory,

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_e^2 \mathbb{I}_{n_i}),$$

a vektor $(\boldsymbol{\varepsilon}_1^\top, \dots, \boldsymbol{\varepsilon}_K^\top)^\top$ je nezávislý s vektorom $(\mathbf{b}_1^\top, \dots, \mathbf{b}_K^\top)^\top$, $\sum_{i=1}^K n_i = N$. Tento model bol navrhnutý v Laird a Ware (1982). Stredná hodnota náhodného vektoru \mathbf{Y}_i je popísaná členom $\mathbb{X}_i \boldsymbol{\beta}$, náhodná štruktúra modelu je popísaná $\mathbb{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$. Náhodná časť má zjavne strednú hodnotu rovnú $\mathbf{0}$ a rozptyl $\mathbb{Z}_i \mathbb{B} \mathbb{Z}_i^\top + \sigma_e^2 \mathbb{I}_{n_i}$.

Za zmienku stojí rozšírenie jednostupňového lineárneho modelu so zmiešanými efektami, v ktorom miesto predpokladu

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_e^2 \mathbb{I}_{n_i})$$

použijeme predpoklad

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_e^2 \boldsymbol{\Lambda}_i),$$

¹angl. *single-level LME model*

²fixné efekty, angl. *fixed effects*

³angl. *random effects, subject specific effects*

kde Λ_i je pozitívne (semi)definitná kovariančná matica reziduálnej zložky. K takémuto lineárnemu modelu sa dostaneme napríklad v prípade, keď uvažujeme *autokorelačnú štruktúru pozorovaní na subjekte*.

1.2.2 Príklady

Uvažujme jednostupňový lineárny model so zmiešanými efektami tvaru

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\mathbf{b} + \boldsymbol{\varepsilon}.$$

Nech $\mathbb{Z} = (1, 1, \dots, 1)^\top$ (\mathbb{Z} je potom aj prvý stĺpec regresnej matice \mathbb{X}), potom modelu hovoríme *model s náhodným počiatkom*⁴ a má tvar

$$Y_{ij} = x_{ij}^\top \boldsymbol{\beta} + b_i^0 + \varepsilon_{ij}, \quad (1.2)$$

kde x_{ij} značí vektor regresorov pre j-te pozorovanie v i-tej skupine/na i-tom subjekte. Nech regresná matica \mathbb{X} obsahuje spojitý regresor v k-tom stĺpci, $k > 1$. Nech $\mathbb{Z} = \mathbb{X}^k$, kde \mathbb{X}^k značí k-ty stĺpec matice \mathbb{X} . Potom model nazývame *modelom s náhodným sklonom*⁵, ktorý je potom tvaru

$$Y_{ij} = x_{ij}^\top \boldsymbol{\beta} + b_i^1 x_{ij}^k + \varepsilon_{ij},$$

kde x_{ij}^k značí hodnotu k-teho regresoru v i-tej skupine/na i-tom subjekte v j-tom pozorovaní. Podobne môžeme skombinovať náhodný počiatok a náhodný sklon a dostaneme ďalší často používaný model

$$Y_{ij} = x_{ij}^\top \boldsymbol{\beta} + b_i^0 + b_i^1 x_{ij}^k + \varepsilon_{ij}.$$

1.3 Odhady parametrov LME modelu

V nasledujúcej podkapitole budeme uvažovať jednostupňový model so zmiešanými efektami popísaný v definícii 1.1. Miesto parametrizácie $\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbb{B})$ budeme ale (aj) kvôli zjednodušeniu zápisu uvažovať parametrizáciu $\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \sigma_e^2 \mathbb{D})$. Teda máme, že

$$\mathbf{Y}_i \sim \mathcal{N}_{n_i}(\mathbb{X}_i \boldsymbol{\beta}, \sigma_e^2 (\mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top + \mathbb{I}_{n_i})). \quad (1.3)$$

Logvierochnosť celého modelu potom môžeme vyjadriť v tvare

$$\begin{aligned} \ell(\boldsymbol{\beta}, \sigma_e^2, \mathbb{D}) = & -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma_e^2) - \frac{1}{2} \sum_{i=1}^K \left\{ \ln(|\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top|) - \right. \\ & \left. - \frac{1}{2\sigma_e^2} (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right\}. \end{aligned} \quad (1.4)$$

Vektor odhadovaných parametrov $\boldsymbol{\xi}^\top = (\boldsymbol{\beta}^\top, \sigma_e^2, \text{vech}(\mathbb{D})^\top)$, kde $\text{vech}(\mathbb{D})$ vyjadruje vektor parametrov určujúcich pozitívne semidefinitnú maticu (naddiagonálne a diagonálne prvky, presný popis nájdeme v Demidenko (2013) strana 669). Dĺžka vektoru parametrov $\boldsymbol{\xi}$ je $p + 1 + q(q + 1)/2$. Parametrický priestor Ξ bude teda

$$\Xi = \left\{ \boldsymbol{\xi}; \boldsymbol{\beta} \in \mathbb{R}^p, \sigma_e^2 > 0, \mathbb{D} \text{ pozitívne semidefinitná} \right\}.$$

⁴náhodným interceptom, angl. random intercept model

⁵angl. random slope

1.3.1 Odhady pevných efektov

GLS

V prípade odhadu pevných efektov β , existuje možnosť použiť tzv. odhad metódou *vážených najmenších štvorcov*,⁶ ktorá zachytáva určitú netriviálnu kovariančnú štruktúru dát. Uvažujme model tvaru

$$\mathbf{Y} = \mathbb{X}\beta + \boldsymbol{\eta},$$

kde \mathbb{X} má N riadkov a p stĺpcov a má plnú hodnosť, β je p -rozmerný vektor koeficientov a pre $\boldsymbol{\eta}$ platí, že $\mathbb{E}\boldsymbol{\eta} = \mathbf{0}$ a $\text{Var } \boldsymbol{\eta} = \sigma^2 \mathbf{W}^{-1}$, $\sigma^2 > 0$ a \mathbf{W}^{-1} pozitívne definitná $N \times N$ známa matica. Keďže \mathbf{W}^{-1} je pozitívne definitná, tak existuje jej *odmocnina*, tj. matica $\mathbf{W}^{-1/2}$, pre ktorú platí $\mathbf{W}^{-1} = \mathbf{W}^{-1/2} \mathbf{W}^{-1/2}$ (Golub a Van Loan (2012), strana 163, Theorem 4.2.7). Tým pádom máme aj $\mathbf{W} = \mathbf{W}^{1/2} \mathbf{W}^{1/2}$. Potom, ak si definujeme $\mathbf{Y}^* = \mathbf{W}^{1/2} \mathbf{Y}$, tak platí, že $\mathbb{E} \mathbf{Y}^* = \mathbf{W}^{1/2} \mathbb{X} \beta = \mathbb{X}^* \beta$ a $\text{Var} \mathbf{Y}^* = \sigma^2 \mathbf{W}^{1/2} \mathbf{W}^{-1} \mathbf{W}^{1/2} = \sigma^2 \mathbf{W}^{1/2} \mathbf{W}^{-1/2} \mathbf{W}^{-1/2} \mathbf{W}^{1/2} = \sigma^2 \mathbb{I}_N$. Teda,

$$\mathbf{Y}^* \sim \mathcal{N}_N(\mathbb{X}^* \beta, \sigma^2 \mathbb{I}_N),$$

a máme znovu situáciu riešiteľnú pomocou metódy najmenších štvorcov. Riešenie je v tvare

$$\hat{\beta}_{GLS} = (\mathbb{X}^T \mathbf{W} \mathbb{X})^{-1} \mathbb{X}^T \mathbf{W} \mathbf{Y}.$$

Aby sme procedúru použili na jednostupňový lineárny model, uvažujme $\boldsymbol{\eta} = \mathbb{Z} \mathbf{b} + \boldsymbol{\varepsilon}$, kde \mathbb{Z} má N riadkov a q^* stĺpcov, $\mathbf{b} \sim \mathcal{N}_N(\mathbf{0}, \sigma_e^2 \mathbb{D}_*)$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_N(\mathbf{0}, \sigma_e^2 \mathbb{I}_N)$, $\text{cov}(\mathbf{b}, \boldsymbol{\varepsilon}) = \mathbf{0}$ a $\sigma^2 \mathbf{W}^{-1} = \Sigma = \text{var } \mathbf{Y} = \sigma_e^2 (\mathbb{Z} \mathbb{D} \mathbb{Z}^T + \mathbb{I}_N)$ a

$$q^* = Kq, \quad N = \sum_{i=1}^K n_i, \quad \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_K \end{pmatrix}, \quad \mathbb{X} = \begin{pmatrix} \mathbb{X}_1 \\ \vdots \\ \mathbb{X}_K \end{pmatrix}, \quad \mathbb{Z} = \begin{pmatrix} \mathbb{Z}_1 & 0 & \dots & 0 \\ 0 & \mathbb{Z}_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mathbb{Z}_K \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_K \end{pmatrix}, \quad \mathbb{D}_* = \begin{pmatrix} \mathbb{D} & 0 & \dots & 0 \\ 0 & \mathbb{D} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mathbb{D} \end{pmatrix}.$$

Potom odhad bude tvaru

$$\hat{\beta}_{GLS} = (\mathbb{X}^T \Sigma^{-1} \mathbb{X})^{-1} \mathbb{X}^T \Sigma^{-1} \mathbf{Y}. \quad (1.5)$$

V prípade odhadu GLS uvažujeme kovariančnú štruktúru známu (tj. matica Σ je známa, respektíve musí byť známa matica $\mathbb{Z} \mathbb{D} \mathbb{Z}^T + \mathbb{I}_N$, pretože σ_e^2 sa zo vzorca (1.5) vykrátí po dosadení za Σ). V prípade, že maticu Σ nepoznáme, je potreba ju odhadnúť.

⁶angl. weighted least squares, WLS, generalized least squares, GLS

Hendersonove rovnice

Predstavme si, že by naším cieľom nebol len odhad pevných efektov, ale zaujímali by nás aj (nejaké) odhady náhodných efektov, napríklad pre porovnanie jednotlivých subjektov/skupín, ktoré vstupujú do štúdie/experimentu alebo overovanie predpokladov modelu (čomu sa budeme ďalej v práci venovať). Vrámci GLS odhadu pevných efektov nedostávame ale žiadnu informáciu o náhodných efektoch \mathbf{b} vstupujúcich do modelu. Z tohto dôvodu prišiel Henderson (Henderson (1984)) s nasledujúcim postupom, ktorý vo veľkej miere využíva metódu odhadu GLS. Uvažujme explicitne model tvaru

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\mathbf{b} + \boldsymbol{\varepsilon},$$

podobne ako v predchádzajúcom prípade a všetky parametre majú zmysel ako vyššie. Nech \mathbb{D}_* a σ_e^2 sú známe. Potom združená hustota (\mathbf{Y}, \mathbf{b}) s vektorovým parametrom $\boldsymbol{\beta}$ je

$$\begin{aligned} f(\mathbf{y}, \mathbf{b}; \boldsymbol{\beta}) &= f(\mathbf{y}|\mathbf{b}; \boldsymbol{\beta})f(\mathbf{b}) = & (1.6) \\ &= \frac{1}{(2\pi)^{N/2}(\sigma_e^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbb{X}\boldsymbol{\beta} - \mathbb{Z}\mathbf{b})^\top(\mathbf{y} - \mathbb{X}\boldsymbol{\beta} - \mathbb{Z}\mathbf{b})\right\} \\ &\times \frac{1}{(2\pi)^{q^*/2}|\sigma_e^2\mathbb{D}_*|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{b}^\top\sigma_e^{-2}\mathbb{D}_*^{-1}\mathbf{b}\right\}. \end{aligned}$$

Ak sa budeme pozerat na združenú hustotu ako na vierohodnosť pre neznáme parametre $(\boldsymbol{\beta}, \mathbf{b})$ a budeme ju maximalizovať, získame odhad tvaru

$$(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}) = \arg \min_{\boldsymbol{\beta}, \mathbf{b}} \begin{pmatrix} \mathbf{Y} - \mathbb{X}\boldsymbol{\beta} - \mathbb{Z}\mathbf{b} \\ \mathbf{b} \end{pmatrix}^\top \begin{pmatrix} \mathbb{I}_N & 0 \\ 0 & \mathbb{D}_*^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y} - \mathbb{X}\boldsymbol{\beta} - \mathbb{Z}\mathbf{b} \\ \mathbf{b} \end{pmatrix}.$$

Tento výraz sa dá prepísať ako

$$\arg \min_{\boldsymbol{\beta}, \mathbf{b}} \left[\begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbb{X} & \mathbb{Z} \\ \mathbf{0} & -\mathbb{I}_{q^*} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{pmatrix} \right]^\top \begin{pmatrix} \mathbb{I}_N & 0 \\ 0 & \mathbb{D}_*^{-1} \end{pmatrix} \left[\begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbb{X} & \mathbb{Z} \\ \mathbf{0} & -\mathbb{I}_{q^*} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{pmatrix} \right].$$

Toto je ale znovu problém vážených najmenších štvorcov (ako vyššie), ktorého riešenie je riešenie sústavy rovníc

$$\begin{pmatrix} \mathbb{X}^\top\mathbb{X} & \mathbb{X}^\top\mathbb{Z} \\ \mathbb{Z}^\top\mathbb{X} & \mathbb{Z}^\top\mathbb{Z} + \mathbb{D}_*^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbb{X}^\top\mathbf{Y} \\ \mathbb{Z}^\top\mathbf{Y} \end{pmatrix}.$$

Ak všetky inverzné matice existujú, tak odhady sú v tvare

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbb{X}^\top\Sigma^{-1}\mathbb{X})^{-1}(\mathbb{X}^\top\Sigma^{-1}\mathbf{Y}), \\ \hat{\mathbf{b}} &= \mathbb{D}_*\mathbb{Z}^\top\Sigma^{-1}(\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}), \end{aligned}$$

kde $\Sigma = (\mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top + \mathbb{I}_N)$. Výhodou Hendersonových rovníc proti odhadu metódou najmenších štvorcov sú odhady náhodných efektov. Tieto odhady $\hat{\mathbf{b}}$ sú v určitom zmysle najlepšie (BLUP, best linear unbiased predictor). Ich vlastnosti nájdeme v Henderson (1984).

Bayesovské empirické odhady

Keďže lineárne zmiešané modely obsahujú určitý typ hierarchie, je prirodzené uvažovať bayesovský prístup pre vyjadrenie žiadaných rozdelení. Konkrétne, pre *aposteriórnu* hustotu náhodných efektov \mathbf{b} máme z Bayesovej vety tvar

$$f(\mathbf{b}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\mathbf{b})f(\mathbf{b})}{\int f(\mathbf{Y}|\mathbf{b})f(\mathbf{b})d\mathbf{b}}.$$

Dá sa ukázať, že aposteriórna hustota \mathbf{b} je hustota mnohorozmerného normálneho rozdelenia so strednou hodnotou $\mathbb{D}_*\mathbb{Z}^\top \Sigma^{-1}(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})$ a kovariančnou maticou $\sigma_e^2(\mathbb{Z}\mathbb{Z}^\top + \mathbb{D}_*^{-1})^{-1}$ (viď Appendix A.1). Potom za odhad $\hat{\mathbf{b}}$ je možné použiť *aposteriórnu strednú hodnotu*, ktorá je rovná strednej hodnote aposteriórneho rozdelenia náhodných efektov \mathbf{b} . Odhad je potom tvaru

$$\hat{\mathbf{b}} = \mathbb{D}_*\mathbb{Z}^\top \Sigma^{-1}(\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}),$$

kde $\hat{\boldsymbol{\beta}}$ je nejaký odhad pevných efektov (napríklad vyššie spomínaný GLS odhad). Potom vidíme, že tento odhad náhodných efektov pomocou bayesovského prístupu je totožný s odhadom skrz Hendersonove rovnice. Dôležité je podotknúť, že všetky odhady pevných aj náhodných efektov využívajú fakt, že matica kovariančnej štruktúry Σ je známa. V prípade neznámej kovariančnej matice Σ je potreba túto maticu odhadnúť. S týmto problémom sa budeme zaoberať v nasledujúcich častiach.

1.3.2 Odhady kovariančnej štruktúry

ML s profilovaním vierohodnosti I

V tejto časti si ukážeme prvý spôsob využívajúci metódu maximálnej vierohodnosti. Princípom odhadu je postupné odhadovanie jednotlivých parametrov, pričom zostávajúce parametre budeme považovať za známe (pevné), takzvané profilovanie vierohodnosti. Oba prístupy maximálnej vierohodnosti sú zamerané na vyjadrenie vierohodnosti v modeli v čo najjednoduchšom tvare (výpočtovo, nie nutne matematicky), ktorý potom vstupuje do optimalizácie (maximalizácie).

Pozrime sa najprv na parameter σ_e^2 . Derivácia logvierohodnosti celého modelu (1.4) je

$$\frac{\partial \ell}{\partial \sigma_e^2} = -\frac{N}{2\sigma_e^2} + \frac{1}{2\sigma_e^4} \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i\boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i\mathbb{D}\mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i\boldsymbol{\beta})$$

a maximum dostávame v

$$\hat{\sigma}_e^2 = \frac{1}{N} \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i\boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i\mathbb{D}\mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i\boldsymbol{\beta}).$$

Po dosadení do (1.4) dostávame

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \sigma_e^2, \mathbb{D}) &= -\frac{N}{2}(\ln(2\pi) + 1 - \ln(N)) \\
&\quad - \frac{1}{2} \left\{ N \ln \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right. \\
&\quad \left. + \sum_{i=1}^K \ln(|\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top|) \right\}
\end{aligned}$$

Ak ignorujeme konštantný člen, ktorý nemá vplyv na maximalizáciu logvierohodnosti, dostaneme *profilovú logvierohodnosť* definovanú ako

$$\begin{aligned}
\ell_p(\boldsymbol{\beta}, \mathbb{D}) &= -\frac{1}{2} \left\{ N \ln \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right. \\
&\quad \left. + \sum_{i=1}^K \left\{ \ln(|\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top|) \right\} \right\}.
\end{aligned}$$

Potom po využití vzorcov na redukciu dimenzie A.2 máme

$$\begin{aligned}
\ell_p(\boldsymbol{\beta}, \mathbb{D}) &= -\frac{1}{2} \left\{ N \ln \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right. \\
&\quad \left. + \sum_{i=1}^K \ln(|\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top|) \right\} \\
&= -\frac{1}{2} \left\{ N \ln \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbb{I}_{n_i} - \mathbb{Z}_i (\mathbb{D}^{-1} + \mathbb{Z}_i^\top \mathbb{Z}_i)^{-1} \mathbb{Z}_i^\top) (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right. \\
&\quad \left. + \sum_{i=1}^K \ln(|\mathbb{I}_q + \mathbb{D} \mathbb{Z}_i^\top \mathbb{Z}_i|) \right\} \\
&= -\frac{1}{2} \left\{ N \ln \sum_{i=1}^K \left((\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right. \right. \\
&\quad \left. \left. - (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top \mathbb{Z}_i (\mathbb{D}^{-1} + \mathbb{Z}_i^\top \mathbb{Z}_i)^{-1} \mathbb{Z}_i^\top (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right) \right. \\
&\quad \left. + \sum_{i=1}^K \ln(|\mathbb{D}| |\mathbb{D}^{-1} + \mathbb{Z}_i^\top \mathbb{Z}_i|) \right\}
\end{aligned}$$

Keďže platí $\ln |\mathbb{D}| = \ln |\mathbb{D}^{-1}|^{-1} = -\ln |\mathbb{D}^{-1}|$, konečne máme *parametrizáciu faktorom presnosti* $\boldsymbol{\Delta}^7$, ktorý definujeme ako $\boldsymbol{\Delta}^\top \boldsymbol{\Delta} = \mathbb{D}^{-1}$,

$$\begin{aligned}
\ell_p(\boldsymbol{\beta}, \boldsymbol{\Delta}) &= -\frac{1}{2} \left\{ N \ln \sum_{i=1}^K \left((\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right. \right. \\
&\quad \left. \left. - (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top \mathbb{Z}_i (\boldsymbol{\Delta}^\top \boldsymbol{\Delta} + \mathbb{Z}_i^\top \mathbb{Z}_i)^{-1} \mathbb{Z}_i^\top (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta}) \right) \right. \\
&\quad \left. + \sum_{i=1}^K \ln |\boldsymbol{\Delta}^\top \boldsymbol{\Delta} + \mathbb{Z}_i^\top \mathbb{Z}_i| - 2K \ln |\boldsymbol{\Delta}| \right\}, \tag{1.7}
\end{aligned}$$

⁷angl. precision factor parametrization, precision matrix

pretože platí $\ln |\mathbf{\Delta}^\top \mathbf{\Delta}| = \ln |\mathbf{\Delta}|^2 = 2 \ln |\mathbf{\Delta}|$. Pre ďalšie zjednodušenie profilovej vierohodnosti (1.7) je možno nahradiť $\boldsymbol{\beta}$ jeho GLS odhadom $\hat{\boldsymbol{\beta}}$ a po netriviálnych úpravách, ktoré nájdeme v Demidenko (2013) strana 54 sa dostaneme k finálnemu tvaru vierohodnosti

$$\ell_p(\mathbf{\Delta}) - \frac{1}{2} \left(N \ln q(\mathbf{\Delta}) + \sum_{i=1}^K \ln |\mathbf{\Delta}^\top \mathbf{\Delta} + \mathbb{Z}_i^\top \mathbb{Z}_i| - 2K \ln |\mathbf{\Delta}| \right),$$

kde $q = q(\mathbf{\Delta}) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^K (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})^\top (\mathbb{I}_{n_i} + \mathbb{Z}_i \mathbf{\Delta}^\top \mathbf{\Delta} \mathbb{Z}_i^\top)^{-1} (\mathbf{Y}_i - \mathbb{X}_i \boldsymbol{\beta})$. Tento tvar je vhodný na optimalizáciu, pretože väčšina hodnôt vstupujúcich do výpočtu sa dá priamo a rýchlo napočítať (Demidenko (2013) strana 55).

ML s profilovaním vierohodnosti II

Najprv sa pozrieme na združenú pravdepodobnosť pre \mathbf{Y} , ktorá slúži ako vierohodnosť modelu. Vďaka nezávislosti jednotlivých subjektov platí

$$L(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\theta}) = f(\mathbf{y}) = \prod_{i=1}^K f(\mathbf{y}_i) = \prod_{i=1}^K \int_{\mathbb{R}^q} f(\mathbf{y}_i | \mathbf{b}_i) \cdot f(\mathbf{b}_i) d\mathbf{b}_i. \quad (1.8)$$

Zameriame sa na $f(\mathbf{b}_i)$. Máme

$$\begin{aligned} f(\mathbf{b}_i) &= \frac{1}{(2\pi)^{q/2} |\sigma_e^2 \mathbb{D}|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{b}_i^\top \sigma_e^{-2} \mathbb{D}^{-1} \mathbf{b}_i \right\} \\ &= \frac{1}{(2\pi \sigma_e^2)^{q/2} \text{abs} |\mathbf{\Delta}|^{-1}} \exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{\Delta} \mathbf{b}_i\|^2 \right\}, \end{aligned} \quad (1.9)$$

kde $\mathbf{\Delta}^\top \mathbf{\Delta} = \mathbb{D}^{-1}$ a $|\mathbb{D}|^{1/2} = ((|\mathbf{\Delta}| |\mathbf{\Delta}|)^{-1})^{1/2} = \text{abs}(|\mathbf{\Delta}|)^{-1}$, ako v predchádzajúcom prípade. Potom ale platí pre exponenciálu vo vierohodnosti (rozpísaný tvar vidíme v (1.6)) pre \mathbf{b}_i

$$\exp \left\{ -\frac{1}{2\sigma_e^2} (\|\mathbf{y}_i - \mathbb{X}_i \boldsymbol{\beta} - \mathbb{Z}_i \mathbf{b}_i\|^2 + \|\mathbf{\Delta} \mathbf{b}_i\|^2) \right\}. \quad (1.10)$$

Po zadefinovaní

$$\tilde{\mathbf{y}}_i = \begin{pmatrix} \mathbf{y}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbb{X}}_i = \begin{pmatrix} \mathbb{X}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbb{Z}}_i = \begin{pmatrix} \mathbb{Z}_i \\ \mathbf{\Delta} \end{pmatrix}$$

máme pre exponenciálu vyjadrenie

$$\exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^K \|\tilde{\mathbf{y}}_i - \tilde{\mathbb{X}}_i \boldsymbol{\beta} - \tilde{\mathbb{Z}}_i \mathbf{b}_i\|^2 \right\}.$$

Pri maximalizovaní vierohodnosti by sme sa dostali práve k minimalizovaniu exponenciály, kde vidíme, že ide o minimalizáciu strednej štvorcovej chyby. Použijeme teda metódu najmenších štvorcov a dostaneme odhad (viď odhad náhodných efektov v Hendersonových rovniciach)

$$\hat{\mathbf{b}}_i = (\tilde{\mathbb{Z}}_i^\top \tilde{\mathbb{Z}}_i)^{-1} \tilde{\mathbb{Z}}_i^\top (\tilde{\mathbf{y}}_i - \tilde{\mathbb{X}}_i \boldsymbol{\beta}).$$

Potom sa po úpravách, ktorých detaily nájdeme v Pinheiro a Bates (2009) (strana 63-64), dostávame k tvaru vierohodnosti

$$L(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Delta}) = \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^K \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\hat{\mathbf{b}}_i\|^2 \right\} \prod_{i=1}^K \frac{\text{abs}|\boldsymbol{\Delta}|}{\sqrt{|\tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i|}}. \quad (1.11)$$

Vyjadrenie (1.11) sa dá priamo použiť v maximalizačných procedúrach, z ktorých dostávame odhady parametrov $\boldsymbol{\beta}$, σ_e^2 , $\boldsymbol{\theta}$, kde $\boldsymbol{\theta}$ označuje parametre určujúce $\boldsymbol{\Delta}$. Maximalizácia je ale jednoduchšia, keď sa navyše zbavíme parametrov $\boldsymbol{\beta}$ a σ_e^2 , tj. vyjadríme si ich odhady $\hat{\boldsymbol{\beta}}$ a $\hat{\sigma}_e^2$ ako funkcie parametra $\boldsymbol{\theta}$. Jedným zo spôsobov je využitie Q-R rozkladu matíc.

Veta 1.4 Q-R rozklad matice

Nech \mathbf{A} je reálna $m \times n$ matica. Potom existuje ortogonálna $m \times m$ matica \mathbf{Q} a horná trojuholníková $m \times n$ matica \mathbf{R} taká, že

$$\mathbf{A} = \mathbf{Q}\mathbf{R}.$$

Vetu môžeme nájsť napríklad v Harville (2008) v časti 6.4d ako dôsledok *Gram-Schmidtovho* ortogonalizačného procesu. Z vyjadrení (1.8), (1.9) a (1.10) s využitím značenia vyššie máme pre jeden subjekt vierohodnosť v tvare

$$L(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Delta}) = \frac{\text{abs}|\boldsymbol{\Delta}|}{(2\pi\sigma_e^2)^{n_i/2}} \cdot \int_{\mathbb{R}^q} \frac{1}{(2\pi\sigma_e^2)^{q/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 \right\} d\mathbf{b}_i.$$

Naším cieľom bude upraviť integrand výrazu. Najprv sa ale pozrieme na Q-R rozklad matice

$$\begin{pmatrix} \tilde{\mathbf{Z}}_i & \tilde{\mathbf{X}}_i & \tilde{\mathbf{y}}_i \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_i & \mathbf{X}_i & \mathbf{y}_i \\ \boldsymbol{\Delta} & \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{Q}^i \begin{pmatrix} \mathbf{R}_Z^i & \mathbf{R}_X^i & \mathbf{c}_1^i \\ \mathbf{0} & \mathbf{R}_0^i & \mathbf{c}_0^i \end{pmatrix}.$$

Potom platí

$$\begin{aligned} \|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2 &= \|\mathbf{Q}^i\|^2 \left\| \begin{pmatrix} \mathbf{c}_1^i \\ \mathbf{c}_0^i \end{pmatrix} - \begin{pmatrix} \mathbf{R}_X^i \\ \mathbf{R}_0^i \end{pmatrix} \boldsymbol{\beta} - \begin{pmatrix} \mathbf{R}_Z^i \\ \mathbf{0} \end{pmatrix} \mathbf{b}_i \right\|^2 \\ &= \|\mathbf{c}_1^i - \mathbf{R}_X^i\boldsymbol{\beta} - \mathbf{R}_Z^i\mathbf{b}_i\|^2 + \|\mathbf{c}_0^i - \mathbf{R}_0^i\boldsymbol{\beta}\|^2, \end{aligned}$$

pretože matica \mathbf{Q}^i je ortogonálna, a teda $\|\mathbf{Q}^i\|^2 = 1$. Integrál je potom tvaru

$$\begin{aligned} &\int_{\mathbb{R}^q} \frac{1}{(2\pi\sigma_e^2)^{q/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \left(\|\mathbf{c}_1^i - \mathbf{R}_X^i\boldsymbol{\beta} - \mathbf{R}_Z^i\mathbf{b}_i\|^2 + \|\mathbf{c}_0^i - \mathbf{R}_0^i\boldsymbol{\beta}\|^2 \right) \right\} d\mathbf{b}_i \quad (1.12) \\ &= \exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{c}_0^i - \mathbf{R}_0^i\boldsymbol{\beta}\|^2 \right\} \int_{\mathbb{R}^q} \frac{\exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{c}_1^i - \mathbf{R}_X^i\boldsymbol{\beta} - \mathbf{R}_Z^i\mathbf{b}_i\|^2 \right\}}{(2\pi\sigma_e^2)^{q/2}} d\mathbf{b}_i. \end{aligned}$$

Keďže zložky \mathbf{R}_Z^i , \mathbf{R}_X^i , \mathbf{R}_0^i sú regulárne matice, môžeme použiť vetu o transformácii. Nech

$$\boldsymbol{\phi}_i = \frac{\mathbf{c}_1^i - \mathbf{R}_X^i\boldsymbol{\beta} - \mathbf{R}_Z^i\mathbf{b}_i}{(\sigma_e^2)^{1/2}}.$$

Potom Jakobián transformácie je rovný $(\sigma_e^2)^{q/2} |\mathbf{R}_Z^i|$ a integrál (1.12) je potom rovný

$$\begin{aligned} & \exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{c}_0^i - \mathbf{R}_0^i \boldsymbol{\beta}\|^2 \right\} \int_{\mathbb{R}^q} \frac{\exp \left\{ -\frac{1}{2} \|\boldsymbol{\phi}_i\|^2 \right\}}{(2\pi)^{q/2}} \frac{1}{|\mathbf{R}_Z^i|} d\boldsymbol{\phi}_i \\ &= \exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{c}_0^i - \mathbf{R}_0^i \boldsymbol{\beta}\|^2 \right\} \frac{1}{|\mathbf{R}_Z^i|} \end{aligned}$$

Dostali sme sa k značnému zjednodušeniu časti vierohodnosti. Vierohodnosť jedného subjektu i má celkovo tvar

$$L(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Delta}) = \frac{1}{(2\pi\sigma_e^2)^{n_i/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{c}_0^i - \mathbf{R}_0^i \boldsymbol{\beta}\|^2 \right\} \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right)$$

a celková vierohodnosť je potom

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\Delta}) &= \prod_{i=1}^K \frac{1}{(2\pi\sigma_e^2)^{n_i/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \|\mathbf{c}_0^i - \mathbf{R}_0^i \boldsymbol{\beta}\|^2 \right\} \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right) \\ &= \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp \left\{ -\sum_{i=1}^K \frac{1}{2\sigma_e^2} \|\mathbf{c}_0^i - \mathbf{R}_0^i \boldsymbol{\beta}\|^2 \right\} \prod_{i=1}^K \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right) \\ &= \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \left\| \begin{pmatrix} \mathbf{c}_0^1 - \mathbf{R}_0^1 \boldsymbol{\beta} \\ \vdots \\ \mathbf{c}_0^K - \mathbf{R}_0^K \boldsymbol{\beta} \end{pmatrix} \right\|^2 \right\} \prod_{i=1}^K \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right) \\ &= \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \left\| \begin{pmatrix} \mathbf{Q}^\top \mathbf{c}_0^1 - \mathbf{Q}^\top \mathbf{R}_0^1 \boldsymbol{\beta} \\ \vdots \\ \mathbf{Q}^\top \mathbf{c}_0^K - \mathbf{Q}^\top \mathbf{R}_0^K \boldsymbol{\beta} \end{pmatrix} \right\|^2 \right\} \prod_{i=1}^K \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right) \\ &= \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \left\| \begin{pmatrix} \mathbf{c}_0 - \mathbf{R}_0 \boldsymbol{\beta} \\ \mathbf{k}_0 - \mathbf{0} \end{pmatrix} \right\|^2 \right\} \prod_{i=1}^K \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right) \\ &= \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp \left\{ -\frac{\|\mathbf{k}_0\|^2 + \|\mathbf{c}_0 - \mathbf{R}_0 \boldsymbol{\beta}\|^2}{2\sigma_e^2} \right\} \prod_{i=1}^K \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right) \end{aligned}$$

vdaka ďalšiemu Q-R rozkladu

$$\begin{pmatrix} \mathbf{R}_0^1 & \mathbf{c}_0^1 \\ \vdots & \vdots \\ \mathbf{R}_0^K & \mathbf{c}_0^K \end{pmatrix} = \mathbf{Q} \begin{pmatrix} \mathbf{R}_0 & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{k}_0 \end{pmatrix}.$$

Maximalizáciu vierohodnosti dosiahneme položením výrazu $\|\mathbf{c}_0 - \mathbf{R}_0 \boldsymbol{\beta}\|^2 = 0$, z čoho dostaneme 'profilový' odhad $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{R}_0^{-1} \mathbf{c}_0$, keďže matica \mathbf{R}_0 je regulárna a 'profilový' odhad $\widehat{\sigma}_e^2(\boldsymbol{\theta})$ ako (obyčajnú) maximalizáciu zvyšku, ako funkcie σ_e^2 , čo nám dá odhad $\widehat{\sigma}_e^2(\boldsymbol{\theta}) = \frac{\|\mathbf{k}_0\|^2}{N}$. Výsledná profilová vierohodnosť je teda

$$L(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \widehat{\sigma}_e^2(\boldsymbol{\theta}), \boldsymbol{\theta}) = \left(\frac{N}{2\pi \|\mathbf{k}_0\|^2} \right)^{N/2} \exp \left\{ -\frac{N}{2} \right\} \prod_{i=1}^K \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_Z^i|} \right).$$

Vierohodnosť, alebo jej zodpovedajúca logvierohodnosť sú potom maximalizované voči parametru $\boldsymbol{\theta}$, na ktorom závisia \mathbf{k}_0 , $\boldsymbol{\Delta}$ a \mathbf{R}_Z^i .

REML

Ďalšou z možností odhadu parametrov kovariančnej štruktúry je *restricted maximum likelihood estimator* (ďalej len REML). Ukážeme si dva spôsoby odvodenia tohto odhadu.

V prvej časti budeme uvažovať všeobecný model a ukážeme jeho aplikáciu pre lineárny model so zmiešanými efektami. Nech všeobecný lineárny model je definovaný ako $\mathbf{Y} \sim \mathcal{N}(\mathbb{X}\boldsymbol{\beta}, \mathbf{V})$, kde \mathbb{X} je $n \times p$ matica plnej hodnoty a \mathbf{V} je $n \times n$ kovariančná matica, určená parametrom $\boldsymbol{\theta}$. Pri REML odhade maximalizujeme logvierochnosť pre reziduálny vektor $\hat{\mathbf{e}} = \mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}$, kde $\hat{\boldsymbol{\beta}} = (\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbf{V}^{-1} \mathbf{Y}$ (zovšeobecnený odhad metódou najmenších štvorcov). Keďže \mathbf{Y} má normálne rozdelenie, $\hat{\boldsymbol{\beta}}$ a $\hat{\mathbf{e}}$ sú lineárnymi funkciami \mathbf{Y} , a teda majú tiež normálne rozdelenie. Ďalej, $\hat{\boldsymbol{\beta}}$ a $\hat{\mathbf{e}}$ sú nezávislé, pretože

$$\begin{aligned} \text{cov}(\mathbb{X}^\top \mathbf{V}^{-1} \mathbf{Y}, \hat{\mathbf{e}}) &= \mathbb{X}^\top \mathbf{V}^{-1} \mathbf{V} (\mathbb{I}_n - \mathbf{V}^{-1} \mathbb{X} (\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top) \\ &= \mathbb{X}^\top - \mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X} (\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X})^{-1} \mathbb{X} = 0 \end{aligned}$$

Teda máme, že logvierochnosť pre \mathbf{Y} je súčinom logvierochností pre $\hat{\mathbf{e}}$ a pre $\hat{\boldsymbol{\beta}}$. Ale $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, (\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X})^{-1})$, a teda logvierochnosť pre reziduálny vektor $\hat{\mathbf{e}}$, až na konštantu, bude

$$\begin{aligned} \ell(\hat{\mathbf{e}}; \boldsymbol{\theta}) &= \ell(\mathbf{Y}, \boldsymbol{\theta}) - \ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}) = -\frac{1}{2} \left(\ln |\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X}| + \ln |\mathbf{V}| \right. \\ &\quad \left. + (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right). \end{aligned}$$

Ale platí

$$\begin{aligned} &(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) \\ &= (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \end{aligned} \tag{1.13}$$

takže môžeme logvierochnosť pre $\hat{\mathbf{e}}$ prepísať ako

$$\ell(\hat{\mathbf{e}}; \boldsymbol{\theta}) = -\frac{1}{2} \left(\ln |\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X}| + \ln |\mathbf{V}| + \hat{\mathbf{e}}^\top \mathbf{V}^{-1} \hat{\mathbf{e}} \right).$$

Maximalizácia logvierochnosti pre $\hat{\mathbf{e}}$ je ekvivalentná maximalizácii

$$\ell_R(\boldsymbol{\beta}; \boldsymbol{\theta}) = -\frac{1}{2} \left(\ln |\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X}| + \ln |\mathbf{V}| + (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) \right),$$

pretože maximalizácia ℓ_R pre $\boldsymbol{\beta}$ nám dá $\hat{\boldsymbol{\beta}}$. Funkcia ℓ_R sa nazýva *reziduálna logvierochnosť* (Demidenko (2013) strana 56). Všimnime si, že ℓ_R sa líši od obvyčajnej logvierochnosti členom $-\frac{1}{2} \ln |\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X}|$. Táto logvierochnosť nie je skutočná vierochnosť a navyše závisí na parametrizácii modelu. Napriek tomu asymptoticky ML a REML odhady splývajú (viď Demidenko (2013), strana 144).

Iný spôsob získania REML odhadu je skrz bayesovský prístup s neinformatívnym apriórnyim rozdelením pre $\boldsymbol{\beta}$. Potom na získanie marginálneho rozdelenia pre σ_e^2 a \mathbb{D} potrebujeme vyintegrovat $\boldsymbol{\beta}$,

$$L_R(\sigma_e^2, \mathbb{D}) = \int_{\beta \in \mathbb{R}^p} L(\beta, \sigma_e^2, \mathbb{D}) d\beta,$$

kde L je štandardná vierohodnosť

$$L(\beta, \sigma_e^2, \mathbb{D}) = (2\pi)^{-N/2} |\mathbf{V}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbb{X}\beta)^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\beta) \right\}.$$

Znovu podľa (1.13) môžeme integrál vyššie prepísať ako

$$\begin{aligned} & (2\pi)^{-N/2} |\mathbf{V}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbb{X}\hat{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\beta}) \right\} \times \\ & \times \int_{\beta \in \mathbb{R}^p} \exp \{ (\hat{\beta} - \beta)^\top \mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X} (\hat{\beta} - \beta) \} d\beta, \end{aligned}$$

ale podľa známeho vzťahu z Graybill (1982) (strana 332) máme po vyintegrovaní vzťah pre L_R v tvare

$$L_R(\sigma_e^2, \mathbb{D}) = (2\pi)^{-(N-p)/2} |\mathbf{V}|^{-1/2} |\mathbb{X}^\top \mathbf{V}^{-1} \mathbb{X}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbb{X}\hat{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\beta}) \right\},$$

čo je ekvivalentné s reziduálnou logvierohodnosťou vyššie. Pre jednostupňový lineárny model so zmiešanými efektami máme $\mathbf{V} = \sigma_e^2 (\mathbb{I}_N + \mathbb{Z}^\top \mathbb{D} \mathbb{Z})$.

MINQUE

Inou možnosťou je kvadratický odhad variančných parametrov⁸ navrhnutý v Rao (2001). Tento odhad je inšpirovaný tvarom odhadu pre σ^2 v klasickom normálnom lineárnom modeli

$$\mathbf{Y} = \mathbb{X}\beta + \varepsilon, \tag{1.14}$$

kde \mathbf{Y} je $n \times 1$ vektor pozorovaní, \mathbb{X} je $n \times p$ regresná matica plnej hodnosti, β je p -rozmerný vektor pevných (fixných) efektov a $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbb{I}_n)$. Odhad pre σ^2 v modeli (1.14) je

$$\hat{\sigma}^2 = \frac{1}{n-p} \mathbf{Y} (\mathbb{I}_n - \mathbb{X}(\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top) \mathbf{Y}. \tag{1.15}$$

Na výraz (1.15) sa dá pozeráť ako na kvadratickú formu v \mathbf{Y} . Rao (2001) navrhol teda odhad σ^2 v tvare

$$\hat{\sigma}_{Rao}^2 = \mathbf{Y}^\top \mathbf{A} \mathbf{Y},$$

kde \mathbf{A} je $n \times n$ symetrická, pozitívne semidefinitná matica, aby $\hat{\sigma}_{Rao}^2 \geq 0$. Miesto minimalizácie rozptylu kvadratickej formy, ktorá vyžaduje vyššie momenty rozdelenia, navrhol minimalizáciu (Frobeniovej) normy matice \mathbf{A} . Výsledný odhad sa nazýva MINQUE a pre niektoré modely s predpokladom normality, vedie tento

⁸angl. minimum norm quadratic unbiased estimation (MINQUE)

postup ku kvadratickému odhadu, ktorý je nestranný a má minimálny rozptyl. Princíp tohoto odhadu si predvedieme práve na obyčajnom lineárnom modeli (1.14).

Uvažujme teda odhad tvaru (1.15). Potom platí po využití známeho vzťahu $\mathbb{E}(\boldsymbol{\varepsilon}^\top \mathbf{A} \boldsymbol{\varepsilon}) = \text{tr}(\mathbf{A} \text{Var}(\boldsymbol{\varepsilon}))$, že

$$\begin{aligned} \mathbb{E}\hat{\sigma}_{Rao}^2 &= \mathbb{E}(\mathbf{Y}^\top \mathbf{A} \mathbf{Y}) = \mathbb{E}(\boldsymbol{\beta}^\top \mathbb{X}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + 2\boldsymbol{\varepsilon}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^\top \mathbf{A} \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta}^\top \mathbb{X}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + \text{tr}(\mathbf{A} \text{Var}(\boldsymbol{\varepsilon})) = \boldsymbol{\beta}^\top \mathbb{X}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + \sigma^2 \text{tr}(\mathbf{A}), \end{aligned}$$

kde $\text{tr}()$ značí stopu matice. Aby bol MINQUE odhad nestranný, potrebujeme aby $\mathbb{X}^\top \mathbf{A} \mathbb{X} = 0$ pre všetky $\boldsymbol{\beta}$ a $\text{tr}(\mathbf{A}) = 1$. Takže budeme hľadať minimum (Frobeniovej) normy matice \mathbf{A} , čo je ekvivalentné s minimalizáciou $\text{tr}(\mathbf{A} \mathbf{A}^\top)$, za podmienok $\mathbb{X}^\top \mathbf{A} \mathbb{X} = 0$ pre všetky $\boldsymbol{\beta}$ a $\text{tr}(\mathbf{A}) = 1$. Na riešenie tohoto problému využijeme maticovú podobu Lagrangeovej funkcie,

$$\mathcal{L}(\mathbf{A}, \mathbf{L}_1, \ell_2) = \frac{1}{2} \text{tr}(\mathbf{A} \mathbf{A}^\top) + \text{tr}(\mathbb{X}^\top \mathbf{A} \mathbb{X} \mathbf{L}_1^\top) + (1 - \text{tr}(\mathbf{A})) \ell_2,$$

kde \mathbf{L}_1 a ℓ_2 je maticový, resp. skalárny Langrangeov multiplikátor. Podľa známych vzťahov pre maticové derivácie v A.3 máme

$$\frac{\partial \mathcal{L}}{\partial \mathbf{A}} = \mathbf{A} + \mathbb{X} \mathbf{L}_1 \mathbb{X}^\top - \ell_2 \mathbb{I}_n = \mathbf{0},$$

z čoho plynie $\mathbf{A} = \ell_2 \mathbb{I}_n - \mathbb{X} \mathbf{L}_1 \mathbb{X}^\top$. Z podmienky $\mathbb{X}^\top \mathbf{A} \mathbb{X} = 0$ máme, že $\mathbf{L}_1 = \ell_2 (\mathbb{X}^\top \mathbb{X})^{-1}$. Ďalej teda platí

$$\mathbf{A} = \ell_2 (\mathbb{I}_n - \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top).$$

$(\mathbb{I}_n - \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top)$ je ale idempotentná matica, a teda jej stopa je rovná hodnosti $\text{tr}(\mathbb{I}_n - \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top) = n - \text{rank}(\mathbb{X}) = n - p$. Teda, ak zoberieme

$$\ell_2 = \frac{1}{\text{rank}(\mathbb{I}_n - \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top)} = \frac{1}{n - p},$$

tak matica \mathbf{A} bude splňovať aj podmienku $\text{tr}(\mathbf{A}) = 1$. Tento odhad je presne rovnaký ako odhad (1.15). Nás ale bude hlavne zaujímať MINQUE odhad kovariančných parametrov v jednostupňovom lineárnom zmiešanom modeli.

Uvažujme jednostupňový zmiešaný model z definície 1.1 (parametrizácia s kovariančnou maticou $\mathbf{b}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbb{B})$). Znovu hľadáme symetrickú, pozitívne semidefinitnú maticu \mathbf{A} . Začneme so strednou hodnotou odhadu

$$\begin{aligned} \mathbb{E}\hat{\sigma}_{Rao}^2 &= \boldsymbol{\beta}^\top \mathbb{X}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + \text{tr}(\mathbf{A} \text{Var}(\mathbb{Z} \mathbf{b} + \boldsymbol{\varepsilon})) \\ &= \boldsymbol{\beta}^\top \mathbb{X}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + \sigma_e^2 \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{A} \mathbb{Z} \mathbb{B} \mathbb{Z}^\top) \\ &= \boldsymbol{\beta}^\top \mathbb{X}^\top \mathbf{A} \mathbb{X} \boldsymbol{\beta} + \sigma_e^2 \text{tr}(\mathbf{A}) + \text{tr}(\mathbb{B} \mathbb{Z}^\top \mathbf{A} \mathbb{Z}). \end{aligned}$$

Aby bol odhad nestranný, potrebujeme $\mathbb{X}^\top \mathbf{A} \mathbb{X} = \mathbf{0}$ a $\mathbb{Z}^\top \mathbf{A} \mathbb{Z} = \mathbf{0}$ a $\text{tr}(\mathbf{A}) = 1$. Zavedieme si maticu $\mathbb{H} = (\mathbb{X}, \mathbb{Z})$ a prvé dve podmienky skombinujeme do jednej

$\mathbb{H}^\top \mathbf{A} \mathbb{H} = \mathbf{0}$. Aby boli podmienky ekvivalentné, potrebujeme aby $\mathbb{Z}^\top \mathbf{A} \mathbb{X} = \mathbf{0}$ ⁹. Lenže, z $\mathbb{X}^\top \mathbf{A} \mathbb{X} = \mathbf{0}$ máme, že $(\mathbf{A}^{1/2} \mathbb{X})^\top (\mathbf{A}^{1/2} \mathbb{X}) = \mathbf{0}$, pretože \mathbf{A} je pozitívne semidefinitná, a teda existuje $\mathbf{A}^{1/2}$. Platí teda $\mathbf{A}^{1/2} \mathbb{X} = \mathbf{0}$. Po vynásobení zľava najprv maticou $\mathbf{A}^{1/2}$, a potom maticou \mathbb{Z}^\top , dostaneme výsledok $\mathbb{Z}^\top \mathbf{A} \mathbb{X} = \mathbf{0}$. Použitím princípu MINQUE odhadu popísaného vyššie, minimalizujeme $\text{tr}(\mathbf{A} \mathbf{A}^\top)$ za podmienok $\mathbb{H}^\top \mathbf{A} \mathbb{H} = \mathbf{0}$ a $\text{tr}(\mathbf{A}) = 1$. Odhad je teda

$$\hat{\sigma}_{Rao}^2 = \frac{1}{n - \text{rank}(\mathbb{H})} \mathbf{Y} (\mathbb{I}_n - \mathbb{H}(\mathbb{H}^\top \mathbb{H})^{-1} \mathbb{H}^\top) \mathbf{Y}.$$

Napríklad pre VARCOMP z 1.1.2 máme odhad σ_e^2 v tvare

$$\hat{\sigma}_{Rao}^2 = \frac{1}{\sum_{i=1}^K (n_i - 1)} \sum_{i=1}^K \sum_{j=1}^{n_i} (Y_{ij} - \bar{\mathbf{Y}}_i)^2,$$

kde $\bar{\mathbf{Y}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$.

Na podobnej myšlienke, ako je založený odhad $\hat{\sigma}_{Rao}^2$, funguje aj odhad pre \mathbb{B} , resp. pre $\text{vec}(\mathbb{B})$. Znovu sa vezme kvadratická forma v \mathbf{Y} , teraz ale pre vektorový parameter $\text{vec}(\mathbb{B})$. Rovnako ako v predchádzajúcom prípade z podmienky nestrannosti vyplynú podmienky pre tvar matice, ktorá 'zabezpečuje' kvadratickú formu (matica \mathbf{A} v predchádzajúcom prípade). Tento postup je ale značne komplikovaný a zdĺhavý, preto samotný odhad nebudeme uvádzať. Jeho tvar nájdeme v Demidenko (2013) (strana 162).

Ďalšie metódy

Medzi ďalšie metódy odhadu kovariančnej štruktúry patria napríklad odhad metódou momentov alebo odhad pomocou metódy najmenších štvorcov, ktoré nájdeme v Demidenko (2013) (strana 166 resp. 171).

Spoločným problémom niektorých spomenutých odhadov je, že nezaručujú pozitívnu definitnosť matice \mathbb{B} , resp. \mathbb{D} . Tento problém je možné riešiť pomocou projekcie matice na priestor pozitívne (semi)definitných matíc, ktorá v určitom zmysle nájde 'najbližšiu' pozitívne (semi)definitnú maticu k nášmu odhadu kovariančnej matice. Detaily nájdeme v Demidenko (2013) strana 176 odsek 3.13.

1.4 Testovanie hypotéz v jednostupňovom zmiešanom modeli

Popri veľkosti jednotlivých efektov (pevných aj náhodných), ktoré dostávame z odhadovacích procedúr, je ale ďalšou otázkou presnosť nášho odhadu. Zaujímá nás, akej veľkej chyby sa dopúšťame tým a tým odhadom. Takisto nás veľakrát zaujíma nulovosť, respektíve signifikantnosť jednotlivých efektov. Navyše, v LME modeloch odhadujeme parametre kovariančnej štruktúry oproti klasickým regresným modelom.

⁹Matica \mathbb{Z} ale často pozostáva z podmnožiny stĺpcov matice \mathbb{X} , tým pádom je podmienka splnená automaticky.

1.4.1 Pevné efekty

Rovnako ako v klasickom lineárnom modeli, aj v lineárnych modeloch so zmiešanými efektami nás zaujímajú testy hypotéz tvaru

$$\mathbf{H}_0 : \beta_j = \beta_j^0 \qquad \mathbf{H}_A : \beta_j \neq \beta_j^0.$$

Často volíme $\beta_j^0 = 0$, či daný regresor patrí do modelu, resp. prislúchajúci pevný efekt β_j je signifikantný. Testy hypotéz o pevných efektoch sa opierajú o teóriu maximálnej vierohodnosti. Označme si odhady získané pomocou maximalizácie vierohodnosti $\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2, \hat{\boldsymbol{\theta}}$. Ak uvažovaný model platí a platia predpoklady regularity (Cramer (1946), strana 500), tak vieme, že odhady $\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2, \hat{\boldsymbol{\theta}}$ sú konzistentné, platí

$$\sqrt{K} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\sigma}_e^2 - \sigma_e^2 \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}^{-1})$$

a Fisherova informačná matica \mathbf{I} je blokovo diagonálna, tj

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_\beta & 0 & 0 \\ 0 & \mathbf{I}_{\sigma_e^2} & \mathbf{I}_{\sigma_e^2, \boldsymbol{\theta}} \\ 0 & \mathbf{I}_{\boldsymbol{\theta}, \sigma_e^2} & \mathbf{I}_\theta \end{pmatrix},$$

čoho dôkaz nájdeme napríklad v Demidenko (2013) strana 121. Potom teda máme

$$\sqrt{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{I}_\beta^{-1}).$$

Ďalej z teórie lineárnych modelov (s netriviálnou kovariančnou štruktúrou) vieme, že

$$\text{cov}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) = \hat{\sigma}_e^2(\boldsymbol{\theta}) \left(\frac{1}{K} \sum_{i=1}^K \mathbb{X}_i^\top (\mathbb{I}_{n_i} + \mathbf{Z}_i \mathbb{D}(\boldsymbol{\theta}) \mathbf{Z}_i^\top)^{-1} \mathbb{X}_i \right)^{-1},$$

kde po nahradení $\boldsymbol{\theta}$ za $\hat{\boldsymbol{\theta}}$ dostaneme konzistentný odhad kovariančnej matice pre $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. Potom pre test hypotézy \mathbf{H}_0 proti alternatíve \mathbf{H}_A máme testovú štatistiku

$$T_j = \sqrt{K} \frac{\beta_j - \beta_j^0}{\sqrt{\text{cov}(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}))_{j,j}}},$$

kde dolný index j, j označuje j -ty diagonálny prvok odhadu kovariančnej matice. Štatistika T_j má asymptoticky rozdelenie $\mathcal{N}(0, 1)$.

1.4.2 Kovariančná štruktúra

Problém ale nastáva pri zisťovaní signifikantnosti náhodných efektov. Pretože predpokladáme, že náhodné efekty \mathbf{b}_i majú nulovú strednú hodnotu, testy o nulovosti týchto efektov (ako náhodných veličín) sú ekvivalentné hypotéze

$$\mathbf{H}_0 : \mathbb{D} = \mathbf{0}. \tag{1.16}$$

Keďže ale predpokladáme, že parametrický priestor obsahuje *pozitívne semidefinitné* matice, test (1.16) testuje hodnotu, ktorá je na kraji parametrického priestoru. V takomto prípade klasické postupy zlyhávajú (Demidenko (2013), strana

134 príklad v strede). V prípade LME modelov sa ale ponúka možnosť použiť upravený F-test.

Nech

$$\mathbf{W} = \begin{pmatrix} \mathbb{X}_1 & \mathbb{Z}_1 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbb{X}_K & \mathbf{0} & \dots & \mathbf{0} & \mathbb{Z}_K \end{pmatrix}$$

a $r = \text{rank}(\mathbf{W})$. Potom za \mathbf{H}_0 v (1.16) platí, že

$$F = \frac{\frac{S_{LS} - S_{min}}{r-p}}{\frac{S_{min}}{N-r}} \sim \mathcal{F}(r-p, N-r), \quad (1.17)$$

kde $\mathcal{F}(a,b)$ značí F-rozdelenie s a a b stupňami voľnosti, $S_{LS} = \sum_{i=1}^K \|\mathbf{y}_i - \mathbb{X}_i \hat{\boldsymbol{\beta}}_{LS}\|^2$ s $\hat{\boldsymbol{\beta}}_{LS}$ odhadom fixných efektov metódou najmenších štvorcov a

$$S_{min} = \min_{\boldsymbol{\beta}, \mathbf{b}_1, \dots, \mathbf{b}_K} \sum_{i=1}^K \|\mathbf{y}_i - \mathbb{X}_i \boldsymbol{\beta} - \mathbb{Z}_i \mathbf{b}_i\|^2 = \min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{W} \boldsymbol{\gamma}\|^2,$$

kde $\boldsymbol{\gamma} = (\boldsymbol{\beta}^\top, \mathbf{b}_1^\top, \dots, \mathbf{b}_K^\top)^\top$ (tj. súčet najmenších štvorcov za prítomnosti náhodných efektov). Dôkaz (1.17) nájdeme v Demidenko (2013) (strana 136, Theorem 15). Spomínané rozdelenie testovej štatistiky v (1.17) je *presné*. V prípadoch, že nás zaujíma prítomnosť konkrétneho náhodného efektu v modeli, nás zaujímajú hypotézy tvaru

$$\mathbf{H}_0 : \mathbb{D} = \begin{pmatrix} \mathbb{D}_{11} & \mathbf{0} \\ \mathbf{0}^\top & \mathbf{0} \end{pmatrix} \quad \mathbf{H}_A : \begin{pmatrix} \mathbb{D}_{11} & \mathbb{D}_{12} \\ \mathbb{D}_{12}^\top & \mathbb{D}_{22} \end{pmatrix}.$$

V týchto prípadoch je možné použiť jednoduché modifikácie testu pomerom vierohodností. Pre najjednoduchšiu hypotézu

$$\mathbf{H}_0 : \mathbb{D} = d_{11} = 0 \quad \mathbf{H}_A : \mathbb{D} = d_{11} > 0$$

o prítomnosti jedného náhodného efektu máme podľa Stram a Lee (1994) rozdelenie klasickej štatistiky testu pomeru vierohodností $L_R \sim \frac{1}{2} \delta(0) + \frac{1}{2} \chi_1^2$, tj, testová štatistika má rozdelenie zložené z váženého súčtu diracovho rozdelenia v bode 0 a chí kvadrát rozdelenia s jedným stupňom voľnosti s rovnakými váhami rovnými 1/2. Podobne pre hypotézu

$$\mathbf{H}_0 : \mathbb{D} = \begin{pmatrix} \mathbb{D}_{11} & \mathbf{0} \\ \mathbf{0}^\top & \mathbf{0} \end{pmatrix} \quad \mathbf{H}_A : \mathbb{D} = \begin{pmatrix} \mathbb{D}_{11} & \mathbf{d}_{12} \\ \mathbf{d}_{12}^\top & d_{22} \end{pmatrix}$$

platí podľa Stram a Lee (1994), že testová štatistika testu pomerom vierohodností L_R má rozdelenie, ktoré sa skladá z váženého súčtu rozdelení χ_q^2 a χ_{q+1}^2 s váhami 1/2, kde q je dĺžka vektoru \mathbf{d}_{12} (celkovo sa testuje $q+1$ parametrov, \mathbf{d}_{12} a d_{22}). Táto hypotéza zodpovedá testovaniu, či daný náhodný efekt s rozptylom d_{22} patrí do modelu za predpokladu, že ostatné náhodné efekty tam patria.

Kapitola 2

Overovanie predpokladov LME modelu

Súčasťou analýzy dát by vždy malo byť *overenie (verifikácia) predpokladov* použitých modelov. Nesplnenie predpokladov (akéhokoľvek) modelu môže viesť k nepresným hodnotám parametrov záujmu, nesprávnej interpretácii a celkovo k mylným záverom analýzy. O efektoch porušenia predpokladov na výstupy z modelov budeme hovoriť v nasledujúcich častiach. Samotný dopad porušenia predpokladov budeme porovnávať a zobrazovať dvomi spôsobmi - grafickými výstupmi a skrz hodnoty odhadov parametrov a testových štatistík. V tejto kapitole sa budeme venovať grafickým metódam, ktoré sú zmienené napríklad v Pinheiro a Bates (2009).

Predpoklady lineárneho zmiešaného modelu si rozdelíme na dve skupiny

- predpoklady o chybovej zložke ε_{ij} ,
- predpoklady o náhodných efektoch \mathbf{b}_i .

V druhej časti sa budeme venovať hodnotám odhadom parametrov a testom, resp. výsledkom testov. Bude nás zaujímať vplyv porušenia predpokladov na presnosť odhadov a závery testov a prípadná kvantifikácia rozdielov na základe porušenia konkrétneho predpokladu.

2.1 Predpoklady a diagnostika

V našej práci sa budeme venovať predpokladom rozdelení ε_{ij} a \mathbf{b}_i a ich porušeniam. V menšej miere sa pozrieme aj na nezávislosť, resp. nekorelovanosť ε_{ij} vrámci subjektu a vplyvu porušenia predpokladov rozdelení na (auto)koreláciu chybovej zložky. Taktiež jedným z porušení predpokladov je autoregresné správanie chybovej zložky ε_{ij} .

Na porovnanie modelov s nesplnenými predpokladmi s modelom so splnenými predpokladmi budeme používať

- histogram - tvar histogramu pomáha určiť resp. graficky overiť symetriu, šikmosť a celkové rozloženie reziduí,

- reziduálny graf (scatterplot) reziduí proti spojitému regresoru (čas) - podobne ako histogram pomáha graficky overiť rozloženie reziduí a navyše je schopný zachytiť trendy vo vývoji reziduí,
- odhad autokorelačnej funkcie - pomáha graficky overiť korelačnú štruktúru reziduí, teda prítomnosť určitej formy závislosti medzi reziduami,
- QQ-graf - kvantil-kvantil graf, ktorý pomáha graficky určiť normalitu reziduí a náhodných efektov, porovnáva kvantily reziduí/náhodných efektov s kvantilmi normálneho rozdelenia a taktiež porovnáva kvantily Mahalanobisovej vzdialenosti s kvantilmi χ_2^2 rozdelenia.

2.2 Simulácie

Pred samotným začiatkom simulácie odozvy Y_{ij} začneme s vytvorením regresných matíc

$$\mathbb{X} = \begin{pmatrix} \mathbb{X}_1 \\ \vdots \\ \mathbb{X}_K \end{pmatrix}$$

a

$$\mathbb{Z} = \begin{pmatrix} \mathbb{Z}_1 \\ \vdots \\ \mathbb{Z}_K \end{pmatrix},$$

kde K je počet subjektov. Jedna matica \mathbb{X}_i má tvar

$$\mathbb{X}_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}.$$

Matice \mathbb{X}_i a \mathbb{Z}_i generujeme pre každý subjekt zvlášť. Najprv je nutné vygenerovať počty pozorovaní pre jednotlivé subjekty. Počet pozorovaní na jednotlivých subjektoch je náhodne vybraný (ale rovnaký pre každé porušenie, tj. každý dataset sa skladá z rovnakého počtu pozorovaní) z diskrétného rovnomerného rozdelenia na 1 až 10 (model s menším počtom pozorovaní) resp. 1 až 118 (model s väčším počtom pozorovaní) zvýšený o 2 (minimálny počet pozorovaní je 3, maximálny 12, resp. 120). Matica \mathbb{X}_i má prvý stĺpec daný prítomnosťou počiatku (interceptu) v modeli, tj. je to jednotkový vektor dĺžky n_i .

Druhý stĺpec reprezentuje spojený regresor čas. Všetky subjekty obsahujú pozorovania v čase 1, zvyšných $n_i - 1$ časov je generovaných z diskrétného rozdelenia na 2 až 12, resp. 2 až 120. Pravdepodobnosti jednotlivých hodnôt klesajú s rastúcou hodnotou času, presnejšie hodnoty časov 2 až 4, resp. 2 až 40 sú trikrát pravdepodobnejšie ako hodnoty časov 5 až 8, resp. 41 až 80 a šesťkrát pravdepodobnejšie ako hodnoty časov 9 až 12, resp. 81 až 120. V princípe nám ide o reflektovanie reality, kedy nie sú vždy k dispozícii merania v každom čase. Navyše, pri opakovaných meraniach je častým sprievodným znakom znižovanie sa počtu meraní s rastúcim časom. V prípade, že subjekty zodpovedajú ľuďom, sa často stáva, že nie každý človek príde na dané meranie/vyšetrovanie vo všetky časy,

ktoré sú predpísané. Tiež sa stáva, že človek pred koncom experimentu zo štúdie vystúpi, prípadne sa jej nedožije. Práve tieto skutočnosti má reflektovať diskrétno rozdelenie času s klesajúcou pravdepodobnosťou s rastúcou hodnotou času.

Maticu \mathbb{Z}_i generujeme ako podmaticu matice \mathbb{X}_i , pretože budeme uvažovať len modely, v ktorom sa s náhodným efektom vyskytuje aj jeho pevná verzia (s počiatkom b_i^0 aj β_0 a s b_i^1 aj β_1). V prípade modelu s náhodným počiatkom je \mathbb{Z}_i prvý stĺpec matice \mathbb{X}_i , náhodným sklonom je \mathbb{Z}_i druhý stĺpec matice \mathbb{X}_i , v prípade kombinácie (náhodný počiatok aj sklon) máme zas $\mathbb{X}_i = \mathbb{Z}_i$.

2.2.1 Model s náhodným počiatkom

Prvý model, ktorý budeme skúmať je model s náhodným počiatkom (*random intercept model*) zadaný v 1.2 s jedným pevným regresorom, predstavujúcim čas. Model bude mať teda tvar

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 \cdot t_{ij} + b_i^0 + \varepsilon_{ij} \\ &= (\beta_0 + b_i^0) + \beta_1 \cdot t_{ij} + \varepsilon_{ij} \end{aligned} \quad (2.1)$$

s $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ a $b_i^0 \sim \mathcal{N}(0, \sigma_e^2 \mathbb{D})$.

Chybová zložka ε_{ij}

Budeme skúmať nasledujúce porušenia predpokladu normality chybovej zložky ε_{ij} v modeli (2.1) proti modelu (2.1) so splnenými predpokladmi¹:

- bimodálne rozdelenie (zmes normálnych rozdelení)²

$$\varepsilon_{ij} \sim H_{ij} \cdot \mathcal{N}(-2, \sigma_e^2) + (1 - H_{ij}) \cdot \mathcal{N}(2, \sigma_e^2),$$

kde $H_{ij} \sim \text{Alt}(0,5)$ ($\text{Alt}(p)$ značí *alternatívne rozdelenie* s pravdepodobnosťou úspechu p),

- Studentovo t rozdelenie s tromi stupňami voľnosti³, ktoré reprezentuje rozdelenie s ťažšími chvostami v porovnaní s normálnym rozdelením

$$\varepsilon_{ij} \sim t_3,$$

- zmes rozdelení, ktorá reprezentuje kontamináciu normálneho rozdelenia odľahlými hodnotami (outliermi)⁴

$$\varepsilon_{ij} \sim H_{ij} \cdot \mathcal{N}(0, \sigma_e^2) + (1 - H_{ij})t_1,$$

kde $H_{ij} \sim \text{Alt}(0,60)$ pre menší model M1 a $H_{ij} \sim \text{Alt}(0,95)$ pre väčší model M2,

¹v grafoch označené aj ako Bez

²v grafoch označené aj ako Bimod

³v grafoch označené aj ako HT

⁴v grafoch označené aj ako Outliers

- Pearsonovo rozdelenie typu IV⁵, ktoré predstavuje zošikmené rozdelenie⁶ ε_{ij} s parametrami $m = \frac{7}{2}$, $\nu = 3$, $a = 1$ a $\lambda = \frac{3}{5}$, ktorého šikmost je rovná $-\frac{8}{\sqrt{34}} \approx -1,37$ (hustotu a parametrizáciu nájdeme v A.4),
- normálne rozdelenie ε_{ij} s časovo premenlivým rozptylom⁷ (v tomto prípade rozptyl rastie s časom)

$$\varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma_e^2 \frac{t_{ij}}{12}\right)$$

v menšom modeli M1 a

$$\varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma_e^2 \frac{t_{ij}}{120}\right)$$

vo väčšom modeli M2,

- autoregresívna (závislá) chybová zložka vrámci subjektu i reprezentovaná AR(1) procesom⁸

$$\varepsilon_{ij} = \frac{4}{5}\varepsilon_{i(j-1)} + U_{ij},$$

kde $U_{ij} \sim \mathcal{N}(0, \sigma_e^2)$.

Reziduá, ktoré môžeme vidieť na nasledujúcich grafoch sú reziduá zodpovedajúce chybovej zložke ε_{ij} , tj. ak $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{b}_i^0$ sú odhadnuté efekty na subjekte i , potom *reziduami* myslíme rozdiel

$$\widehat{\varepsilon}_{ij} = y_{ij} - \widehat{\beta}_0 - \widehat{\beta}_1 \cdot t_{ij} - \widehat{b}_i^0.$$

Prezentované budú výsledky pre dva rôzne veľké modely M1 a M2:

- M1 s nastavením parametrov
 - pevné efekty: $\beta_0 = 5, \beta_1 = 0,5, \sigma_e^2 = 1$,
 - náhodné efekty: $\sigma_e^2 \mathbb{D} = 0,5$,
 - počet subjektov $K = 20$,
 - celkový počet pozorovaní je $N = 137$,
 - počty subjektov s daným počtom pozorovaní nájdeme v tabuľke 2.1.

Počet pozorovaní	3	4	6	7	8	9	10	12
Počet subjektov	3	3	3	2	4	2	1	2

Tabuľka 2.1: Tabuľka s počtom subjektov s daným počtom pozorovaní.

- M2 s nastavením parametrov

– pevné efekty: $\beta_0 = 5, \beta_1 = 0,5, \sigma_e^2 = 1$,

⁵v grafoch označené aj ako Skewed

⁶normálne rozdelenie má šikmost rovnú 0

⁷v grafoch označené aj ako Heterosked

⁸v grafoch označené aj ako AR(1)

	Min.	1.Kvartil	Medián	Priemer	3.Kvartil	Max.
Odozva	2,3	5,9	7,4	7,6	9,2	12,4
Čas	1,0	2,0	5,0	5,5	8,0	12,0

Tabuľka 2.2: Popisné štatistiky datasetu pre model M1 so splnenými predpokladmi.

- náhodné efekty: $\sigma_e^2 \mathbb{D} = 0,5$,
- počet subjektov $K = 200$,
- počet pozorovaní n_i na subjektoch sa pohybuje od 3 do 120 s priemerným počtom pozorovaní 63,6, resp. mediánom rovným 63,
- celkový počet pozorovaní $N = 12726$.

Popisné štatistiky nájdeme v tabuľke 2.3.

	Min.	1.Kvartil	Medián	Priemer	3.Kvartil	Max.
Odozva	2,3	16,0	27,5	30,4	43,5	67,7
Čas	1,0	22,0	45,0	50,7	77,0	120,0

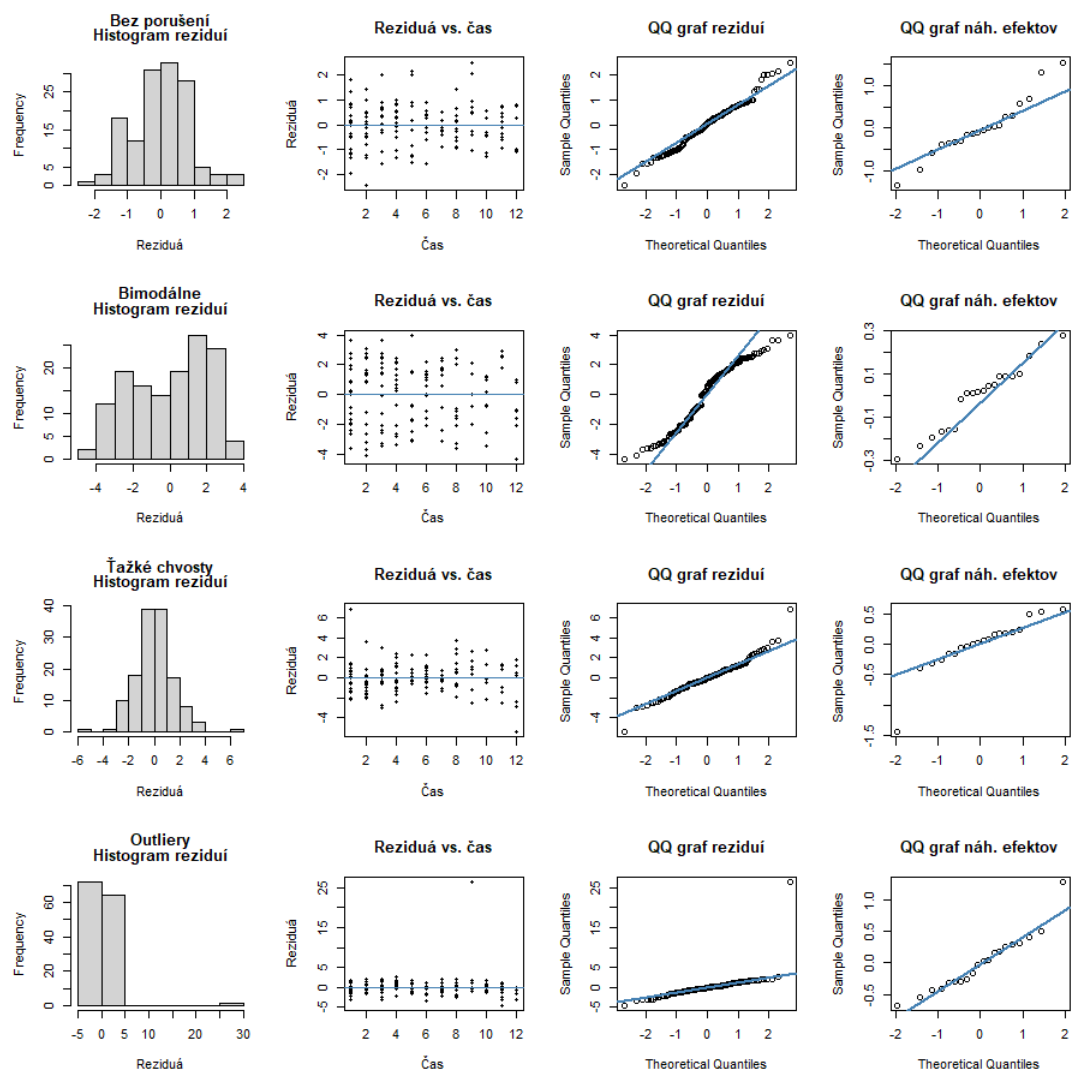
Tabuľka 2.3: Popisné štatistiky datasetu pre model M2 so splnenými predpokladmi.

Na nasledujúcich stránkach uvidíme diagnostické grafy modelu s náhodným počiatkom pre malý (M1) aj veľký (M2) dataset. Porušenia predpokladov sa týkajú len chybovej zložky ε_{ij} . V prvom riadku všetkých obrázkov bude vždy model so splnenými predpokladmi (malý M1/veľký M2 model/dataset), aby bolo jednoduché vidieť rozdiely. V druhom až štvrtom, resp. piatom riadku budú modely s porušeniami. Konkrétny typ porušenia je vždy nad názvom histogramu, tj. v prvom stĺpci a príslušnom riadku. Na prvom obrázku bude podrobne vysvetlený princíp porovnania a hodnotenia platnosti predpokladov s možnými riešeniami porušení predpokladov. V ďalších obrázkoch bude cieľom zdôrazniť len konkrétne grafy, ktoré môžu byť v danom prípade nápomocné a smerodajné.

Na prvom obrázku 2.1 vidíme v prvom riadku zľava histogram, scatterplot, QQ-graf reziduí a QQ-graf náhodných efektov. Ako môžeme pozorovať, histogram má očakávaný 'zvonový' tvar, ktorý pripomína normálne rozdelenie. Histogram nie je úplne symetrický, čo je najpravdepodobnejšie dôsledkom menšieho rozsahu výberu. V prípade scatterplotu nie je vidieť žiadne značné odchýlenie od náhodného vzoru (reziduá sú rozložené nad aj pod nulovou hodnotou náhodne bez viditeľného trendu), ktorý by sme v prípade splnenia predpokladov očakávali. Na treťom grafe zľava vidíme QQ-graf, na ktorom kvantily reziduí veľmi presne kopírujú diagonálnu úsečku, ktorá značí zhodu kvantilov reziduí a normálneho rozdelenia. Na rovnakom základe by sme posúdili štvrtý graf, QQ-graf náhodných efektov, ktorý, aj keď iba s malým počtom bodov ($K = 20$), uspokojivo kopíruje kvantily normálneho rozdelenia.

V druhom riadku sa nachádzajú diagnostické grafy menšieho modelu M1 s porušením predpokladu normality - bimodálnym rozdelením reziduí 2.2.1. Hneď na prvom grafe, histograme, je vidieť, že reziduá majú dve lokálne maximá. Jednou z možností je, že reziduá teda pochádzajú zo zmesi dvoch (alebo viacerých) rozdelení. V ďalšom grafe, scatterplote, nie sú výrazné znaky akéhokoľvek trendu.

Reziduá vyzerajú byť znovu náhodne rozmiestnené nad aj pod nulovou hodnotou. Na QQ-grafe reziduí ale môžeme jasne vidieť odchýlku na oboch koncoch. Na poslednom grafe nie je vidieť výrazné odlišenie od očakávaného vzoru, resp. v porovnaní s QQ-grafom náhodných efektov v modeli bez porušenia predpokladov nie je vidieť výrazná zmena. Na odhalenie bimodálneho rozdelenie chybovej zložky najlepšie poslúži histogram a QQ-graf, prípadne scatterplot v prípade, že 'centrá' (oblasti s najvyššou hodnotou hustoty) rozdelení, z ktorých sa zmes skladá, sú od seba dostatočne vzdialené. S bimodálnym rozložením reziduí je možné sa stretnúť napríklad v prípade, keď sa populácia skúmaných subjektov skladá z dvoch subpopulácií (napríklad muži/ženy). V prípade, že je sa skutočne skladá o populáciu zloženú z dvoch menších, pomôže zaradenie ďalšieho regresoru, indikátoru, ktorý bude práve túto skutočnosť zachytávať.



Obr. 2.1: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov chybovej zložky - bimodálne rozdelenie, ťažké chvosty a outliery - v modeli s menším počtom pozorovaní (M1).

V treťom riadku sa nachádzajú grafy menšieho modelu M1 s chybovou zlož-

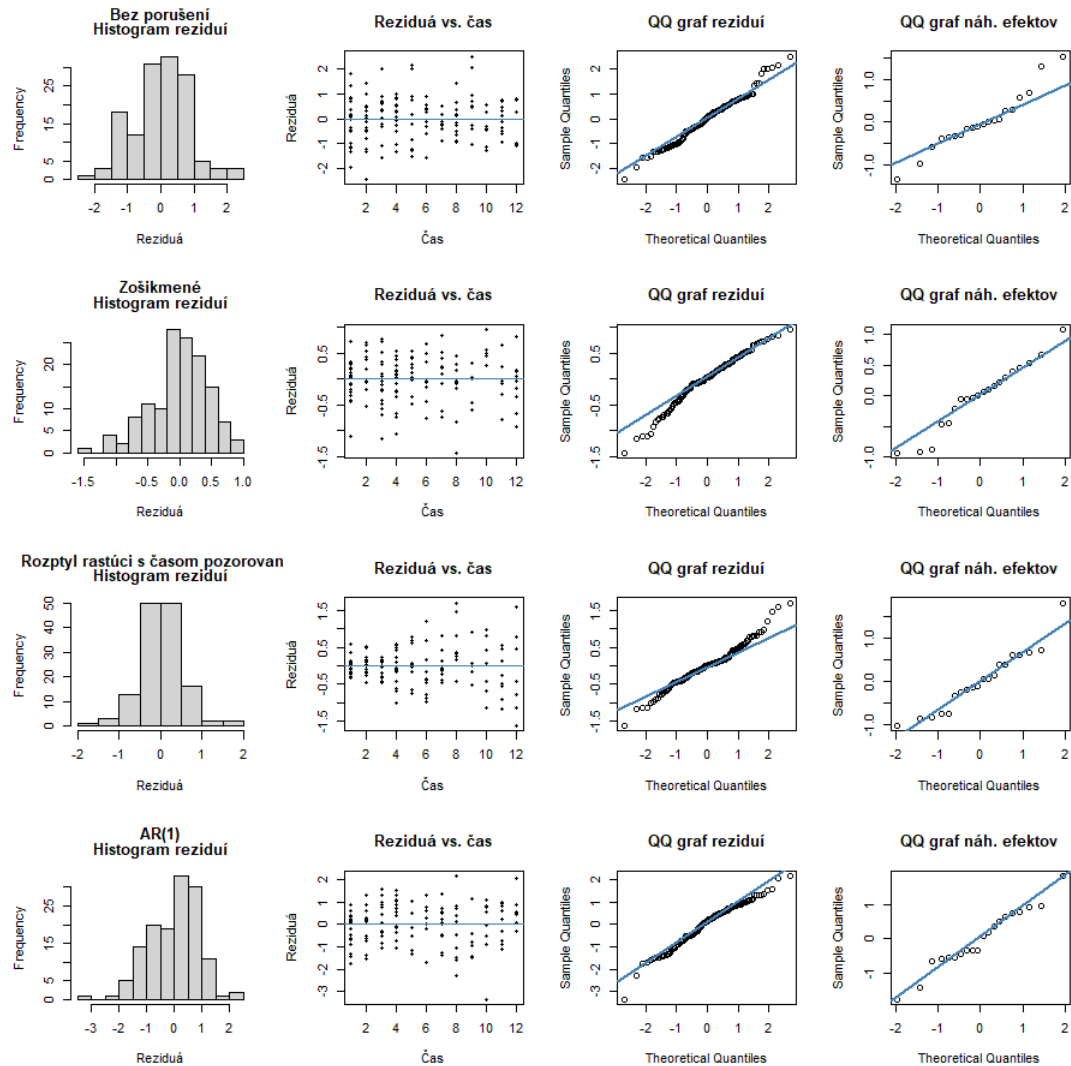
kou sledujúcou t_3 rozdelenie. Toto rozdelenie má ťažšie chvosty ako normálne rozdelenie. Túto vlastnosť môžeme vidieť ako na histograme, tak na scatterplote reziduí. Je potreba podotknúť, že grafy na obrázku 2.1 majú rôzne rozsahy osí, ako horizontálnej tak aj vertikálnej. Reziduá majú väčšiu absolútnu hodnotu v porovnaní s bez porušenia predpokladov modelom, teda sú viac vzdialené od nulovej hodnoty. Rozdelenia sú ale symetrické, takže samotný histogram by normalitu reziduí nevyvrátil. Keď sa ale pozrieme na príslušný QQ-graf, je vidieť, že chvosty (extrémne hodnoty) sa odchyľujú od kvantilov normálneho rozdelenia. V prípade náhodných efektov sa nič nemení, ako v predchádzajúcich prípadoch, vzhľadom na počet subjektov nie je zjavná odchýlka od kvantilov normálneho rozdelenia. V prípade podozrenia na ťažšie chvosty rozdelenia je najviac nápomocný QQ-graf reziduí.

V štvrtom riadku máme normálne rozdelenie reziduí kontaminované Cauchyho (Studentovo t_1) rozdelením. Jednoducho povedané, reziduá obsahujú odlahlé hodnoty (outliery). V takomto prípade histogram aj scatterplot málokedy naznačujú výrazné porušenia predpokladu normality, avšak náš dataset obsahuje veľmi vzdialený outlier, ktorý je vidieť ako na histograme, tak aj na QQ-grafe reziduí. V QQ-grafe náhodných efektov, aj keď je horný koniec mierne vzdialený, vzhľadom na počet subjektov je možné považovať predpoklad normality náhodných efektov za splnený.

Na ďalšom obrázku 2.2 vidíme pokračovanie porovnaní menšieho modelu s porušeniami predpokladov reziduí a modelu so splnenými predpokladmi (1. riadok). V druhom riadku máme porušenie predpokladu normality - 'zošikmené' rozdelenie chybovej zložky (rozdelenie s nenulovou, v tomto prípade zápornou, šikmosťou). Na histograme vidíme, že rozdelenie reziduí nie je symetrické, väčšia časť reziduí je napravo od 0 a naľavo sú reziduá väčšie v absolútnej hodnote. Presne tieto skutočnosti je vidieť aj na scatterplote. Scatterplot ale sám o sebe nemusí hneď navodzovať dojem, že máme 'nenormálne' reziduá. Najjasnejšie je ale vidieť šikmosť reziduí na QQ-grafe reziduí, kde je 'dolný chvost' značne odlišný od kvantilov normálneho rozdelenia. Na QQ-grafe náhodných efektov vidíme, že zošikmené rozdelenie chybovej zložky nemá zásadný vplyv na normalitu náhodných efektov. V prípade podozrenia na zošikmené rozdelenie je vhodné použiť kombináciu grafov histogram, scatterplot a QQ-graf, ktoré ľahko odhalia tento typ porušenia predpokladov.

V treťom riadku sa nachádza model M1 s nehomogénnym rozptylom, ktorý rastie s časom. Na histograme nie je badať odchýlku od normality. Keď sa ale pozornejšie pozrieme na scatterplot, je zľahka vidieť zľava doprava zväčšujúce sa vzdialenosti reziduí (veľkostí v absolútnej hodnote) od nulovej priamky. Tento trend reziduí je jasným znakom nehomogenity rozptylu chybovej zložky. Taktiež porušenie predpokladu normality môžeme vidieť aj na QQ-grafe reziduí, kde sa oba konce odchyľujú od kvantilov normálneho rozdelenia. QQ-graf ale nenaznačuje, aký predpoklad je porušený. QQ-graf náhodných efektov sa v zásade neodlišuje od modelu bez porušenia predpokladov. V prípade nehomogénneho rozptylu chybovej zložky je ideálne použiť scatterplot.

V štvrtom riadku máme autoregresívnu chybovú zložku. Ani na jednom z grafov nie je vidieť zásadný rozdiel v porovnaní s grafmi modelu so splnenými predpokladmi.

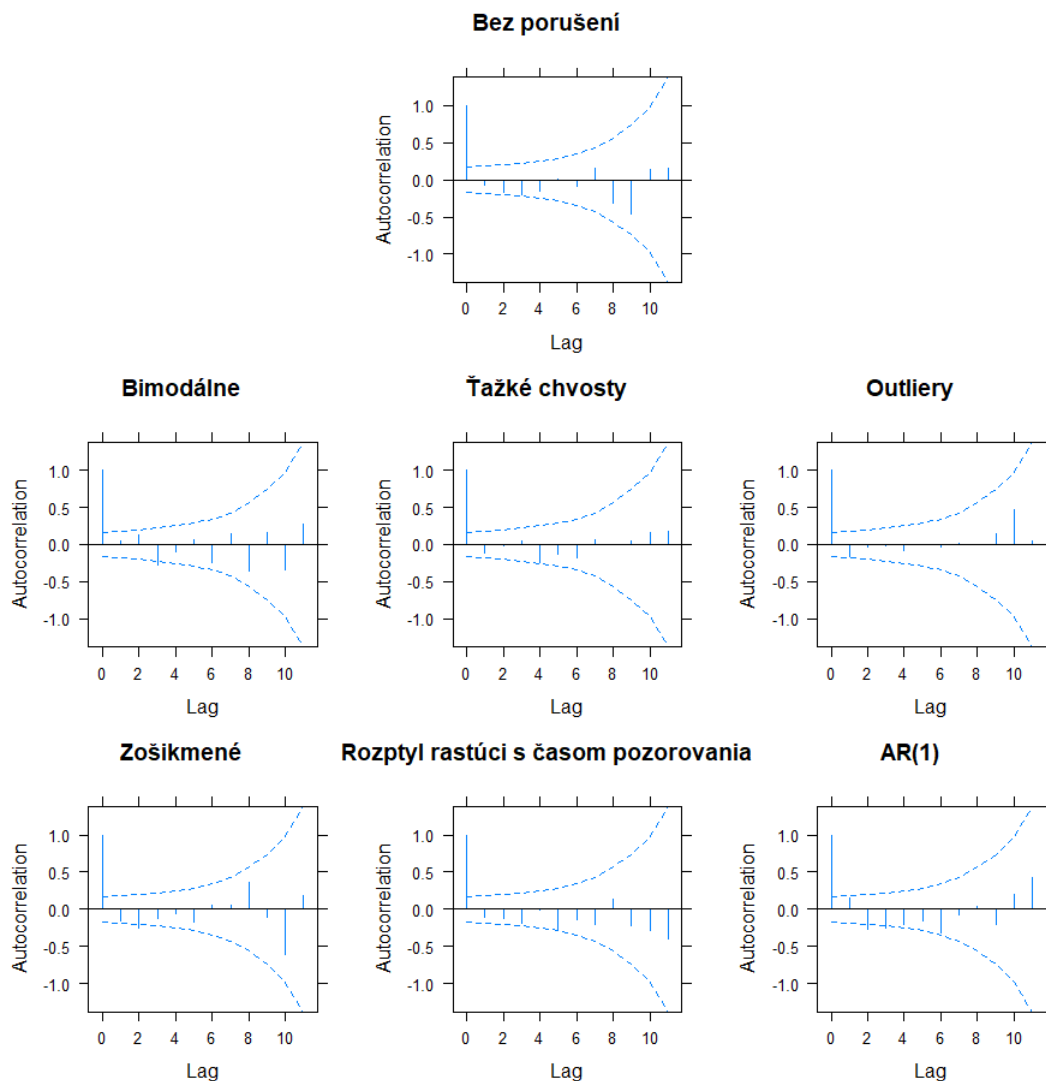


Obr. 2.2: Pokračovanie porovnania grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre ďalšie porušenia predpokladov chybovej zložky - zošíkmené rozdelenie, rozptyl rastúci s časom, AR(1) závislosť - v modeli s menším počtom pozorovaní (M1).

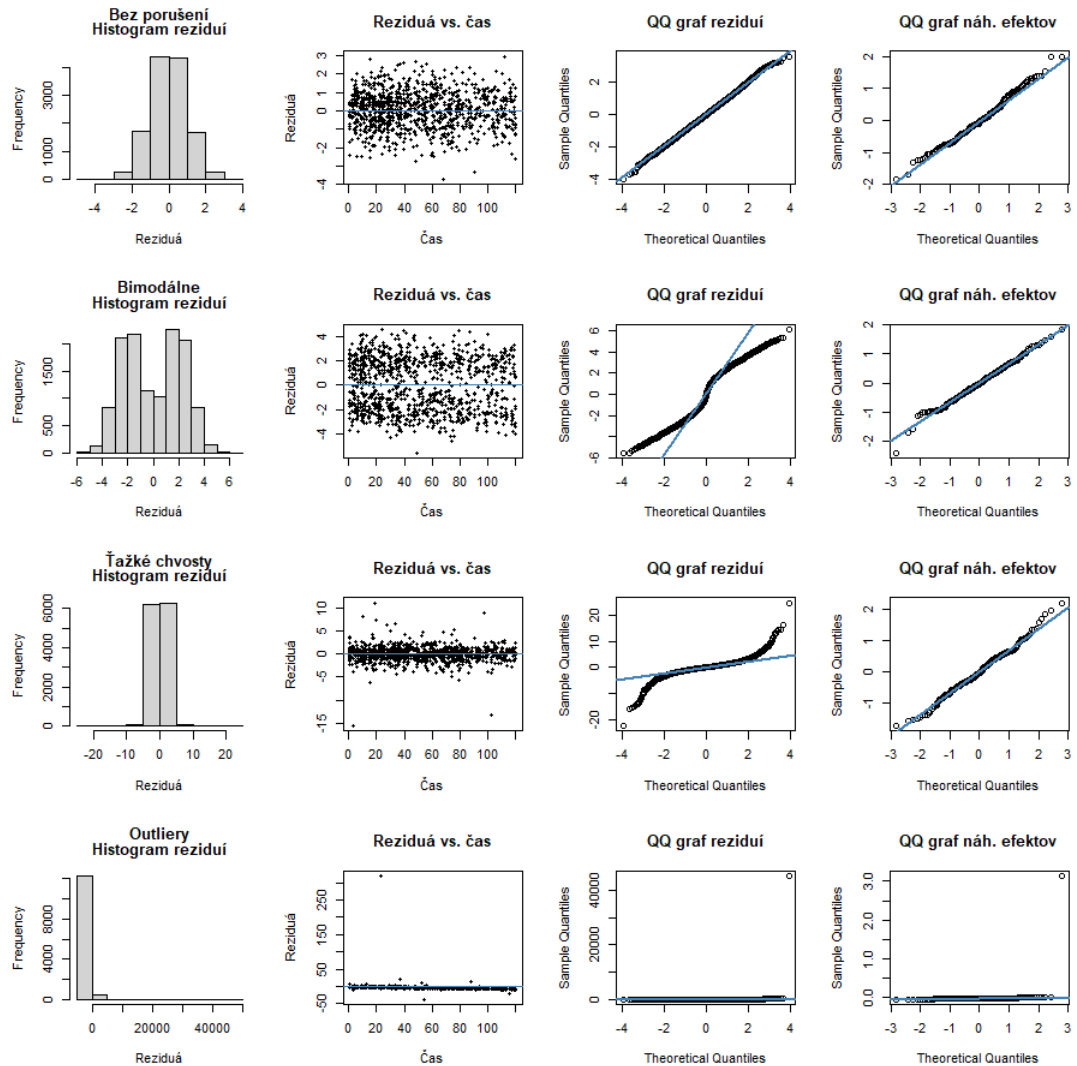
Na overenie platnosti predpokladu o nezávislosti chybovej zložky vrámci subjektu použijeme graf autokorelačnej funkcie. Na vertikálnej osi máme odhadnutú koreláciu, na horizontálnej osi je oneskorenie l ⁹. Korelácia je odhadnutá pre každý časový rozdiel $|t_{ij} - t_{ik}|$, $j, k = 1, \dots, n_i$ a $i = 1, \dots, K$. Tiež sú do grafu doplnené kritické hodnoty $\pm u_{1-\alpha/2} / \sqrt{N(l)}$, kde u_p je p-percentný kvantil normálneho rozdelenia a $N(l)$ je počet reziduálnych dvojíc s oneskorením l (detailsy nájdeme v Chatfield (1980) strana 60-63). Tieto hodnoty sú spojené a vyznačené prerušovanými čiarami. Budeme ich používať ako pravidlo palca na rozhodnutie o splnení či nesplnení predpokladov nezávislosti chybovej zložky vrámci subjektu. Kritické hodnoty sú aproximáciou 95% intervalov spoľahlivosti pre odhady jednotlivých korelácií.

⁹angl. lag

Budeme sa sústrediť na samostatné porušenia ε_{ij} pre malé modely M1. Tieto porušenia budú reprezentovať správanie sa zvyšných neuvedených porušení, pretože v malých modeloch M1 sa autokorelačná štruktúra neprejavuje na odhadoch autokorelačných funkcií. Keď sa pozrieme na obrázok 2.3 je vidieť, že všetky odhady korelácií sú medzi kritickými hodnotami. Navyše, z obrázkov nič nenasvedčuje tomu, že by nejaká autokorelačná štruktúra v dátach bola, v jednotlivých odhadoch korelácií nie je vidieť nejaký trend. Problémom tejto metódy je, že je málo citlivá pri malých rozsahoch výberu. Užitočnosť sa prejaví až na prípade veľkých modelov M2.



Obr. 2.3: Odhady autokorelačných funkcií pre modely s porušeniami predpokladov pre ε_{ij} v malom modeli M1. Grafy obsahujú odhady korelácií v rôznych oneskoreniach l a kritické hodnoty (prerušovaná čiara).

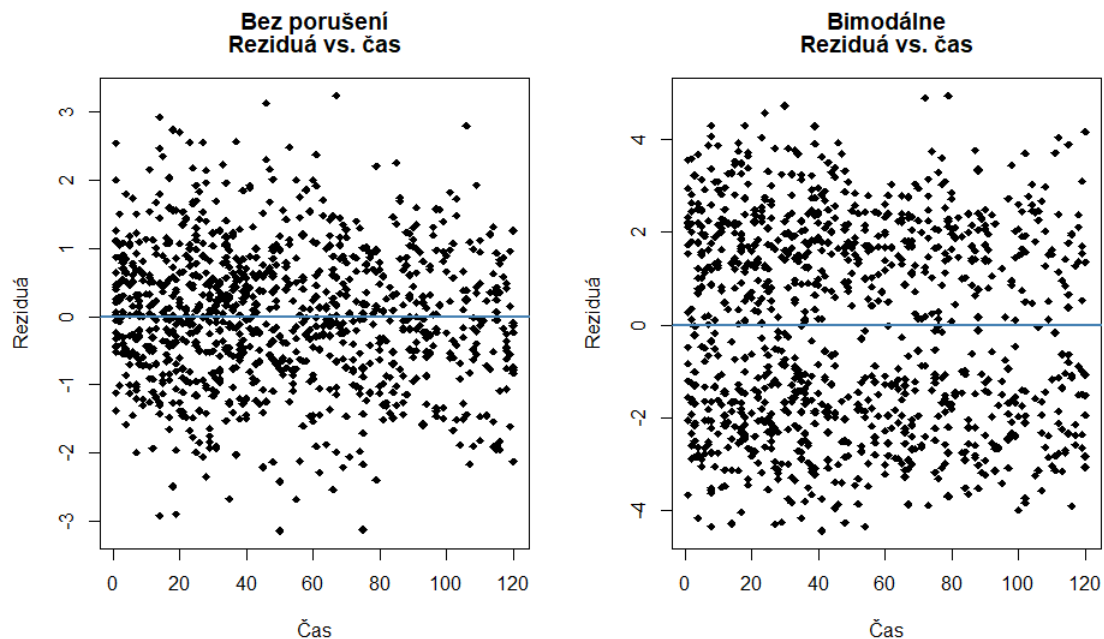


Obr. 2.4: Porovnanie grafov (zľava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov chybovej zložky - bimodálne rozdelenie, ťažké chvosty a outliery - v modeli s väčším počtom pozorovaní (M2).

Na obrázkoch 2.4 (hore) a 2.6 (dole) budú skúmané tie isté porušenia ako na obrázkoch 2.1 a 2.2 avšak na väčšom modeli M2. Väčšina porušení je výraznejšia vo väčšom rozsahu dát. Čo sa ale týka scatterplotu, ten je veľkým množstvom dát zahltený, preto je vybraná len náhodná podmnožina reziduí veľkosti 1000, ktorú môžeme vidieť na všetkých scatterplotoch reziduí vo veľkom modeli M2.

Na obrázku 2.4 v druhom riadku znovu nachádzame bimodálne reziduá v prvých troch grafoch zľava. Histogram jasne naznačuje dve maximá, QQ-graf reziduí vykazuje výraznú odlišnosť kvantilov reziduí a normálneho rozdelenia. Keď sa pozrieme bližšie na scatterplot 2.5, vidíme v porovnaní s modelom so splnenými predpokladmi výrazný rozdiel. Reziduá na grafe modelu s bimodálnou chybovou zložkou sa zhľukujú okolo -2 a 2 .

V treťom a štvrtom riadku máme porušenia s ťažkými chvostami, ktoré sa správajú veľmi podobne. Histogramy sú zhustené do relatívne malého priestoru



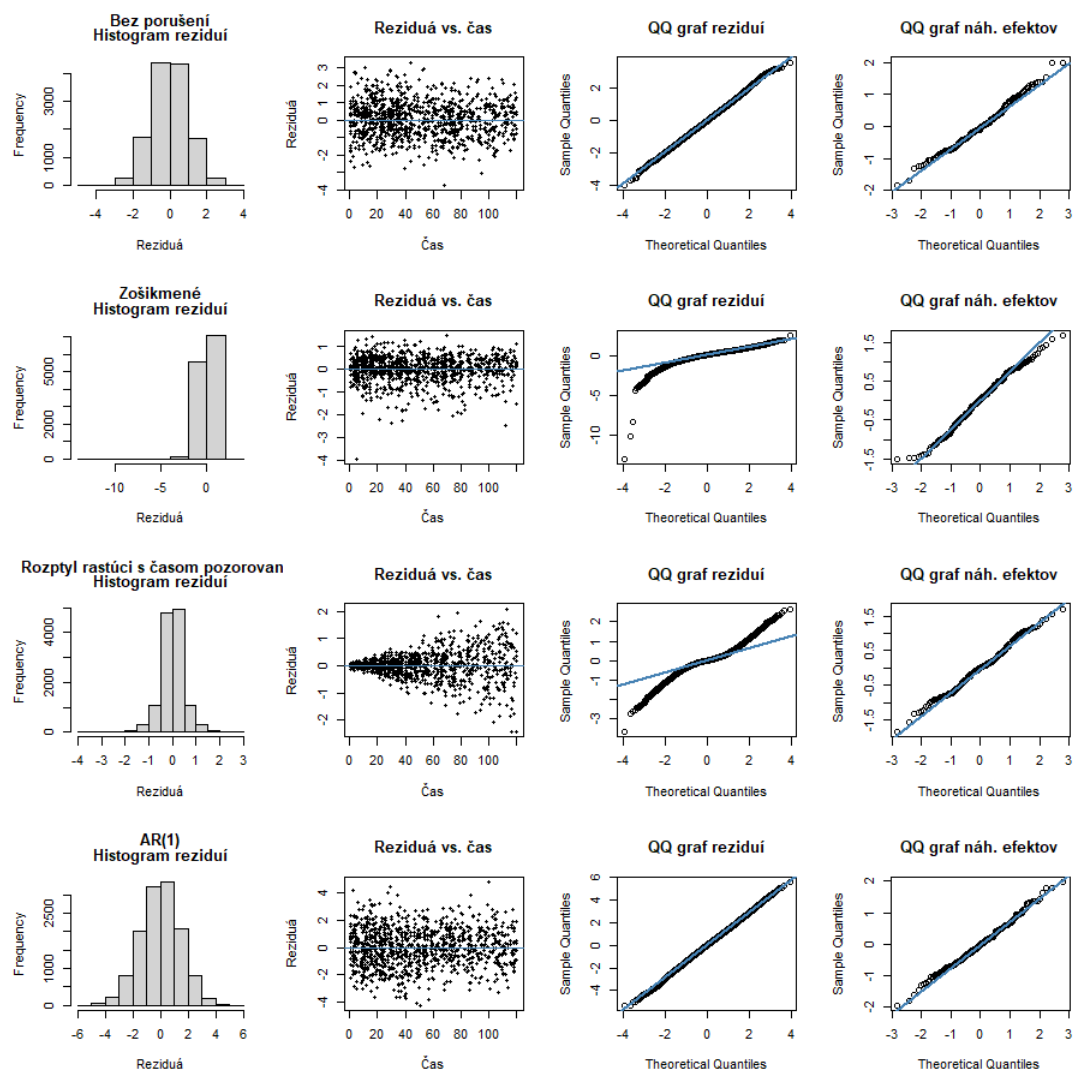
Obr. 2.5: Porovnanie scatterplotov reziduí modelu s bimodálnou chybovou zložkou a modelom so splnenými predpokladmi.

ale existuje množstvo reziduí, ktoré sú od centra viac či menej vzdialené. Podobný pohľad poskytuje scatterplot. V oboch prípadoch je treba brať ohľad na rozsahy osí. Obe porušenia majú netriviálne množstvá reziduí podstatne vzdialené od 0, vo štvrtom riadku (outliery) sú reziduá podstatne viac rozptýlené ako v treťom riadku (ťažké chvosty). Túto istú skutočnosť potvrdzuje aj QQ-graf reziduí. Avšak na QQ-graf náhodných efektov nemajú tieto porušenia zásadný vplyv.

Na obrázku 2.6 vidíme zase v druhom riadku grafy prislúchajúce zošikmenému rozdeleniu chybovej zložky. Ako v predchádzajúcom prípade je porušenie vo väčšom modeli M2 oveľa výraznejšie. Môžeme to pozorovať na histograme, ktorý je nesymetrický (doprava naklonený), scatterplot, ktorého väčšina bodov je okolo 0 a v záporných hodnotách bez kladných protihodnôt. Na QQ-grafe reziduí zase vidíme nesúlad na dolnom chvoste, ktorý predstavuje práve zošikmenie rozdelenia. QQ-graf náhodných efektov je znovu imúnny voči porušeniu predpokladu a rozdelenie sa dá považovať za normálne.

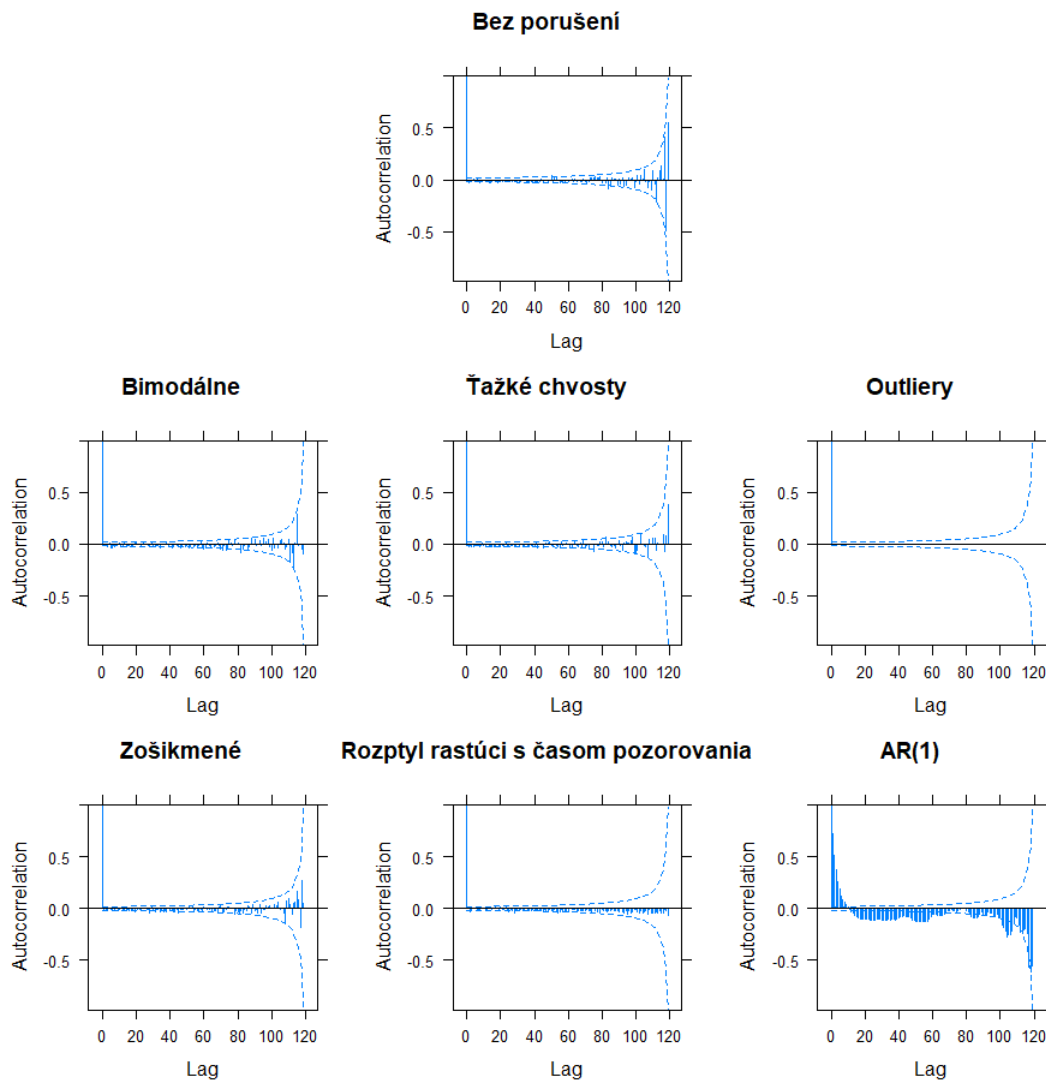
V treťom riadku máme znovu porušenie predpokladu normality chybovej zložky rozptylom rastúcim s časom. Z histogramu nie je patrné nič, avšak zo scatterplotu je zjavné, že rozptyl reziduí rastie s časom. Na QQ-grafe tiež vidieť veľkú odchýlku kvantilov reziduí od kvantilov normálneho rozdelenia. V tomto prípade je teda ideálne použiť na overenie scatterplot, prípadne začať s QQ-grafom a potom pokračovať scatterplotom, ktorý spresní možnosti porušenia predpokladov.

V štvrtom riadku máme autoregresívnu chybovú zložku, ktorá znovu na skúmaných grafoch nevykazuje žiadne porušenia predpokladov.



Obr. 2.6: Pokračovanie porovnania grafov (zľava, histogram reziduí, reziduá proti času, QQ graf reziduí a QQ graf náhodných efektov) pre ďalšie porušenia predpokladov chybovej zložky - zošíkmené rozdelenie, rozptyl rastúci s časom, AR(1) závislosť - v modeli s väčším počtom pozorovaní (M2).

Ďalej, na obrázku 2.7 vidíme odhady autokorelačnej funkcie pre jednotlivé porušenia predpokladov pre ε_{ij} vo veľkom modeli M2. Na grafe AR(1) sa viditeľne prejavuje porušenie predpokladu o nezávislej chybovej zložke vrámci subjektov vysokými hodnotami autokorelačnej funkcie. Hlavne v prvej polovici grafu je vidieť, že odhadnutá korelácia je oveľa väčšia ako kritické hodnoty. V prípade podobného správania autokorelačného grafu by sme mali skontrolovať predpoklad nezávislosti chybovej zložky vrámci subjektu a prípadne uvažovať netriviálnu závislosťnú štruktúru ε_{ij} pri fite modelu. Viac o závislostných štruktúrach nájdeme napríklad v Pinheiro a Bates (2009) strana 266. Vo zvyšných prípadoch porušení predpokladov sa prejavuje autokorelácia na grafoch autokorelačnej funkcie len vtedy, keď jedno z porušení je AR(1) porušenie. Z tohto dôvodu grafy nebudeme uvádzať.



Obr. 2.7: Odhady autokorelačných funkcií pre modely s porušeniami predpokladov pre ε_{ij} vo veľkom modeli M2. Grafy obsahujú odhady korelácií v rôznych oneskoreniach l a kritické hodnoty (prerušovaná čiara).

Dohromady máme v tabuľke 2.4 porovnanie účinností metód na odhaľovanie porušení predpokladov s výnimkou odhadu autokorelačnej funkcie, ktorý je relevantný len v prípade AR(1) porušenia.

Porušenia / Grafy	Histo.	Scatter.	QQ-graf reziduí	QQ-graf n. ef.
Bimodálne	účinné	účinné	účinné	neúčinné
Ťažké chvosty	účinné	účinné	účinné	neúčinné
Outliery	účinné	účinné	účinné	neúčinné
Zošikmené	účinné	účinné	účinné	neúčinné
Heterogénny rozpt.	neúčinné	účinné	účinné	neúčinné
AR(1)	neúčinné	neúčinné	neúčinné	neúčinné

Tabuľka 2.4: Tabuľka účinností jednotlivých metód na odhaľovanie porušení predpokladov chybovej zložky.

Náhodné efekty \mathbf{b}_i

Pre náhodné efekty \mathbf{b}_i budeme uvažovať nasledujúce porušenia predpokladu normality:

- bimodálne rozdelenie

$$b_i^0 \sim H_{ij} \cdot \mathcal{N}(-4, \sigma_e^2 \mathbb{D}) + (1 - H_{ij}) \cdot \mathcal{N}(4, \sigma_e^2 \mathbb{D}),$$

kde $H_{ij} \sim \text{Alt}(0,5)$,

- rozdelenie s nehomogénnym rozptylom, resp. zmes normálnych rozdelení s rôznymi rozptylmi

$$b_i^0 \sim H_{ij} \cdot \mathcal{N}(0, \sigma_e^2 \mathbb{D}) + (1 - H_{ij}) \mathcal{N}(0, 10\sigma_e^2 \mathbb{D}),$$

kde $H_{ij} \sim \text{Alt}(0,5)$,

- rozdelenie

$$b_i^0 \sim H_{ij} \cdot \mathcal{N}(0, \sigma_e^2 \mathbb{D}) + (1 - H_{ij}) t_1,$$

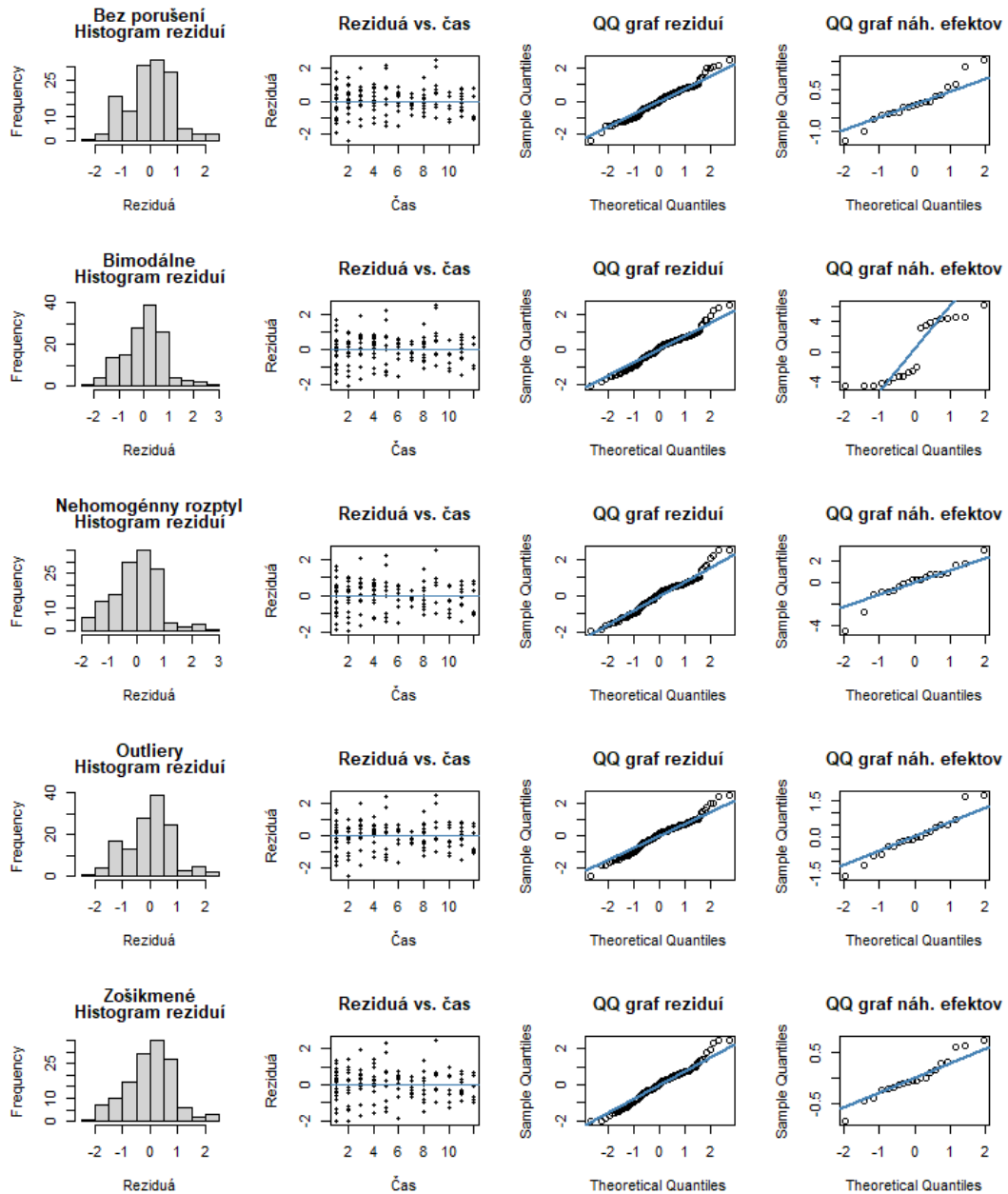
kde $H_{ij} \sim \text{Alt}(0,90)$ pre menší model M1 a $H_{ij} \sim \text{Alt}(0,95)$ pre väčší model M2,

- Pearsonovo rozdelenie b_i^0 typu IV s parametrami $m = \frac{7}{2}$, $\nu = 3$, $a = 1$ a $\lambda = \frac{3}{5}$, ktorého šikmost' je rovná $-\frac{8}{\sqrt{34}} \approx -1,37$ (hustotu a parametrizáciu nájdeme v A.4).

Rovnako ako v časti s chybovou zložkou ε_{ij} budeme za reziduá považovať

$$\widehat{\varepsilon}_{ij} = y_{ij} - \widehat{\beta}_0 - \widehat{\beta}_1 \cdot t_{ij} - \widehat{b}_i^0.$$

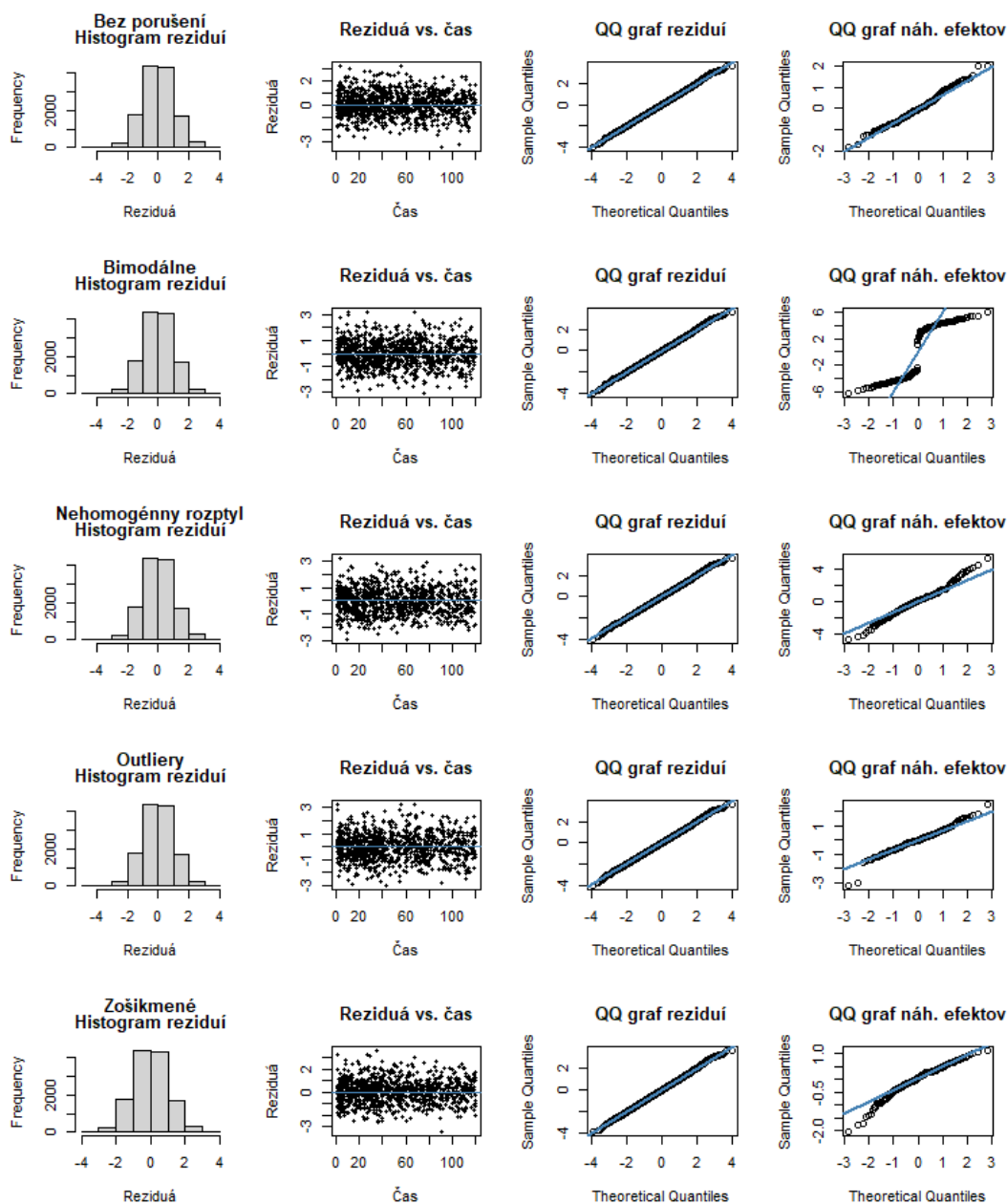
Taktiež využijeme popísané modely M1 a M2 a tiež ten istý typ grafov ako doteraz (histogram reziduí, scatterplot reziduí proti času, QQ-graf reziduí a QQ-graf náhodných efektov).



Obr. 2.8: Porovnanie grafov (zľava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliery a zošíkmené rozdelenie - v modeli s menším počtom pozorovaní (M1).

Keď sa pozrieme na grafy týkajúce sa reziduí v modeli s menším počtom pozorovaní M1 na obrázku 2.8 (prvé tri stĺpce grafov), nič nenasvedčuje tomu, že by mal model porušené predpoklady. Očakávane je vplyv vidieť na QQ-grafoch náhodných efektov. Pre bimodálne rozdelenie náhodných efektov máme podobné správanie QQ-grafu ako v prípade, keď sme mali QQ-graf bimodálneho porušenia na chybovej zložke. Podobne je to aj pre ostatné porušenia. Pri nehomogénnom rozptyle sa dolný chvost QQ-grafu náhodných efektov odchyľuje od normálnych kvantilov, pri porušení outliermi sa nám odchyľuje dosť výrazne posledný bod a nakoniec pri zošíkmenom rozdelení nepozorujeme prakticky žiadnu odchýlku.

V prípade modelov s väčším počtom pozorovaní M2 sú závery veľmi podobné. Jednotlivé grafy na obrázku 2.9 týkajúce sa reziduí nevykazujú žiadne známky porušenia predpokladov. V poslednom stĺpci je znovu vidieť charakteristické črty bimodálneho porušenia (výrazná odlišnosť od kvantilov normálneho rozdelenia), outlierov (veľmi vzdialené kvantily od očakávaných) a zošikmeného rozdelenia náhodných efektov (dolný chvost sa výrazne odlišuje do očakávaného priebehu).



Obr. 2.9: Porovnanie grafov (zľava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliery a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).

Dohromady máme v tabuľke 2.5 účinnosti metód na odhaľovanie porušení predpokladov

Porušenia / Grafy	Histo.	Scatter.	QQ-graf reziduí	QQ-graf n. ef.
Bimodálne	neúčinné	neúčinné	neúčinné	účinné
Heterogénny rozpt.	neúčinné	neúčinné	neúčinné	účinné
Outliery	neúčinné	neúčinné	neúčinné	účinné
Zošikmené	neúčinné	neúčinné	neúčinné	účinné

Tabuľka 2.5: Tabuľka účinností jednotlivých metód na odhaľovanie porušení predpokladov náhodných efektov.

Kombinácia chybovej zložky ε_{ij} a náhodných efektov \mathbf{b}_i

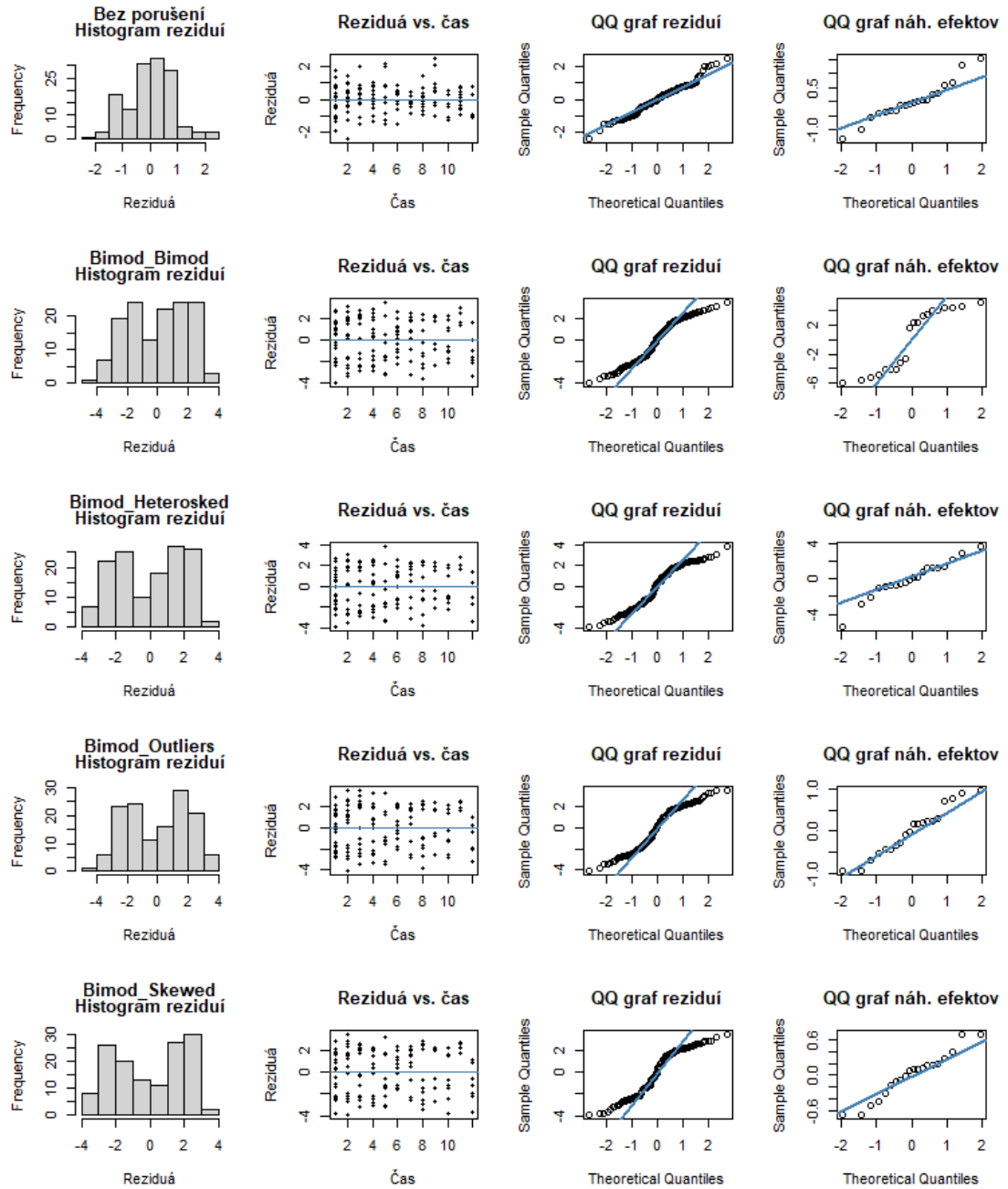
V nasledujúcej časti budeme uvažovať kombinácie porušení predpokladov pre chybovú zložku, aj pre náhodné efekty. V súčte ich je 24 pre malý model M1 a rovnaký počet pre veľký model M2. V zásade sa ale nelíšia od grafov už spomínaných. Predvedieme si to na príklade kombinácii porušení z tabuľky 2.6.

ε_{ij}	\mathbf{b}_i
Bimodálne	Bimodálne
Bimodálne	Heterogénny rozpt.
Bimodálne	Outliery
Bimodálne	Zošikmené

Tabuľka 2.6: Tabuľka kombinácií porušení.

Na obrázkoch 2.10 a 2.11 môžeme vidieť diagnostické grafy pre jednotlivé porušenia. V druhom riadku môžeme vidieť histogram jednoznačne ukazujúci dve maximá rozdelenia reziduí, podobne QQ-graf reziduí je zatočený do typického 'S', ktoré sme mohli vidieť v grafoch vyššie. Rovnaké 'S' vidíme aj v prípade QQ-grafu náhodných efektov. Oba QQ-grafy teda vykazujú nesplnenie predpokladov ako v chybovej zložke, tak v náhodných efektoch. V prípade odhadu autokorelačnej funkcie nie je badať žiadne významné odchýlky, preto tieto grafy nebudeme uvádzať.

Zvyšné grafy nájdeme v časti A.5 Dohromady môžeme konštatovať, že porušenia predpokladov sa správajú v určitom zmysle aditívne, neinterferujú medzi sebou. V prípade kombinácie porušení chybovej zložky aj náhodného efektu sú diagnostické grafy vo väčšine rovnaké, ako v prípade, že je porušený predpoklad len v jednej časti.

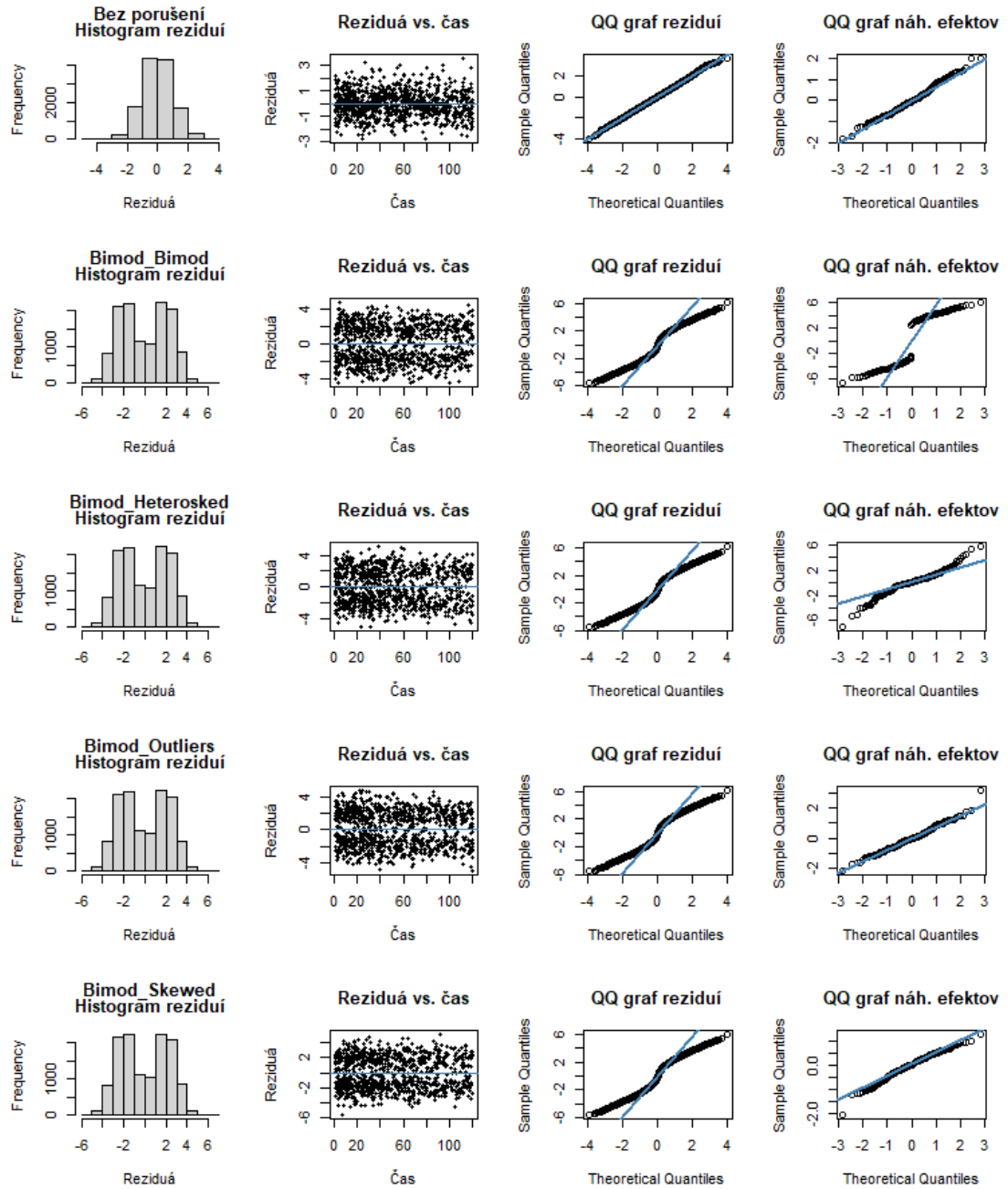


Obr. 2.10: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov popísané v tabuľke 2.6 v modeli s menším počtom pozorovaní (M1).

2.2.2 Model s náhodným sklonom

Model je tvaru

$$\begin{aligned}
 Y_{ij} &= \beta_0 + \beta_1 \cdot t_{ij} + b_i^1 \cdot t_{ij} + \varepsilon_{ij} \\
 &= \beta_0 + (\beta_1 + b_i^1) \cdot t_{ij} + \varepsilon_{ij}
 \end{aligned}$$



Obr. 2.11: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov popísané v tabuľke 2.6 v modeli s väčším počtom pozorovaní (M2).

s $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ a $b_i^1 \sim \mathcal{N}(0, \sigma_e^2 \mathbb{D})$. Model sa líši od modelu (2.1) tým, že má miesto náhodného počiatku náhodný sklon. Pri skúmaní vplyvu porušení na diagnostické grafy boli použité rovnaké predpoklady o porušeníach chybovej zložky ε_{ij} i náhodných efektov \mathbf{b}_i ako v prípade modelu (2.1). Vzhľadom na to, že závery a komentáre k diagnostickým grafom sú totožné ako v prípade modelu (2.1), nebudeme ich tu znovu uvádzať. Grafy sú k nahliadnutiu v externej prílohe.

2.2.3 Model s náhodným počiatkom a sklonom

V tejto časti budeme skúmať kombináciu vyššie popísaných modelov s náhodným počiatkom a náhodným sklonom. Model bude mať tvar

$$\begin{aligned} Y_{ij} &= \beta_0 + b_i^0 + \beta_1 \cdot t_{ij} + b_i^1 \cdot t_{ij} + \varepsilon_{ij} \\ &= (\beta_0 + b_i^0) + (\beta_1 + b_i^1) \cdot t_{ij} + \varepsilon_{ij}, \end{aligned} \quad (2.2)$$

s $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ a $\begin{pmatrix} b_i^0 \\ b_i^1 \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, \sigma_e^2 \mathbb{D})$.

V modeli (2.2) budeme skúmať rovnaké porušenia predpokladov chybovej zložky ako v prípade modelu (2.1). Rozdiel je ale samozrejme v porušeníach predpokladov náhodných efektov. Snahou bolo preniesť vyššie prezentované porušenia do dvoch rozmerov, pretože náhodný efekt \mathbf{b}_i má teraz dvojrozmerné normálne rozdelenie v prípade splnenia predpokladov.

Porušenia predpokladov náhodných efektov \mathbf{b}_i

Budeme uvažovať nasledujúce porušenia predpokladov

- bimodálne rozdelenie

$$\begin{pmatrix} b_i^0 \\ b_i^1 \end{pmatrix} \sim H_{ij} \cdot \mathcal{N}_2\left(\begin{pmatrix} -4 \\ -4 \end{pmatrix}, \sigma_e^2 \mathbb{D}\right) + (1 - H_{ij}) \cdot \mathcal{N}_2\left(\begin{pmatrix} 4 \\ 4 \end{pmatrix}, \sigma_e^2 \mathbb{D}\right),$$

kde $H_{ij} \sim \text{Alt}(0,5)$,

- rozdelenie s nehomogénnym rozptylom, resp. zmes normálnych rozdelení s rôznymi rozptylmi

$$\begin{pmatrix} b_i^0 \\ b_i^1 \end{pmatrix} \sim H_{ij} \cdot \mathcal{N}_2(\mathbf{0}, \sigma_e^2 \mathbb{D}) + (1 - H_{ij}) \mathcal{N}_2(\mathbf{0}, 10\sigma_e^2 \mathbb{D}),$$

kde $H_{ij} \sim \text{Alt}(0,5)$,

- rozdelenie

$$\begin{pmatrix} b_i^0 \\ b_i^1 \end{pmatrix} \sim H_{ij} \cdot \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbb{D}) + (1 - H_{ij}) MVT(\mathbf{0}, 2, \sigma_e^2 \mathbb{D}),$$

kde $H_{ij} \sim \text{Alt}(0,95)$ a $MVT(\boldsymbol{\mu}, \nu, \Sigma)$ je mnohorozmerné Studentovo t rozdelenie s parametrom polohy $\boldsymbol{\mu}$, počtom stupňov voľnosti ν a škálovou maticou Σ . Hustotu nájdeme v A.6,

-

$$\begin{pmatrix} b_i^0 \\ b_i^1 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} X_1 \\ \rho X_1 + \sqrt{1 - \rho^2} X_2 \end{pmatrix},$$

kde X_1 a X_2 sú nezávislé a majú Pearsonove rozdelenie typu IV s parametrami $m_1 = \frac{7}{2}$, $\nu_1 = 3$, $a_1 = 1$ a $\lambda_1 = \frac{3}{5}$ a $m_2 = \frac{7}{2}$, $\nu_2 = 3$, $a_2 = 1$ a $\lambda_2 = -\frac{3}{5}$ a $\rho = 0,75$. Naša snaha bola dosiahnuť zošikmené rozdelenie s nenulovou koreláciou medzi zložkami.

Taktiež sa budeme rovnako pozerat' na dva rôzne veľké modely M1 a M2 s nastavením parametrov

- M1 s nastavením parametrov
 - pevné efekty: $\beta_0 = 5, \beta_1 = 0,5, \sigma_e^2 = 1,$
 - náhodné efekty: $\sigma_e^2 \mathbb{D} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$
 - počet subjektov $K = 20,$
 - celkový počet pozorovaní je $N = 137,$
 - počty subjektov s daným počtom pozorovaní nájdeme v tabuľke 2.7.

Počet pozorovaní	3	4	6	7	8	9	10	12
Počet subjektov	3	3	3	2	4	2	1	2

Tabuľka 2.7: Tabuľka s počtom subjektov s daným počtom pozorovaní pre menší model M1.

	Min.	1.Kvartil	Medián	Priemer	3.Kvartil	Max.
Odozva	-6,5	2,4	4,2	5,0	7,2	28,4
Čas	1,0	2,0	5,0	5,3	8,0	12,0

Tabuľka 2.8: Popisné štatistiky datasetu pre model M1 so splnenými predpokladmi.

- M2 s nastavením parametrov
 - pevné efekty: $\beta_0 = 5, \beta_1 = 0,5, \sigma_e^2 = 1,$
 - náhodné efekty: $\sigma_e^2 \mathbb{D} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$
 - počet subjektov $K = 200,$
 - celkový počet pozorovaní je $N = 12726,$
 - počet pozorovaní na subjektoch sa pohybuje od 3 do 120 s priemerným počtom pozorovaní 63,6, resp. mediánom rovným 63.

	Min.	1.Kvartil	Medián	Priemer	3.Kvartil	Max.
Odozva	-207,5	-0,3	21,9	35,5	62,6	376,4
Čas	1,0	22,0	45,0	50,8	77,0	120,0

Tabuľka 2.9: Popisné štatistiky datasetu pre model M2 so splnenými predpokladmi.

Porušenia predpokladov chybovej zložky ε_{ij} sa prejavujú na všetkých typoch grafov veľmi podobne ako v prípade modelu (2.1), preto ich nebudeme uvádzať. Zaujímavé sú v tomto prípade ale diagnostické grafy náhodných efektov. Budeme používať QQ-grafy na posúdenie normality jednotlivých (marginálnych) rozdelení náhodných efektov \mathbf{b}_i . Taktiež budeme používať QQ-graf tzv. Mahalanobisovej vzdialenosti voči kvantilom χ^2 rozdelenia s dvomi stupňami voľnosti. Nakoniec sa pozrieme aj na scatterplot odhadov jednotlivých zložiek proti sebe (\widehat{b}_i^1 proti \widehat{b}_i^0), teda na vizualizáciu združeného rozdelenia odhadov náhodných efektov $\widehat{\mathbf{b}}_i$.

Mahalanobisova vzdialenosť

V tejto časti si vysvetlíme v jednoduchosti princíp QQ-grafu pre mnohorozmerné normálne rozdelenie. Nech $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ s $\boldsymbol{\mu} \in \mathbb{R}^p$ a Σ pozitívne definitná. Potom vieme, že

$$(\mathbf{X} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2.$$

Ďalej uvažujme náhodný výber \mathbf{X}_1 až \mathbf{X}_n z $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. *Mahalanobisova vzdialenosť* vektoru \mathbf{X}_i od vektoru (priemeru) $\bar{\mathbf{X}}$ je potom definovaná ako

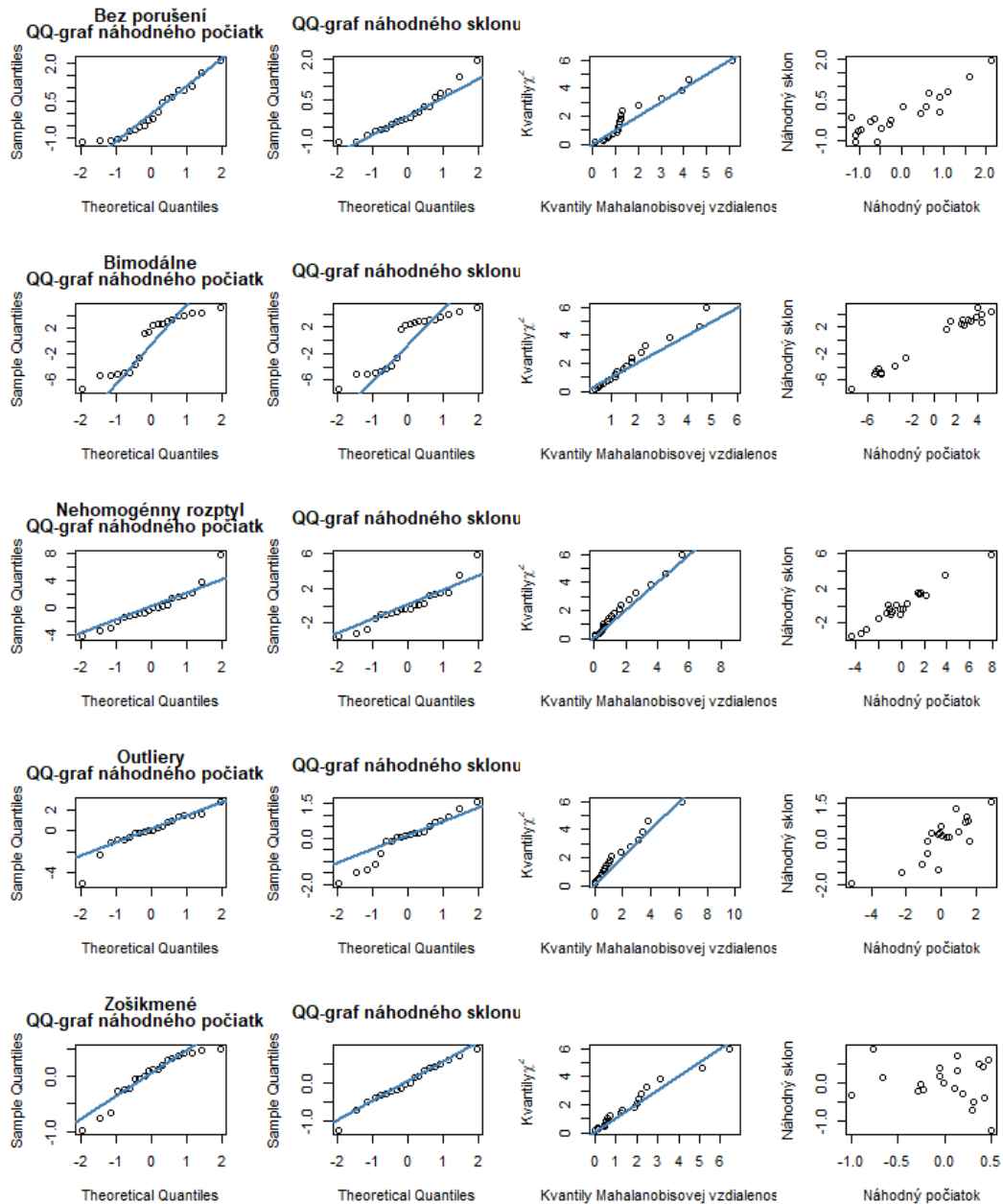
$$MD(\mathbf{X}_i) = \sqrt{(\mathbf{X}_i - \bar{\mathbf{X}})^\top \widehat{\Sigma}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})},$$

kde $\widehat{\Sigma}$ je odhad kovariančnej matice Σ . V prípade náhodného výberu z mnohorozmerného normálneho rozdelenia ale vieme, že druhá mocnina Mahalanobisovej vzdialenosti týchto vektorov od ich (vektorového) priemeru má približne χ^2 rozdelenie s p stupňami voľnosti. Teda, v prípade diagnostiky mnohorozmerného normálneho rozdelenia, môžeme použiť QQ-graf Mahalanobisových vzdialeností od priemeru a príslušného χ^2 rozdelenia, v našom prípade s dvomi stupňami voľnosti. Ďalšie informácie o Mahalanobisovej vzdialenosti a jej praktickom využití nájdeme napríklad v De Maesschalck a kol. (2000), Ghorbani (2019) a Fauconnier a Haesbroeck (2009).

Diagnostické grafy náhodných efektov \mathbf{b}_i

Na obrázku 2.12 vidíme diagnostické grafy s porušením predpokladov pre \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliery a zošíkmené rozdelenie. V prípade bimodálneho rozdelenia vidíme na QQ-grafoch silné odchýlenie od normálnych kvantilov. Pri Mahalanobisovej vzdialenosti sa jemne odchyľuje od kvantilov χ^2 rozdelenia s dvomi stupňami voľnosti. Na scatterplote nie je možné posúdiť normalitu pre malý počet bodov. V zvyšných troch porušeníach vidíme len malé porušenia na QQ-grafoch, takže by sa predpoklad normality dal považovať za splnený. Akurát v prípade outlierov sa kvantily líšia v dolnom chvoste v porovnaní s normálnych rozdelením, ale vzhľadom na počet subjektov $K = 20$ by normalita nemusela byť nutne zamietnutá. Na grafoch Mahalanobisovej vzdialenosti nie je vidieť odchýlenie od očakávaných kvantilov χ^2 rozdelenia s dvomi stupňami voľnosti. Na scatterplotoch nie je možné posúdiť splnenie alebo nesplnenie predpokladu normality, keďže s počtom subjektov $K = 20$ je bodov taktiež len 20, a v takom prípade sú scatterploty veľmi riedke. Podobne vyzerajú aj grafy pre porušenia predpokladov ε_{ij} a tiež kombinácie porušení predpokladov

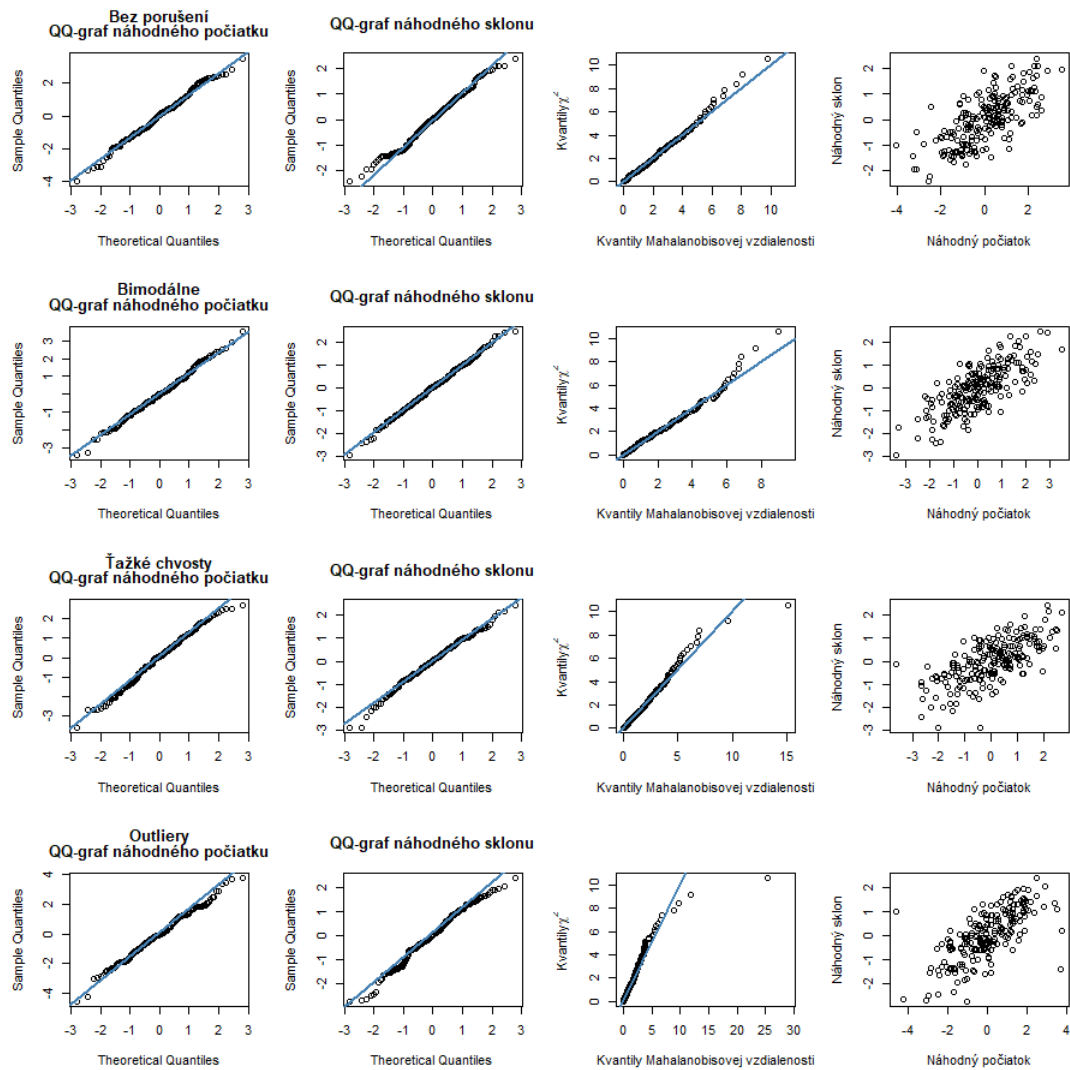
pre ε_{ij} a \mathbf{b}_i . Odhad autokorelačnej funkcie je veľmi podobný s odhadom v prípade modelu s náhodným počiatkom. Vzhľadom na túto skutočnosť nebudeme uvádzať tieto grafy. Pre malé počty subjektov sú použité diagnostické grafy málopravné.



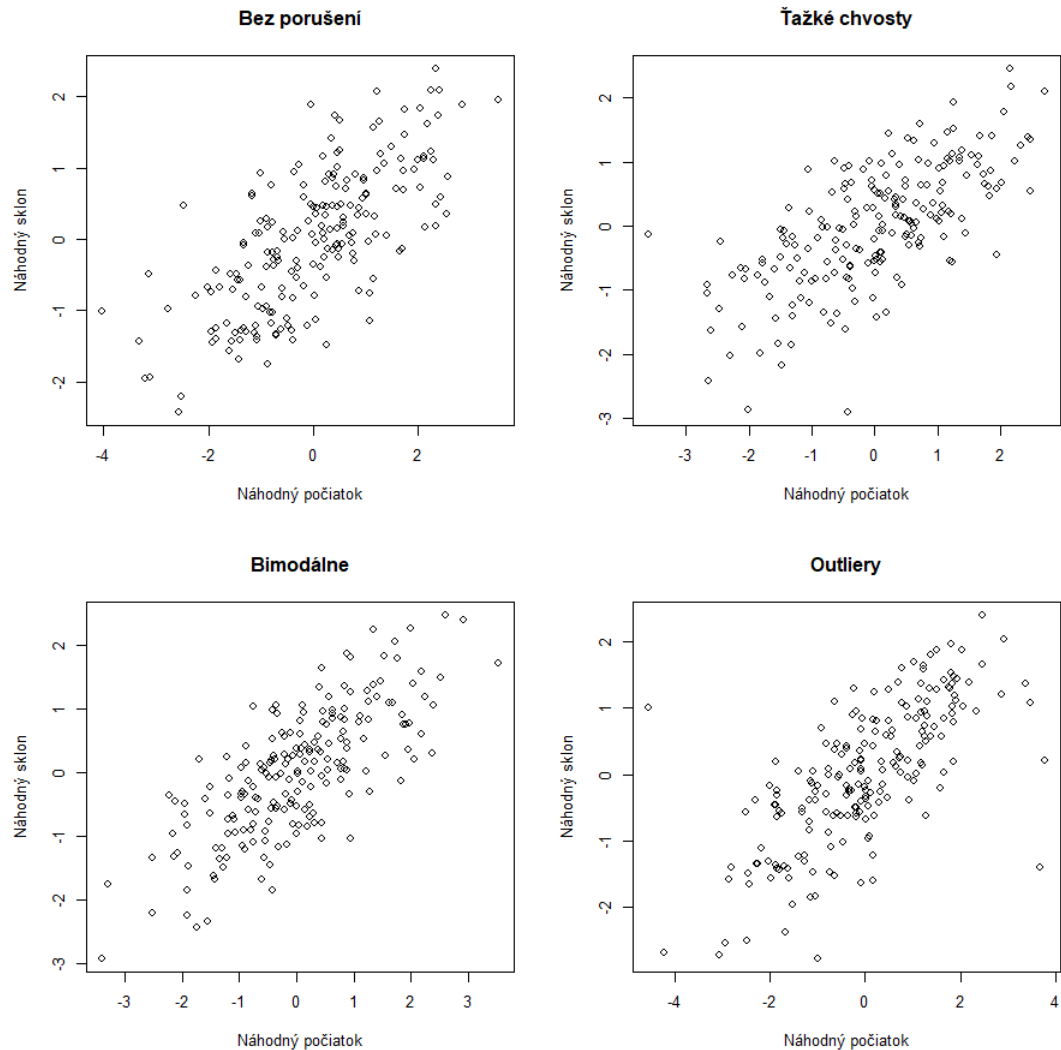
Obr. 2.12: Zľava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku v menšom modeli M1 s porušeniami predpokladov normality \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliery a zošikmené rozdelenie.

Z tohto dôvodu sa presunieme rovno k modelu M2 s väčším počtom pozorovaní, kde sa porušenia prejavujú výraznejšie. Začneme obrázkom, v ktorom sú porušené predpoklady len o ε_{ij} - bimodálne rozdelenie, rozdelenie s ťažkými chvostami a outliery. Na obrázku 2.13 môžeme vidieť podobne, ako v predchádzajúcich prípadoch - prvý riadok, model so splnenými predpokladmi, proti grafom s po-

rušeniami v riadkoch 2 až 4. V prvých dvoch stĺpcoch, v ktorých sa nachádzajú QQ-grafy náhodného počiatku a náhodného sklonu, nie je badať žiadnu odchýlku od normality. Podobne v prípade tretieho stĺpca - QQ-grafu druhej mocniny Mahalanobisovej vzdialenosti. Za jedínú odchýlku by sa dal považovať prípad outlierov, ktorý kazí QQ-graf. V poslednom stĺpci vidíme scatterploty náhodných efektov. Tie môžeme bližšie vidieť na ďalšom obrázku 2.14. Nie je na ňom vidieť veľké odchýlky od modelu bez porušenia predpokladov. Zvyšné porušenia predpokladov ε_{ij} majú veľmi podobné diagnostické grafy, preto ich nebudeme uvádzať. Porušenia predpokladov ε_{ij} teda nemajú vplyv na diagnostické grafy náhodných efektov vo väčšom súbore dát M2.



Obr. 2.13: Zľava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku vo väčšom modeli M2 s porušeniami predpokladov normality ε_{ij} - bimodálne rozdelenie, rozdelenie s ťažkými chvostami a outliery.



Obr. 2.14: Scatterploty náhodných efektov vo väčšom modeli M2 s porušeniami predpokladov normality ε_{ij} - bimodálne rozdelenie, rozdelenie s ťažkými chvostami a outliery.

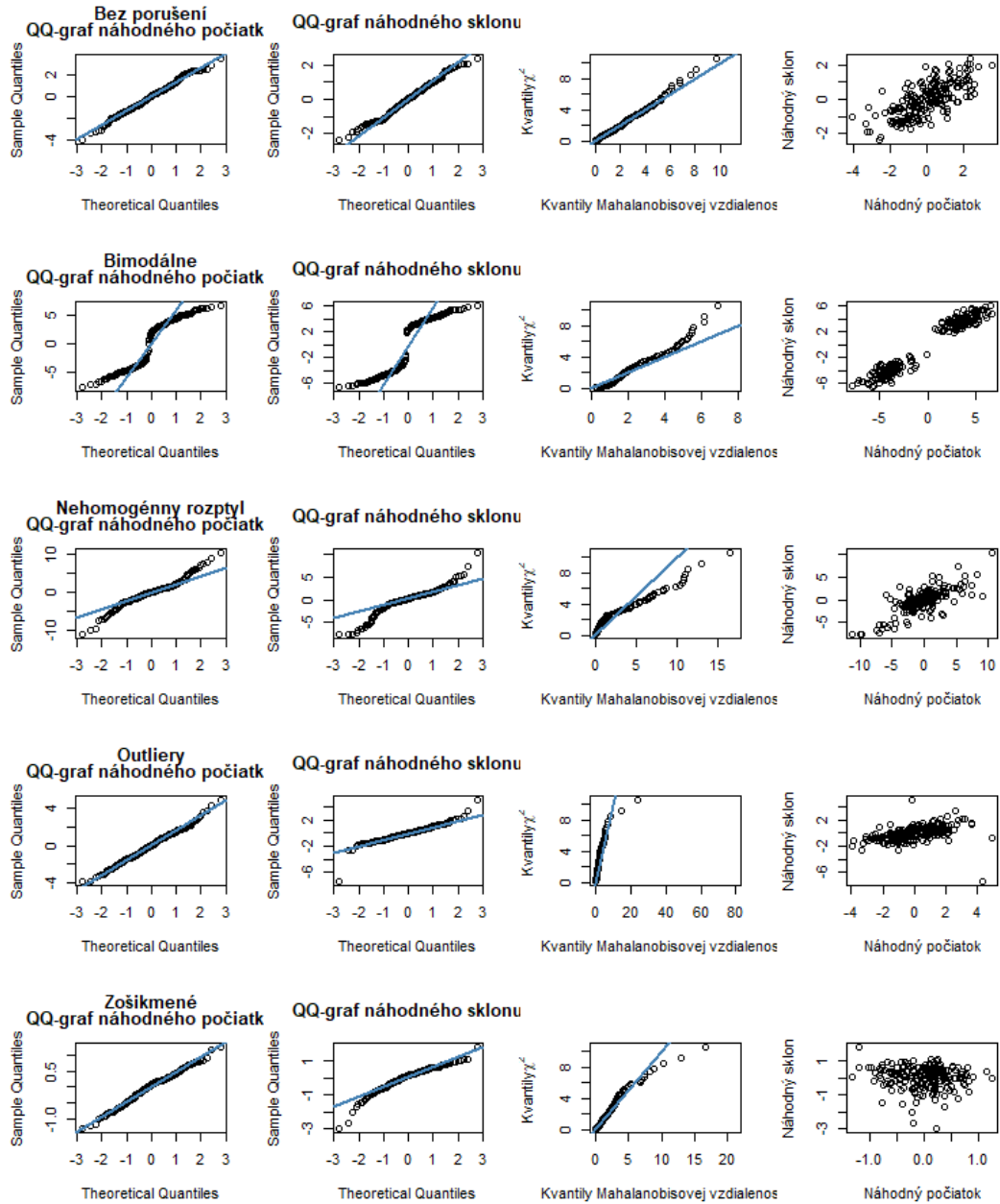
Ďalší obrázok, na ktorý sa pozrieme, bude obsahovať porušenia predpokladov o náhodných efektoch \mathbf{b}_i opísané v 2.2.1. Na obrázku 2.15 môžeme vidieť znovu v prvom riadku model bez porušenia predpokladov. V druhom riadku máme porušenie predpokladu normality náhodných efektov \mathbf{b}_i a to bimodálne rozdelenie zložené z dvoch mnohorozmerných normálnych rozdelení s rôznymi strednými hodnotami. Na marginálnych QQ-grafoch je vidieť už známe 'S', ktoré sme mohli vidieť pri bimodalite v predchádzajúcich prípadoch. QQ-graf štvorca Mahalanobisovej vzdialenosti tiež výrazne vybočuje od kvantilov χ_2^2 rozdelenia. Na scatterplote náhodných efektov vidíme dve oddelené zoskupenia bodov, ktoré práve reprezentujú porušenie bimodalitou, resp. zmesou rozdelení. Lepšie je to vidieť na detaily scatterplotov spolu s ostatnými porušeniami 2.16.

V druhom riadku máme porušenie predpokladu normality náhodných efektov \mathbf{b}_i - nehomogénny rozptyl, resp. zmes normálnych rozdelení s rôznymi rozptylmi. Na marginálnych QQ-grafoch je vidieť odchylenie sa od kvantilov normálneho

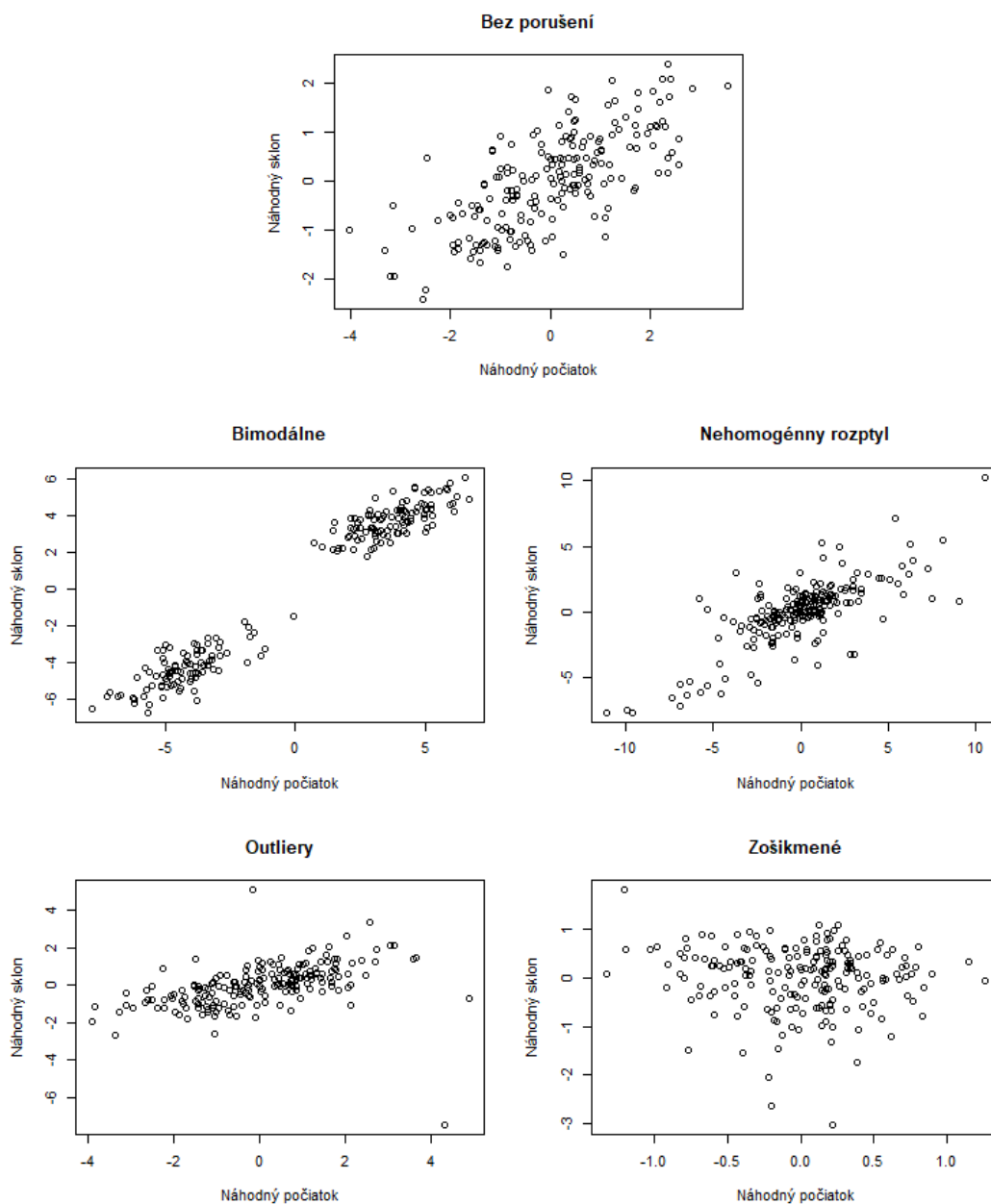
rozdelenia na oboch koncoch, na QQ-grafe štvorca Mahalanobisovej vzdialenosti je tiež výrazná odchýlka od χ_2^2 kvantilov. Na scatterplote, ani na jeho detaile 2.16 by združená normalita s veľkou pravdepodobnosťou nebola vyvrátená. Najlepšie v tejto situácii obstoja QQ-grafy.

V ďalšom riadku sa nachádza porušenie predpokladu normality outliermi, odľahlými hodnotami. Na druhom a treťom QQ-grafe zľava je vidieť malú odchýlku od kvantilov normálneho, resp. χ_2^2 rozdelenia. Na scatterplote je vidieť vzdialenejšie hodnoty výraznejšie. Na odhalenie outlierov je ideálny scatterplot v kombinácii s QQ-grafmi.

V poslednom riadku máme zošikmené rozdelenie náhodných efektov. Prvý QQ-graf zľava nevykazuje žiadne známky porušenia predpokladov, druhý už ale áno, dolný chvost sa odchyľuje. Tiež tretí QQ-graf sa odchyľuje podstatne od kvantilov χ_2^2 rozdelenia. Na scatterplote taktiež vidieť výraznú 'nenormalitu', resp. asymetriu. Na odhalenie zošikmenia sú vhodné všetky vyobrazené diagnostické grafy.



Obr. 2.15: Zľava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku vo väčšom modeli M2 s porušeniami predpokladov normality \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliery a zošíkmené rozdelenie.



Obr. 2.16: Scatterploty náhodných efektov vo väčšom modeli M2 s porušeniami predpokladov normality \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliery a zošíkmené rozdelenie.

Na záver sa pozrieme na jeden z prípadov kombinácie porušenia predpokladov ε_{ij} a \mathbf{b}_i . Kombinácie vidíme v tabuľke 2.10.

ε_{ij}	\mathbf{b}_i
Outliery	Bimodálne
Outliery	Heterogénny rozpt.
Outliery	Outliery
Outliery	Zošikmené

Tabuľka 2.10: Tabuľka kombinácií porušení.

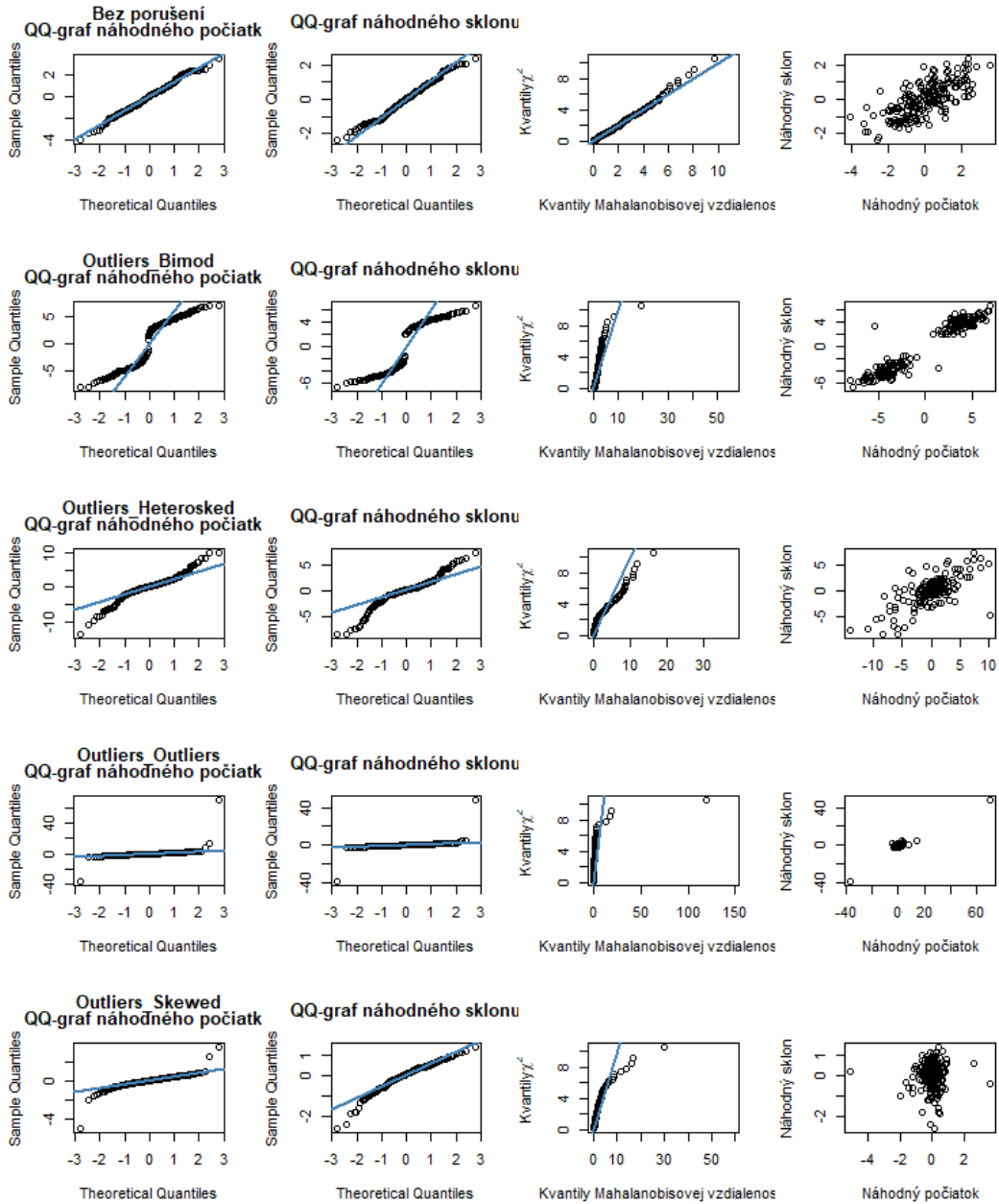
Na obrázkoch 2.17 a 2.18 vidíme diagnostické grafy a detail scatterplotov. V prvom riadku je model so splnenými predpokladmi, v druhom máme kombináciu outliery-bimodálne. Na marginálnych QQ-grafoch je vidieť znovu tvar písmena 'S', ktorý je znakom bimodality. Tiež na scatterplote vidieť oddelené skupiny bodov, ktoré bimodalitu len podporujú.

V treťom riadku máme kombináciu outliery - nehomogénny rozptyl, kde je znovu vidieť odchýlky na marginálnych QQ-grafoch, ako sme už mohli vidieť pri porušení výlučne predpokladov náhodných efektov \mathbf{b}_i .

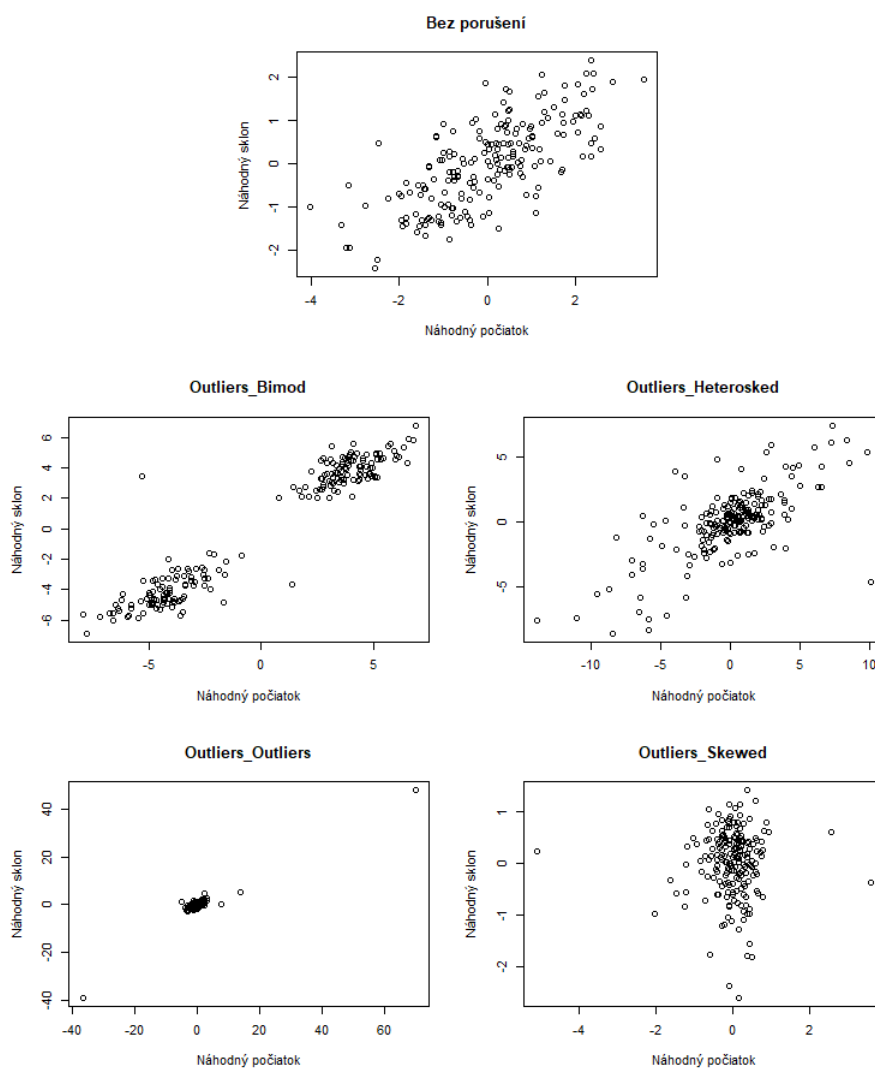
V štvrtom riadku vidíme kombináciu outliery-outliery, ktorá sa prejavuje na všetkých grafoch veľmi vzdialenými bodmi. Na marginálnych QQ-grafoch je vidieť, že sa chvosty odlišujú, na QQ-grafe štvorca Mahalanobisovej vzdialenosti taktiež, najvyššie body sa odchyľujú od kvantilov χ_2^2 . Na scatterplote vidíme tiež jeden veľmi odľahlý bod, ktorý je zjavne outlier.

V poslednom riadku máme kombináciu porušení predpokladov outliery - zošikmené rozdelenie. Znovu sa prejavuje šikmosť rozdelenia na marginálnych QQ-grafoch, tiež na QQ-grafe štvorca Mahalanobisovej vzdialenosti - vo všetkých sa chvosty odlišujú od očakávaných kvantilov - a aj na scatterplote, kde je asymetria tiež badateľná.

Súhrnne, QQ-grafy fungujú na odhaľovanie porušení predpokladov veľmi dobre, dokonca s podobnými až rovnakými vzormi ako pre rovnaké porušenia predpokladov v predchádzajúcich modeloch. Pre viacrozmerné dáta nám pribudla možnosť scatterplotu náhodných efektov \mathbf{b}_i a QQ-grafu Mahalanobisovej vzdialenosti, ktoré sú ale účinné len pre väčšie počty subjektov.



Obr. 2.17: Zľava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku vo väčšom modeli M2 s porušeniami predpokladov z tabuľky 2.10.



Obr. 2.18: Scatterploty náhodných efektov vo väčšom modeli M2 s porušeniami predpokladov z tabuľky 2.10

Kapitola 3

Vplyv porušení predpokladov na hodnoty odhadov a závery testov hypotéz

V tejto kapitole sa budeme venovať vplyvu porušení predpokladov LME modelov pre ε_{ij} a \mathbf{b}_i na hodnoty odhadov parametrov LME modelov a na závery testov hypotéz o týchto parametroch. Porušenia predpokladov budú totožné s porušeniami v kapitole 2, takisto modely, ktoré budú skúmané, sú rovnaké ako v kapitole 2. Budeme sa venovať najzložitejšiemu modelu (2.2), teda modelu s náhodným počiatkom aj sklonom

$$\begin{aligned} Y_{ij} &= \beta_0 + b_i^0 + \beta_1 \cdot t_{ij} + b_i^1 \cdot t_{ij} + \varepsilon_{ij} \\ &= (\beta_0 + b_i^0) + (\beta_1 + b_i^1) \cdot t_{ij} + \varepsilon_{ij} \end{aligned} \quad (3.1)$$

s $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ a $\begin{pmatrix} b_i^0 \\ b_i^1 \end{pmatrix} \sim \mathcal{N}_2(\mathbf{0}, \sigma_e^2 \mathbb{D})$.

Parametre záujmu sú β_0, β_1 a σ_e^2 v prípade fixných efektov a \mathbb{D} v prípade náhodných efektov. Uvažované porušenia predpokladov normality pre ε_{ij} sú

- bimodálne rozdelenie,
- rozdelenie s ťažkými chvostami (Studentovo t_3 rozdelenie),
- normálne rozdelenie s odľahlými hodnotami (outliermi),
- zošikmené rozdelenie (Pearsonovo typu IV),
- rozdelenie s časovo premenlivým rozptylom,
- AR(1) model pre chybovú zložku ε_{ij} .

Detaily týchto porušení nájdeme v 2.2.1. Porušenia predpokladov normality pre náhodné efekty \mathbf{b}_i budú uvažované

- bimodálne rozdelenie,
- rozdelenie s nehomogénnym rozptylom,
- rozdelenie s odľahlými hodnotami (outliermi),

- zošikmené rozdelenie náhodných efektov \mathbf{b}_i .

Details porušení predpokladov \mathbf{b}_i sú popísané v 2.2.3.

Samotné modely s porušenými predpokladmi sú generované rovnako ako v kapitole 2. Každá simulácia modelu je opakovaná 1000 krát a na základe výsledkov z jednotlivých 1000 modelov sú zozbierané sumárne štatistiky.

Pre parametre β_0 a β_1 je to priemer odhadov, priemer smerodajných odchýlok odhadov, empirická smerodajná odchýlka odhadov a pravdepodobnosť pokrytia¹ 95% intervalov spoľahlivosti. Tiež pre oba parametre je odhadnutá sila testu

$$\mathbf{H}_0 : \beta_j = 0 \quad \text{proti} \quad \mathbf{H}_A : \beta_j \neq 0 \quad j = 0,1. \quad (3.2)$$

Navyše pre β_1 je odhadnutá hladina testu (3.2), ktorá je získaná upravením simulácií (nastavenie parametrov týchto simulácií je totožné s parametrami definovanými v kapitole 2 s výnimkou parametra β_1 , ktorý je nastavený na 0 aby bolo možné odhadnúť hladinu testu).

Pre parameter σ_e^2 štatistiky obsahujú priemer odhadov a pravdepodobnosť pokrytia 95% intervalov spoľahlivosti. Pre parameter \mathbb{D} , resp. parametre d_{11} , d_{22} a d_{12} sú to priemery odhadov.

3.1 Odhady parametrov

V tabulke 3.1 a 3.2 vidíme sumárne štatistiky odhadov $\widehat{\beta}_0$ a $\widehat{\beta}_1$ spolu s pravdepodobnosťou pokrytia pre malý (M1) aj veľký model (M2). V prvom stĺpci tabuľky 3.1 vidíme jednotlivé odhady $\widehat{\beta}_0$ pre malé modely M1, ktoré, až na porušenie predpokladov odľahlými hodnotami, sú veľmi blízko skutočnej hodnote $\beta_0 = 5$. Pri porušení predpokladov odľahlými pozorovaniami vidíme mierne vychýlený odhad $\widehat{\beta}_0 = 5,20$. V prípade veľkých modelov M2 sú odhady prakticky rovné skutočnej hodnote β_0 . V druhom stĺpci vidíme priemernú smerodajnú odchýlku $\widehat{\beta}_0$. Výrazne sa znovu odlišujú len outliery, kde je táto smerodajná odchýlka pomerne veľmi veľká (4,46 v prípade M1, 0,33 v prípade M2). Empirická smerodajná odchýlka v treťom stĺpci sa len minimálne odlišuje od priemernej, s výnimkou outlierov, kde je viac ako štvornásobná pre malý model M1 a skoro trojnásobná pre veľký model M2. Vo štvrtom stĺpci môžeme vidieť pravdepodobnosti pokrytia skutočnej hodnoty parametru intervalmi spoľahlivosti. Hodnoty sú v rozmedzí 0,93 až 0,98. Pri porušení outliermi vidíme CP vyššie, ako pri ostatných porušeníach, čo bude najpravdepodobnejšie spôsobené veľkou šírkou intervalov spoľahlivosti, ktoré potom pokrývajú aj skutočnú hodnotu β_0 .

V prípade odhadu $\widehat{\beta}_1$ sú závery podobné. Odhady sa približne zhodujú so skutočnou hodnotou $\beta_1 = 0,5$, s najväčším vychýlením a smerodajnou odchýlkou v prípade outlierov. Taktiež malý model M1 s odľahlými hodnotami má zhruba štvornásobne väčšiu smerodajnú odchýlku, ako ostatné verzie. Je vidieť, že s rastúcim počtom subjektov sa zvyšuje presnosť odhadov, tj. znižuje sa priemerná smerodajná odchýlka jednotlivých odhadov. Empirická smerodajná odchýlka je taktiež veľmi blízka priemernej smerodajnej odchýlke, okrem prípadu malého modelu M1 outlierov, kde je násobne väčšia. CP kolíše medzi 0,93 a 0,96.

¹angl. coverage probability, CP

	Model	$\widehat{\beta}_0$	$\overline{\text{sd}(\widehat{\beta}_0)}$	Empirická sd	CP
Bez porušení	M1	4,97	0,35	0,36	0,93
	M2	5,00	0,10	0,10	0,94
Bimodálne	M1	4,99	0,49	0,47	0,96
	M2	5,00	0,11	0,11	0,95
Ťažké chvosty	M1	4,99	0,42	0,44	0,94
	M2	5,00	0,11	0,10	0,96
Outliery	M1	5,20	4,46	20,22	0,98
	M2	4,99	0,33	0,89	0,97
Zošíkmené	M1	4,99	0,33	0,32	0,95
	M2	5,00	0,10	0,10	0,95
Rozptyl rastúci s časom pozorovania	M1	5,01	0,32	0,32	0,93
	M2	5,00	0,10	0,10	0,95
AR(1)	M1	5,01	0,43	0,43	0,95
	M2	5,00	0,12	0,12	0,94

Tabuľka 3.1: Tabuľka štatistík pre modely s porušeniami predpokladov normality pre ε_{ij} . Tabuľka obsahuje priemer odhadov $\widehat{\beta}_0$, ich smerodajnú odchýlku ($\overline{\text{sd}}$), empirickú smerodajnú odchýlku a pravdepodobnosť pokrytia (CP). Skutočná hodnota parametra $\beta_0 = 5$.

	Model	$\widehat{\beta}_1$	$\overline{\text{sd}(\widehat{\beta}_1)}$	empirická sd	CP
Bez porušení	M1	0,48	0,22	0,23	0,93
	M2	0,50	0,07	0,07	0,94
Bimodálne	M1	0,50	0,23	0,23	0,94
	M2	0,50	0,07	0,07	0,94
Ťažké chvosty	M1	0,51	0,23	0,22	0,95
	M2	0,50	0,07	0,07	0,96
Outliery	M1	0,37	0,84	2,98	0,95
	M2	0,50	0,07	0,07	0,96
Zošíkmené	M1	0,50	0,22	0,23	0,94
	M2	0,50	0,07	0,07	0,96
Rozptyl rastúci s časom pozorovania	M1	0,51	0,22	0,22	0,93
	M2	0,50	0,07	0,07	0,95
AR(1)	M1	0,51	0,23	0,23	0,95
	M2	0,50	0,07	0,07	0,95

Tabuľka 3.2: Tabuľka štatistík pre modely s porušeniami predpokladov normality pre ε_{ij} . Tabuľka obsahuje priemer odhadov $\widehat{\beta}_1$, ich smerodajnú odchýlku ($\overline{\text{sd}}$), empirickú smerodajnú odchýlku a pravdepodobnosť pokrytia (CP). Skutočná hodnota parametra $\beta_1 = 0,5$.

V tabulke 3.3 vidíme odhady $\widehat{\sigma}_e^2$ spolu s pravdepodobnosťou pokrytia a odhady $\widehat{\mathbb{D}}$, resp. \widehat{d}_{11} , \widehat{d}_{22} a \widehat{d}_{12} . Vidíme, že na odhady $\widehat{\sigma}_e^2$ majú porušenia predpokladov značný vplyv. V prípade bimodálneho rozdelenia je odhad rozptylu $\widehat{\sigma}_e^2$ rovný 2,21 pre malý model M1 a 2,24 pre veľký model M2. Odhad rozptylu rozdelenia vstupujúceho do porušenia sa pohybuje okolo hodnoty 3. Procedúra odhadov LME modelu teda neodhaduje sa odlišuje aj od rozptylu rozdelenia vstupujúceho do porušenia, aj od hodnoty σ_e^2 modelu bez porušenia predpokladov.

V prípade ťažkých chvostov je to $\widehat{\sigma}_e^2 = 1,67$, resp. $\widehat{\sigma}_e^2 = 1,73$, v prípade outlierov $\widehat{\sigma}_e^2 = 21,88$, resp. $\widehat{\sigma}_e^2 = 13,34$. V týchto prípadoch je ale teoretický rozptyl nekonečný (pre t-rozdelenia, ktoré vstupujú do porušenia, platí, že ich druhé momenty sú rovné $+\infty$) Pre bližšiu predstavu o rozložení odhadov σ_e^2 pri porušení predpokladov outliermi sú k dispozícii kvantily v tabulke 3.4. Keď si porovnáme priemer a medián odhadov $\widehat{\sigma}_e^2$ pri porušení predpokladov odľahlými hodnotami, vidíme, že priemer je silne ovplyvnený vysokými hodnotami - outliermi. Z tohto dôvodu je priemer odhadov mnohonásobne väčší. Avšak medián je stále veľmi vzdialený (4,97 pre malý model M1, 5,41 pre veľký model M2) od skutočnej hodnoty $\sigma_e^2 = 1$ pre model bez porušenia predpokladov.

V ďalších prípadoch, ktorými sú zošikmené rozdelenie, rozptyl rastúci s časom a AR(1) modely chybovej zložky, je zase odhad rozptylu nižší ako skutočná hodnota. Zaujímavá je časť s pravdepodobnosťou pokrytia, kde vidieť, že pri porušení AR(1) malého modelu M1 je pravdepodobnosť pokrytia nenulová, rovná 0,71. Pri ostatných porušeníach interval spoľahlivosti pokrýval skutočnú hodnotu v menej ako desiatich prípadoch. Je dôležité spomenúť, že zväčšujúci sa počet subjektov a pozorovaní nepriblížil odhad rozptylu k jeho skutočnej hodnote.

	Model	$\widehat{\sigma}_e^2$	CP	\widehat{d}_{11}	\widehat{d}_{22}	\widehat{d}_{12}
Bez porušenia	M1	1,00	0,94	1,99	0,99	0,99
	M2	1,00	0,95	2,00	1,00	0,99
Bimodálne	M1	2,21	0,00	2,31	1,02	0,99
	M2	2,24	0,00	2,01	1,00	1,00
Ťažké chvosty	M1	1,67	0,00	2,19	1,00	0,97
	M2	1,73	0,00	2,01	1,00	1,00
Outliery	M1	21,88	0,00	3376,37	102,31	-395,75
	M2	13,34	0,00	114,28	1,03	0,39
Zošikmené	M1	0,58	0,00	1,99	0,99	0,99
	M2	0,58	0,00	1,99	1,00	1,00
Rozptyl rastúci s časom pozorovania	M1	0,53	0,00	1,91	0,99	1,01
	M2	0,50	0,00	1,99	1,00	1,00
AR(1)	M1	0,92	0,71	3,22	1,02	0,90
	M2	1,40	0,00	2,51	1,01	1,00

Tabuľka 3.3: Tabuľka štatistík pre modely s porušeniami predpokladov normality pre ε_{ij} . Tabuľka obsahuje priemer odhadov $\widehat{\sigma}_e^2$, pravdepodobnosť pokrytia (CP) pre $\widehat{\sigma}_e^2$ a priemery odhadov \widehat{d}_{11} , \widehat{d}_{22} a \widehat{d}_{12} . Skutočná hodnota parametrov $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$.

V tabulke 3.3 sa tiež nachádzajú odhady $\widehat{\mathbb{D}} = \begin{pmatrix} \widehat{d}_{11} & \widehat{d}_{12} \\ \widehat{d}_{12} & \widehat{d}_{22} \end{pmatrix}$. Vidíme, že odhady

Outliery	Model	Priemer	Medián	75%	95%	99%	99,5%	Max
$\widehat{\sigma}_e^2$	M1	21,88	4,97	10,05	54,02	379,02	551,04	2871,75
	M2	13,34	5,41	9,59	44,50	150,08	185,02	714,41

Tabuľka 3.4: Priemer a kvantily odhadov $\widehat{\sigma}_e^2$ pre porušenie predpokladu normality pre ε_{ij} odlahlými hodnotami.

sa výraznejšie odlišujú od skutočnej hodnoty $\mathbb{D} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ pri porušení predpokladov odlahlými hodnotami a pri AR(1) chybovej zložke. Zvyšujúci sa počet subjektov a pozorovaní na subjektoch sice priblížil odhady skutočným hodnotám, no v prípade outlierov je stále odhad \widehat{d}_{11} extrémne vzdialený. Celkovo na odhady parametrov modelu najviac vplývajú odlahlé pozorovania, preto je potrebné pred budovaním modelu si uvedomiť, s akými dátami pracujeme a či práve takéto odlahlé hodnoty neobsahujú.

Prejdeme na porušenia predpokladov \mathbf{b}_i s predpokladmi pre ε_{ij} splnenými. V prípade efektu β_0 nie je na odhade badať žiadny vplyv porušenia predpokladov. Veľkosť modelu sa prejavuje na priemernej smerodajnej odchýlke (väčší model menšia odchýlka) a na pravdepodobnosti pokrytia, ktorá je pre väčšie modely zhruba o jednu stotinu vyššia. Hodnoty nájdeme v appendixe v tabuľke A.1.

V prípade efektu β_1 taktiež porušenia predpokladov pre \mathbf{b}_i veľmi nevplývajú na odhady. Výnimkou je ale zošikmené rozdelenie \mathbf{b}_i , kde je odhad v oboch modeloch (M1 aj M2) rovný $\widehat{\beta}_1 = -0,29$. Detaily nájdeme taktiež v appendixe v tabuľke A.1. V prípade odhadu $\widehat{\sigma}_e^2$ taktiež nie je badať žiadny vplyv porušenia predpokladov na hodnotu odhadu, či pravdepodobnosť pokrytia.

	Model	\widehat{d}_{11}	\widehat{d}_{22}	\widehat{d}_{12}
Bez porušenia	M1	1,99	0,99	0,99
	M2	2,00	1,00	0,99
Bimodálne	M1	17,99	16,94	16,96
	M2	18,01	16,97	16,98
Nehomogénny rozptyl	M1	10,89	5,43	5,42
	M2	10,99	5,51	5,51
Outliery	M1	243,57	67,38	-55,56
	M2	182,68	23,93	25,52
Zošikmené	M1	0,36	0,34	-0,01
	M2	0,34	0,34	0,00

Tabuľka 3.5: Tabuľka štatistík pre modely s porušeniami predpokladov normality pre \mathbf{b}_i . Tabuľka obsahuje priemer odhadov $\widehat{\mathbb{D}}$. Skutočná hodnota parametra $\mathbb{D} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

Nakoniec v tabuľke 3.5 môžeme vidieť odhady $\hat{\mathbb{D}}$. Pri porušení predpokladov bimodálnym rozdelením sú odhady \hat{d}_{11} , \hat{d}_{22} a \hat{d}_{12} zhruba na úrovni 18,17,17 pre malý aj veľký model. Odhadnutá kovariančná matica rozdelenia vstupujúceho do porušení je na úrovni $\hat{\mathbb{D}} = \begin{pmatrix} 10 & 9 \\ 9 & 9 \end{pmatrix}$. Vidíme teda rozdiel medzi odhadom vyplývajúcim z LME procedúry, odhadom kovariančnej štruktúry rozdelenia vstupujúceho do predpokladov a hodnotou \mathbb{D} pre prípad bez porušení predpokladov.

Pri nehomogénnom rozptyle sa zase zhruba udržiava pomer $\frac{d_{11}}{d_{12}}$, resp. $\frac{d_{11}}{d_{22}}$, ktorý je 2 ku 1. Tiež si môžeme všimnúť, že odhadnutá matica kovariačnej štruktúry $\hat{\mathbb{D}}$ vyplývajúca z LME procedúry je približne rovná priemeru (prvok po prvku) kovariančných matíc, ktoré vstupovali do modelu. Porušenie je zmes normálnych rozdelení s kovariančnými maticami $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ a $\begin{pmatrix} 20 & 10 \\ 10 & 10 \end{pmatrix}$ a váhami $\frac{1}{2}$. Priemer matíc je $\begin{pmatrix} 11 & 5,5 \\ 5,5 & 5,5 \end{pmatrix}$. Takisto odhadnutá kovariančná matica rozdelenia vstupujúceho do porušení je blízka hodnote spomínaného priemeru kovariančných matíc. V tomto prípade teda LME procedúra správne odhaduje kovariančnú štruktúru rozdelenia vstupujúceho do porušenia, ale líši sa od hodnôt \mathbb{D} pre prípad so splnenými predpokladmi.

Pri outlieroch sú odhady veľmi vzdialené od skutočnosti (sú mnohonásobne väčšie). Pri zošíkmenom rozdelení taktiež, ale odhady sú menšie ako skutočné hodnoty. Je zjavné, že porušenia predpokladov \mathbf{b}_i majú priamy a silný dôsledok na odhady kovariančnej štruktúry náhodných efektov \mathbf{b}_i .

Hodnoty odhadov pre kombinácie porušení predpokladov pre ε_{ij} a \mathbf{b}_i , nájdeme v appendixe v tabuľkách A.1, A.2 a A.3. Vplyv porušení je veľmi podobný ako v prípadoch samostatných (nekombinovaných) porušení.

3.2 Sila a hladina testu

V tejto časti preskúmame silu a hladinu testu (3.2) s nulovou hypotézou $\mathbf{H}_0 : \beta_i = 0$. Silou testu myslíme pravdepodobnosť zamietnutia nulovej hypotézy za predpokladu, že neplatí. Na hodnoty sily testu sa pozrieme pri rovnakých nastaveniach parametrov a porušeníach predpokladov ako doteraz. Silu testu odhadneme ako počet zamietnutých nulových hypotéz voči celkovému počtu simulácií. Hypotézu zamietame na hladine $\alpha = 0,05$. Vzhľadom na hodnoty počiatku ($\beta_0 = 5$) a smerodajných odchýlok je ale sila testu (3.2) vždy rovná 1, až na prípad outlierov, kde pre malý model M1 je rovná 0,80 a pre veľký model M2 je rovná 0,98. Z rovnakého dôvodu je pre všetky veľké modely M2 sila testu pre parameter β_1 rovná 1. Zaujímavým prípadom sú ale malé modely M1 a testy hypotézy (3.2) pre parameter β_1 . Pre tieto prípady sa sila pohybuje na úrovni 0,60 s výnimkou outlierov, kde je podstatne nižšia, rovná 0,37. Konkrétne hodnoty sily testu (3.2) môžeme vidieť v tabuľke 3.6.

Hladinou testu myslíme pravdepodobnosť zamietnutia nulovej hypotézy za predpokladu, že nulová hypotéza platí.² Na to, aby sme mohli overiť dodržanie hladiny testu, musíme zabezpečiť aby platila nulová hypotéza pre jeden z parametrov β_0 alebo β_1 . Častejšou situáciou v praxi je nulovosť sklonu, preto sme sa rozhodli, že

²angl. type I error, chyba prvého druhu

	Model	Sila testu, $\beta_1 = 0,5$	Hladina testu, $\beta_1 = 0$
Bez porušení	M1	0,58	0,07
	M2	1,00	0,05
Bimodálne	M1	0,57	0,06
	M2	1,00	0,05
Ťažké chvosty	M1	0,61	0,06
	M2	1,00	0,06
Outliery	M1	0,37	0,04
	M2	1,00	0,06
Zošíkmené	M1	0,60	0,06
	M2	1,00	0,04
Rozptyl rastúci s časom pozorovania	M1	0,63	0,07
	M2	1,00	0,05
AR(1)	M1	0,62	0,05
	M2	1,00	0,05

Tabuľka 3.6: Sila testu hypotézy (3.2) pre parameter $\beta_1 = 0,5$ pri jednotlivých porušeníach predpokladov pre ε_{ij} . V druhom stĺpci sa nachádza hladina testu pri nastavení parametru $\beta_1 = 0$.

budeme skúmať hladinu práve pre parameter β_1 . Z tohto dôvodu sme nasimulovali znovu modely spomínané v kapitole 2 (malé M1 aj veľké M2) s tým, že parameter záujmu β_1 sme položili rovný 0. Za odhad dosiahnutej hladiny považujeme pomer zamietnutých nulových hypotéz voči celkovému počtu simulácií. Hladina testu (3.2), na základe ktorej zamietame \mathbf{H}_0 , je rovná 0,05. V tabuľke 3.6 vidíme odhadnuté hladiny testov pre jednotlivé porušenia. Zhruba sa všetky verzie modelov držia hladiny 0,05 s najmenšou hodnotou 0,04 v prípade malého modelu M1 outlierov a veľkého modelu M2 so zošíkmeným rozdelením. Najvyššiu odhadnutú hladinu 0,07 majú malé modely M1 v prípadoch bez porušení a v prípade s rozptylom rastúcim s časom pozorovania.

Sila testov pri porušených predpokladoch pre \mathbf{b}_i je pre β_0 taktiež prakticky všade rovná 1 s výnimkou outlierov, kde je drobná odchýlka (0,98 pre M1, 0,99 pre M2). Vzhľadom na skutočnú hodnotu efektu $\beta_1 = 0,5$ a veľkosti smerodajných odchýlok, je sila testu (3.2) oveľa nižšia pre malé modely M1 (od 0,09 v prípade bimodálneho rozdelenia po 0,60 v prípade zošíkmeného rozdelenia, prípad bez porušení 0,58) a aj pre niektoré veľké modely M2 (0,40 pre bimodálne rozdelenie, 0,86 pre nehomogénny rozptyl a 0,88 pre outlierov) v porovnaní so silami testov pri porušených predpokladoch pre ε_{ij} .

Hladinu testy dodržujú s malými odchýlkami, podobne ako v prípade porušení predpokladov pre ε_{ij} . Väčšina väčších modelov má dosiahnutú hladinu testu blízku 0,05 (outliery 0,04). Výnimkou je však zošíkmené rozdelenie. Toto porušenie má hladinu testu rovnú 1 (správne sme nezamietli nulovú hypotézu v menej ako 10 prípadoch). Odhad $\widehat{\beta}_1$ je v tomto prípade $\widehat{\beta}_1 = -0,78$ so smerodajnou odchýlkou 0,13 pre malý model M1 a $\widehat{\beta}_1 = -0,79$ so smerodajnou odchýlkou 0,04 pre veľký model M2, čoho dôsledkom je aj jednotková hladina testu v oboch prípadoch. Všetky hodnoty nájdeme v tabuľke 3.7.

Silu a hladinu testu pre zostávajúce kombinácie porušení predpokladov pre ε_{ij}

	Model	Sila testu, $\beta_1 = 0,5$	Hladina testu, $\beta_1 = 0$
Bez porušení	M1	0,58	0,07
	M2	1,00	0,05
Bimodálne	M1	0,09	0,06
	M2	0,40	0,05
Nehomogénny rozptyl	M1	0,20	0,06
	M2	0,86	0,05
Outliery	M1	0,46	0,06
	M2	0,88	0,04
Zošikmené	M1	0,60	1,00
	M2	1,00	1,00

Tabuľka 3.7: Sila testu hypotézy (3.2) pre parameter $\beta_1 = 0,5$ pri jednotlivých porušeníach predpokladov pre \mathbf{b}_i . V druhom stĺpci sa nachádza hladina testu pri nastavení parametru $\beta_1 = 0$.

a \mathbf{b}_i nájdeme v appendixe v tabuľkách A.4 a A.5.

Záver

V našej práci sme sa venovali lineárnym modelom so zmiešanými efektami, ich diagnostike a dôsledkom nesplnenia predpokladov.

V prvej kapitole sme uviedli potrebné definície lineárnych zmiešaných modelov, metódy odhadov jednotlivých parametrov (fixné efekty, kovariančná štruktúra náhodných efektov) a testy hypotéz. Testy sa týkajú nulovosti fixných efektov $\mathbf{H}_0 : \beta_j = 0$, celkovej prítomnosti náhodných efektov $\mathbf{H}_0 : \mathbb{D} = \mathbf{0}$ a prítomnosti jednotlivých náhodných efektov $\mathbf{H}_0 : d_{22} = 0$.

V druhej kapitole sme sa venovali grafickej diagnostike modelov s porušenými predpokladmi. Porušenia predpokladov sa týkajú ε_{ij} a \mathbf{b}_i (buď samostatne alebo v kombinácii). Najsilnejšou zbraňou na diagnostiku odchýlení od predpokladov sa javí QQ-graf - či už QQ-graf reziduí, náhodných efektov, či QQ-graf Mahalanobisovej vzdialenosti. V prípade porušenia predpokladov pre ε_{ij} je QQ-graf citlivý vo všetkých možnostiach (tvar 'S' pre bimodálne rozdelenie, odchýlené nízke, resp. vysoké kvantily pre porušenia predpokladov ťažkými chvostami, odlahlými pozorovaniami, zošikmeným rozdelením a s rozptylom rastúcim s časom) s výnimkou AR(1) chybovej zložky. Pre veľké modely M2 bol užitočný aj scatterplot reziduí, hlavne pre porušenia bimodálnym rozdelením, outliermi a rozptylom rastúcim s časom. Na histogram mali vplyv bimodálne rozdelenie, outlieri a zošikmené rozdelenie. Porušenia predpokladov pre ε_{ij} nemali prakticky žiadny vplyv na QQ-grafy náhodných efektov. Na odhalenie závislosti reziduí vrámci subjektu je najviac vhodný graf odhadu autokorelačnej štruktúry. Avšak porušenie predpokladu nezávislosti je viditeľné len v prípade veľkých modelov M2. Celkovo, všetky porušenia predpokladov ε_{ij} sa väčšmi prejavujú vo väčšom modeli M2.

Pre porušenia predpokladov pre \mathbf{b}_i platí pre QQ-grafy reziduí to isté, čo pre porušenia predpokladov pre ε_{ij} . Taktiež je viditeľný tvar 'S' pre bimodálne rozdelenie a odchyľujúce sa chvosty pre ostatné porušenia predpokladov pre \mathbf{b}_i . Pre prípad modelu s náhodným počiatkom aj sklonom na QQ-grafoch Mahalanobisovej vzdialenosti je viditeľná odchýlka na hornom chvoste pre všetky porušenia. Navyše je vidieť jasné porušenia predpokladov v scatterplotoch znovu hlavne pre väčšie modely M2. Dva oddelené zhluky bodov pre bimodálne rozdelenie, nerovnomerne rozdelené body pre heterogénny rozptyl, veľmi vzdialené body pre outlieri a nesymetrický oblak bodov pre zošikmené rozdelenie.

Pre kombinácie porušenia predpokladov sa jednotlivé porušenia pre ε_{ij} a \mathbf{b}_i správajú 'aditívne', tj. porušenia vykazujú rovnaké známky nedodržania predpokladov v príslušných diagnostických grafoch pre ε_{ij} a \mathbf{b}_i .

V tretej kapitole sme sa zaoberali vplyvu porušenia predpokladov na hodnoty odhadov, silu a hladinu testov pre β_0 a β_1 . V prípade porušenia predpokladov pre ε_{ij} sú odhady parametrov $\widehat{\beta}_0$ a $\widehat{\beta}_1$ veľmi blízko skutočnej hodnote. Najväčšiu odchýlku od skutočnej hodnoty má porušenie odlahlými pozorovaniami.

Pre odhady parametru $\widehat{\sigma}_e^2$ sú ale výsledky diametrálne odlišné. Vo všetkých porušeníach sa odhad výrazne odchyľuje od skutočnej hodnoty parametra σ_e^2 . Dôležité je ale podotknúť, že rozdelenia, ktoré vstupujú do porušení predpokladov nemajú nutne rozptyl $\sigma_e^2 = 1$. Pre prípad parametrov d_{11} , d_{22} a d_{12} sú odhady relatívne blízke skutočným hodnotám parametrov, hlavne pre veľké modely M2, s výnimkou outlierov, kde sú tieto odhady veľmi nepresné. Prítomnosť odľahlých hodnôt môže silne zmeniť odhady parametrov.

V prípade porušení predpokladov pre \mathbf{b}_i sú odhady parametrov $\widehat{\beta}_0$, $\widehat{\beta}_1$ a $\widehat{\sigma}_e^2$ prakticky rovné skutočnej hodnote. Očakávané majú tieto porušenia predpokladov silný dopad na odhady kovariančnej štruktúry. Najviac sa odchyľujú outlieri. Znovu ale treba dodať, že kovariančná štruktúra, ktorá vstupuje do porušení, nemá nutne totožnú kovariančnú štruktúru s prípadom so splnenými predpokladmi ($d_{11} = 2$ a $d_{12} = d_{22} = 1$).

Kombinácia porušení predpokladov má znovu akúsi 'aditívnu' povahu. Porušenia, ktoré mali vplyv na parametre β_0 , β_1 a σ_e^2 , majú vplyv aj v tomto prípade a naopak.

Ďalej sme sa venovali testu hypotézy (3.2) na hladine 0,05. Vzhľadom na veľkosť fixného efektu β_0 a príslušných smerodajných odchýlok je ale sila testu vždy rovná 1. Pre parameter β_1 je sila testu pre porušenia predpokladov ε_{ij} pre malé modely zhruba na úrovni 0,6, pre veľké modely je znovu rovná 1. Hladinu testu (3.2) sme skúmali na modeloch s upraveným parametrom β_1 rovným 0. Hladina testu je pre malé modely vyššia zhruba o 1 až 2 percentá s výnimkou outlierov, kde je hladina nižšia. V prípade veľkých modelov je hladina rovná 0,05 okrem porušenia odľahlých hodnôt, kde je rovná 0,06.

Pre porušenia predpokladov \mathbf{b}_i je sila testu výrazne nižšia, ako pre porušenia ε_{ij} . Hladina testu sa drží na úrovni porušení ε_{ij} okrem zošikmeného rozdelenia, kde je hladina rovná 1, a teda sme správne nezamietli nulovosť efektu v menej ako 10 prípadoch.

Celkovo, najväčší vplyv na diagnostické grafy, odhady parametrov a závery testov hypotézy (3.2) majú odľahlé hodnoty, preto je potrebné pred modelovaním urobiť dôkladnú analýzu dát, či práve takéto hodnoty neobsahujú.

Zoznam použitej literatúry

- CHATFIELD, C. (1980). *The Analysis of Time Series: An Introduction*. Springer US. ISBN 978-0-412-22460-7,978-1-4899-2923-5.
- CRAMER, H. (1946). *Mathematical Methods of Statistics*. ISBN 0691005478,9780691005478.
- DE MAESSCHALCK, R., JOUAN-RIMBAUD, D. a MASSART, D. (2000). The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, **50**(1), 1–18. ISSN 0169-7439. doi: [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7). URL <https://www.sciencedirect.com/science/article/pii/S0169743999000477>.
- DEMIDENKO, E. (2013). *Mixed Models: Theory and Applications with R*. Wiley Series in Probability and Statistics. Wiley. ISBN 9781118593066.
- FAUCONNIER, C. a HAESBROECK, G. (2009). Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology*, **6**(4), 363–379. ISSN 1572-3127. doi: <https://doi.org/10.1016/j.stamet.2008.12.005>. URL <https://www.sciencedirect.com/science/article/pii/S1572312708000853>.
- GHORBANI, H. (2019). Mahalanobis distance and its application for detecting multivariate outliers. *Facta Universitatis Series Mathematics and Informatics*, **34**, 583. doi: 10.22190/FUMI1903583G.
- GOLUB, G. H. a VAN LOAN, C. F. (2012). *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, fourth edition. ISBN 1421407949,9781421407944.
- GRAYBILL, F. A. (1982). *Matrices with Applications in Statistics*. Brooks/Cole, 2nd revised edition edition. ISBN 0534980384,9780534980382.
- HARVILLE, D. (2008). *Matrix Algebra From a Statistician's Perspective*.
- HENDERSON, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph. ISBN 9780889550308.
- KOTZ, S. a NADARAJAH, S. (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press.
- LAIRD, N. M. a WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**(4), 963–974.

- LINDLEY, D. V. a SMITH, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34** (1), 1–41. ISSN 00359246.
- NAGAHARA, Y. (1999). The pdf and cf of pearson type iv distributions and the ml estimation of the parameters. *Statistics & Probability Letters*, **43** (3), 251–264. ISSN 0167-7152. doi: [https://doi.org/10.1016/S0167-7152\(98\)00265-X](https://doi.org/10.1016/S0167-7152(98)00265-X). URL <https://www.sciencedirect.com/science/article/pii/S016771529800265X>.
- PINHEIRO, J. C. a BATES, D. M. (2009). *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer, New York, NY. ISBN 9780387989570.
- RAO, C. R. (2001). *Linear Statistical Inference and its Applications*. Wiley Series in Probability and Statistics. Wiley, second edition. ISBN 978-0-471-21875-3.
- SMITH, A. F. M. (1973). A general bayesian linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **35**(1), 67–75. ISSN 00359246.
- STRAM, D. O. a LEE, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, **50**(4), 1171–1177.
- WAND, M. P. (2002). Vector differential calculus in statistics. *The American Statistician*, **56**(1), 55–62. ISSN 00031305.

Zoznam obrázkov

2.1	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov chybovej zložky - bimodálne rozdelenie, ťažké chvosty a outliery - v modeli s menším počtom pozorovaní (M1).	26
2.2	Pokračovanie porovnania grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre ďalšie porušenia predpokladov chybovej zložky - zošikmené rozdelenie, rozptyl rastúci s časom, AR(1) závislosť - v modeli s menším počtom pozorovaní (M1).	28
2.3	Odhady autokorelačných funkcií pre modely s porušeniami predpokladov pre ε_{ij} v malom modeli M1. Grafy obsahujú odhady korelácií v rôznych oneskoreniach l a kritické hodnoty (prerušovaná čiara).	29
2.4	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov chybovej zložky - bimodálne rozdelenie, ťažké chvosty a outliery - v modeli s väčším počtom pozorovaní (M2).	30
2.5	Porovnanie scatterplotov reziduí modelu s bimodálnou chybovou zložkou a modelom so splnenými predpokladmi.	31
2.6	Pokračovanie porovnania grafov (zlava, histogram reziduí, reziduá proti času, QQ graf reziduí a QQ graf náhodných efektov) pre ďalšie porušenia predpokladov chybovej zložky - zošikmené rozdelenie, rozptyl rastúci s časom, AR(1) závislosť - v modeli s väčším počtom pozorovaní (M2).	32
2.7	Odhady autokorelačných funkcií pre modely s porušeniami predpokladov pre ε_{ij} vo veľkom modeli M2. Grafy obsahujú odhady korelácií v rôznych oneskoreniach l a kritické hodnoty (prerušovaná čiara).	33
2.8	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliery a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).	35

2.9	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre porušenia predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).	36
2.10	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov popísané v tabuľke 2.6 v modeli s menším počtom pozorovaní (M1).	38
2.11	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov popísané v tabuľke 2.6 v modeli s väčším počtom pozorovaní (M2).	39
2.12	Zlava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku v menšom modeli M1 s porušeniami predpokladov normality \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliers a zošikmené rozdelenie.	43
2.13	Zlava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku vo väčšom modeli M2 s porušeniami predpokladov normality ε_{ij} - bimodálne rozdelenie, rozdelenie s ťažkými chvostami a outliers.	44
2.14	Scatterploty náhodných efektov vo väčšom modeli M2 s porušeniami predpokladov normality ε_{ij} - bimodálne rozdelenie, rozdelenie s ťažkými chvostami a outliers.	45
2.15	Zlava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku vo väčšom modeli M2 s porušeniami predpokladov normality \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliers a zošikmené rozdelenie.	47
2.16	Scatterploty náhodných efektov vo väčšom modeli M2 s porušeniami predpokladov normality \mathbf{b}_i - bimodálne rozdelenie, rozdelenie s nehomogénnym rozptylom, outliers a zošikmené rozdelenie.	48
2.17	Zlava QQ-grafy náhodného počiatku a náhodného sklonu, QQ-graf štvorca Mahalanobisovej vzdialenosti a scatterplot náhodného sklonu proti náhodnému počiatku vo väčšom modeli M2 s porušeniami predpokladov z tabuľky 2.10.	50
2.18	Scatterploty náhodných efektov vo väčšom modeli M2 s porušeniami predpokladov z tabuľky 2.10	51

A.1	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - ťažké chvosty - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).	72
A.2	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - ťažké chvosty - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).	73
A.3	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - outliersy - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).	74
A.4	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - outliersy - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).	75
A.5	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - zošikmené rozdelenie - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).	76
A.6	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - zošikmené rozdelenie - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).	77
A.7	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - rozptyl rastúci s časom - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).	78
A.8	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - rozptyl rastúci s časom - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliersy a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).	79

A.9	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov)pre kombinácie porušení predpokladov chybovej zložky - AR(1) závislosť - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliery a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).	80
A.10	Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - AR(1) závislosť - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliery a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).	81

Zoznam tabuliek

2.1	Tabuľka s počtom subjektov s daným počtom pozorovaní.	24
2.2	Popisné štatistiky datasetu pre model M1 so splnenými predpokladmi.	25
2.3	Popisné štatistiky datasetu pre model M2 so splnenými predpokladmi.	25
2.4	Tabuľka účinností jednotlivých metód na odhaľovanie porušení predpokladov chybovej zložky.	33
2.5	Tabuľka účinností jednotlivých metód na odhaľovanie porušení predpokladov náhodných efektov.	37
2.6	Tabuľka kombinácií porušení.	37
2.7	Tabuľka s počtom subjektov s daným počtom pozorovaní pre menší model M1.	41
2.8	Popisné štatistiky datasetu pre model M1 so splnenými predpokladmi.	41
2.9	Popisné štatistiky datasetu pre model M2 so splnenými predpokladmi.	41
2.10	Tabuľka kombinácií porušení.	48
3.1	Tabuľka štatistík pre modely s porušeniami predpokladov normality pre ε_{ij} . Tabuľka obsahuje priemer odhadov $\widehat{\beta}_0$, ich smerodajnú odchýlku (\overline{sd}), empirickú smerodajnú odchýlku a pravdepodobnosť pokrytia (CP). Skutočná hodnota parametra $\beta_0 = 5$	54
3.2	Tabuľka štatistík pre modely s porušeniami predpokladov normality pre ε_{ij} . Tabuľka obsahuje priemer odhadov $\widehat{\beta}_1$, ich smerodajnú odchýlku (\overline{sd}), empirickú smerodajnú odchýlku a pravdepodobnosť pokrytia (CP). Skutočná hodnota parametra $\beta_1 = 0,5$	54
3.3	Tabuľka štatistík pre modely s porušeniami predpokladov normality pre ε_{ij} . Tabuľka obsahuje priemer odhadov $\widehat{\sigma}_e^2$, pravdepodobnosť pokrytia (CP) pre $\widehat{\sigma}_e^2$ a priemery odhadov \widehat{d}_{11} , \widehat{d}_{22} a \widehat{d}_{22} . Skutočná hodnota parametrov $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$	55
3.4	Priemer a kvantily odhadov $\widehat{\sigma}_e^2$ pre porušenie predpokladu normality pre ε_{ij} odlahlými hodnotami.	56
3.5	Tabuľka štatistík pre modely s porušeniami predpokladov normality pre \mathbf{b}_i . Tabuľka obsahuje priemer odhadov $\widehat{\mathbb{D}}$. Skutočná hodnota parametra $\mathbb{D} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$	56

3.6	Sila testu hypotézy (3.2) pre parameter $\beta_1 = 0,5$ pri jednotlivých porušeníach predpokladov pre ε_{ij} . V druhom stĺpci sa nachádza hladina testu pri nastavení parametru $\beta_1 = 0$	58
3.7	Sila testu hypotézy (3.2) pre parameter $\beta_1 = 0,5$ pri jednotlivých porušeníach predpokladov pre \mathbf{b}_i . V druhom stĺpci sa nachádza hladina testu pri nastavení parametru $\beta_1 = 0$	59
A.1	Tabuľka štatistík odhadov parametrov modelov pre prípady porušenia ε - bez porušení a bimodálne rozdelenie so všetkými porušeniami pre \mathbf{b}_i . Skutočná hodnota parametrov je $\beta_0 = 5$, $\beta_1 = 0,5$, $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$	83
A.2	Tabuľka štatistík odhadov parametrov modelov pre prípady porušenia ε - ťažké chvosty, outliery, a zošikmené rozdelenie so všetkými porušeniami pre \mathbf{b}_i . Skutočná hodnota parametrov je $\beta_0 = 5$, $\beta_1 = 0,5$, $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$	84
A.3	Tabuľka štatistík odhadov parametrov modelov pre prípady porušenia ε - rozptyl rastúci s časom a AR(1) chybová zložka so všetkými porušeniami pre \mathbf{b}_i . Skutočná hodnota parametrov je $\beta_0 = 5$, $\beta_1 = 0,5$, $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$	85
A.4	Sila testov $\mathbf{H}_0 : \beta_1 = 0$ pre všetky kombinácie porušení predpokladov pre ε_{ij} a \mathbf{b}_i bez jednoduchých, nekombinovaných porušení. V riadkoch sú porušenia pre ε_{ij} , v stĺpcoch pre \mathbf{b}_i	86
A.5	Hladina testov $\mathbf{H}_0 : \beta_1 = 0$ pre všetky kombinácie porušení predpokladov pre ε_{ij} a \mathbf{b}_i bez jednoduchých, nekombinovaných porušení. V riadkoch sú porušenia pre ε_{ij} , v stĺpcoch pre \mathbf{b}_i	86

Dodatok A

Appendix

A.1 Odvodenie aposteriórneho rozdelenia náhodných efektov

Uvažujme jednostupňový zmiešaný lineárny model

$$\begin{aligned}\mathbf{Y}|\mathbf{b} &\sim \mathcal{N}(\mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\mathbf{b}, \sigma_e^2\mathbb{I}_N) \\ \mathbf{b} &\sim \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbb{D}_*).\end{aligned}$$

Potom, ak si označíme $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbb{X}\boldsymbol{\beta}$, máme

$$\begin{aligned}\tilde{\mathbf{Y}}|\mathbf{b} &\sim \mathcal{N}(\mathbb{Z}\mathbf{b}, \sigma_e^2\mathbb{I}_N) \\ \mathbf{b} &\sim \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbb{D}_*).\end{aligned}$$

Pre aposteriórne rozdelenie náhodných efektov potom platí (viď Lindley a Smith (1972) strana 4, Lemma)

$$\mathbf{b}|\tilde{\mathbf{Y}} \sim \mathcal{N}(\mathbf{C}\mathbf{c}, \mathbf{C}),$$

kde

$$\begin{aligned}\mathbf{C}^{-1} &= \frac{1}{\sigma_e^2}\mathbb{Z}^\top\mathbb{Z} + \frac{1}{\sigma_e^2}\mathbb{D}_*^{-1}, \\ \mathbf{c} &= \frac{1}{\sigma_e^2}\mathbb{Z}^\top\tilde{\mathbf{Y}}.\end{aligned}$$

Stredná hodnota aposteriórneho rozdelenia je rovná

$$\mathbf{C}\mathbf{c} = \sigma_e^2(\mathbb{Z}^\top\mathbb{Z} + \mathbb{D}_*^{-1})^{-1}\frac{1}{\sigma_e^2}\mathbb{Z}^\top\tilde{\mathbf{Y}}. = (\mathbb{Z}^\top\mathbb{Z} + \mathbb{D}_*^{-1})^{-1}\mathbb{Z}^\top\tilde{\mathbf{Y}}$$

Potom pomocou maticových vzťahov (1) a (4) z Smith (1973) (strana 67) dostávame

$$\begin{aligned}(\mathbb{Z}^\top\mathbb{Z} + \mathbb{D}_*^{-1})^{-1}\mathbb{Z}^\top &= (\mathbb{D}_* - \mathbb{D}_*\mathbb{Z}^\top(\mathbb{I}_N + \mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top)^{-1}\mathbb{Z}\mathbb{D}_*)\mathbb{Z}^\top \\ &= \mathbb{D}_*\mathbb{Z}^\top - \mathbb{D}_*\mathbb{Z}^\top(\mathbb{I}_N + \mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top)^{-1}\mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top \\ &= \mathbb{D}_*\mathbb{Z}^\top - \mathbb{D}_*\mathbb{Z}^\top(\mathbb{I}_N - (\mathbb{I}_N + \mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top)^{-1}) \\ &= \mathbb{D}_*\mathbb{Z}^\top - \mathbb{D}_*\mathbb{Z}^\top + \mathbb{D}_*\mathbb{Z}^\top(\mathbb{I}_N + \mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top)^{-1}.\end{aligned}$$

Stredná hodnota aposteriórneho rozdelenia je potom

$$\mathbb{E}[\mathbf{b}|\mathbf{Y}] = \mathbb{D}_* \mathbf{Z}^\top (\mathbb{I}_N + \mathbf{Z} \mathbb{D}_* \mathbf{Z}^\top)^{-1} (\mathbf{Y} - \mathbb{X} \boldsymbol{\beta}).$$

A.2 Vzorce na redukciu dimenzie problému

Uvažujme maticu $\Sigma_i = \mathbb{I}_{n_i} + \mathbf{Z}_i \mathbb{D} \mathbf{Z}_i^\top$. Potom vďaka (3) z Smith (1973) (strana 67) dostávame pre inverznú maticu pre Σ_i vzťah

$$\Sigma_i^{-1} = (\mathbb{I}_{n_i} + \mathbf{Z}_i \mathbb{D} \mathbf{Z}_i^\top)^{-1} = \mathbb{I}_{n_i} - \mathbf{Z}_i (\mathbb{D}^{-1} + \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathbf{Z}_i^\top.$$

Tiež, pre determinant platí

$$|\mathbb{I}_{n_i} + \mathbf{Z}_i \mathbb{D} \mathbf{Z}_i^\top| = |\mathbb{I}_q + \mathbb{D} \mathbf{Z}_i^\top \mathbf{Z}_i|.$$

A.3 Vzťahy pre maticové derivácie

$$\frac{\partial \text{tr}(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{I} \quad \frac{\partial \text{tr}(\mathbf{A} \mathbf{A}^\top)}{\partial \mathbf{A}} = 2\mathbf{A} \quad \frac{\partial \text{tr}(\mathbf{C} \mathbf{A} \mathbf{B})}{\partial \mathbf{A}} = \mathbf{C}^\top \mathbf{B}^\top$$

Prehľad ďalších vzťahov nájdeme v Wand (2002), resp. Demidenko (2013) strana 670.

A.4 Pearsonove rozdelenie typu IV

Personove rozdelenie typu IV je rozdelenie s hustotou

$$f(x) = \left| \frac{\Gamma(m + \frac{\nu}{2i})}{\Gamma(m)} \right|^2 \frac{1}{a \cdot B(m - \frac{1}{2}, \frac{1}{2})} \left(1 + \left(\frac{x - \lambda}{a} \right)^2 \right)^{-m} \exp \left\{ -\nu \cdot \arctan \frac{x - \lambda}{a} \right\},$$

kde $\Gamma(\cdot)$ značí (komplexnú) *Gamma* funkciu a $B(\cdot, \cdot)$ značí *Beta* funkciu. Hustota je definovaná pre hodnoty parametrov $a > 0$, $m > 1/2$, $\nu \neq 0$ a $\lambda \in \mathbb{R}^1$. Definíciu nájdeme napríklad v Nagahara (1999) Theorem 2.1 s voľbou parametrov $m = b$, $\nu = -2b\delta$, $a = \tau$ a $\lambda = \mu$. Pre centrálné momenty rozdelenia platí

$$\mu_2 = \frac{a^2}{r^2(r-1)}(r^2 + \nu^2),$$

kde $r = 2(m-1)$ a

$$\mu_3 = -\frac{4a^2\nu(r^2 + \nu^2)}{r^3(r-1)(r-2)}$$

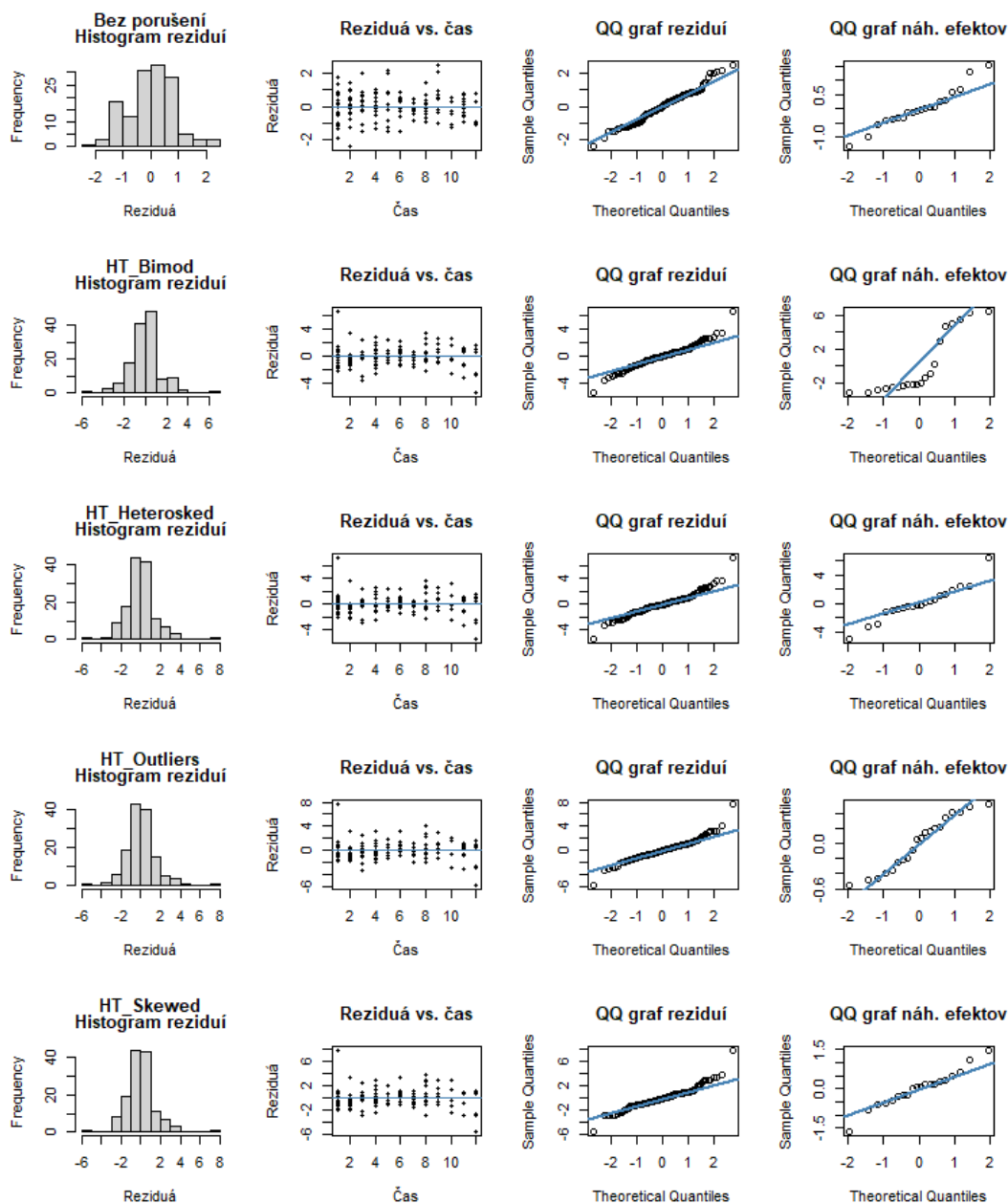
ako môžeme vidieť v Nagahara (1999) Theorem 2. Šikmost rozdelenia je potom rovná

$$\gamma_3 = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{-\frac{4a^2\nu(r^2 + \nu^2)}{r^3(r-1)(r-2)}}{\left(\frac{a^2}{r^2(r-1)}(r^2 + \nu^2)\right)^{3/2}} = \frac{4\nu}{r-2} \sqrt{\frac{r-1}{r^2 + \nu^2}}.$$

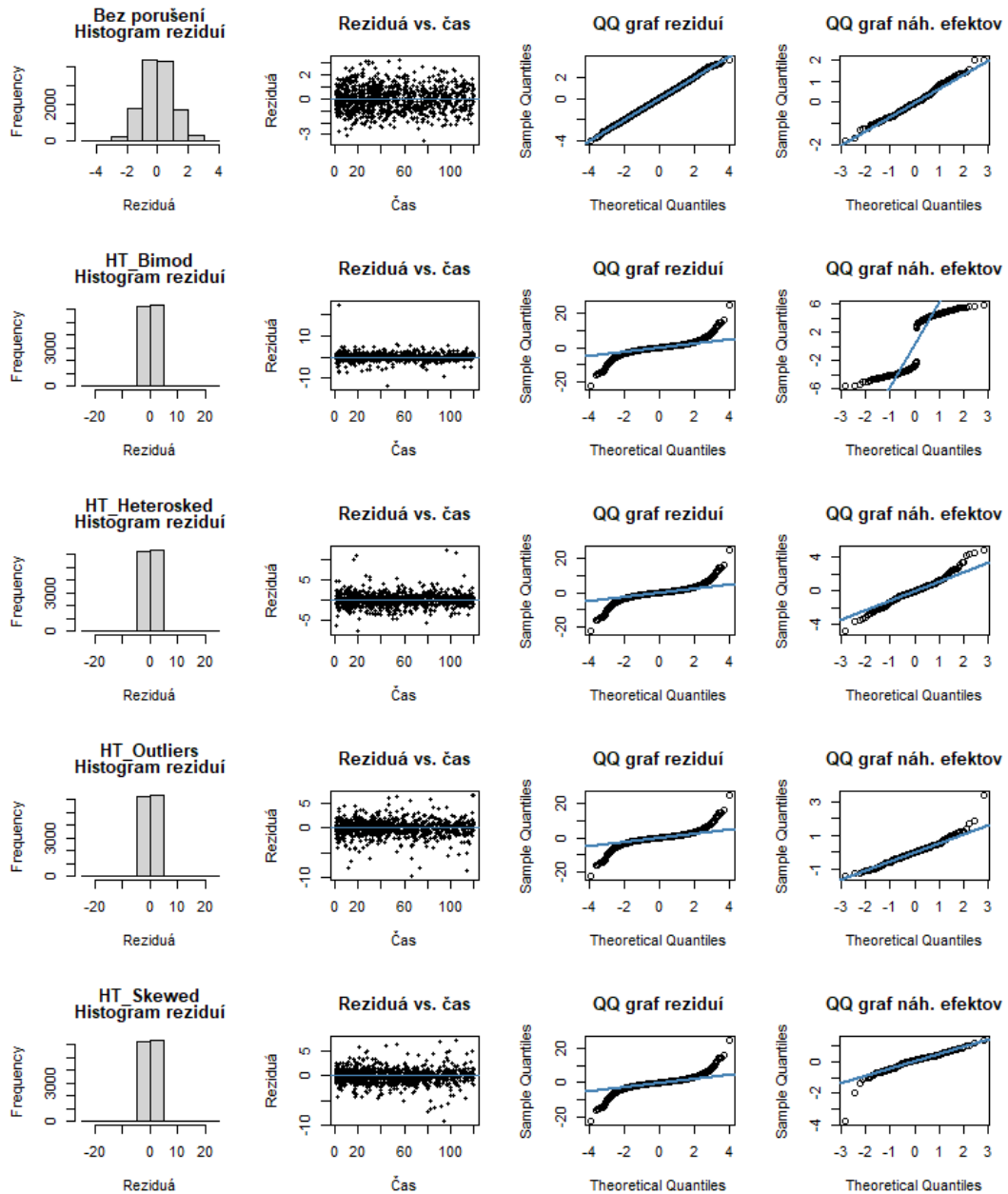
¹ ν nemusí byť nutne nenulové, ale šikmost rozdelenia s $\nu = 0$ je rovná 0)

A.5 Grafy modelov s náhodným počiatkom s kombináciou porušení predpokladov

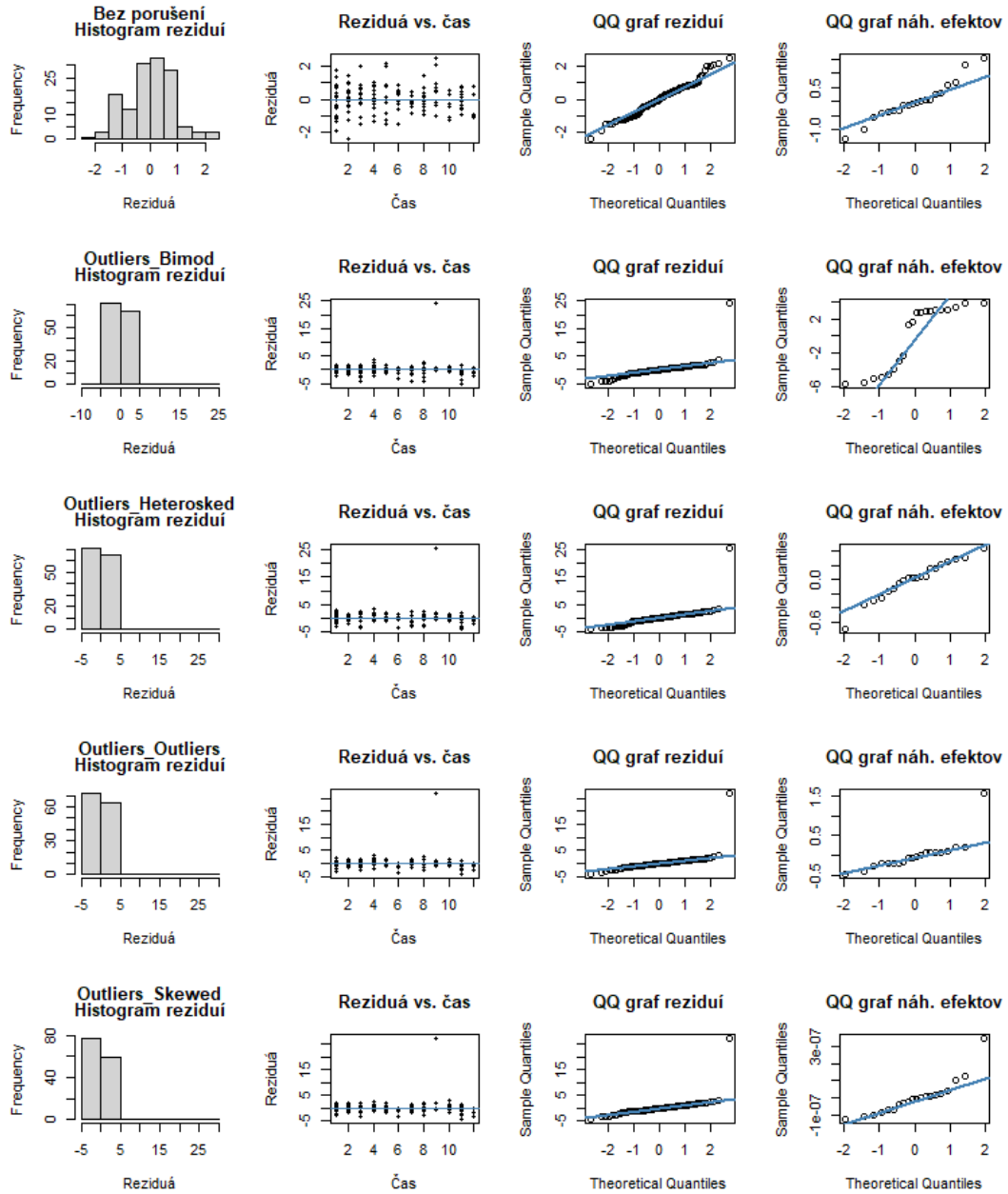
Jednotlivé kombinácie sú identifikované názvami predpokladov nad jednotlivými histogramami.



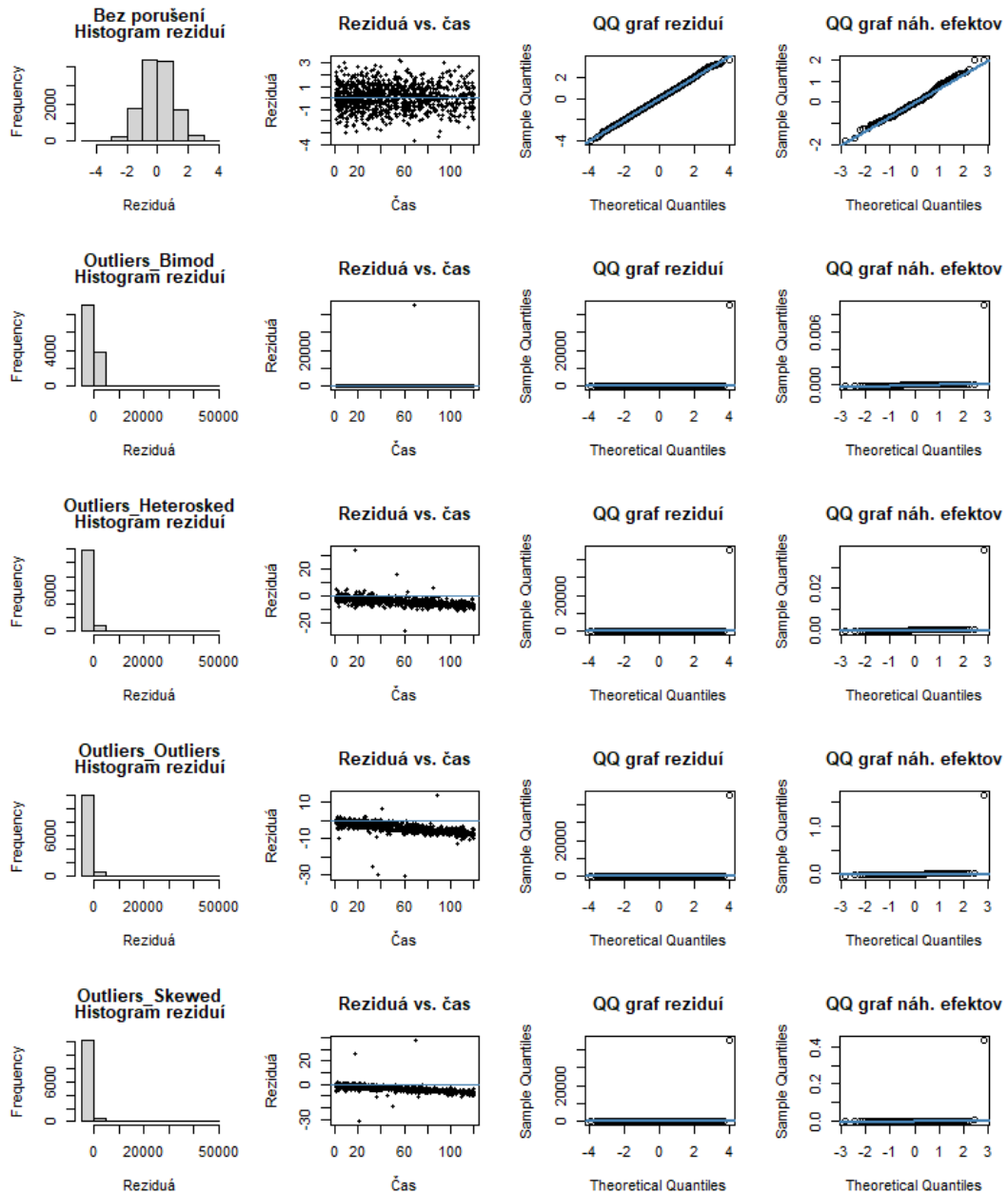
Obr. A.1: Porovnanie grafov (zľava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - ťažké chvosty - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).



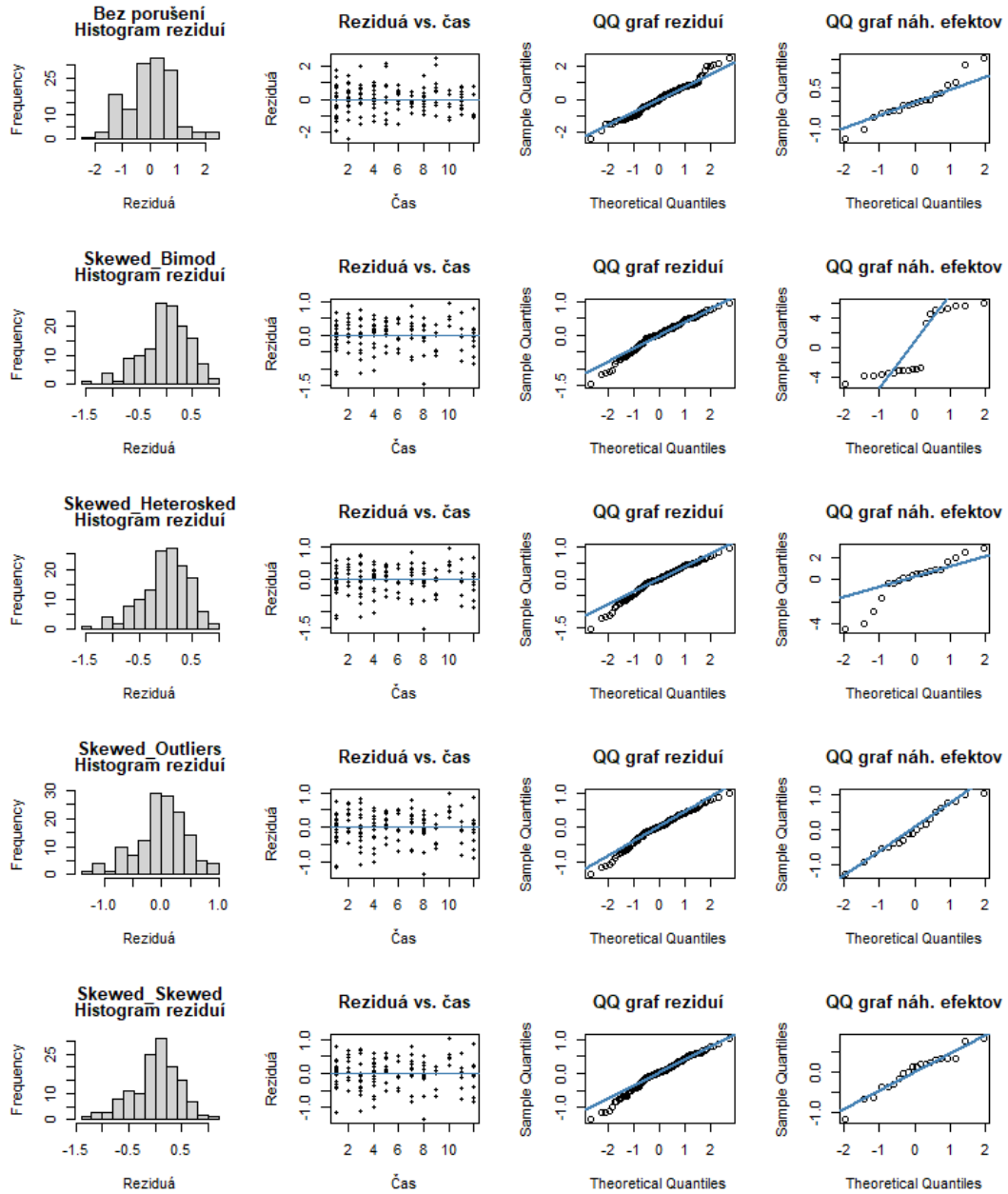
Obr. A.2: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - ťažké chvosty - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).



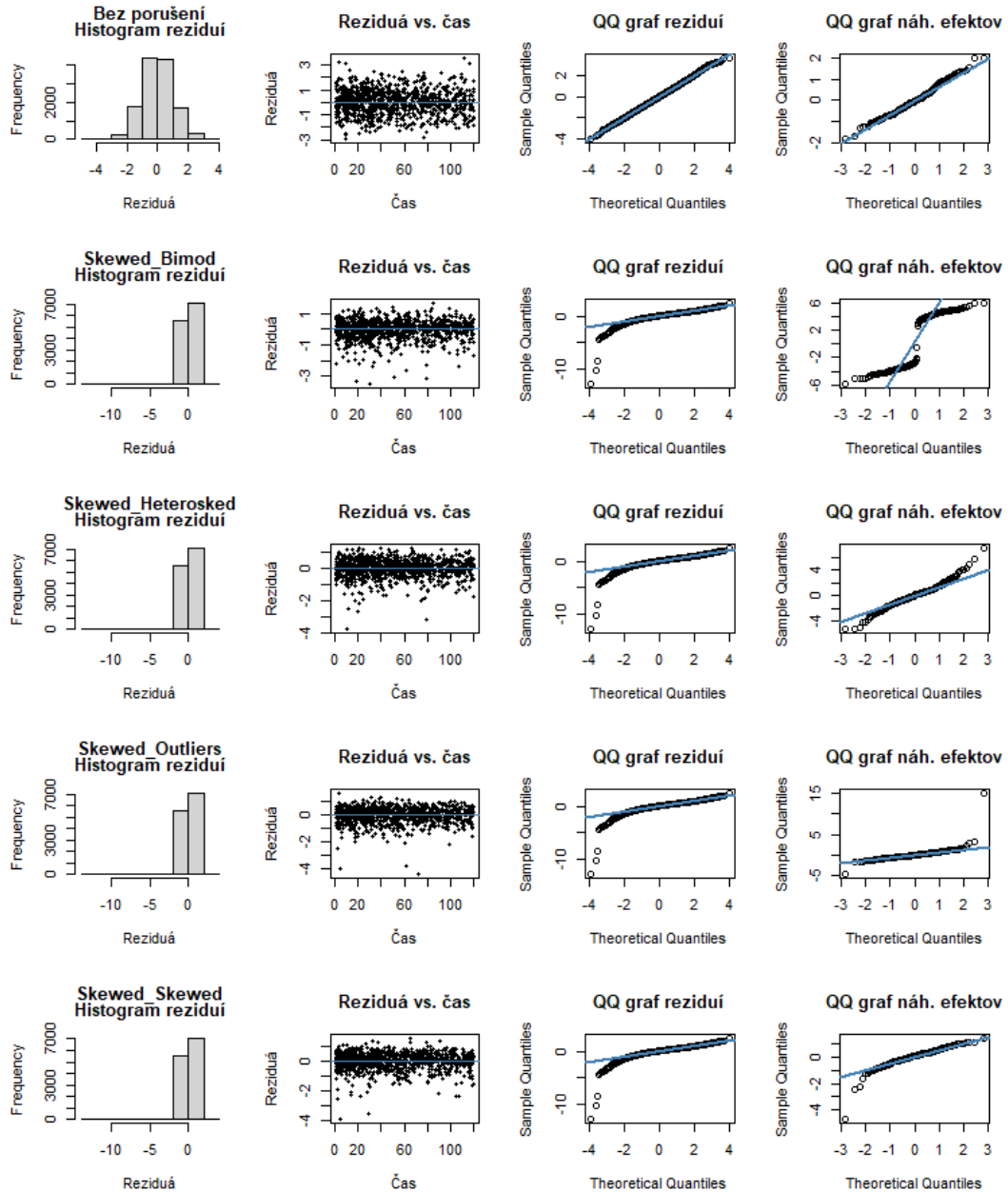
Obr. A.3: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - outliers - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).



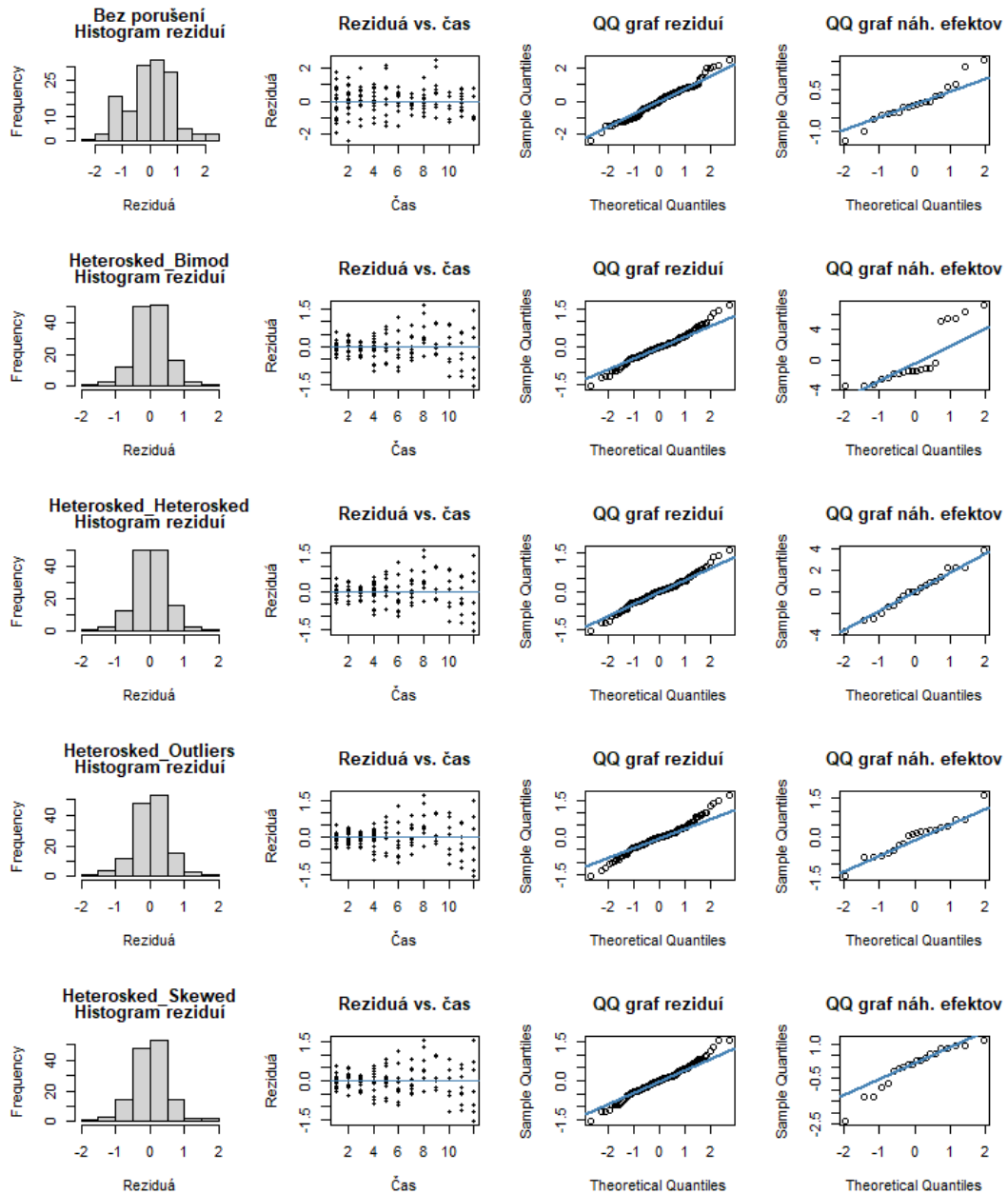
Obr. A.4: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - outliers - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).



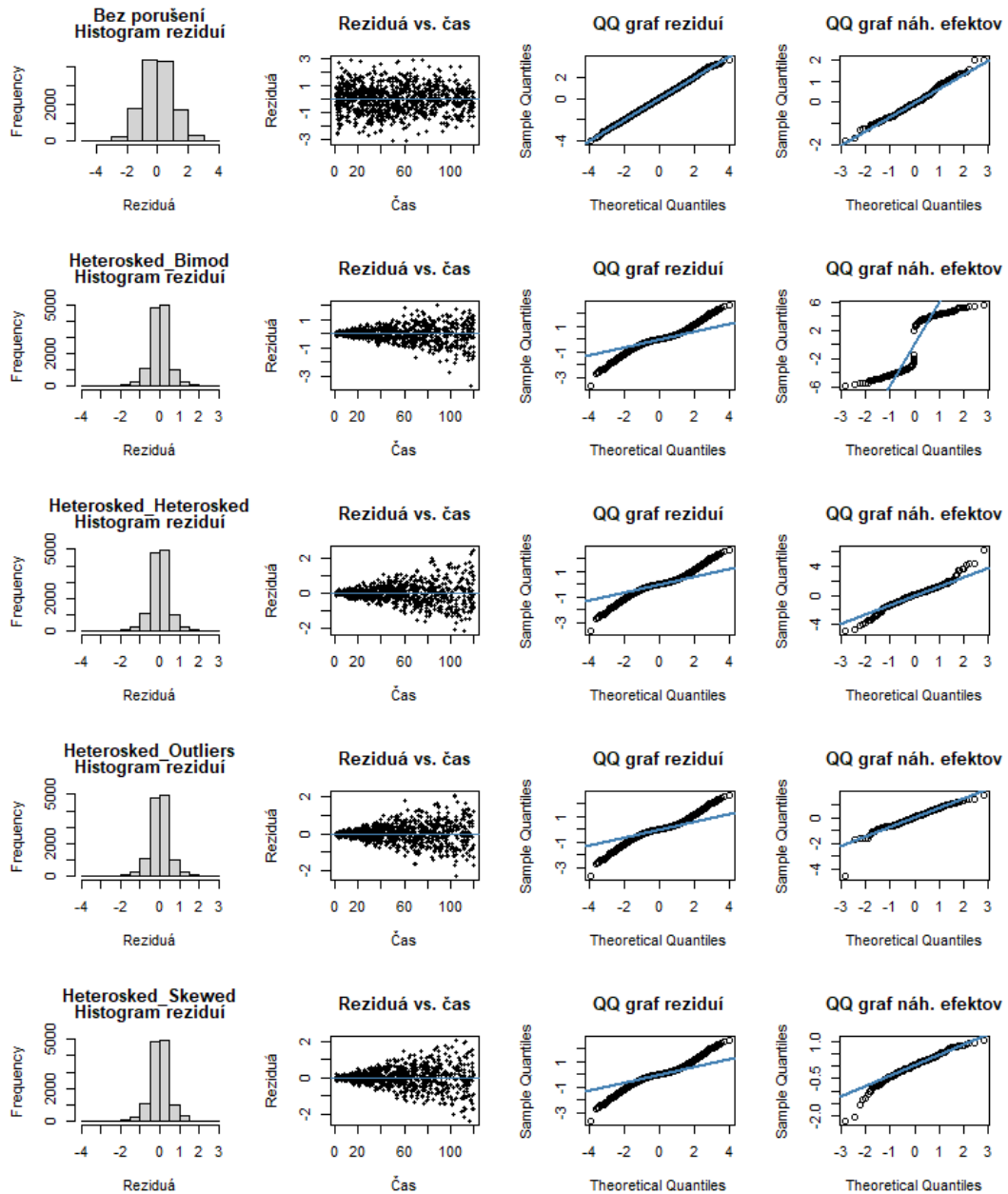
Obr. A.5: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - zošikmené rozdelenie - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).



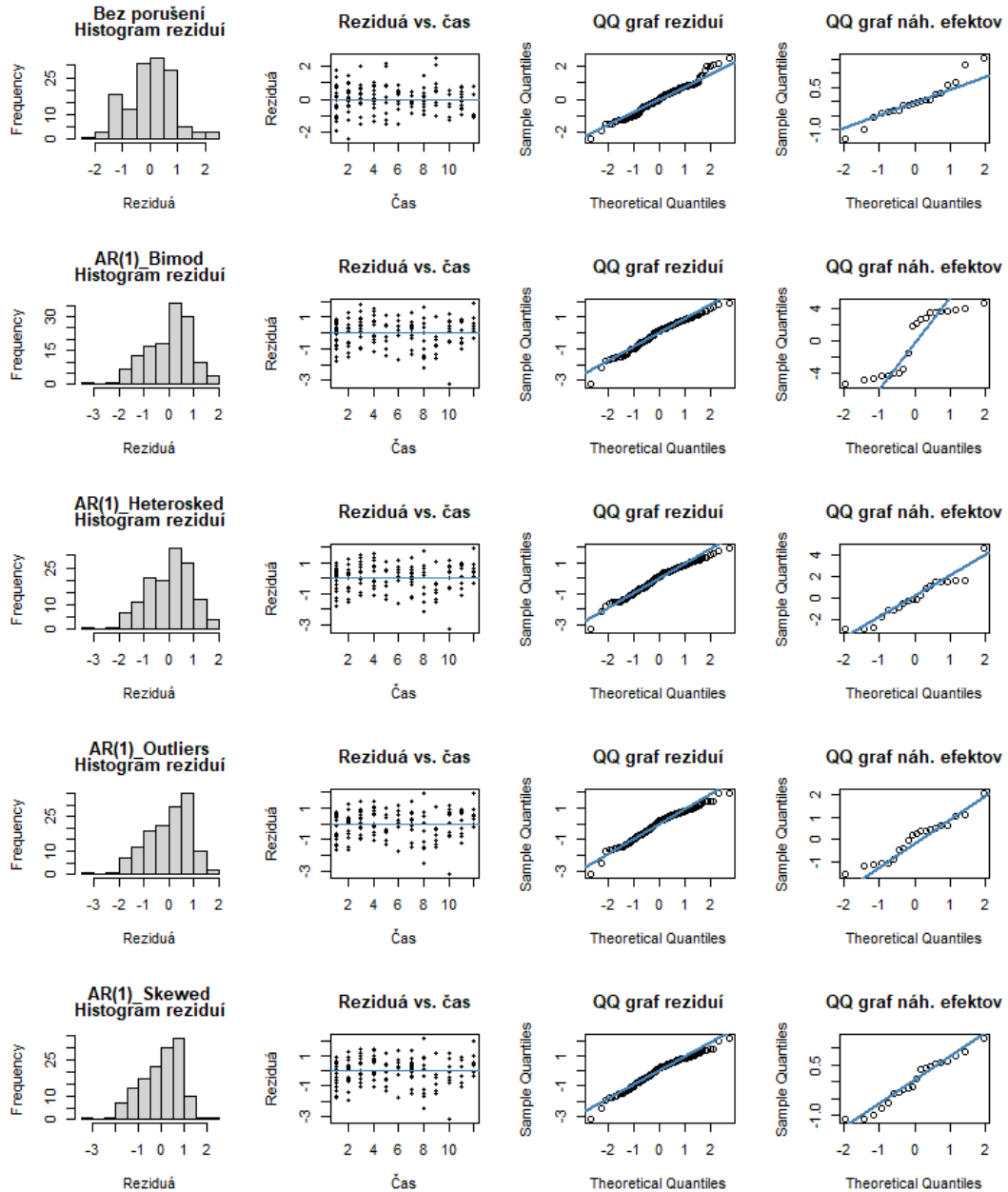
Obr. A.6: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - zošikmené rozdelenie - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).



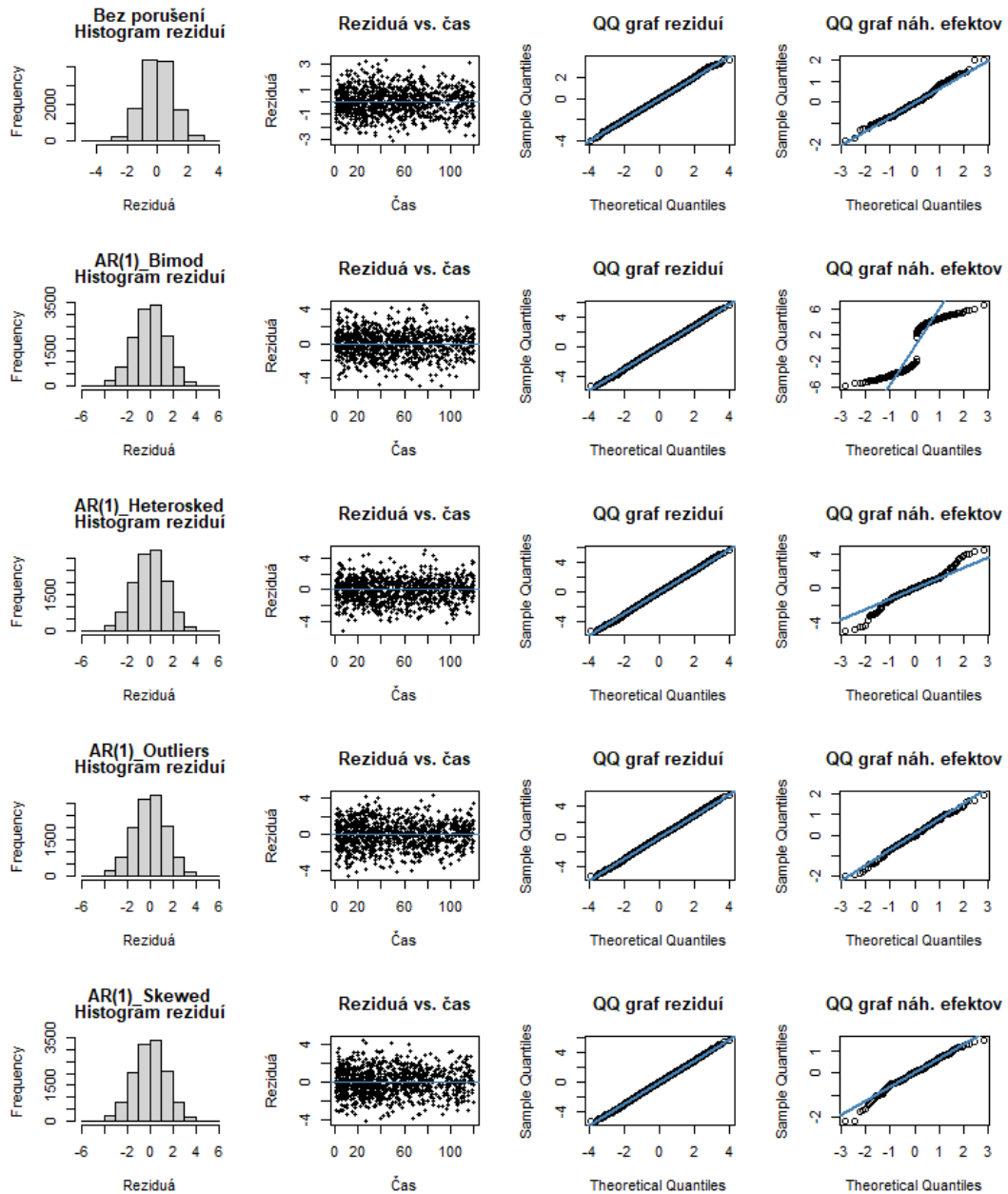
Obr. A.7: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - rozptyl rastúci s časom - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).



Obr. A.8: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - rozptyl rastúci s časom - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).



Obr. A.9: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov)pre kombinácie porušeni predpokladov chybovej zložky - AR(1) závislost - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošikmené rozdelenie - v modeli s menším počtom pozorovaní (M1).



Obr. A.10: Porovnanie grafov (zlava, histogram reziduí, scatterplot reziduí proti času, QQ graf reziduí a QQ graf náhodných efektov) pre kombinácie porušení predpokladov chybovej zložky - AR(1) závislosť - a predpokladov náhodných efektov - bimodálne rozdelenie, nehomogénny rozptyl, outliers a zošíkmené rozdelenie - v modeli s väčším počtom pozorovaní (M2).

A.6 Mnohorozmerné Studentovo rozdelenie

Mnohorozmerné Studentovo rozdelenie $MVT(\boldsymbol{\mu}, \nu, \Sigma)$ je rozdelenie náhodného vektoru \mathbf{X} dĺžky p s hustotou

$$f(\mathbf{X}) = \frac{\Gamma((\nu + p)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)|\Sigma|^{1/2}} \left(1 + \frac{1}{\nu}(\mathbf{X} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})\right)^{-(\nu+p)/2},$$

kde $\boldsymbol{\mu}$ je centralizačný parameter, Σ je škálová matica a ν sú stupne voľnosti. Vlastnosti nájdeme v Kotz a Nadarajah (2004).

A.7 Výsledky simulácii odhadov parametrov

Model	$\widehat{\beta}_0$	$sd(\widehat{\beta}_0)$	CP	$\widehat{\beta}_1$	$sd(\widehat{\beta}_1)$	CP	$\widehat{\sigma}_e^2$	CP	\widehat{d}_{11}	\widehat{d}_{22}	\widehat{d}_{12}
Bez porušení - Bez porušení	M1	4,97	0,35	0,93	0,48	0,22	0,93	0,94	1,99	0,99	0,99
	M2	5,00	0,10	0,94	0,50	0,07	0,94	0,95	2,00	1,00	0,99
Bez porušení - Bimodálne	M1	4,98	0,96	0,93	0,47	0,92	0,94	0,95	17,99	16,94	16,96
	M2	4,99	0,30	0,95	0,49	0,29	0,95	0,95	18,01	16,97	16,98
Bez porušení - Nehomogénny rozptyl	M1	4,99	0,74	0,94	0,50	0,51	0,94	0,94	10,89	5,43	5,42
	M2	5,00	0,23	0,95	0,50	0,17	0,95	0,94	10,99	5,51	5,51
Bez porušení - Outliery	M1	5,07	0,77	0,95	0,55	0,47	0,94	0,95	243,57	67,38	-55,26
	M2	5,00	0,30	0,96	0,51	0,17	0,96	0,96	182,68	23,93	25,52
Bez porušení - Zošíkmené	M1	5,00	0,21	0,94	-0,29	0,13	0,00	0,99	0,36	0,34	-0,01
	M2	5,00	0,05	0,96	-0,29	0,04	0,00	1,00	0,34	0,34	0,00
Bimodálne - Bimodálne	M1	5,02	1,01	0,93	0,52	0,92	0,93	0,00	18,02	16,92	16,86
	M2	4,99	0,30	0,94	0,49	0,29	0,94	0,00	18,04	16,98	17,01
Bimodálne - Heterogénny rozptyl	M1	4,99	0,82	0,94	0,51	0,52	0,94	0,00	11,10	5,51	5,59
	M2	4,99	0,24	0,95	0,50	0,17	0,95	0,00	10,93	5,50	5,49
Bimodálne - Outliery	M1	5,16	0,89	0,95	0,64	0,61	0,95	0,00	278,30	531,38	279,05
	M2	4,97	0,29	0,96	0,49	0,17	0,96	0,00	224,01	42,03	77,96
Bimodálne - Zošíkmené	M1	5,00	0,40	0,96	-0,30	0,14	0,00	0,00	0,73	0,34	-0,05
	M2	5,00	0,06	0,95	-0,29	0,04	0,00	0,00	0,34	0,34	0,00

Tabulka A.1: Tabulka štatistík odhadov parametrov modelov pre prípady porušení ε - bez porušení a bimodálne rozdelenie so všetkými porušeniami pre \mathbf{b}_i . Skutočná hodnota parametrov je $\beta_0 = 5$, $\beta_1 = 0,5$, $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$.

Model	$\widehat{\beta}_0$	$\overline{\text{sd}(\widehat{\beta}_0)}$	CP	$\widehat{\beta}_1$	$\overline{\text{sd}(\widehat{\beta}_1)}$	CP	$\widehat{\sigma}_e^2$	CP	\widehat{d}_{11}	\widehat{d}_{22}	\widehat{d}_{12}
Ťažké chvosty - Bimodálne	M1	5,00	0,99	0,94	0,92	0,95	1,68	0,00	18,67	16,98	16,89
	M2	5,01	0,30	0,95	0,22	0,96	1,73	0,00	18,01	16,99	16,99
Ťažké chvosty - Heterogénny rozptyl	M1	4,96	0,78	0,94	0,51	0,94	1,65	0,00	11,25	5,50	5,50
	M2	5,01	0,24	0,95	0,17	0,95	1,73	0,00	11,01	5,50	5,50
Ťažké chvosty - Outliery	M1	4,88	0,85	0,96	0,46	0,96	1,67	0,00	361,51	22,69	-1,22
	M2	5,03	0,27	0,95	0,17	0,95	1,72	0,00	131,87	22,14	28,56
Ťažké chvosty - Zošíkmené	M1	5,01	0,32	0,94	0,14	0,00	1,65	0,01	0,63	0,36	-0,05
	M2	5,00	0,05	0,96	0,04	0,00	1,72	0,00	0,34	0,34	0,00
Outliery - Bimodálne	M1	2,38	5,21	0,96	1,48	0,94	14,44	0,00	76010,08	1787,11	-11566,83
	M2	4,85	0,71	0,96	0,48	0,96	25,96	0,00	389,23	17,07	12,88
Outliery - Heterogénny rozptyl	M1	5,62	4,12	0,96	0,99	0,93	17,99	0,00	4457,20	115,21	-626,84
	M2	5,69	1,65	0,96	0,19	0,95	81,57	0,00	31592,23	71,36	-1351,78
Outliery - Outliery	M1	4,93	3,25	0,98	0,82	0,96	14,11	0,00	724,39	56,75	-132,25
	M2	5,04	0,88	0,97	0,18	0,96	24,43	0,00	6124,58	20,08	-40,11
Outliery - Zošíkmené	M1	4,80	4,74	0,98	0,82	0,23	22,54	0,00	9429,50	126,91	-967,56
	M2	4,99	0,84	0,97	0,05	0,01	34,47	0,00	3531,86	1,29	-54,47
Zošíkmené - Bimodálne	M1	4,95	0,95	0,94	0,46	0,94	0,58	0,00	18,05	17,02	17,03
	M2	5,01	0,30	0,95	0,29	0,95	0,58	0,00	17,99	16,98	16,98
Zošíkmené - Heterogénny rozptyl	M1	4,99	0,73	0,94	0,50	0,95	0,58	0,00	11,11	5,46	5,53
	M2	5,00	0,23	0,95	0,16	0,95	0,58	0,00	10,94	5,47	5,45
Zošíkmené - Outliery	M1	4,99	0,60	0,96	0,41	0,94	0,58	0,00	29,19	15,49	16,01
	M2	5,02	0,31	0,97	0,19	0,96	0,58	0,00	240,55	43,03	57,72
Zošíkmené - Zošíkmené	M1	5,01	0,16	0,94	0,13	0,00	0,58	0,00	0,33	0,34	0,00
	M2	5,00	0,04	0,93	0,04	0,00	0,58	0,00	0,34	0,34	0,00

Tabulka A.2: Tabulka štatistík odhadov parametrov modelov pre prípady porušení ε - ťažké chvosty, outliery, a zošíkmené rozdelenie so všetkými porušeniami pre \mathbf{b}_i . Skutočná hodnota parametrov je $\beta_0 = 5$, $\beta_1 = 0,5$, $\sigma_e^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$.

Model	$\widehat{\beta}_0$	$\widehat{\text{sd}}(\widehat{\beta}_0)$	CP	$\widehat{\beta}_1$	$\widehat{\text{sd}}(\widehat{\beta}_1)$	CP	$\widehat{\sigma}_\varepsilon^2$	CP	\widehat{d}_{11}	\widehat{d}_{22}	\widehat{d}_{12}
Rozptyl rastúci s časom pozorovania - Bimodálne	M1 4,98	0,95	0,95	0,48	0,92	0,94	0,52	0,00	17,96	17,09	17,08
	M2 4,99	0,30	0,95	0,49	0,29	0,95	0,50	0,00	17,99	17,01	17,00
Rozptyl rastúci s časom pozorovania - Heterogénny rozptyl	M1 4,99	0,71	0,94	0,50	0,50	0,94	0,53	0,00	10,54	5,31	5,32
	M2 5,02	0,23	0,95	0,51	0,17	0,96	0,50	0,00	10,99	5,51	5,51
Rozptyl rastúci s časom pozorovania - Outliery	M1 5,11	0,65	0,95	0,51	0,41	0,95	0,53	0,00	99,39	11,37	9,43
	M2 4,98	0,23	0,96	0,50	0,15	0,96	0,50	0,00	9,37	19,12	10,47
Rozptyl rastúci s časom pozorovania - Zošikmené	M1 4,99	0,14	0,94	-0,29	0,13	0,00	0,52	0,00	0,26	0,34	0,01
	M2 5,00	0,04	0,95	-0,29	0,04	0,00	0,50	0,00	0,33	0,34	0,00
AR(1) - Bimodálne	M1 4,96	0,99	0,92	0,47	0,92	0,93	0,92	0,68	19,12	17,05	16,89
	M2 4,99	0,31	0,94	0,49	0,29	0,93	1,40	0,00	18,52	17,02	17,00
AR(1) - Heterogénny rozptyl	M1 5,03	0,77	0,95	0,51	0,51	0,93	0,92	0,71	11,97	5,54	5,35
	M2 5,00	0,24	0,95	0,49	0,17	0,95	1,40	0,00	11,55	5,56	5,55
AR(1) - Outliery	M1 4,93	0,74	0,95	0,54	0,47	0,95	0,93	0,71	112,46	35,08	-9,27
	M2 4,96	0,30	0,97	0,48	0,18	0,95	1,40	0,00	323,90	89,86	130,14
AR(1) - Zošikmené	M1 4,99	0,31	0,92	-0,29	0,14	0,00	0,92	0,72	1,50	0,36	-0,10
	M2 5,00	0,07	0,95	-0,29	0,04	0,00	1,41	0,00	0,86	0,34	-0,01

Tabuľka A.3: Tabuľka štatistík odhadov parametrov modelov pre prípady porušení ε - rozptyl rastúci s časom a AR(1) chybová zložka so všetkými porušeniami pre \mathbf{b}_i . Skutočná hodnota parametrov je $\beta_0 = 5$, $\beta_1 = 0,5$, $\sigma_\varepsilon^2 = 1$, $d_{11} = 2$ a $d_{12} = d_{22} = 1$.

Sila testu, $\beta_1 = 0,5$	Model	Bimodálne	Heterogénny rozptyl	Outliery	Zošikmené
Bimodálne	M1	0,10	0,19	0,47	0,53
	M2	0,40	0,85	0,88	1,00
Ťažké chvosty	M1	0,09	0,20	0,45	0,58
	M2	0,42	0,86	0,89	1,00
Outliery	M1	0,09	0,15	0,30	0,27
	M2	0,35	0,84	0,85	0,99
Zošikmené	M1	0,09	0,19	0,48	0,62
	M2	0,41	0,85	0,88	1,00
Rozptyl rastúci s časom pozorovania	M1	0,09	0,20	0,49	0,64
	M2	0,38	0,86	0,89	1,00
AR(1)	M1	0,09	0,22	0,45	0,58
	M2	0,39	0,84	0,88	1,00

Tabuľka A.4: Sila testov $\mathbf{H}_0 : \beta_1 = 0$ pre všetky kombinácie porušení predpokladov pre ε_{ij} a \mathbf{b}_i bez jednoduchých, nekombinovaných porušení. V riadkoch sú porušenia pre ε_{ij} , v stĺpcoch pre \mathbf{b}_i .

Hladina testu, $\beta_1 = 0$	Model	Bimodálne	Heterogénny rozptyl	Outliery	Zošikmené
Bimodálne	M1	0,07	0,06	0,05	1,00
	M2	0,06	0,05	0,03	1,00
Ťažké chvosty	M1	0,05	0,06	0,04	1,00
	M2	0,04	0,05	0,05	1,00
Outliery	M1	0,06	0,04	0,04	0,77
	M2	0,06	0,06	0,04	1,00
Zošikmené	M1	0,06	0,05	0,06	1,00
	M2	0,05	0,05	0,05	1,00
Rozptyl rastúci s časom pozorovania	M1	0,06	0,06	0,05	1,00
	M2	0,05	0,04	0,04	1,00
AR(1)	M1	0,07	0,07	0,05	1,00
	M2	0,07	0,05	0,04	1,00

Tabuľka A.5: Hladina testov $\mathbf{H}_0 : \beta_1 = 0$ pre všetky kombinácie porušení predpokladov pre ε_{ij} a \mathbf{b}_i bez jednoduchých, nekombinovaných porušení. V riadkoch sú porušenia pre ε_{ij} , v stĺpcoch pre \mathbf{b}_i .