

UNIVERZITA PALACKÉHO V OLMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

## DIPLOMOVÁ PRÁCE

Ekonometrické modely



Vedoucí diplomové práce:  
**Mgr. Jaroslav Marek, Ph.D.**  
Rok odevzdání: 2011

Vypracovala:  
**Bc. Hana Marková**  
AME, II. ročník

## **Prohlášení**

Prohlašuji, že jsem tuto diplomovou práci vytvořila samostatně za vedení Mgr. Jaroslava Marka, Ph.D. a že jsem v seznamu použité literatury uvedla všechny zdroje použité při zpracování práce.

V Olomouci dne 29. března 2011

Hana Marková

## **Poděkování**

Ráda bych na tomto místě upřímně poděkovala vedoucímu mé diplomové práce panu Mgr. Jaroslavu Markovi, Ph.D. za obětavou spolupráci i za čas, který mi věnoval při konzultacích. Dále si zaslouží poděkování pan RNDr. Miloslav Závodný za to, že mě naučil pracovat s typografickým systémem  $\text{\TeX}$  a poskytl k němu pomocné soubory. Také bych ráda poděkovala své rodině a partnerovi za jejich podporu a pomoc v době mého studia.

# Obsah

Úvod	4
1 Motivace	5
2 Předmět ekonometrické analýzy	7
2.1 Konstrukce ekonometrického modelu	7
2.2 Typy dat	8
2.3 Oblasti využití ekonometrických modelů	9
3 Klasický model lineární regrese	10
3.1 Motivace	10
3.2 Metoda nejmenších čtverců (MNČ)	11
3.3 Vlastnosti odhadu metodou nejmenších čtverců	12
3.4 Normální model	14
3.5 Statistická indukce v klasickém lineárním modelu	15
3.5.1 Testování významnosti odhadnutých parametrů	15
3.5.2 Intervaly spolehlivosti odhadnutých parametrů	16
3.5.3 Kritéria shody odhadnutého modelu s daty	17
4 Ekonometrická zobecnění lineární regrese	20
4.1 Zobecněný model lineární regrese, ZMNČ	21
4.2 Heteroskedasticita	22
4.3 Autokorelovanost reziduí	26
4.4 Multikolinearita	29
4.5 Kritéria pro výběr modelu	34
5 Model simultánních rovnic	36
5.1 Maticové vyjádření MSR	37
5.2 Identifikace strukturního tvaru MSR	39
5.3 Rekurzivní systémy	41
5.4 Dvoustupňová MNČ	42
6 Ekonometrické prognózování	45
7 Příklad faktorového modelu (modelování vývoje cen akcií)	48
8 Příklad simultánních rovnic (modelování prodejnosti bytů v developerských projektech)	65
Závěr	72
Literatura	74

# Úvod

Ekonometrie je kvantitativní ekonomická disciplína, která se zabývá měřením ekonomických veličin na základě analýzy reálných dat pomocí ekonometrických modelů a metod. Předpokladem úspěšného zvládnutí této problematiky je osvojení znalostí z lineární algebry, pravděpodobnosti, statistiky a z ekonomické teorie.

Hlavním cílem mé práce je prostudovat dostupnou ekonometrickou literaturu a představit postupy, které lze v rámci současné ekonometrie použít a na příkladech demonstrovat užití vybraných úloh z finanční ekonometrie.

Celá práce je rozdělena do osmi kapitol. V první části uvádím motivační příklad, zabývající se poptávkovou funkcí po alkoholických nápojích v ČSSR. Účelem této motivace je získání představy o aplikačních možnostech v rámci reálné ekonometrické analýzy. Dále se seznámíme s konstrukcí ekonometrického modelu a oblastmi jeho využití. Postupně se budu zabývat klasickým a zobecněným modelem lineární regrese a poté modelem simultánních rovnic. Ve všech třech případech se zaměřím na metody odhadu neznámých parametrů a na postupy při porušení podmínek jejich aplikace. Ukážeme si také jak se postupuje při ekonometrickém prognózování v těchto modelech. Poslední dvě kapitoly jsou věnovány použití všech poznatků z ekonometrické teorie v podobě příkladů zaměřených na praktické aktuální problémy z oblasti ekonomie a finančnictví. Nejprve se budu zabývat faktorovými modely pro predikci vývoje cen akcií na burze, poté poptávkou a nabídkou po nově stavěných bytech (developerských projektech), modelovanou současně tzv. simultánními rovnicemi.

# 1 Motivace

## Poptávková funkce a degenerovaná nabídka - výlet do minulosti

Modelování poptávky a nabídky je pro ekonomy velmi zajímavé. Např. v článku Jana Mládka [62] s názvem *Ekonometrická analýza alkoholických nápojů v ČSSR* autor hledá vhodnou poptávkovou funkci.

V článku uvádí *Snížení poptávky obyvatelstva po alkoholických nápojích bylo nepochybně hlavním cílem značných změn v maloobchodních cenách alkoholických nápojů v minulých letech. Zejména výrazná byla úprava cen piva 15. října 1984. Zintenzívnění boje za snížení spotřeby alkoholických nápojů v osmdesátých letech nevylučuje, že frekvence cenových změn za snížení naší stále vysoké spotřeby alkoholických nápojů se v budoucnosti zvýší. Není tedy pochyb, že otázka kvantitativního vyjádření vlivu změny cen alkoholických nápojů na poptávku obyvatelstva po této komoditě je z hospodářsko-politického hlediska zajímavá.*

V článku je hledán odhad poptávkových funkcí pro pivo, révové víno, destiláty a lihoviny. Například pivo autor modeluje s indexem determinace 0,959 jako

$$P = 0,9252 \text{ ICR}^{0,1270} \text{ PRIMLI}^{-0,0812} \text{ P}_{-1}^{1,0584},$$

kde ICR je index ceny lihovin,  $P_{-1}$  je spotřeba piva před 15. 10. 1984 a PRIMLI jsou čisté reálné příjmy na obyvatele v Kčs. Statickou poptávkovou funkci po vínu odhadl autor tímto způsobem

$$V = 2,4184 + 1,9036 \text{ ICR} - 4,2525 \text{ ICV} + 0,6116 \text{ PRIMLI},$$

kde ICV je index ceny vína.

Nabídkové funkce autor nevyužívá a uvádí, že *v podmínkách socialistické ekonomiky jsou maloobchodní ceny stanoveny plánovacím orgánem současně s více-méně přesně stanoveným množstvím zboží dodávaného na trh. Nabídková funkce tak degeneruje do podoby bodové nabídky.*

Při odhadech autor vychází z toho, že na trhu alkoholických nápojů existuje převaha nabídky nad poptávkou a tudíž je cena pro průměrného spotřebitele cenou rovnovážnou, tedy že kvantum nakupovaného zboží odpovídá ceně. Není ale jasné, zda toto plně koresponduje se skutečností.

Podle článků

- Klaus, V., Rudlovčák, V.: *Nerovnovážné modely v ekonomii. Ekonomicko-matematický obzor (2) 18, 1982,*
- Dlouhý, V.: *Ekonometrický model československé obchodní bilance, Academia, 12, 1985,*
- Dyba, K.: *Odhad poptávkové funkce na nerovnovážném trhu. Politická ekonomie, 12, 1973,*

na našem vnitřním trhu existuje globální rovnováha, ale neexistuje rovnováha strukturální. Není vůbec jisté, zda by kvantum nakupovaného alkoholu bylo nižší, kdyby u nás fungovala i strukturální rovnováha. Může se ale stát, že nedostatek zboží na jiných trzích (byty, zahraniční rekreace, kvalitní zboží dlouhodobé spotřeby) způsobí větší nákupy na trhu alkoholu.

Některé zatím otevřené otázky ohledně konzumace alkoholu má dále řešit Dyba, K. a Klaus, V.

V praktické části mé diplomové práce budu postupovat podobně jako autor v článku [62]. Budu také hledat vhodné regresní modely, i když se omezím pouze na lineární. Předmětem mého zájmu nebude zkoumání poptávky po alkoholu, ale modelování cen akcií na burze s využitím faktorových modelů a modelování poptávky po nově stavěných bytech v ČR pomocí simultánních rovnic. Doufám, že zvolené úlohy budou pro čtenáře zajímavé.

## 2 Předmět ekonometrické analýzy

V literatuře [8] je *ekonometrie* popsána jako věda, která vychází ze spojení matematiky, statistiky, ekonomie a v poslední době i informatiky. Vznikla za účelem vyhledávání, měření a matematické formulace ekonomických vztahů a zákonitostí. Pomocí matematických metod vymezuje kvantitativní závislosti hospodářského života.

*Finanční ekonometrie* je v [3] termín, který se dnes používá pro jakoukoli kvantitativní (statistickou) analýzu finančních dat. Tj. popisná statistická analýza, stat. identifikace, odhad a verifikace příslušného modelu, stat. testování hypotéz a konstrukce předpovědí.

### 2.1 Konstrukce ekonometrického modelu

Nalezení vhodného ekonometrického modelu není jednoduché. Bývá doporučováno, např. viz [3, 5, 9], použít algoritmus, jehož jednotlivé kroky lze obecně charakterizovat následujícím způsobem:

**Krok 1:** Jednoznačná formulace (finančního) problému, který chceme řešit, tj. vymezení příslušného teoretického rámce a vyhledání případných předchozích pokusů o řešení tohoto či podobného problému.

**Krok 2:** Formulace teoretického (ekonomického) modelu adekvátního teorii z předchozího kroku (většinou je nutné se spokojit s určitým stupněm aproximace).

**Krok 3:** Vyhledání vhodných (finančních) dat.

**Krok 4:** Specifikace ekonomického modelu, neboli formulace základní hypotézy. Po vhodné statistické specifikaci stochastických vlivů je výsledkem ekonometrický model. Ten popisuje základní hypotézu jednou či více rovnicemi, které jsou zpravidla vzájemně závislé.

**Krok 5:** Kvantifikace ekonometrického modelu vyjadřuje volbu vhodné odhadové metody (většinou je ovlivněna softwarem, který máme k dispozici) včetně ověření



předpokladů vyžadovaných zvolenou metodou. V další fázi probíhá kvantifikace intenzity a směru vzájemného působení do modelu zahrnutých proměnných, měřené odhadnutými parametry ekonometrického modelu.

**Krok 6:** Je-li ekonometrický model odhadnut, následuje jeho statistická verifikace (ověření). Jde o ověření statistickým testem, zda statistický model je kompatibilní s použitými daty, tzn. zda jsou odhadnuté parametry v souladu s výchozími teoretickými předpoklady. Pokud tomu tak není, je nutné revidovat kroky 2, 3, 4 a 5.

**Krok 7:** Interpretace modelu je věcné posouzení, zda odhadnutý model je prakticky kompatibilní s realitou.

**Krok 8:** Praktické využití modelu pro řešení problému z kroku 1. Často je výsledkem doporučení určité akce, pokud určité veličiny překročí jisté mezní hodnoty.

Praktické úlohy zahrnují více či méně jakýsi faktor nejistoty. Tato nejistota je obvykle dána ve formě tzv. náhodných složek (chyb), vyskytujících se v řadě ekonometrických vztahů. Zde tkví rozdíl od ekonomické teorie a totiž v tom, že vztahy mezi různými ekonomickými veličinami nejsou dány přesně, ale respektuje se jejich stochastický (pravděpodobnostní) charakter.

Správně specifikovaný model, z hlediska matematického tvaru závislosti a stochastických předpokladů o rozdělení pravděpodobnosti náhodných složek, umožňuje vhodně zvolenými statistickými a ekonometrickými metodami kvantifikovat měřené parametry modelu.

## 2.2 Typy dat

V rámci ekonometrie lze klasifikovat analyzovaná data do tří skupin (podrobněji viz [3, 10]):

1. Časová data – jedná se o data ve tvaru *časových řad*, tj. o hodnoty určité veličiny pozorované v časovém intervalu s určitou frekvencí,
2. Průřezová data – jedná se o data ve tvaru *průřezového výběru*, tj. o hodnoty určité veličiny jednotlivých subjektů pozorované ve stejném období, tj. k ur-

čitému časovému okamžiku; není důležité uspořádání dat (lze je libovolně přerovnávat),

3. Panelová data – kombinace časových a průřezových dat, která vznikají opakovaným šetřením souboru subjektů v různých obdobích; ekonometrické modely pro panelová data mají řadu specifík.

Ekonometrická data lze klasifikovat i podle jiných hledisek, např. na *kvantitativní data* (ceny, kursy, počty) a *kvalitativní data* (rating).

## 2.3 Oblasti využití ekonometrických modelů

Ekonometrická analýza nachází uplatnění jak při kvantifikaci a verifikaci ekonomických hypotéz na makro-úrovni, tak i při zkoumání vztahů mezi poptávkou, důchody a cenami nebo spotřebou a důchody na mikro-úrovni.

Oblastí aplikace jednorovnicových ekonometrických modelů je především studium *poptávkových funkcí*, které souvisí s domácnostmi a dále *produkčních a nákladových funkcí*, jejichž předmětem zkoumání jsou firmy.

Z makroekonomických modelů se jedná především o *spotřební funkce* nebo *investiční funkce*. Také při ověřování *funkcí poptávky po penězích* v souvislosti s analýzou efektivnosti monetární a fiskální politiky představuje ekonometrické modelování důležitý nástroj. Ekonometrie se také zabývá, jak již bylo řečeno, *analýzou vývoje nebo chování* zkoumaného systému v období pozorování, nazývanou také *aplikací modelu ex post*. Spočívá v interpretaci a testování významu odhadnutých parametrů i modelu jako celku. Dále se nabízí možnost *předpovědi* budoucích hodnot vysvětlovaných proměnných, neboli *aplikace ex ante*. Slouží k odhadu endogenních proměnných pomocí odhadnutého modelu v období mimo interval pozorování. Předpokladem je znalost očekávaných vysvětlujících proměnných v období predikce. V neposlední řadě ekonometrické modely umožňují *volbu optimální hospodářské politiky*, spočívající ve výběru nejlepší varianty ekonomických nástrojů řízení. Podrobněji jsou zájmové oblasti a cíle ekonometrie popsány např. v literaturách [3, 8, 10, 13, 23].

## 3 Klasický model lineární regrese

### 3.1 Motivace

*Regresní analýza* jako nejdůležitější ekonometrický nástroj slouží pro kvantitativní popis vztahu mezi ekonomickými a finančními veličinami označovanými jako *proměnné*. Úkolem regrese je vysvětlit změny hodnot jedné proměnné změnami hodnot jiných proměnných. *Vysvětlovaná proměnná* (závisle proměnná) se obvykle značí  $y$  a *vysvětlující proměnné* (nezávisle proměnné nebo také regresory)  $x_0, x_1, \dots, x_k$ . Ve vícerovnicových ekonometrických modelech se ovšem často vysvětluje několik vysvětlovaných proměnných  $y_1, \dots, y_m$  najednou. Regrese je studována ve všech učebnicích matematické statistiky, např. viz [1, 2, 16] a také ve všech učebnicích ekonometrie, např. v [3, 6, 20, 24]. Z uvedených pramenů budu v této kapitole čerpat.

Formálně lze zapsat *lineární regresní model* jako

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, \quad t = 1, \dots, n, \quad (1)$$

kde  $t$  je většinou časový index;  $y_t$  je hodnota vysvětlované proměnné  $y$  pozorovaná v čase  $t$  a  $x_{t0}, x_{t1}, \dots, x_{tk}$  jsou hodnoty vysvětlujících proměnných  $x_0 \equiv 1, x_1, \dots, x_k$  pozorované v čase  $t$  ( $x_{t0}$  lze specifikovat jako zvláštní (umělou) proměnnou, která nabývá ve všech pozorováních hodnoty rovné jedné).  $\beta_0, \beta_1, \dots, \beta_k$  jsou neznámé (regresní) parametry modelu, přičemž  $\beta_0$  je tzv. absolutní člen;  $\varepsilon_t$  je reziduální složka modelu (nebo také náhodná složka, reziduum). Reziduální složka v sobě zahrnuje souhrn vlivů, které nejsou v modelu explicitně uvedeny, chyby v měření, nekorektní volbu regresního vztahu nebo některé jevy náhodného charakteru. Tento formální zápis se pak při konkrétních aplikacích vhodně modifikuje. Pro náhodnou složku platí, že  $E(\varepsilon_t) = 0$ ,  $t = 1, \dots, n$ , potom očekávanou hodnotou  $y_t$  můžeme vyjádřit jako

$$E(y_t) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk}, \quad t = 1, \dots, n.$$

Tento deterministický vztah se také nazývá *regresní funkce základního souboru*. Regresní koeficienty měří změnu  $E(y_t)$ , odpovídající jednotkové změně jedné vy-

světlující proměnné, přičemž ostatní vysvětlující proměnné zůstávají neměnné.

Daleko přehlednější je však *maticový zápis* lineárního regresního modelu

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

kde

$$\mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1k} \\ x_{20} & x_{21} & \dots & x_{2k} \\ \vdots & \ddots & \vdots & \vdots \\ x_{n0} & x_{n1} & \dots & x_{nk} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

přičemž  $\mathbf{X}$  je matice ( $n \times p$ ) pozorování hodnot vysvětlujících proměnných,  $p = k + 1$  je počet vysvětlujících proměnných v modelu s absolutním členem,  $\boldsymbol{\beta}$  je sloupcový vektor ( $p \times 1$ )  $p$  neznámých proměnných a  $\boldsymbol{\varepsilon}$  je sloupcový vektor ( $n \times 1$ )  $n$  hodnot nepozorovatelné náhodné složky. Rozdíl počtu pozorování  $n$  a počtu odhadovaných parametrů  $p$  je *počet stupňů volnosti*, přičemž musí platit nerovnost  $n > p$ .

Zatím se zabýváme lineárním modelem, což znamená, že model je *lineární v parametrech*, ale nemusí být *lineární v proměnných*. Příkladem je *exponenciální regresní model* s neznámým parametrem v exponentu mocninné funkce, který však může být převeden na lineární regresní model logaritmickou transformací, viz [1, 3, 16].

### 3.2 Metoda nejmenších čtverců (MNČ)

K odhadu parametrů lineárního regresního modelu (1) se nejčastěji používá přístup založený na metodě nejmenších čtverců.

*Metoda nejmenších čtverců* hledá odhady parametrů  $\boldsymbol{\beta}$ , tak že se vzhledem k těmto parametrům minimalizuje součet čtverců

$$S = \sum_{t=1}^n (y_t - (\beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk}))^2 = \sum_{t=1}^n (y_t - \mathbf{x}_t \boldsymbol{\beta})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (3)$$

V (3) se minimalizuje součet druhých mocnin vertikálních vzdáleností hodnot vysvětlované proměnné  $y_t$  od *regresní nadroviny* (v nejjednodušším případě se jedná

o regresní přímku). Snažíme se tak o co nejlepší proložení nadroviny množinou pozorovaných bodů.

Optimalizační úloha, kdy minimalizujeme (3) přes parametry  $\beta$ , má řešení

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (4)$$

Odhad  $\hat{\beta}$  parametrů  $\beta$  se obvykle nazývá *bodovým odhadem metodou nejmenších čtverců* a označuje zkratkou *MNČ-odhad*. V souvislosti s tím se zavádí další dva důležité pojmy

- *výběrová regresní funkce s MNČ-hodnotami*

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}, \text{ resp.}$$

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \dots + \hat{\beta}_k x_{tk}, \quad t = 1, \dots, n,$$

jsou tzv. *vyrovnané hodnoty*  $y_t$ ;

- *MNČ-rezidua*

$$\hat{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}. \quad (5)$$

### 3.3 Vlastnosti odhadu metodou nejmenších čtverců

Odvození vlastností MNČ-odhadu je možné jen v případě, že model splňuje určité předpoklady. Tyto předpoklady charakterizující tzv. *klasický model lineární regrese* (KMLR) se často uvádějí v následujícím tvaru:

**(P1):**  $E(\varepsilon_t) = 0$ , tj. střední hodnota reziduální složky je nulová pro všechna  $t$ ;

**(P2):**  $\text{var}(\varepsilon_t) = \sigma^2 < \infty$ , tj. rozptyl reziduální složky je konstantní a konečný pro všechna  $t$  (předpoklad konstantního rozptylu reziduálních složek se označuje jako *homoskedasticita*);

**(P3):**  $\text{cov}(\varepsilon_s, \varepsilon_t) = 0$  pro  $s \neq t$ , tj. reziduální složky jsou navzájem nekorelované pro všechna  $s \neq t$ ;

(P4):  $\text{cov}(x_{ti}, \varepsilon_t) = 0$ , tj. regresory jsou ve stejném čase nebo pro stejnou průřezovou jednotku nekorelované s reziduální složkou pro všechna  $i$  a  $t$ ;

(P4'):  $h(\mathbf{X}) = p$ , tj. nenáhodná matice  $\mathbf{X}$  má lineárně nezávislé sloupce (uplatní se pouze při nenáhodných regresorech).

Lze je zapsat také maticově:

(P1):  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ;

(P2) a (P3):  $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$  (tzv. kovarianční matice);

(P4):  $E(\mathbf{X}'\boldsymbol{\varepsilon}) = \mathbf{0}$ ;

(P4'):  $h(\mathbf{X}) = p$ .

Kromě parametrů  $\boldsymbol{\beta}$  obsahuje KMLR ještě neznámý parametr  $\sigma^2 > 0$ . V rámci metodiky nejmenších čtverců se za jeho MNČ-odhad obvykle bere

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n \hat{\varepsilon}_t^2}{n-p} = \frac{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}{n-p} = \frac{\text{RSC}}{n-p} = \frac{1}{n-p} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (6)$$

Odhad se nazývá *nestranný (nevychýlený)*, jestliže jeho střední hodnota je rovna hodnotě odhadovaného parametru (na odhady pohlížíme jako na náhodné veličiny). V KMLR lze snadno ukázat, že

- MNČ-odhad  $\hat{\boldsymbol{\beta}}$  je nestranný odhad parametrů  $\boldsymbol{\beta}$ , tj.

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta};$$

- MNČ-odhad  $\hat{\sigma}^2$  je nestranný odhad parametru  $\sigma^2$ , tj.

$$E(\hat{\sigma}^2) = \sigma^2.$$

Odhad se nazývá *konzistentní*, jestliže při rostoucím rozsahu výběru  $n$  konverguje podle pravděpodobnosti ke skutečné hodnotě odhadovaného parametru (konzistence je asymptotická vlastnost odhadu). V praxi pro konzistenci odhadu stačí např. ověřit, že tento odhad je (asymptoticky) nestranný a zároveň, že jeho

rozptyl při rostoucím rozsahu výběru  $n$  konverguje k nule. V klasickém modelu lineární regrese lze ukázat, že za platnosti rutinních předpokladů

- MNČ-odhad  $\hat{\beta}$  je konzistentní odhad parametrů  $\beta$ ;
- MNČ-odhad  $\hat{\sigma}^2$  je konzistentní odhad parametru  $\sigma^2$ .

Odhad se nazývá *eficientní* vůči jinému odhadu téhož parametru, jestliže má menší rozptyl (přesněji: jestliže nemá větší rozptyl). Pro libovolný lineární nestranný odhad  $\hat{\beta}$  parametrů  $\beta$  platí

$$\text{var}(\hat{\beta}_i) \leq \text{var}(\hat{\beta}'_i) \quad \forall i, \text{ kde}$$

$$\text{var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

MNČ-odhad  $\hat{\beta}$  v KMLR je dokonce *nejlepší nestranný lineární odhad (BLUE)* parametrů  $\beta$  (tzv. *Gaussova-Markovova<sup>1</sup> věta*). To znamená, že je současně lineární funkcí hodnot vysvětlovaných proměnných  $y_1, y_2, \dots, y_n$ , nestranným odhadem parametrů  $\beta$  a eficientním odhadem vůči každému lineárnímu nestrannému odhadu parametrů  $\beta$ .

Více o MNČ a odvození odhadů jednotlivých parametrů a jejich vlastností lze nalézt např. v [3, 5, 8, 9, 15, 23].

### 3.4 Normální model

O normálním modelu se mluví v situaci, kdy k předpokladům **(P1)** – **(P4)** klasického modelu lineární regrese přidáme další předpoklad

**(P5):**  $\varepsilon_t \sim N(0, \sigma^2)$ , tj. reziduální složky jsou normálně rozdělené pro všechna  $t$ , resp. platí  $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ .

Normalita je nejjednodušším předpokladem v situaci, kdy je nutné specifikovat pravděpodobnostní rozdělení reziduální složky (k testování různých hypotéz

---

<sup>1</sup>Carl Friedrich Gauss (1777-1855), německý matematik a fyzik  
Andrej Andrejevič Markov (1856-1922), ruský matematik

v modelu, konstruování spolehlivostních a předpovědních intervalů, nalezení maximálních věrohodných odhadů parametrů aj.).

Prostředky matematické statistiky lze ukázat, že v normálním modelu mimo jiné platí:

- MNČ-odhad má vícerozměrné normální rozdělení

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}); \quad (7)$$

- MNČ-odhad  $\hat{\sigma}^2$  parametru  $\sigma^2$  má chí-kvadrát rozdělení

$$\frac{n-p}{\sigma^2} \hat{\sigma}^2 \sim \chi^2(n-p);$$

- MNČ-odhady  $\hat{\boldsymbol{\beta}}$  a  $\hat{\sigma}^2$  jsou navzájem nezávislé.

### 3.5 Statistická indukce v klasickém lineárním modelu

#### 3.5.1 Testování významnosti odhadnutých parametrů

Protože bodová odhadová funkce regresních parametrů  $\hat{\boldsymbol{\beta}}$  poskytuje odhady  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  na základě jednoho výběru pozorování ze základního souboru, je třeba testovat jejich statistickou významnost.

Kdybychom znali konstantní rozptyl náhodných složek  $\sigma^2$ , mohli bychom jako východisko k testování hypotéz o skutečných hodnotách jednotlivých regresních parametrů použít (7). Ve skutečnosti ale  $\sigma^2$  neznáme, proto při testování významnosti regresních parametrů vycházíme z jeho nestranného MNČ-odhadu (6). Známe-li  $\hat{\sigma}^2$ , určíme dále i nestranné odhady rozptylů odhadnutých parametrů  $\hat{\beta}_j$  na základě

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Odmocniny odhadů rozptylů  $\hat{\sigma}^2 x^{jj}$  na diagonále kovarianční matice  $\widehat{\text{var}}(\hat{\boldsymbol{\beta}})$  jsou odhady standardních chyb bodových odhadů  $\hat{\beta}_j$ . Nediagonální prvky  $\widehat{\text{var}}(\hat{\boldsymbol{\beta}})$  představují odhadnuté kovariance dvojic bodových odhadů, tj.

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \hat{\sigma}^2 x^{ij}, \quad i \neq j.$$



Protože podíl  $\frac{(\hat{\beta}_j - E(\hat{\beta}_j))}{\sqrt{\text{var}(\hat{\beta}_j)}}$  má standardizované normální rozdělení, potom poměr

$$t_j = \frac{(\hat{\beta}_j - \beta_j)}{\hat{\sigma}\sqrt{x^{jj}}}, \quad j = 0, \dots, k \quad (8)$$

má pro každé  $j$  **Studentovo rozdělení t** o  $n - p$  stupních volnosti, tj.  $t_j \sim t_{n-p}$ . Tato testovací statistika je vhodná především pro malé výběry, tzn.  $n \leq 30$ . Testovací kritérium (8) umožňuje testovat hypotézy týkající se skutečné hodnoty libovolného jednoho regresního parametru  $\beta_j$ . Většinou se volí hladina významnosti  $\alpha = 0,05$  při použití oboustranného t-testu  $t_{\frac{\alpha}{2}}$ . Pokud  $|t_j| \leq t_{\frac{\alpha}{2}}$ , pak příslušnou nulovou hypotézu na zvolené hladině významnosti akceptujeme, v opačném případě ji zamítáme ve prospěch alternativní hypotézy.

V ekonometrické analýze se velmi často testuje nulová hypotéza, že regresní parametr  $\beta_j = 0$ . Tedy, zda bodový odhad  $\hat{\beta}_j$  je nebo není statisticky významný. Testovací kritérium

$$t_j = \frac{|\hat{\beta}_j|}{\hat{\sigma}\sqrt{x^{jj}}}, \quad j = 0, \dots, k \quad (9)$$

se nazývá **t – poměr**. Pokud nulovou hypotézu akceptujeme, s pravděpodobností  $100(1 - \alpha)\%$  usuzujeme, že odhad  $\hat{\beta}_j$  není statisticky významný.

Pro libovolné  $n - p > 30$  se  $t$ -rozdělení téměř neliší od normálního rozdělení, takže můžeme ignorovat počet stupňů volnosti a položit pro hladinu významnosti kritickou hodnotu  $t_{n-p}(1 - \frac{\alpha}{2}) = u(1 - \frac{\alpha}{2}) = 1,96$ .

### 3.5.2 Intervaly spolehlivosti odhadnutých parametrů

Odmítnutí nulové hypotézy v průběhu testování statistické významnosti odhadnutých parametrů lineárního regresního modelu ještě neznamená, že MNČ-odhady  $\hat{\beta}_j$  jsou přesnými odhady skutečných hodnot parametrů  $\beta_j$  v základním souboru. Abychom mohli určit stupeň shody odhadnuté a skutečné hodnoty příslušného parametru, musíme pro jeho skutečnou hodnotu stanovit *interval spolehlivosti (konfidenční interval)*. Jedná se tedy o nalezení mezí, uvnitř kterých se při opakovaných výběrech nalézá skutečná hodnota parametru s urči-

tým stupněm spolehlivosti, s požadovanou pravděpodobností, která se volí předem a nazýváme ji hladinou spolehlivosti.

Při malém rozsahu výběru vycházíme z  $t$ -rozdělení, použijeme testovací statistiku (8) pro  $j$ -tý odhadnutý regresní parametr a zvolíme hladinu spolehlivosti  $\alpha$ , potom

$$P(-t_{\frac{\alpha}{2}} < t_j < t_{\frac{\alpha}{2}}) = 1 - \alpha.$$

Dosažením za  $t_j$  pro  $\alpha = 0,05$  dostáváme

$$P(\hat{\beta}_j - t_{\frac{\alpha}{2}} \cdot \hat{\sigma} \sqrt{x^{jj}} < \beta_j < \hat{\beta}_j + t_{\frac{\alpha}{2}} \cdot \hat{\sigma} \sqrt{x^{jj}}) = 0,95.$$

95%-ní interval spolehlivosti pro  $j$ -tý parametr lze tedy vyjádřit ve tvaru

$$\beta_j = \hat{\beta}_j \pm t_{0,025} \cdot \hat{\sigma} \sqrt{x^{jj}}, \quad j = 0, \dots, k. \quad (10)$$

Odhadnutý koeficient  $\hat{\beta}_j$  se významně liší od nuly, tj. je statisticky významný, neobsahuje-li interval spolehlivosti (10) hodnotu nula a opačně.

Pro dva nebo více regresních parametrů dohromady se konstruuje tzv. *oblast spolehlivosti*. Potom místo  $t$ -poměru použijeme **F** – **poměr** se stupni volnosti  $p$  a  $n - p$ . Pomocí  $F$ -poměru lze testovat také libovolné hypotézy týkající se všech parametrů  $\beta_j$ . Obecně pro  $p$  parametrů je simultánní oblast spolehlivosti určena  $p$ -rozměrným elipsoidem spolehlivosti. Více viz [2, 8, 16].

### 3.5.3 Kritéria shody odhadnutého modelu s daty

Poté, co model lineární regrese odhadneme, je nutné posoudit, zda je skutečně kompatibilní s použitými daty. To lze provést pomocí statistických testů, ale také orientačně na základě velikosti tzv. *vícenásobného koeficientu determinace*.

Jako jednoduchý nástroj se v tomto kontextu především nabízí *reziduální (nevysvětlený) součet čtverců*

$$\text{RSC} = \sum_{t=1}^n \hat{\varepsilon}_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2, \quad (11)$$

na jehož minimalizaci je MNČ-metodika založena: čím je jeho nezáporná hodnota menší, tím je přijatelnější zkonstruovaný model. Dalším používaným typem

součtu čtverců je *úplný (celkový) součet čtverců*

$$\text{USČ} = \sum_{t=1}^n (y_t - \bar{y})^2 \quad (12)$$

a *vysvětlený součet čtverců*

$$\text{VSČ} = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2. \quad (13)$$

Pro tyto součty čtverců platí

$$\text{USČ} = \text{VSČ} + \text{RSČ}.$$

*Koeficient (index) determinace* se potom definuje jako

$$R^2 = \frac{\text{VSČ}}{\text{USČ}} = 1 - \frac{\text{RSČ}}{\text{USČ}}, \quad (14)$$

hodnota  $R^2$  se pohybuje od nuly do jedné, pokud je index blízký jedné, model daným datům padne dobře. Obsahuje-li výběrová regresní funkce úrovnovou konstantu, lze koeficient vícenásobné determinace  $R^2$  vyjádřit jako

$$R^2 = 1 - \frac{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}{\mathbf{y}' \mathbf{y}} = \frac{\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n \bar{y}^2}{\mathbf{y}' \mathbf{y} - n \bar{y}^2}. \quad (15)$$

Pokud jsou všechny regresní koeficienty nulové, pak  $\text{RSČ} = \text{USČ}$ , takže z celkového součtu čtverců není vysvětleno nic a  $R^2 = 0$ . Důležitý je také fakt, že hodnota koeficientu determinace nikdy neklesne přidáním dalších vysvětlujících proměnných do modelu.

Koeficient determinace  $R^2$  je proto vhodný k testování statistické významnosti modelu jako celku. Tento test probíhá pomocí  $F$ -poměru

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - p}{k}, \quad (16)$$

přičemž proměnné jsou centrovány, tzn. s nulovými průměry ( $\bar{y} = 0$ ). Podíl má tedy  $F$ -rozdělení s  $k, n - p$  stupni volnosti. Je-li tento poměr větší než tabulková

hodnota  $F_{k,n-p}$ , potom na zvolené hladině významnosti odmítneme nulovou hypotézu o nevýznamnosti  $R^2$  ve prospěch alternativní hypotézy, že hodnota  $R^2$  a tudíž i shoda odhadnutého modelu s daty, jsou statisticky významné. Nebo-li akceptujeme předpoklad, že ne všechny regresní parametry  $\beta_j$  jsou nulové pro  $j = 0, \dots, k$ . V opačném případě vysvětlující proměnné významně neovlivňují vysvětlovanou proměnnou a všechny regresní parametry  $\beta_j$  jsou nulové,  $j = 0, \dots, k$ .

Pro centrované proměnné lze hypotézu, že všechny skutečné hodnoty regresních parametrů  $\beta_j$  jsou nulové nebo rovny určitým hodnotám, testovat pomocí modifikovaného F poměru (16) jako

$$F = \frac{(\hat{\beta} - \beta)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta)}{\hat{\epsilon}' \hat{\epsilon}} \cdot \frac{n-p}{k}, \quad (17)$$

který má opět rozdělení  $F_{k,n-p}$ .

**Poznámka 3.1** Při statistické verifikaci lineárního regresního modelu  $t$  nebo  $F$  testy či koeficientem determinace nemusíme vždy dospět k jednoznačnému závěru. Může se stát, že  $R^2$  je signifikantní (významný), ale některé nebo všechny parametry  $\beta_j$  nikoliv (viz testovací kritérium (9)) nebo naopak. V takové situaci musíme rozhodnout, zda přiřadíme větší váhu  $R^2$  nebo standardním chybám odhadnutých regresních parametrů modelu. Obvykle se řídí pravidlem, že kritériu  $R^2$  se dává přednost tehdy, je-li odhadnutý model určen především pro prognózování, zatímco standardní chyby mají větší význam při vysvětlení a verifikaci či analýze zkoumaných ekonomických vztahů.

## 4 Ekonometrická zobecnění lineární regrese

V ekonometrické praxi dochází k častému porušování předpokladů **(P1)** – **(P5)** pro klasický model lineární regrese, neboť konkrétně práce s ekonomickými a finančními daty má svá specifika porušující standardní omezení pro všeobecné aplikace regrese ve statistice. V takovém případě je nutné odpovědět na otázku, jaký vliv má nedodržení jednotlivých předpokladů na vlastnosti MNČ-odhadů regresních parametrů, popř. jak postup modifikovat, aby kvalita odhadů zůstala zachována.

V této kapitole probereme zobecnění či rozšíření těchto předpokladů vyžadovaná v ekonomické a finanční praxi a budeme se zabývat otázkami, které se týkají tohoto porušování:

- jak lze taková porušení detekovat?
- jaké jsou jejich nejčastější příčiny v praxi?
- jaké jsou důsledky jejich případného ignorování?

V souvislosti s poslední otázkou se mohou objevit následující negativní důsledky:

- parametry jsou špatně odhadnuty;
- směrodatné odchylky těchto odhadů jsou špatně odhadnuty;
- pro testové statistiky jsou použita neadekvátní pravděpodobnostní rozdělení.

Řešení v případě porušení některého z předpokladů může být hledáno tak, že

- je použit jiný model, jehož předpoklady jsou splněny;
- je použita alternativní odhadová či testová statistika, která je vůči porušení předpokladů necitlivá.

## 4.1 Zobecněný model lineární regrese, ZMNČ

Teoretickým základem pro naznačený postup je tzv. *zobecněný model lineární regrese*

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}^*, \quad (18)$$

který předpoklady **(P2)** a **(P3)**, tj.  $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$ , zobecňuje do tvaru

**(P2')** a **(P3')** :  $\text{var}(\boldsymbol{\varepsilon}^*) = \boldsymbol{\Sigma}^* = \sigma^2 \boldsymbol{\Omega}$ , kde  $\boldsymbol{\Omega}$  je pozitivně definitní matice ( $\boldsymbol{\Omega} > 0$ ), tzn. že rozptyl reziduální složky nemusí být konstantní a reziduální složky nemusí být navzájem nekorelované. Dále také může být porušen předpoklad **(P4')** o lineární nezávislosti sloupců matice  $\mathbf{X}$ . Ostatní předpoklady klasického modelu lineární regrese jsou neměnné.

*Nejlepší nestranný lineární odhad (BLU-odhad)* parametrů  $\boldsymbol{\beta}$  lineárního modelu, který vyhovuje obecnější podmínce uvedené výše, je nejjednodušší pomocí *zobecněné metody nejmenších čtverců (ZMNČ)*, nazývané podle svého autora také *Aitkenovým<sup>2</sup> odhadem* a je tvaru

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{y} \quad (19)$$

s rozptylovou (kovarianční) maticí tvaru

$$\text{var}(\tilde{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \quad (20)$$

a s nestranným odhadem parametru  $\sigma^2 > 0$  tvaru

$$\tilde{\sigma}^2 = \frac{\tilde{\boldsymbol{\varepsilon}}' \boldsymbol{\Omega}^{-1} \tilde{\boldsymbol{\varepsilon}}}{n - p}, \quad \text{kde } \tilde{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\beta}}, \quad (21)$$

takže kovarianční matici (20) lze nestranně odhadnout jako

$$\widetilde{\text{var}}(\tilde{\boldsymbol{\beta}}) = \tilde{\sigma}^2 (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}.$$

---

<sup>2</sup>Alexander Craig Aitken (1895-1967), novozélandský matematik

**Poznámka 4.1** Lze ukázat, že pokud bychom v zobecněném modelu lineární regrese použili MNČ-odhady, pak

- MNČ-odhad  $\hat{\beta}$  podle (4) zůstává nestranným odhadem parametrů  $\beta$ , ale není už obecně nejlepší mezi nestrannými lineárními odhady parametrů  $\beta$ ;
- MNČ-odhad  $\hat{\sigma}^2$  podle (6) není obecně nestranným odhadem parametru  $\sigma^2$ ;

**Poznámka 4.2** V případě normality  $\varepsilon$  mají normální rozdělení i náhodné složky zobecněného modelu lineární regrese. V téhle situaci lze pro zobecněný model použít standardní testy statistické významnosti a stanovit intervaly spolehlivosti pro jednotlivé parametry vektoru  $\beta$ .

Protože však matici  $\Omega$  obvykle neznáme, konstruujeme ji až po MNČ-odhadu modelu (2) na základě spočtených reziduí, přičemž způsob transformace (2) na (18) je odlišný v případě heteroskedasticity než při existenci autokorelace náhodných složek.

## 4.2 Heteroskedasticita

O heteroskedasticitě se mluví v případě porušení předpokladu *homoskedasticity*, tj. předpokladu

$$(P2): \text{var}(\varepsilon_t) = \sigma^2 < \infty.$$

Jinými slovy, jestliže reziduální složky nemají konstantní rozptyl (tj. jestliže množství náhodnosti obsažené ve výstupu  $y_t$  může být pro každé pozorování různé), pak se označují jako *heteroskedastické*. Typická je heteroskedasticita pro modelování finančních průřezových dat, kdy dochází k velkým změnám v hodnotách vysvětlujících proměnných. Mnohem méně se heteroskedasticita vyskytuje při odhadu modelu z údajů časových řad.

Jako příklady měnícího se rozptylu náhodných složek můžeme uvést poptávkové či úsporové funkce, kdy s rostoucími disponibilními příjmy domácností roste

variabilita jejich výdajů či úspor. V případě produkční funkce se mění rozptyl objemu produkce s počtem zkoumaných firem nebo s jejich velikostí.

Formálně v rámci zobecněného modelu lineární regrese představuje heteroskedasticita případ, kdy

$$\text{var}(\varepsilon^*) = \sigma^2 \mathbf{\Omega} = \sigma^2 \mathbf{diag}\{k_1, \dots, k_n\}, \quad \sigma^2 > 0; k_1, \dots, k_n > 0,$$

tj. reziduální složky  $\text{var}(\varepsilon_t^*)$  (a při deterministických regresorech také hodnoty vysvětlované proměnné  $y_t^*$ ) mají nekonstantní rozptyl  $\sigma^2 k_t$  s neznámými kladnými hodnotami  $k_t$  a jsou vzájemně nekorelované.

**Poznámka 4.3** *V praxi může heteroskedasticita vzniknout různými způsoby. Často se např. údaje, které by byly původně zvládnutelné klasickým (tj. homoskedastickým) modelem lineární regrese, z nejrůznějších příčin průměrují přes určité skupiny dat (např. z důvodu ochrany firemních či osobních dat). Dalším důvodem může být chybná specifikace modelu, spočívající ve vynechání některé podstatné vysvětlující proměnné, značně rozdílné hodnoty dat způsobující, že rozptyl vysvětlované proměnné je funkcí některé vysvětlující proměnné nebo kumulace výskytů chyb měření, čímž se zvětšuje rozptyl vysvětlované proměnné i rozptyl reziduí.*

Důsledky, které se projeví při ignoraci heteroskedasticity modelu a použití klasického MNČ-odhadu:

- MNČ-odhad  $\hat{\beta}$  zůstává nestranným a konzistentním odhadem parametrů  $\beta$ , ale není už eficientní ani obecně nejlepší (tj. BLU-odhad) mezi nestrannými lineárními odhady parametrů  $\beta$ ;
- MNČ-odhad  $\hat{\sigma}^2$  není obecně nestranným odhadem parametru  $\sigma^2$ ;
- nelze použít standardní postup pro výpočet rozptylové matice odhadu  $\hat{\beta}$ , může to vést ke zcela chybným závěrům.



Řešení heteroskedasticity je jednoduché, pokud známe její příčiny a tyto příčiny lze navíc modelově interpretovat. Nejčastější formou heteroskedasticity je závislost rozptylu náhodných složek na velikosti pozorování jedné nebo více vysvětlujících proměnných. Nechť má heteroskedasticita obecný tvar

$$\text{var}(\varepsilon_t) = \sigma_t^2 = h(\mathbf{x}_t' \boldsymbol{\gamma}), \quad t = 1, \dots, n,$$

kde  $h$  je známá funkce,  $\mathbf{x}_t = (1, x_{2t}, \dots, x_{pt})'$  je vektor  $p$  pozorovatelných proměnných, které mají vliv na variabilitu modelu a  $\boldsymbol{\gamma}$  je neznámý vektor parametrů. Přitom se rozlišuje

- *aditivní heteroskedasticita*

$$h(\mathbf{x}_t' \boldsymbol{\gamma}) = \mathbf{x}_t' \boldsymbol{\gamma}; \quad (22)$$

- *multiplikativní heteroskedasticita*

$$h(\mathbf{x}_t' \boldsymbol{\gamma}) = e^{\mathbf{x}_t' \boldsymbol{\gamma}}. \quad (23)$$

Platí-li aditivita (22), tj. rozptyl normálně rozdělené náhodné složky je lineární funkcí exogenních proměnných modelu, lze psát odhadovou funkci ZMNČ ve tvaru

$$\tilde{\boldsymbol{\beta}}_1 = (\mathbf{X}' \boldsymbol{\Omega}_1^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}_1^{-1} \mathbf{y},$$

přičemž  $\boldsymbol{\Omega}_1$  je diagonální matice s  $t$ -tým prvkem  $\mathbf{x}_t' \boldsymbol{\gamma}$ . Vektor parametrů  $\boldsymbol{\gamma}$  ale neznáme, proto musí být odhadnut. Goldfeld a Quandt navrhli konzistentní odhadovou funkci MNČ pro odhad  $\boldsymbol{\gamma}$  jako

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon}^2.$$

Odpovídající odhadová funkce parametrů  $\boldsymbol{\beta}$  při použití ZMNČ a odhadu  $\boldsymbol{\gamma}$  má pak tvar

$$\tilde{\tilde{\boldsymbol{\beta}}}_1 = (\mathbf{X}' \hat{\boldsymbol{\Omega}}_1^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\boldsymbol{\Omega}}_1^{-1} \mathbf{y}, \quad (24)$$

kde  $\hat{\boldsymbol{\Omega}}_1$  je diagonální matice s  $t$ -tým prvkem  $(\mathbf{x}_t' \hat{\boldsymbol{\gamma}})^2$ .

V případě multiplikativity (23), lze psát odhadovou funkci ZMNČ ve tvaru

$$\tilde{\beta}_2 = (\mathbf{X}'\Omega_2^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega_2^{-1}\mathbf{y},$$

přičemž  $\Omega_2$  je diagonální matice s prvky  $\exp(\mathbf{x}'_t\gamma)$ . Odhadovou funkci MNČ pro odhad  $\gamma$  můžeme vyjádřit jako

$$\hat{\gamma} = \left(\sum_{t=1}^n \mathbf{x}_t\mathbf{x}'_t\right)^{-1} \sum_{t=1}^n \mathbf{x}_t \ln \varepsilon_t^2.$$

Odpovídající odhadová funkce parametrů  $\beta$  vypadá následovně

$$\tilde{\beta}_2 = (\mathbf{X}'\hat{\Omega}_2^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}_2^{-1}\mathbf{y}, \quad (25)$$

která je konzistentní a pro normálně rozdělené náhodné složky je i asymptoticky eficientní (totéž platí i pro případ aditivity).

V praxi ovšem, jak už jsem zmínila, obvykle příčiny heteroskedasticity neznáme. Pro tento případ byly navrženy procedury, které nabízejí teoreticky podložená řešení. Testem heteroskedasticity, projevující se lineární závislostí směrodatné odchylky náhodných složek modelu na některé z vysvětlujících proměnných, je *Spearmanův<sup>3</sup> test korelace pořadí*, viz [8]. Tato metoda je bez příslušného softwaru numericky náročná. Postupuje se tak, že nejprve vypočteme MNČ-rezidua (5) a bez ohledu na znaménka uspořádáme vzestupně nebo sestupně absolutní hodnoty těchto reziduí a pozorování příslušné vysvětlující proměnné. Poté spočteme koeficient korelace pořadí ze vzorce

$$r_P = 1 - \frac{6 \sum_{t=1}^n d_t^2}{n(n^2 - 1)}, \quad (26)$$

kde  $d_t$  jsou difference v pořadí odpovídajících dvojic  $\hat{\varepsilon}_t$  a  $X_t$ . Následně testujeme významnost spočtených koeficientů korelace pořadí mezi rezidui a všemi vysvětlujícími proměnnými na zvolené hladině významnosti pomocí statistiky

$$t = r_P \frac{\sqrt{n-p}}{\sqrt{1-r_P^2}}. \quad (27)$$

---

<sup>3</sup>Charles Spearman (1863-1945), anglický matematik a psycholog

Je-li hodnota  $t$  větší než kritická tabelovaná hodnota pro Spearmanův test, akceptujeme hypotézu heteroskedasticity. Článek, ve kterém Spearman poprvé test korelace pořadí zveřejnil lze nalézt v [59].

Pro případ, kdy směrodatná odchylka je monotónní funkcí některé vysvětlující proměnné, byl sestaven *Goldfeld-Quandtův test*, viz [6, 8]. Předpokladem jeho aplikace je normalita a sériová nezávislost náhodných složek a dostatečně velký výběr, alespoň dvojnásobek pozorování než je počet odhadovaných regresních parametrů. Postupuje se následovně. Podle velikosti se vzestupně uspořádají všechna  $n$  pozorování příslušné vysvětlující proměnné. Poté se určitý počet prostředních pozorování, většinou  $m = n/4$ , vynechá a zbytek  $n-m$  pozorování (sudý počet) rozdělíme do dvou skupin. Pro každou skupinu o rozsahu  $(n-m)/2 > p$  odhadneme klasickou MNČ parametry modelu a reziduální součty čtverců (11)  $S_1$  a  $S_2$ . Pro existenci heteroskedasticity je hodnota statistiky

$$F = \frac{S_2}{S_1} \quad (28)$$

větší než kritická hodnota  $F_{(n-m-2p)/2, (n-m-2p)/2}(\alpha)$ . Čím větší je hodnota statistiky  $F$  ve vztahu ke kritické hodnotě, tím více je narušen předpoklad konstantního rozptylu náhodných složek. Tento test byl poprvé zveřejněn v článku [58].

### 4.3 Autokorelovanost reziduí

K porušení předpokladu *nekorelovaných reziduí*

**(P3):**  $\text{cov}(\varepsilon_s, \varepsilon_t) = 0$  pro  $s \neq t$

dochází často tak, že regresní model je kvantifikovaný především pomocí dat ve tvaru časových řad (s kratšími intervaly pozorování) a vykazuje tzv. *autokorelovanost reziduí*, kdy reziduální složka  $\varepsilon_t$  je korelovaná se svými zpožděnými a budoucími hodnotami  $\varepsilon_{t+k}$  ( $k \neq 0$ ). Pro veličiny s časovým uspořádáním je totiž korelovanost v čase poměrně obvyklá, přičemž předpona „auto“ se používá proto, že se tato korelovanost odehrává v rámci jedné proměnné (tzn. závislost

nikoli mezi dvěma nebo několika proměnnými, ale pouze mezi posloupností hodnot jedné proměnné). Kdybychom autokorelovanost ignorovali, mohli bychom dospět k neeficientnímu odhadu příslušného modelu lineární regrese. Obecně jsou důsledky autokorelace náhodných složek, pokud jde o vlastnosti MNČ-odhadů, obdobné jako v případě heteroskedasticity.

**Poznámka 4.4** *Příčin a důsledků autokorelace existuje více. Jednou z nich je setrvačnost údajů časových řad ekonomických veličin, kdy pozorování nejsou nezávislá, ale sériově zkorelovaná. Dále to může být chybná specifikace matematické formy modelu spočívající v hrubé aproximaci různých funkčních závislostí lineárním vztahem. Autokorelovanost může dále zapříčinit zahrnutí chyby měření vysvětlované proměnné do náhodné složky modelu, výskyt zpožděných vysvětlujících proměnných nebo data obsahující zprůměrované, vyrovnané či inter(extra)polované údaje.*

Vůbec nejjednodušší typ autokorelovanosti spočívá v modelování reziduální složky  $\varepsilon_t$  pomocí tzv. *autoregresního modelu prvního řádu* označovaného jako AR(1):

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \quad (29)$$

kde  $|\rho| < 1$  je *koeficient autokorelace prvního řádu* a  $u_t$  je tzv. *bílý šum*, což je často používané označení pro časovou řadu *normálně rozdělených navzájem nekorelovaných náhodných chyb* (reziduální složky  $\{\varepsilon_t\}$  v klasickém modelu lineární regrese tvoří zřejmě přímo bílý šum, tj.  $\rho = 0$ ). Velmi důležitou roli v autoregresním vztahu (29) hraje znaménko parametru  $\rho$ , neboť kladné  $\rho > 0$  (tzv. *pozitivní autokorelovanost*) indikuje setrvačnost ve znaménku sousedních hodnot  $\varepsilon_t$ , zatímco záporné  $\rho < 0$  (tzv. *negativní autokorelovanost*) naopak předznamenává časté změny ve znaménku sousedních hodnot  $\varepsilon_t$ .

Lze se přesvědčit, že pro příslušné nenulové kovariance  $\text{cov}(\varepsilon_s, \varepsilon_t)$  pro  $s \neq t$  vzhledem k (29) obecně platí

$$\text{cov}(\varepsilon_s, \varepsilon_t) = \rho^{t-s}\sigma^2, \quad \text{neboli} \quad \text{cov}(\varepsilon_t, \varepsilon_{t-1}) = \rho\sigma^2, \quad t = 1, \dots, n. \quad (30)$$

Vidíme tedy, že kovariance je pouze funkcí délky zpoždění  $t - s$  a nezávisí na období pozorování (předpoklad stacionarity).

Model AR(1), který je jedním z nejobvyklejších modelů časových řad, umožňuje použít *Durbin-Watsonův*<sup>4</sup> test autokorelovanosti reziduí (viz [3, 5, 6, 8]). Ve své základní podobě je schopen testovat jen autokorelaci prvního řádu (tj. jen nenulovost korelačního koeficientu  $\rho$  mezi sousedními reziduálními hodnotami  $\varepsilon_{t-1}$  a  $\varepsilon_t$ ) s nulovou hypotézou  $H_0 : \rho = 0$ . Tzv. *DW statistika*  $d$  je definována vztahem

$$d = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}. \quad (31)$$

Přesná výběrová rozdělení statistiky  $d$  nelze obecně odvodit pro všechny možné množiny pozorování. Lze však stanovit jakousi dolní ( $d_D$ ) a horní ( $d_H$ ) mez statistiky  $d$ , které jsou nezávislé na konkrétních hodnotách vysvětlujících proměnných a jejich hodnoty jsou pro různé stupně volnosti tabelovány. Nyní se postupuje následujícím způsobem:

1. pokud  $0 < d < d_D$ , jedná se o významnou kladnou autokorelaci prvního řádu,
2. pro  $4 - d_D < d < 4$  akceptujeme alternativní hypotézu záporné autokorelace prvního řádu,
3. je-li  $d_H < d < 4 - d_H$ , nezamítáme nulovou hypotézu o sériové nezávislosti,
4. v případech, kdy  $d_D \leq d \leq d_H$ , nebo  $4 - d_D \geq d \geq 4 - d_H$ , výsledky testu jsou neprůkazné.

Existuje i *modifikovaná DW statistika*, specifikovaná pro čtvrtletní data:

$$d_4 = \frac{\sum_{t=5}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-4})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}. \quad (32)$$

Její významnost se testuje stejně jako původní DW statistika. Pro aplikaci DW procedur je ale podmínkou existence absolutního členu v regresním modelu.

---

<sup>4</sup>James Durbin (1923), britský statistik a ekonometrik  
Geoffrey S. Watson (1921-1998), australský matematik

Durbin-Watsonův test byl poprvé zveřejněn v článku [57].

Později byly navrženy také obecnější testy umožňující vyšetřit i autokorelace vyšších řádů. Dnes se v příslušných ekonometrických softwarech nabízí především *Breusch-Godfreyův test* autokorelovanosti reziduí (viz [23]), který při modelování reziduální složky  $\varepsilon_t$  vychází z *autoregresního modelu vyššího řádu* AR(p),  $p \geq 1$

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \dots + \rho_p\varepsilon_{t-p} + u_t. \quad (33)$$

Testuje se nulová hypotéza  $H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$  oproti  $H_1 : \rho_1 \neq 0$  nebo  $\rho_2 \neq 0 \dots$  nebo  $\rho_p \neq 0$ . BG test potom odhadne pomocný model

$$\hat{\varepsilon}_t = \gamma_0 + \gamma_1x_{t1} + \dots + \gamma_kx_{tk} + \rho_1\hat{\varepsilon}_{t-1} + \rho_2\hat{\varepsilon}_{t-2} + \dots + \rho_p\hat{\varepsilon}_{t-p} + u_t. \quad (34)$$

A potom platí: pokud  $(n-p) \cdot R^2 \geq \chi_p^2(\alpha)$ , kde  $R^2$  je index determinace v pomocném modelu (34), zamítáme nulovou hypotézu ve prospěch alternativy o existenci autokorelace. Doporučení pro volbu řádu  $p$  autoregresního modelu spočívá ve frekvenci dat. Tj. např.  $p = 4$  pro čtvrtletní data,  $p = 12$  pro měsíční pozorování apod. Článek zveřejňující tento test lze vyhledat v [60].

Jakmile je potvrzen určitý typ korelovanosti reziduí, lze použít odhady pro zobecněný model lineární regrese. Přitom velmi rozšířená v praxi byla původně např. *Cochran-Orcuttova metoda* [3], a to zvláště pokud Durbin-Watsonův test potvrdil autokorelovanost reziduí prvního řádu. Tato metoda a další příbuzné metody mají tu nevýhodu, že kladou velmi silná omezení na strukturu modelu.

#### 4.4 Multikolinearita

Posledním předpokladem klasického modelu lineární regrese, jehož porušením jsem se v rámci uvažovaných zobecnění ještě nezabývala, je předpoklad

**(P4')**:  $h(\mathbf{X}) = p$  (při nenáhodných regresorech).

Předpoklad **(P4')** souvisí s jevem, který se v ekonometrii označuje jako *multikolinearita*.

V některých případech ekonometrické praxe mohou být sloupce regresní matice  $\mathbf{X}$  (tzv. matice pozorování) lineárně závislé. Taková situace se označuje jako *perfektní multikolinearita*. Mnohem častější je ovšem případ (*téměř*) *multikolinearity* spočívající v (*téměř*) lineární závislosti sloupců regresní matice  $\mathbf{X}$ . To znamená, že matice  $\mathbf{X}'\mathbf{X}$  má determinant blízký nule, takže ji lze numericky jen obtížně invertovat. To má důsledek pro příslušný regresní model, ve kterém pak buď vůbec nelze prakticky zkonstruovat základní MNČ-odhad, nebo odhadnuté parametry mají tak vysoké rozptyly, že jsou nepoužitelné.

Ze statistického hlediska není multikolinearita nic jiného než vysoká vzájemná korelovanost regresorů. Proto nejjednodušším příznakem multikolinearity je velká hodnota (kladná nebo záporná) výběrového korelačního koeficientu mezi dvěma regresory. Hůře se však rozpoznává multikolinearita způsobená korelovaností mezi více regresory. Pro jistotu je zapotřebí podotknout, že korelovanost mezi vysvětlovanou proměnnou a regresorem se již v žádném případě za multikolinearitu nepovažuje.

Podstatou zkoumání multikolinearity je především zjistit intenzitu závislosti mezi dvěma nebo více vysvětlujícími proměnnými a ne pouze konstatovat, zda existuje či ne. Multikolinearita je ve finanční praxi bohužel hodně častá, protože veličiny používané ve skupinách jako vysvětlující faktory mají nezřídka velice podobný (nebo naopak zcela opačný) vývojový trend (např. úroková míra a inflace nebo výnosy a riziko). Výskyt multikolinearity má (jak již bylo naznačeno) velmi negativní vliv v ekonometrických modelech, např. v situaci, kdy má model s multikolinearitou vyšší  $R^2$  (čili „vypadá dobře“ jako celek), ale jednotlivé odhadnuté parametry mohou být nevýznamné s širokými intervaly spolehlivosti.

**Poznámka 4.5** *Nejčastější příčinou existence silné multikolinearity vysvětlujících proměnných je tendence časových řad, zejména makroúdajů jako jsou HDP, spotřeba, investice, důchody, úspory apod., vyvíjet se stejným směrem a vykazovat i podobné přírůstky. Významnou multikolinearitu způsobují průřezová data (např. odhadujeme-li spotřební funkci, vysvětlující proměnné disponibilní příjem a výše jmění jsou silně zkorelovány) nebo zpožděné hodnoty zahrnuté do množiny vysvět-*

lujících proměnných, zpravidla jsou také zkorelovány. Perfektní multikolinearita se může také objevit při nula-jednotkových umělých vysvětlujících proměnných.

Vysoký stupeň multikolinearity se projevuje především snižováním přesnosti odhadnutých parametrů v důsledku chyb odhadové funkce MNČ. Odhady však zůstávají nestranné a eficientní. Pouze při opakovaných výběrech stejného rozsahu se MNČ odhadnuté parametry v různých výběrech významně liší. Tato odlišnost odhadů parametrů má za následek velkou citlivost odhadové funkce MNČ i na velmi malé změny v matici pozorování.

Narozdíl od problémů heteroskedasticity a autokorelace je multikolinearita jevem, se kterým se setkáváme nikoli v základním souboru, ale v jednom výběru pozorování, čili mluvíme o zjišťování existence a významnosti multikolinearity v konkrétním výběru. Nejjednodušší metodou zjišťování multikolinearity je tzv. *index podmíněnosti*  $\kappa$  matice  $\mathbf{X}'\mathbf{X}$ . Ten je definován jako

$$\kappa^2 = \frac{\lambda_1}{\lambda_p}, \quad (35)$$

kde  $\lambda_1$  a  $\lambda_p$  je největší a nejmenší vlastní číslo matice  $\mathbf{X}'\mathbf{X}$ . Pokud  $\kappa > 20$ , jedná se o mírnou kolinearitu, pokud je však  $\kappa > 100$ , nestačí pouze metody na potlačení kolinearity, ale je potřeba některé faktory vypustit či sloučit.

**Poznámka 4.6** *V následujících odstavcích není nutné jednotlivé charakteristiky a kritéria vyšetřovat pro  $j = 0$ , neboť pro nultý regresor (absolutní člen) platí  $x_0 \equiv 1$ .*

Pro rozpoznání multikolinearity existuje také kritérium, kdy se posuzuje velikost nediagonálních prvků *výběrové korelační matice regresorů*  $R$ . Tato korelační matice má prvky

$$r_{jl} = \frac{s_{jl}}{s_j s_l}, \quad j, l = 1, \dots, k, \quad (36)$$

kde  $s_{jl}$  jsou prvky výběrové kovarianční matice  $S$  (viz [21]) definované jako

$$s_{jl} = \frac{1}{n-1} \sum_{t=1}^n (X_{tj} - \bar{x}_j)(X_{tl} - \bar{x}_l). \quad (37)$$



Je-li  $|r_{jl}| > 0,8$ , pak dvojice regresorů  $j$  a  $l$  zřejmě způsobuje multikolinearitu a můžeme zkusit jeden z nich vynechat.

Pokud je vysvětlujících proměnných více, nabízí se také metoda *pomocných regresí* [8]. Postup je takový, že pro každý  $j$ -tý faktor,  $j = 1, \dots, k$ , spočteme dílčí koeficienty vícenásobné determinace  $R_j^2$ , pro které platí

$$R_j^2 = 1 - \frac{R_{j1}^2}{R_{j2}^2}, \text{ kde} \quad (38)$$

$$R_{j1}^2 = (\mathbf{x}_j - \mathbf{X}^* \hat{\beta}_j^*)' (\mathbf{x}_j - \mathbf{X}^* \hat{\beta}_j^*),$$

$$R_{j2}^2 = (\mathbf{x}_j - \bar{\mathbf{x}}_j)' (\mathbf{x}_j - \bar{\mathbf{x}}_j),$$

$$\hat{\beta}_j^* = (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{x}_j$$

a  $\mathbf{X}^*$  je matice všech regresorů bez  $j$ -tého regresoru. Je-li některý z těchto koeficientů blízky jedné, usuzujeme na silnou multikolinearitu. Na základě vztahu (16) lze významnost dílčích koeficientů determinace  $R_j^2$  ověřit pomocí statistiky  $F$  ve tvaru

$$F_j = \frac{R_j^2}{1 - R_j^2} \frac{n - p + 1}{k - 1}, \quad j = 1, \dots, k. \quad (39)$$

Je-li některá hodnota  $F_j$  větší jak kritická hodnota  $F_{k-1, n-p+1}(\alpha)$ , zamítáme nulovou hypotézu o úplné nezávislosti regresorů ve prospěch existence lineární závislosti  $j$ -tého faktoru na zbývajících faktorech. Nevýhodou této techniky je, že vzájemnou kolinearitu mezi nezávisle proměnnými nelze zjistit v případě, když mezi nimi existuje více než jedna lineární závislost.

Pomocí metody zv. *Scottova testační charakteristika* nebo také *Scottovo kritérium multikolinearity* (viz [18]) můžeme posoudit stupeň multikolinearity. Platí

$$M_T = \frac{T_S - 1}{T_S + 1}, \text{ kde} \quad (40)$$

$$T_S = \frac{\sum_{j=1}^k T_j^2}{k}$$

a  $T_j$  jsou hodnoty  $t$ -poměrů z testu nenulovosti regresních parametrů (9) pro  $j = 1, \dots, k$ . Pokud

- a)  $M_T < 0,33$ , model není ovlivněn multikolinearitou a není třeba ho upravovat,
- b)  $0,33 \leq M_T \leq 0,8$ , model je málo vyhovující a je vhodná jeho úprava,
- c)  $M_T > 0,8$ , model je nevhovující a je zapotřebí provést úpravu.

Doporučovaných postupů pro modely s multikolinearitou je celá řada, žádný z nich ale nevede k úplnému řešení problému. Níže je uvedeno několik z nich.

- (a) *Ignorování multikolinearity*: Někdy totiž může být regresní model adekvátní i v případě multikolinearity. Multikolinearita nezhodnotí BLU-vlastnosti MNČ-odhadu, neboť předpoklady **(P1)** – **(P4)** klasického modelu lineární regrese zůstávají v platnosti. S tím souvisí schopnost multikolinearitu vůbec rozpoznat, pro tento účel jsou k dispozici různé prostředky:
  - *orientační kritéria pro rozpoznání multikolinearity* (posouzení velikosti nediagonálních prvků výběrové korelační matice regresorů nebo velikosti výběrového koeficientu mnohonásobné korelace mezi jedním a zbývajících regresory);
  - *test založený na výběrovém koeficientu mnohonásobné korelace*, že regresor je nekorelovaný se zbývajících;
  - *test založený na výběrovém parciálním korelačním koeficientu*, že dva různé regresory jsou navzájem nekorelované s vyloučením vlivů zbývajících regresorů.
- (b) *Vynechání vysvětlujících proměnných způsobujících multikolinearitu*: Tento postup může někdy narušit finančně-ekonomickou interpretaci modelu.

- (c) *Transformace některých vysvětlujících proměnných*: Jedná se o úpravy vysvětlujících proměnných jako např. centrování odečtením výběrového průměru, normování centrovaných veličin, nahrazení dvojice silně korelovaných regresorů jejich poměrem aj.
- (d) *Rozšíření datového souboru*: Použití rozsáhlejšího souboru dat např. zvýšením frekvence pozorování, zkombinováním časových a průřezových dat apod.
- (e) *Použití apriorní informace*: Někdy jsou k dispozici dodatečné informace o konstruovaném modelu.
- (f) *Použití metody hlavních komponent nebo hřebenové regrese*: Neobjektivnější způsob, který umožní přejít k malému počtu takových lineárních kombinací původních regresorů, které jsou navzájem ortogonální, takže využívají co nejefektivněji původní informaci o vysvětlující proměnné.

Podrobněji o jednotlivých postupech a uvedených metodách viz např. [1, 3, 8, 16].

## 4.5 Kritéria pro výběr modelu

Existují objektivní kritéria a postupy pro výběr modelu, především pro stanovení optimálního výběru a počtu vysvětlujících proměnných. Patří sem i tzv. informační kritéria, která dostatečně penalizují narůstající počet regresorů. Optimální počet regresorů tak můžeme zjistit pomocí *kritéria AIC - Akaikeho*<sup>5</sup> *informačního kritéria (Akaike information criterion)*. Hodnoty těchto kritérií závisí na počtu regresorů  $p$  následovně

$$AIC(p) = \ln(s_p^2) + \frac{2p}{n}, \text{ kde} \quad (41)$$

$s_p^2 = \frac{n-p}{n} \hat{\sigma}^2$  je ML-odhad rozptylu reziduální složky (viz [3]), přičemž indexem je vyznačen odpovídající počet regresorů. Optimální počet regresorů se pak hledá

---

<sup>5</sup>Hiroto Akaikeho (1927-2009), japonský statistik

minimalizací daného kritéria přes  $p$ . Poprvé zveřejněná teorie o AIC kritériu je obsažena v článku [61].

V případě, že máme k dispozici soubor potenciaálních regresorů, je možné pro získání nejlepšího tvaru regresního modelu (2) postupovat iteračně pomocí  $t$ -poměrů (9). Metoda zv. *kroková regrese (stepwise regression)* [3, 21] postupuje tak, že se začne s nejmenším modelem obsahujícím pouze absolutní člen. Při každém přidaném regresoru se potom pomocí testu významnosti regresních parametrů vyzkouší, zda nelze některý ze stávajících regresorů vyřadit.

Mezi tzv. iterační selekční metody patří kromě krokové regrese také *metoda eliminace zpět (backward elimination)* [3]. V tomto případě se začíná naopak s největším modelem obsahujícím všech  $p$  potenciaálních regresorů a vyřadíme z nich ten, který má nejmenší  $t$ -poměr. To opakujeme tak dlouho, až každý z dosud nevyřazených regresorů je významný. Další selekční metody jsou popsány v [3].

Více o ekonometrických zobecněních lineární regrese lze dohledat např. v literaturách [3, 5, 6, 8, 9, 23].

## 5 Model simultánních rovnic

V předchozích kapitolách jsme se zabývali závislostí jediné vysvětlované proměnné na konečné množině nestochastických vysvětlujících proměnných a na náhodné složce. V ekonomické praxi však chování některých makro- i mikroekonomických veličin nevysvětlíme v rámci jedné rovnice, ale pouze soustavou vzájemně závislých vztahů. V tomto případě je typické, že proměnné vystupují v některých rovnicích jako vysvětlující, v jiných jako vysvětlované. To znamená, že některá z endogenních proměnných může být kromě exogenních proměnných určena i některou z ostatních endogenních proměnných, v modelu tedy mají některé vysvětlující proměnné stochastický charakter. Je-li několik endogenních proměnných (jak vysvětlované tak i vysvětlující) současně determinovány soustavou lineárních či nelineárních vztahů, jedná se o *model simultánních rovnic* (MSR).

V rámci MSR obsahuje alespoň jedna rovnice více než jednu endogenní proměnnou, přičemž celkový počet endogenních proměnných je roven počtu lineárně nezávislých simultánních rovnic. Další rozdíl od jednorovnicových modelů spočívá v tom, že v MSR nelze odhadnout parametry jednotlivých rovnic bez ohledu na informace obsažené v ostatních rovnicích modelu. V MSR tedy nelze použít klasickou MNČ, neboť není splněna jedna ze základních podmínek pro tuto aplikaci: nezávislost všech vysvětlujících proměnných na náhodné složce. MNČ potom neposkytuje nestranné a konzistentní odhady. Podrobněji se o MSR lze dočíst např. v [3, 5, 8, 9, 10, 11].

Obecně se v MSR rozlišují dva typy kauzálních vztahů, proto lze rozlišovat

- *interdependentní soustavy simultánních rovnic*, kde mezi endogenními proměnnými existují přímé či nepřímé zpětné vazby,
- *rekurzivní systémy*, v nichž se mezi endogenními proměnnými vyskytují pouze jednostranné (příčinné) vazby.

V následujícím textu se prozatím budeme zabývat pouze prvním typem MSR.

Interdependentní soustava simultánních rovnic, obsahující vzájemně závislé vztahy, může být zapsána jako

$$y_{tj} = \sum_{i=1, i \neq j}^q \gamma_{ij} y_{ti} + \sum_{k=0}^p \beta_{kj} x_{tk} + \varepsilon_{tj}, \quad j = 1, \dots, q, \quad t = 1, \dots, n. \quad (42)$$

Soustava obsahuje rovnici pro každou z  $q$  endogenních proměnných a nazývá se simultánní jen v případě, že  $\gamma_{ij} \neq 0$  pro nějaké  $i \neq j$ . Kdyby všechny parametry  $\gamma_{ij}$  byly nulové, pak by se (42) zredukovala na SUR soustavu (zdánlivě nesouvisející rovnice), viz [3].

**Poznámka 5.1** *Proměnná  $y_{tj}$  je endogenní proměnnou, pokud alespoň v jedné rovnici dané soustavy figuruje jako vysvětlovaná, v ostatních rovnicích ale může klidně být vysvětlující. Exogenní proměnná může být pouze vysvětlující, v žádné rovnici dané soustavy nesmí vystupovat jako vysvětlovaná. Protože ale v MSR mohou být vysvětlujícími proměnnými i endogenní proměnné, rozlišují se ještě tzv. predeterminované proměnné, tj. vysvětlující proměnné nekorelované na náhodných složkách jednotlivých rovnic.*

## 5.1 Maticové vyjádření MSR

Obsahuje-li lineární MSR (interdependentního typu) při  $n$  pozorováních celkem  $q$  endogenních proměnných  $\mathbf{y}_1, \dots, \mathbf{y}_q$ , dále  $p$  predeterminovaných proměnných  $\mathbf{x}_1, \dots, \mathbf{x}_p$  a  $q$  náhodných složek  $\varepsilon_1, \dots, \varepsilon_q$ , jeho *strukturní tvar* lze získat maticovým vyjádřením soustavy rovnic (42) jako

$$\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B} = \mathbf{E}, \quad (43)$$

kde  $\mathbf{Y} = (y_{tj})$  je  $n \times q$  matice endogenních proměnných,  $t = 1, \dots, n$ ,  
 $j = 1, \dots, q$ ,

$\mathbf{\Gamma} = (\gamma_{ij})$  je  $q \times q$  regulární matice strukturních parametrů endogenních proměnných,  $i, j = 1, \dots, q$ ,

$\mathbf{X} = (x_{tk})$  je  $n \times p$  matice predeterminovaných proměnných,  $t = 1, \dots, n$ ,

$$k = 0, \dots, p,$$

$\mathbf{B} = (\beta_{kj})$  je  $p \times q$  matice strukturních parametrů predeterminovaných proměnných,  $k = 0, \dots, p$ ,  $j = 1, \dots, q$ ,

$\mathbf{E} = (\varepsilon_{tj})$  je  $n \times q$  matice náhodných složek strukturního tvaru.

Stochastické předpoklady jsou obdobné jako v klasickém lineárním regresním modelu a pro  $t = 1, \dots, n$  platí

$$\mathbf{E}(\varepsilon_t) = \mathbf{0},$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t'}) = \mathbf{0}, \quad t \neq t',$$

$$\varepsilon_t \sim N(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma}_{q \times q} \text{ p. d. kovarianční matice.}$$

Pokud se ve strukturních rovnicích vyskytují intercepty, přisoudíme dále v každé strukturní rovnici pomocí *normovacího pravidla* jednomu z nenulových strukturních parametrů endogenních proměnných konkrétní hodnotu, zpravidla rovnu jedné.

Za předpokladu regularity čtvercové matice strukturních parametrů  $\mathbf{\Gamma}$  lze řešením celého systému (soustavy) strukturních rovnic (43) dospět k *redukovanému tvaru* v maticové podobě

$$\mathbf{Y} = \mathbf{X}\mathbf{\Pi} + \mathbf{F}, \quad \text{kde} \quad (44)$$

$$\mathbf{\Pi}_{p \times q} = -\mathbf{B}\mathbf{\Gamma}^{-1}, \quad \mathbf{F}_{n \times q} = \mathbf{E}\mathbf{\Gamma}^{-1}. \quad (45)$$

Výraz (44) se označuje jako *neomezený redukovaný tvar*, na rozdíl od *omezeného redukovaného tvaru*, vyjádřeného jako

$$\mathbf{Y} = -\mathbf{X}\mathbf{B}\mathbf{\Gamma}^{-1} + \mathbf{E}\mathbf{\Gamma}^{-1}. \quad (46)$$

Zatímco strukturní rovnice obsahuje zpravidla více endogenních proměnných, v rovnici redukovaného tvaru můžeme na endogenní proměnné pohlížet jako na výstupy dané soustavy simultánních rovnic. V tomto případě je vysvětlena pouze jedna endogenní proměnná a to jako lineární funkce všech predeterminovaných proměnných a všech náhodných složek modelu. Pro strukturu nezávislých chyb  $\mathbf{f}_t = \varepsilon_t \mathbf{\Gamma}^{-1}$  obdobně platí

$$\mathbf{f}_t \sim N(\mathbf{0}, \mathbf{\Gamma}^{-1} \mathbf{\Sigma} (\mathbf{\Gamma}^{-1})'). \quad (47)$$

Prvky matice parametrů redukovaného tvaru  $\mathbf{\Pi} = (\Pi_{jk})$  jsou obecně nelineárními funkcemi strukturních parametrů a lze je vyjádřit pomocí parciální derivace jako

$$\frac{\partial y_{tj}}{\partial x_{tk}} = \Pi_{jk}, \quad j = 1, \dots, q, \quad k = 1, \dots, p.$$

Protože redukovaný tvar MSR obsahuje jako vysvětlující proměnné pouze preterminované proměnné, o kterých předpokládáme, že jsou nezávislé na náhodných složkách jednotlivých rovnic, lze konzistentní odhady parametrů redukovaného tvaru, tj. matice  $\mathbf{\Pi}$ , získat i klasickou MNČ.

Z dynamického MSR, který obsahuje i zpožděné endogenní proměnné, lze odvodit *konečný tvar*, vyjadřující jednotlivé nezpožděné endogenní proměnné jako funkce jejich hodnot ve výchozím období, běžných a zpožděných hodnot exogenních proměnných a náhodných složek. Více k této problematice např. viz [8].

## 5.2 Identifikace strukturního tvaru MSR

Protože redukovaný tvar MSR lze konzistentně odhadnout klasickou MNČ, nabízí se zde otázka, zda je možno získat z odhadů parametrů redukovaného tvaru modelu odhady strukturních parametrů tohoto modelu. V případě existence matice  $\mathbf{\Gamma}^{-1}$  lze vždy z matic parametrů strukturního tvaru určit matice parametrů redukovaného tvaru pomocí (45) a (47). Zda je však možné jednoznačně vyjádřit odhady strukturních parametrů modelu z odhadů parametrů jeho redukovaného tvaru, to závisí právě na jejich identifikaci.

Rozlišují se strukturní rovnice

— *identifikované*

a) *přesně identifikované* – všechny strukturní parametry modelu lze jednoznačně určit z parametrů neomezeného redukovaného tvaru (44)

b) *předidentifikované* – je-li počet neomezených koeficientů větší než počet strukturních koeficientů, neexistuje jediné řešení

— *podidentifikované* (neidentifikované) – nepostačuje-li počet neomezených koeficientů k určení všech koeficientů strukturního tvaru.



Pokud je některá ze strukturních rovnic podidentifikovaná, lze dosáhnout identifikace této rovnice tím, že do ostatních rovnic modelu zahrneme jednu nebo více zpravidla predeterminovaných proměnných, které se v podidentifikované rovnici nevyskytují.

V praktické ekonometrické analýze se při ověřování identifikovatelnosti simultánních strukturních rovnic používají tzv. *kritéria identifikace*, která umožňují snadně a rychle určit příslušnou identifikaci.

Nutnou a zároveň postačující podmínkou identifikace normované simultánní strukturní rovnice je tzv. *hodnostní podmínka identifikace*. V modelu s  $q$  endogenními proměnnými a  $p$  predeterminovanými proměnnými tato podmínka vyžaduje, aby hodnost matice vytvořené ze strukturních parametrů těchto dvou typů proměnných, které se nevyskytují ve zkoumané rovnici, ale v ostatních rovnicích modelu, byla rovna  $q - 1$ . V praxi se postupuje tak, že vytvoříme matici o rozměrech  $(q - 1) \times (q - 1)$  z parametrů endogenních i predeterminovaných proměnných, které se nevyskytují ve zkoumané rovnici a pokud je alespoň jeden její determinant nenulový, hodnost matice je rovna  $q - 1$  a strukturní rovnice je identifikovaná. V případě, že matice vytvořit nelze a nebo je její determinant nulový, rovnice je podidentifikovaná.

*Řádová podmínka identifikace* je nutnou, nikoliv však postačující podmínkou, která po normovaných strukturních rovnicích vyžaduje, aby

$$p - p_j \geq q_j - 1 \quad j = 1, \dots, q. \quad (48)$$

Tj.  $p$  je celkový počet predeterminovaných proměnných v modelu a  $q_j$ , resp.  $p_j$  je počet endogenních proměnných, resp. predeterminovaných proměnných ve zkoumané  $j$ -té normované strukturní rovnici. Pokud v (48) platí opačná nerovnost, strukturní rovnice není identifikovatelná, pokud rovnost, jedná se o přesnou identifikaci a pokud  $p - p_j > q_j - 1$ , rovnice vyhovuje nutné podmínce přeidentifikace.

V praxi se potom postupuje tak, že nejprve ověříme hodnostní a poté řádové kritérium identifikace. Při hodnostní podmínkou zjištěné neidentifikaci se musí před odhadem provést vhodným způsobem nulová omezení parametrů (tj. doda-

tečné přidání jedné či více predeterminovaných proměnných do modelu) s cílem dosáhnout jejich identifikace. Zjištění, zda strukturní rovnice je přesně identifikovaná nebo předidentifikovaná, má význam při volbě odhadové funkce, protože některé metody odhadu nelze při předidentifikaci použít.

### 5.3 Rekurzivní systémy

Rekurzivní systémy jsou zvláštním typem MSR, které neobsahují žádné zpětné vazby mezi endogenními proměnnými ani vzájemně závislé náhodné složky. Endogenní proměnné strukturního tvaru rekurzivního modelu jsou hierarchicky uspořádány, matice strukturních parametrů endogenních proměnných  $\mathbf{\Gamma}$  je tedy trojúhelníková a kovarianční matice náhodných složek  $\mathbf{\Sigma}$  je diagonální (nulové kovariance). Simultánní role v rekurzivním systému spočívá v tom, že žádná strukturní rovnice neobsahuje vysvětlované endogenní proměnné, které jsou zahrnuty v následujících rovnicích. Po normování diagonálních prvků trojúhelníkové matice  $\mathbf{\Gamma}$  má rekurzivní soustava pro  $q$  endogenních proměnných a  $p$  predeterminovaných proměnných obecný tvar

$$\begin{aligned}
 y_{t1} &= \sum_{k=0}^p \beta_{k1} x_{tk} + \varepsilon_{t1}, \\
 y_{t2} &= \gamma_{12} y_{t1} + \sum_{k=0}^p \beta_{k2} x_{tk} + \varepsilon_{t2}, \\
 &\vdots \\
 y_{tq} &= \sum_{i=1}^{q-1} \gamma_{iq} y_{ti} + \sum_{k=0}^p \beta_{kq} x_{tk} + \varepsilon_{tq}.
 \end{aligned} \tag{49}$$

Z (49) vidíme, že k řešení libovolné  $j$ -té endogenní proměnné, vyjádřené  $j$ -tou rovnicí, je zapotřebí pouze prvních  $j$  strukturních rovnic rekurzivní soustavy ( $j = 1, \dots, q$ ).

Rekurzivní model je vždy přesně identifikován, neboť z koeficientů redukovaného tvaru lze jednoznačně určit všechny nenulové strukturní parametry. Protože

příslušné vysvětlující endogenní proměnné v  $j$ -té rovnici jsou nekorelované na náhodné složce této rovnice, je splněna podmínka nekorelovanosti náhodných složek rovnic a pro odhad parametrů lze použít klasickou MNČ. Výsledkem aplikace MNČ jsou konzistentní a asymptoticky eficientní odhady parametrů.

## 5.4 Dvoustupňová MNČ

Z kap. 5.2 víme, že podmínkou odhadu strukturních parametrů MSR je identifikovatelnost všech jeho stochastických strukturních rovnic. S výjimkou strukturních rovnic rekurzivních systémů nelze pro odhad parametrů použít klasickou MNČ. V případě interdependentního lineárního MSR mají odhady strukturních parametrů charakter *metod omezené informace*, umožňující odhad každé strukturní rovnice zvlášť. Tyto metody využívají informace o omezeních strukturních parametrů dané rovnice, nikoliv už ostatních rovnic systémů. Druhou možností jsou *metody úplné informace*, které odhadují parametry všech rovnic najednou. Tady berou metody v úvahu všechna apriorní omezení MSR a využívají tak veškerou informaci obsaženou ve výběru pozorování.

V praktické ekonometrické analýze se nejčastěji k odhadu strukturních parametrů interdependentního MSR používá *metoda dvoustupňových nejmenších čtverců* (M2NČ), která je použitelná jak pro přesně identifikované tak i pro přeidentifikované rovnice. Její postup spočívá v opakované aplikaci klasické MNČ, přičemž se prvně odhadnou vyrovnané hodnoty vysvětlujících endogenních proměnných rovnice na základě neomezeného redukovaného tvaru (44), poté dochází k vlastnímu odhadu strukturních parametrů příslušné rovnice. Podstatou M2NČ je tedy nahrazení stochastických vysvětlujících endogenních proměnných jejich vyrovnanými nestochastickými hodnotami, které již nejsou korelované s náhodnou složkou rovnice a podmínka pro aplikaci MNČ je splněna.

Uvažujme  $j$ -tou strukturní rovnici normovaného MSR s  $q$  endogenními a  $p$  predeterminedovanými proměnnými

$$\mathbf{y}_t = \boldsymbol{\gamma}_t \mathbf{Y}_t + \boldsymbol{\beta}_t \mathbf{X}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, n, \quad (50)$$

kde  $\mathbf{y}_t$  je vektor pozorování vysvětlované endogenní proměnné,  $\mathbf{Y}_t$  je matice pozorování  $q - 1$  vysvětlujících endogenních proměnných a  $\mathbf{X}_t$  je matice pozorování  $p$  predeterminovaných vysvětlujících proměnných v příslušné rovnici. Vektor náhodných složek je označen jako  $\boldsymbol{\varepsilon}_t$  a  $\boldsymbol{\gamma}_t$ , resp.  $\boldsymbol{\beta}_t$  je vektor strukturních parametrů  $q - 1$  vysvětlujících endogenních proměnných, resp. vektor strukturních parametrů  $p$  predeterminovaných proměnných v  $j$ -té rovnici.

Postupujeme tak, že z neomezeného redukováného tvaru

$$\mathbf{Y}_t = \mathbf{X}\boldsymbol{\Pi}_t + \mathbf{F}_t \quad (51)$$

získáme pomocí MNČ vyrovnané hodnoty vysvětlujících endogenních proměnných jako

$$\hat{\mathbf{Y}}_t = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_t. \quad (52)$$

Substitucí vyrovnaných nestochastických hodnot  $\hat{\mathbf{Y}}_t$  za  $\mathbf{Y}_t$  v (50) získáme regresi, kterou můžeme označit jako

$$\mathbf{y}_t = \hat{\mathbf{Z}}_t\boldsymbol{\delta}_t + \boldsymbol{\varepsilon}_t, \quad (53)$$

kde

$$\hat{\mathbf{Z}}_t = [\hat{\mathbf{Y}}_t, \mathbf{X}_t]. \quad (54)$$

Protože jsou již všechny vysvětlující proměnné nekorelované na náhodné složce, lze získat odhady strukturních parametrů rovnice (50) aplikací MNČ na (53) následovně

$$\hat{\boldsymbol{\delta}}_t = (\hat{\mathbf{Z}}_t'\hat{\mathbf{Z}}_t)^{-1}\hat{\mathbf{Z}}_t'\mathbf{y}_t. \quad (55)$$

Dostáváme vyrovnané hodnoty vysvětlované endogenní proměnné jako

$$\hat{\mathbf{y}}_t = \hat{\mathbf{Z}}_t\hat{\boldsymbol{\delta}}_t. \quad (56)$$

Odhadová funkce M2NČ (55) poskytuje nikoliv nestranné, ale konzistentní odhady strukturních parametrů, které jsou i asymptoticky vydatné. MNČ sice nedává konzistentní odhady MSR, ale pokud se index determinace v odhadnutých rovnicích blíží jedné, vyrovnané hodnoty vysvětlujících endogenních proměnných

jsou velmi dobrými „náhražkami“ jejich skutečných hodnot. Mezi další metody omezené informace patří *metoda nepřímých nejmenších čtverců*, použitelná pouze pro přesně identifikované strukturní rovnice a dále *metoda maximální věrohodnosti s omezenou informací*, která se aplikuje při podmínce normálně rozdělených náhodných složek, navíc sériově nezávislých. Obě metody poskytují stejné vlastnosti odhadů strukturních parametrů jako M2NČ. Více o těchto metodách lze nalézt např. v [3, 8, 9].

Simultánní odhadové postupy jsou ve srovnání s metodami omezené informace výpočetně složitější a poskytují rovněž konzistentní odhady, ale při splnění určitých podmínek jsou asymptoticky vydatnější. Patří sem *metoda třístupňových nejmenších čtverců* či *metoda maximální věrohodnosti s úplnou informací*, více viz [3, 9].

## 6 Ekonometrické prognózování

Jedním z hlavních cílů ekonometrického modelování je *prognózování* (také *předpověď* nebo *predikce*) hodnot vysvětlovaných endogenních proměnných mimo interval pozorování. Předpověď je kvantitativním odhadem pravděpodobnosti budoucí hodnoty konkrétní ekonomické veličiny pomocí minulé i současné informace, která je reprezentována odhadnutým ekonometrickým modelem.

Hovoříme o *předpovědi bodové*, spočívající v odhadu jedné budoucí hodnoty predikované proměnné, nebo o *předpovědi intervalové*, která je obdobou intervalu spolehlivosti odhadu, obsahujícího skutečnou hodnotu predikované proměnné s předem požadovanou pravděpodobností. Platí, že předpověď vysvětlovaných veličin nezískáme zcela přesně, ale s určitou chybou. Odchylka této předpovědi od skutečné hodnoty predikované endogenní proměnné představuje *chybu předpovědi*. Její velikost a variabilita může být ovlivněna náhodným charakterem modelu, náhodnou chybou nebo chybnou specifikací odhadnutého modelu v období predikce kvůli jeho nestabilitě v čase. Předpovědi tedy mají stochastický charakter. Cílem ale je, aby chyba předpovědi měla např. minimální rozptyl. Při použití lineárního regresního modelu nebo rekurzivních MSR lze získat optimální nevychýlené předpovědi. Přesnost ekonometrických předpovědí je určována statistickými a ekonometrickými testy. Často se používají i simulační postupy, více o této problematice viz [12].

Pokud chceme dospět k optimálním vlastnostem předpovědí, musí platit, že odhadnutý model je v čase stabilní, tzn. že jeho specifikace, charakter rozdělení náhodné složky i odhadnuté parametry zůstávají v platnosti i v období predikce. Splnění těchto požadavků se před prognostikou testuje pomocí chyby a přesnosti předpovědí.

Vycházíme z KMLR (2) a z jeho nevychýleného MNČ-odhadu (4). Máme-li k dispozici nová pozorování nebo odhady v období předpovědi, např. v  $t + 1$ , lze stanovit bodovou předpověď konkrétní hodnoty vysvětlované proměnné v období

$t + 1$  pomocí funkce *předpovědi ex ante* jako

$$\hat{y}_{t+1} = \mathbf{x}'_{t+1} \hat{\boldsymbol{\beta}}, \quad (57)$$

kde  $\hat{y}_{t+1}$  je skalár, neboť  $\mathbf{x}'_{t+1}$  je  $1 \times p$  řádkový vektor. Obsahuje-li KMLR všechny vysvětlující proměnné zpožděné o jedno období, pak vektor  $\mathbf{x}'_{t+1}$  tvoří skutečná pozorování. Nejsou-li však některé nebo všechny vysvětlující proměnné takto posunuté v čase, obsahuje tento vektor také předpovědi nebo odhady, takže *ex ante* předpověď  $\hat{y}_{t+1}$  je *podmíněná* vzhledem k  $\mathbf{x}'_{t+1}$ , což je nejčastější případ. Pro libovolné období předpovědi  $P$  můžeme přepsat funkci (57) jako

$$\hat{y}_P = \mathbf{x}'_P \hat{\boldsymbol{\beta}}, \quad (58)$$

kde  $\hat{y}_P$  je podmíněná bodová předpověď vysvětlované proměnné v období  $P$ .

Označíme-li skutečnou hodnotu  $y$  v období predikce jako  $y_P$ , pak *chyba ex ante předpovědi individuální hodnoty* vysvětlované proměnné je dána vztahem

$$\hat{\varepsilon}_P = \hat{y}_P - y_P = \mathbf{x}'_P \hat{\boldsymbol{\beta}} - \mathbf{x}'_P \boldsymbol{\beta} - \varepsilon_P, \quad (59)$$

neboť  $y_P = \mathbf{x}'_P \boldsymbol{\beta} + \varepsilon_P$ . Podmíněná předpověď hodnoty vysvětlované proměnné  $y_P$  je nevychýlená v tom smyslu, že  $E(\hat{y}_P) = E(y_P)$ . Pro rozptyl chyby předpovědi hodnoty  $\hat{y}_P$  potom platí

$$\hat{\sigma}_P^2 = E(\hat{\varepsilon}_P^2) = \sigma^2 [1 + \mathbf{x}'_P (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_P] = \sigma^2 + \mathbf{x}'_P \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_P, \quad (60)$$

kde  $\sigma^2$  je rozptyl náhodné složky KMLR a  $\text{var}(\hat{\boldsymbol{\beta}})$  je kovarianční matice odhadové funkce  $\hat{\boldsymbol{\beta}}$ . Pokud rozptyl  $\sigma^2$  neznáme, nahradíme ho jeho nestranným MNČ-odhadem  $\hat{\sigma}^2$  a poté lze vzorcem

$$\hat{\hat{\sigma}}_P = \hat{\sigma} \sqrt{1 + \mathbf{x}'_P (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_P} \quad (61)$$

vyjádřit *odhad standardní chyby bodové předpovědi individuální hodnoty* vysvětlované proměnné  $y_P$ .

Jsou-li náhodné složky KMLR normálně rozděleny, má chyba předpovědi

i funkce předpovědi  $\hat{y}_P$  také normální rozdělení a pro *standardizovanou chybu předpovědi* platí

$$\frac{\hat{y}_P - y_P}{\hat{\sigma}_P} \sim N(0, 1). \quad (62)$$

V případě nahrazení neznámé hodnoty  $\hat{\sigma}_P$  jejím nevychýleným odhadem  $\hat{\hat{\sigma}}_P$  získáme statistiku Studentova rozdělení, viz [8].

Odvození předchozích vzorců nebo např. intervaly spolehlivosti předpovědí lze nalézt v [8, 9, 10, 11, 23].

Východiskem při ekonometrickém prognózování na základě MSR je redukováný nebo konečný tvar modelu, strukturní tvar MSR není k předpovědím vhodný. Podrobná teorie k této oblasti lze vyhledat v literaturách [8-12]. Obecně mají ekonometrické předpovědi na základě KMLR a MSR určité výhody oproti předpovědím získaným jinými kvantitativními, ale i kvalitativními metodami. Hlavní předností ekonometrických předpovědí je schopnost určit jejich přesnost, i když jen přibližnou. Další výhodou je, že ekonometrické predikce umožňují brát v úvahu například minulé hodnoty predikovaných vysvětlovaných proměnných, minulé i očekávané hodnoty vysvětlujících proměnných, nebo i stochastické vlivy.



## 7 Příklad faktorového modelu (modelování vývoje cen akcií)

Nejdříve stručně o faktorových modelech. Budu se zabývat cennými papíry, jejichž výnosnost (cena) je citlivá na chování určitých faktorů. Pro zjišťování této závislosti se konstruují tzv. *faktorové modely*. Výnosnost portfolia (skladby cenných papírů) neovlivňuje pouze pohyb tržního portfolia (každá z investic na kapitálovém trhu se svou tržní hodnotou), ale i další ekonomické či jiné faktory (vysvětlující proměnné). Mým cílem tedy bude zjistit tyto faktory a jejich závislost na ceně (kurzu) cenných papírů a dále potom predikovat ceny těchto cenných papírů na následující období, např. pro možnost investice.

Nejdříve musím vytvořit správný model lineární regrese. Faktorů ovlivňujících cenu akcie je celá řada. Já jsem za nejvýznamnější zvolila faktor inflace, HDP a nezaměstnanosti a jako další možné vysvětlující proměnné jsem přidala kurzy měn CZK/USD a CZK/EUR. Závislost těchto faktorů budu zkoumat na cenách akcií 8 společností. Pro každou z osmi akcií budu uvažovat lineární regresní model ve tvaru

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + \varepsilon_t, \quad t = 1, \dots, 23, \quad (63)$$

kde  $y_t$ ...cena akcie společnosti v čase  $t$ ,

$x_{t1}$ ...míra inflace (vyj. přírůstkem indexu spotřebitelských cen ke stejnému měsíci předchozího roku) v čase  $t$ ,

$x_{t2}$ ...meziroční růst HDP v čase  $t$ ,

$x_{t3}$ ...obecná míra nezaměstnanosti v čase  $t$ ,

$x_{t4}$ ...kurz měny CZK/USD v čase  $t$ ,

$x_{t5}$ ...kurz měny CZK/EUR v čase  $t$ .

Hodnoty faktorů  $x_{t1}$  a  $x_{t3}$  jsou uvedeny jako průměrné hodnoty těchto vysvětlujících proměnných za předchozí tři měsíce, zdroje dat postupně viz [27] a [29]. Hodnoty  $x_{t2}$  jsou dány přímo jako data čtvrtletní, viz [28]. Kurzy měn jsou

určeny ke konkrétnímu datu uvedenému v tabulce, zdroje viz [30]. Ve zvolené úloze tedy použijeme čtvrtletní data z tabulky 1.

Datum	inflace (%)	HDP (%)	nezam. (%)	CZK/USD	CZK/EUR
1. 4. 05	1,6	4,4	8,4	23,173	30,04
1. 7. 05	1,6	5,1	7,8	24,817	30
1. 10. 05	1,9	4,9	7,8	24,537	29,55
1. 1. 06	2,4	6,9	7,8	24,588	29,005
1. 4. 06	2,8	7,4	8	23,623	28,595
1. 7. 06	2,9	6,2	7,1	22,413	28,495
1. 10. 06	2,9	5,8	7,1	22,375	28,33
1. 1. 07	1,5	5,8	6,6	20,876	27,495
1. 4. 07	1,6	6,1	6,1	21,023	28
1. 7. 07	2,5	6	5,3	21,258	28,715
1. 10. 07	2,5	6	5,2	19,353	27,54
1. 1. 08	4,8	6,6	4,9	18,078	26,62
1. 4. 08	7,4	5,3	4,7	16,084	25,185
1. 7. 08	6,8	4,6	4,3	15,102	23,825
1. 10. 08	6,7	4,2	4,3	17,41	24,51
1. 1. 09	4,7	0,7	4,4	19,346	26,93
1. 4. 09	2,2	-3,4	5,8	20,489	27,15
1. 7. 09	1,4	-5,5	6,4	18,295	25,775
1. 10. 09	0,2	-4,1	7,4	17,473	25,41
1. 1. 10	0,4	-3,1	7,3	18,368	26,465
1. 4. 10	0,7	1,1	8,2	18,84	25,38
1. 7. 10	1,2	2,4	7,2	20,96	25,76
1. 10. 10	1,9	3,0	7,2	17,8	24,43

Tabulka 1: Hodnoty faktorů od 1. 4. 2005 do 1. 10. 2010

V následující tabulce jsou uvedeny ceny akcií jednotlivých společností v korunách. Zdroje dat viz [32].

Datum	ČEZ	TELEFÓNICA O2	UNIPETROL	PHILIP MORRIS
1. 4. 05	406,4	442,9	153,1	18786,3
1. 7. 05	474,8	460,2	140,74	17456
1. 10. 05	739,3	491,6	238,6	18951
1. 1. 06	736,3	257	234,3	18166
1. 4. 06	819,2	501,3	274,8	16072
1. 7. 06	751,7	478,7	198,8	12285
1. 10. 06	790,8	442	194,02	9725
1. 1. 07	991	479,7	238,7	11036
1. 4. 07	935,6	548,3	235,5	9735
1. 7. 07	1096	594,3	286,2	11026
1. 10. 07	1224	535,2	315,2	9725
1. 1. 08	1373	549	329,1	7855
1. 4. 08	1234	513,3	259,1	7125
1. 7. 08	1365	482,7	244,5	4418
1. 10. 08	1107	413,6	192,25	5426
1. 1. 09	804	425,5	146	6426
1. 4. 09	734,8	397,5	116,05	5673
1. 7. 09	835	424	113	6348
1. 10. 09	938	434	138,2	8800
1. 1. 10	864	418	139,5	8796
1. 4. 10	900	443	170	9890
1. 7. 10	865	418	190	8401
1. 10. 10	812	382	205	9320

  

Datum	ORCO	KOMERČNÍ BANKA	CETV	ERSTE GROUP BANK
1. 4. 05	1333,6	3335,1	1193,5	1213,1
1. 7. 05	1403	3164	1213	1252
1. 10. 05	1714	3586	1287	1305
1. 1. 06	1818	3456	1413	1399
1. 4. 06	2500	3285	1614	1389
1. 7. 06	2389	3262	1394	1268
1. 10. 06	2791	3311	1504	1382
1. 1. 07	2755	3119	1474	1638
1. 4. 07	3525	3586	1859	1636
1. 7. 07	3393	3977	2060	1660
1. 10. 07	2931	4478	1812	1492
1. 1. 08	2163	4354	2070	1283
1. 4. 08	1452	3935	1412	1045
1. 7. 08	881,8	3383	1320	907
1. 10. 08	407,1	4036	1148	850
1. 1. 09	175,75	3054	408,5	421
1. 4. 09	103,24	2144	242,6	341,5
1. 7. 09	150,5	2595	364	497
1. 10. 09	194	3486	586	735,5
1. 1. 10	170,34	3929	446,9	698,5
1. 4. 10	189	3940	568,9	794,8
1. 7. 10	122	3492	436	689
1. 10. 10	191	3928	448	716

Tabulka 2: Ceny akcií od 1. 4. 2005 do 1. 10. 2010

Většinu výpočtů budu realizovat v programu MATLAB, jednotlivé m-fily se objeví v příloze na webových stránkách [63]. Ze všeho nejdříve provedu na lineárním regresním modelu (63) odhad regresních parametrů a příslušných charakteristik pomocí MNČ dle (4), (5), (6), (11) a pomocí (15), viz příloha 1 [63]. Výsledky pro jednotlivé akcie jsou uvedeny v následující tabulce.

Název akcie	RSČ	$\hat{\sigma}$	index determinace $R^2$
ČEZ	242324,01	119,39	0,82
TELEFÓNICA O2	48231,41	53,27	0,55
UNIPETROL	19752,63	34,09	0,77
PHILIP MORRIS	43581508,19	1601,13	0,90
ORCO	5088195,48	547,09	0,84
KOMERČNÍ BANKA	2742770,31	401,67	0,56
CETV	1012202,35	244,01	0,86
ERSTE GROUP BANK	430515,70	159,14	0,88

Tabulka 3: Charakteristiky odhadnutého modelu jednotlivých akcií

Z tabulky vidíme, že indexy determinace jsou s dvěma výjimkami dostatečné. Rozptyly, resp. směrodatné odchylky nejsou příliš vysoké, ale je zde určitá rezerva pro jejich snížení. Model tedy není zvolen optimálně a regresní parametry tak nejsou dostatečně kvalitně odhadnuty (resp. mohly by být lepší). Na první pohled se nabízí existence multikolinearity, to znamená, že některé faktory v modelu budou zřejmě závislé. V následujícím tedy nejprve zjistím, zda je mé mínění správné a pokud bude potvrzeno, navrhuji způsoby jak tuto multikolinearitu odstranit či alespoň potlačit.

K tomu abych rozpoznala multikolinearitu, nám v první řadě poslouží tzv. *index podmíněnosti*  $\kappa$  matice  $\mathbf{X}'\mathbf{X}$  (35). Z výsledku  $\kappa \doteq 312,9$  usuzuji silnou multikolinearitu. To znamená, že některý regresor musím vypustit či některé regresory sloučit.

Multikolinearitu lze také rozpoznat pomocí faktu, že index determinace je sice vysoký, ale některé regresní parametry jsou nevýznamné. Testuje se nulová hypotéza  $H_0 : \beta_j = 0$  oproti alternativní hypotéze  $H_1 : \beta_j \neq 0$  *testu významnosti (nenulovosti) regresních parametrů* (8), resp. (9). Například pro akcii TELEFÓ-

NICA O2 platí

$$\beta_0 : \frac{|137,56 - 0|}{\sqrt{97685,9}} \doteq 0,44,$$

podobně potom pro  $\beta_1 : 1,69$ ,  $\beta_2 : 2,58$ ,  $\beta_3 : 1,16$ ,  $\beta_4 : 2,62$ ,  $\beta_5 : 2,51$ . Pro kritickou hodnotu Studentova rozdělení platí  $t_{n-p}(\frac{\alpha}{2}) = t_{17}(0,025) = 2,458$ , tudíž regresní parametry  $\beta_0$ ,  $\beta_1$  a  $\beta_3$  jsou statisticky nevýznamné a parametry  $\beta_2$ ,  $\beta_4$  a  $\beta_5$  statisticky významné. Pro akcii PHILIP MORRIS vyšly podle testu nevýznamné parametry  $\beta_1$ ,  $\beta_2$  a  $\beta_4$ . Podobně to vypadá u všech ostatních vysvětlovaných proměnných, pro které jsou vždy alespoň dva regresní parametry nevýznamné. Multikolinearita je opět potvrzena.

Následující metody mají dva výstupy. Jednak nám potvrdí přítomnost multikolinearity v modelu a také naznačí, které faktory tuto multikolinearitu způsobují (resp. které jsou navzájem lineárně závislé). Nejdříve využijí *metodu pomocných regresí* (38). Budu postupovat tak, že pro každou  $j$ -tou vysvětlující proměnnou a zbývajících  $k - 1$  proměnných vypočítám dílčí koeficienty vícenásobné determinace  $R_j^2$ ,  $j = 1, \dots, k$ . Dostávám

$$\begin{aligned} R_1^2 &= 0,7271, \\ R_2^2 &= 0,5181, \\ R_3^2 &= 0,7445, \\ R_4^2 &= 0,9018, \\ R_5^2 &= 0,9997. \end{aligned}$$

Protože koeficienty  $R_4^2$  a  $R_5^2$  jsou blízké jedné, usuzuji na silnou multikolinearitu. Tento závěr mohu potvrdit pomocí F statistiky (39). Protože  $F_1 \doteq 12,17$ ,  $F_2 \doteq 4,88$ ,  $F_3 \doteq 12,81$ ,  $F_4 \doteq 40,50$  i  $F_5 \doteq 14995,5$  jsou větší než  $F_{4,18}(0,05) = 2,93$ , zamítám nulovou hypotézu o nezávislosti regresorů. Všimněme si, že nejvyšší hodnoty dílčích koeficientů determinace se týkají posledních dvou faktorů - kurzů měn.

Aby bylo mé rozhodnutí ohledně závislých faktorů potvrzené, vyzkouším ještě postup s *výběrovou korelační maticí*. Ze vztahů (36) a (37) pro výběrovou korelační a kovarianční matici dostávám hodnoty jednotlivých výběrových koeficientů, které jsou znázorněny v tabulce 4.

výběrové kovariance $s_{jl}$	výběrové odchyly $s_j$	výběrové korelace $r_{jl}$
$s_{12} = 0,202$		$r_{12} = 0,366$
$s_{13} = -2,100$	$s_1 = 2,019$	$r_{13} = -0,765$
$s_{14} = -2,474$		$r_{14} = -0,439$
$s_{15} = -1,333$	$s_2 = 3,861$	$r_{15} = -0,359$
$s_{23} = -0,277$		$r_{23} = -0,053$
$s_{24} = 4,197$	$s_3 = 1,359$	$r_{24} = 0,390$
$s_{25} = 2,894$		$r_{25} = 0,408$
$s_{34} = 2,405$	$s_4 = 2,788$	$r_{34} = 0,634$
$s_{35} = 1,173$		$r_{35} = 0,469$
$s_{45} = 4,698$	$s_5 = 1,839$	$r_{45} = 0,916$

Tabulka 4: Hodnoty výběrových kovariančních a korelačních koeficientů

Z tabulky vidíme, že mé podezření o lineární závislosti mezi faktory CZK/USD a CZK/EUR bylo potvrzeno.

Nyní nastává otázka, co s modelem lineární regrese (63) se silnou multikolinearitou provést, abych dosáhla vhodného modelu pro co nejlepší odhad regresních parametrů. Z předešlého se jako nejrozumnější jeví odstranění některého z faktorů, nejspíše kurzu měny CZK/USD. Toto rozhodnutí by měly potvrdit metody *stepwise regression* a *backward elimination* (kap. 4.5). Napovědět nám může i kritérium AIC pro zjištění optimálního počtu regresorů v modelu.

Než přejdu ke zmíněným metodám, mohu ještě posoudit stupeň potvrzené multikolinearity. K tomu mi poslouží tzv. *Scottova testační charakteristika*. K výpočtu mohu využít výsledky testu nenulovosti regresních parametrů. Pro akcii TELEFÓNICA O2 podle vztahu (40) dostávám

$$M_T = \frac{4,81 - 1}{4,81 + 1} = 0,66$$

a pro akcii PHILIP MORRIS podle stejného vztahu

$$M_T = \frac{7,00 - 1}{7,00 + 1} = 0,75.$$

Podle podmínek tohoto testu je pro obě akcie model nevyhovující a je vhodná jeho úprava. Ostatní akcie mají výsledky stejné, u některých je úprava dokonce nutná.

Nyní se pokusím zjistit optimální počet regresorů v modelu pomocí kritéria AIC (absolutní člen v tomto případě do výpočtů zahrnovat nemusím). Např. pro akcii CETV ze vztahu (41) postupně v Matlabu dostávám

faktor č.	AIC(1)	RSČ	dvojice č.	AIC(2)	RSČ	trojice č.	AIC(3)	RSČ
1	16,05	0,10	(1,2)	14,95	0,70	(1,2,3)	14,69	0,77
			(1,3)	16,05	0,10	(1,2,4)	14,90	0,71
2	14,95	0,70	(1,4)	15,60	0,43	(1,2,5)	13,57	0,92
			(1,5)	15,05	0,67	(1,3,4)	15,56	0,45
3	16,11	0,04	(2,3)	14,86	0,73	(1,3,5)	14,66	0,78
			(2,4)	14,92	0,71	(1,4,5)	14,76	0,75
4	15,96	0,18	(2,5)	14,61	0,79	(2,3,4)	14,68	0,77
			(3,4)	15,66	0,39	(2,3,5)	14,46	0,82
5	15,51	0,48	(3,5)	14,94	0,70	(2,4,5)	14,36	0,83
			(4,5)	14,94	0,70	(3,4,5)	14,75	0,75

  

čtveřice č.	AIC(4)	RSČ	pětice	AIC(5)	RSČ
(1,2,3,4)	14,59	0,79			
(1,2,3,5)	13,56	0,93			
(1,2,4,5)	13,57	0,92	<b>(1,2,3,4,5)</b>	<b>13,56</b>	<b>0,93</b>
(1,3,4,5)	14,57	0,79			
(2,3,4,5)	14,32	0,84			

Tabulka 5: Hodnoty kritéria AIC pro různé kombinace faktorů

Vidíme, že nejmenší hodnota AIC kritéria je pro  $p = 5$ , tj. pro všechny regresory modelu. Optimální počet regresorů v modelu by tedy měl být spolu s absolutním členem 6. U ostatních akcií vychází stejný závěr, tabulky pro všechny akcie viz příloha 2 [63]. Na tento výsledek budu myslet při závěrečném sestavování nejlepšího tvaru regresního modelu.

Nyní se tedy pokusím použít metodu stepwise regression. Tato metoda je velice zdlouhavá, musím vyzkoušet všechny kombinace regresních parametrů  $\beta_0, \beta_1, \dots, \beta_5$ . Např. pro akcii ERSTE GROUP BANK začnu s tímto modelem

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2}$$

a vyzkouším, zda lze stávající regresor  $x_{t1}$  na základě (9) z modelu vyloučit jako

nevýznamný (resp. s nevýznamným regresním parametrem  $\beta_1$ ). Platí

$$\beta_1 : \frac{|-73,84|}{\sqrt{498,31}} \doteq 3,308.$$

Nutno dodat, že odhady parametrů se provádí znovu a znovu z aktuálních sestavených modelů, nikoliv z původního modelu o 6-ti neznámých regresních parametrech. Vidíme, že tato hodnota je větší než  $t_{20}(0,025) = 2,4231$ . Zamítám nulovou hypotézu o nevýznamnosti parametru  $\beta_1$ , to znamená, že regresor  $x_{t1}$  v modelu ponechám a přidám další

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3}.$$

Nyní budu vyšetřovat už dva parametry  $\beta_1$  a  $\beta_2$ , pro které platí

$$\beta_1 : \frac{|-125,34|}{\sqrt{1262,71}} \doteq 3,527 \text{ a}$$

$$\beta_2 : \frac{108,45}{\sqrt{143,68}} \doteq 9,048.$$

Obě hodnoty jsou větší než kritická hodnota  $t_{19}(0,025) = 2,4334$ , ponechám tedy oba regresory a přidám další. Stejně jsem postupovala pro všechny možné kombinace parametrů, jako např. pro model  $y_t = \beta_0 + \beta_3 x_{t3} + \beta_5 x_{t5}$  nebo  $y_t = \beta_0 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5}$  aj. Ze všech modelů se dostanu k jednomu výslednému a to

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_4 x_{t4} + \beta_5 x_{t5}.$$

Dále např. pro akcii KOMERČNÍ BANKY je výsledný model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_4 x_{t4}$$

a pro akcii UNIPETROL vychází

$$y_t = \beta_0 + \beta_2 x_{t2} + \beta_4 x_{t4}.$$

Abych mohla výsledky metody stepwise regression porovnat a vyvodit z nich důsledky, využiji ještě další z iteračních selekčních metod a to metody backward



elimination. Tato metoda je o poznání jednodušší. Tady naopak začínám s největším modelem, pro akcii ERSTE GROUP BANK tedy platí

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5},$$

kde jednotlivé  $t$ -poměry jsou:  $\beta_1 : 3,234$ ,  $\beta_2 : 8,241$ ,  $\beta_3 : 0,618$ ,  $\beta_4 : 2,589$  a  $\beta_5 : 3,012$ . Parametr  $\beta_3$  je jediný nevýznamný, proto z modelu odstraním regresor  $x_{t3}$ . Pokračuji s modelem

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_4 x_{t4} + \beta_5 x_{t5},$$

pro který platí  $\beta_1 : 3,769$ ,  $\beta_2 : 8,364$ ,  $\beta_4 : 3,474$  a  $\beta_5 : 3,596$ . Všechny parametry už jsou významné, proto je tento model i výsledným. U akcie KOMERČNÍ BANKY postupuji stejně. Zde je první nejmenší regresor  $x_{t3}$  a potom regresor  $x_{t5}$ , čili výsledný model pro tuto akcii je

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_4 x_{t4}.$$

A pro akcii UNIPETROL vyšel závěrečný model dle metody backward elimination následovně

$$y_t = \beta_0 + \beta_2 x_{t2} + \beta_4 x_{t4}.$$

Vidíme, že nám pro všechny tři akcie vyšly stejné výsledné modely jak z metody stepwise regression, tak i z metody backward elimination. Tyto modely by měly být ty nejvhodnější pro nejlepší odhady neznámých regresních parametrů. Pokud ale provedu odhady parametrů podle těchto modelů pro jednotlivé akcie, zjistím, že pro akcii EGB a KB se nepatrně snížila směrodatná odchylka, jinak jsou ostatní charakteristiky pro všechny tři akcie horší. Otázkou je, zda mi tyto výsledné modely vůbec k něčemu jsou a jestli pomohly k nalezení nejlepšího regresního modelu.

Shrneme si dosavadní výsledky. Podle metod stepwise regression a backward elimination se jako nejvhodnější jeví odstranit z modelu 3. faktor nezaměstnanost. Podle předchozích metod to vypadalo spíše na vyjmutí 4. faktoru CZK/USD. Pokud zkusím z původního modelu vyjmout faktor nezaměstnanosti a provedu

nový odhad regresních parametrů, charakteristiky pro modely dvou akcií zůstanou stejné, pro ostatní akcie jsou jen horší. Po odstranění faktoru CZK/USD dopadne výsledek velmi podobně. Z toho plyne, že ani jedna z možností se nejeví jako optimální, protože lepší model než původní jsem zatím nenašla.

Další možností je vyšetřit v modelu existenci autokorelovanosti reziduí, která by mi mohla být nápomocna k nalezení nejvhodnějšího modelu. Pro tyto účely použiji *Durbin-Watsonovu statistiku*  $d$ . Budu testovat nulovou hypotézu o lineární nezávislosti reziduí. Pro  $n = 23$  a  $p = 6$  zjistím z tabulek (viz příloha 3 [63]) horní a dolní mez, tj.  $d_D = 0,89$  a  $d_H = 1,92$ . Pro  $d$  statistiku podle vztahu (31) platí pro jednotlivé akcie

$$\begin{aligned} d_{\text{UNIP}} &= 1,311, & d_{\text{O2}} &= 2,680, \\ d_{\text{CETV}} &= 1,665, & d_{\text{ČEZ}} &= 1,230, \\ d_{\text{EGB}} &= 1,304, & d_{\text{KB}} &= 1,573, \\ d_{\text{ORCO}} &= 0,974, & d_{\text{PM}} &= 1,269, \end{aligned}$$

Vidíme, že všechny  $d$  statistiky patří do 4. typu intervalů, tzn. že ani pro jednu akcii není výsledek průkazný.

Protože jsem pomocí DW testu autokorelovanost reziduí v modelu nepotvrdila ani nevyvrátila, nabízí se zde možnost využít *modifikované DW statistiky*, která v sobě zahrnuje čtvrtletní data, tj. přímo můj případ. Vypočtu tedy modifikované statistiky (32) pro jednotlivé akcie. S výjimkou  $d_{\text{KB}} = 2,002$  (sériová nezávislost) skončily pro ostatní akcie výsledky testu opět neprůkazně.

*Breusch-Godfreyův test* by mohl konečně říci něco více. Nejprve musím sestavit pomocný model (34), kde pro všechna  $t = 1, \dots, n$  postupně vznikne modifikovaná matice  $\mathbf{X}$  a vektor odhadnutých reziduí, čili modifikovaný vektor  $\mathbf{y}$  pro odhad nových parametrů. Tak např. pro akcii PHILIP MORRIS a  $t = 1$  (ukázka v příloze 4 [63]) dostávám

$$774,7 = \gamma_0 + \gamma_1 \cdot 1,600 + \gamma_2 \cdot 4,400 + \dots + \gamma_5 \cdot 30,04 + \rho_1 \cdot 0 + \dots + \rho_5 \cdot 0$$

a dále stejným způsobem pro ostatní  $t$ . Potom  $R_{\text{PM}}^2 = 0,34$  a pro testovací

statistiku platí

$$(23 - 4) \cdot 0,34 = 6,46 < \chi_4^2(0,05) = 9,49.$$

Autokorelovaná rezidua tedy v modelu akcie PHILIP MORRIS nejsou. Pro akcii KOMERČNÍ BANKY je  $R_{KB}^2 = 0,35$  a testovací statistika  $19 \cdot 0,35 = 6,65 < 9,49$ , opět nulovou hypotézu o nekorelovanosti reziduí nezamítám. Stejně to dopadne pro všechny ostatní modely akcií. Autokorelovanost tedy v mém modelu není přítomna.

Poslední alternativou, kterou jsem ještě neprověřila, je možnost porušení konstantních rozptylů reziduálních složek, tzn. existence heteroskedasticity. Nejprve provedu *Spearmanův test korelace pořadí*. Dle (26) spočtu jednotlivé koeficienty korelace pořadí a dosadím do statistiky  $t$  (27). Např. pro model akcie CETV platí

$$t_1 = 0,493, \quad t_2 = 0,199, \quad t_3 = 0,534, \quad t_4 = 0,029, \quad t_5 = 0,155.$$

Tabelovaná kritická hodnota pro Spearmanův test je  $t_{17}(0,05) = 0,4853$ . Jelikož jsou některé hodnoty statistik větší a některé menší než kritická hodnota, nelze jednoznačně rozhodnout, zda jde o heteroskedasticitu či ne. Dále pak např. pro akcii UNIPETROL jsou všechny statistiky větší než kritická hodnota a je zde potvrzena existence heteroskedasticity. Pro akcii ORCO se hodnoty statistik opět střídají, pro PHILIP MORRIS to samé. Výsledky tohoto testu jsou tedy velice odlišné a nejednoznačné, proto nemohu podle tohoto postupu o heteroskedasticitě rozhodnout.

*Goldfeldův-Quandtův test* mi snad poskytne konkrétnější závěry. V mém případě je  $m = 5$ , čili mám dvě skupiny po 9-ti pozorováních. Pro každou skupinu a jednotlivé faktory a akcie jsem spočetla reziduální součty čtverců. Např. pro akcii ČEZu platí

$$X_1 : S_1 = 47094 \text{ a } S_2 = 19794, \text{ tj. } F \doteq 0,420.$$

Hodnota statistiky  $F$  je menší než kritická hodnota  $F_{3,3}(0,05) = 9,28$ , to znamená, že nulovou hypotézu o homoskedasticitě nezamítám. Pro ostatní faktory

modelu vyjdou všechny statistiky  $F$  menší než kritická hodnota. Mám tedy výsledek, že v modelu akcie ČEZ se heteroskedasticita nevyskytuje. Stejně závěry vyjdou poté i pro všechny ostatní akcí. Podle této metody jsem se tedy dopracovala k jednoznačnému výsledku: předpoklad homoskedasticity v modelu není porušen.

Dosud jsem nenašla jednoznačné a nejlepší řešení optimálního tvaru regresního modelu. Existenci multikolinearity mám potvrzenou několika metodami, autokorelovanost reziduí či heteroskedasticita potvrzena nebyla. Čili nyní se nabízí postup odhadu regresních parametrů ZMNČ (kap. 4.1) a to pro původní model (63). Druhou možností je jít na to selským rozumem. Odstranit z původního modelu multikolinearitu (nebo ji alespoň zmírnit), tzn. odstranit faktor nezaměstnanosti nebo faktor CZK/USD a použít MNČ (kap. 3.2). Ale protože by se mi v tomto případě automaticky zhoršily vlastnosti odhadnutých modelů, nahradím odstraněný faktor novým faktorem, který by mohl v modelu do určité míry ovlivnit ceny akcií a porovnáám, zda budou jednotlivé charakteristiky lepší či horší.

Nejprve zkusím dát na ten selský rozum. Otázkou je, jaký přidat do modelu nový faktor. Podle studie [26] jsou nejdůležitějším fundamentálním makroekonomickým faktorem úrokové sazby. Základní úrokovou sazbou v ČR je 14-ti denní reposazba, kterou stanovuje ČNB a od které se odvíjí vývoj všech ostatních úrokových sazeb v ekonomice. Zkusím tedy do mého modelu zapojit faktor úrokové sazby, zdroje dat viz [31]. Jednotlivé hodnoty jsou vyznačeny v následující tabulce.

Datum	úroková sazba (%)
1. 4. 05	2,000
1. 7. 05	1,750
1. 10. 05	1,750
1. 1. 06	2,000
1. 4. 06	2,000
1. 7. 06	2,000
1. 10. 06	2,500
1. 1. 07	2,500
1. 4. 07	2,500
1. 7. 07	2,750
1. 10. 07	3,250
1. 1. 08	3,500
1. 4. 08	3,750
1. 7. 08	3,750
1. 10. 08	3,500
1. 1. 09	2,250
1. 4. 09	1,750
1. 7. 09	1,500
1. 10. 09	1,250
1. 1. 10	1,000
1. 4. 10	1,000
1. 7. 10	0,750
1. 10. 10	0,750

Tabulka 6: Hodnoty úrokových sazeb od 1. 4. 2005 do 1. 10. 2010

Jako první odstráním z modelu faktor nezaměstnanosti a místo něj přidám úrokové sazby (ozn. model A). Výsledky jsou uvedeny v následující tabulce, m-file s jednotlivými výpočty v příloze 5 [63].

Název akcie	RSC	$\hat{\sigma}$	index determinace $R^2$
ČEZ	210586,77	111,30	0,84
TELEFÓNICA O2	51471,51	55,02	0,52
UNIPETROL	20162,76	34,44	0,76
PHILIP MORRIS	86701419,33	2258,34	0,80
ORCO	4057655,75	488,55	0,87
KOMERČNÍ BANKA	2642706,20	394,28	0,58
CETV	530193,77	176,60	0,93
ERSTE GROUP BANK	207679,81	110,53	0,94

Tabulka 7: Charakteristiky odhadnutého nového modelu A

Nyní zkusím odstranit z modelu faktor CZK/USD a přidat úrokové sazby (ozn. model B), viz příloha 6 [63]. Výsledky pro tento postup jsou následující

Název akcie	RSČ	$\hat{\sigma}$	index determinace $R^2$
ČEZ	188736,38	105,37	0,86
TELEFÓNICA O2	61716,32	60,25	0,42
UNIPETROL	19882,76	34,20	0,77
PHILIP MORRIS	44417819,30	1616,42	0,90
ORCO	3788463,16	472,07	0,88
KOMERČNÍ BANKA	3576430,05	458,67	0,43
CETV	557763,66	181,13	0,93
ERSTE GROUP BANK	193974,22	106,82	0,95

Tabulka 8: Charakteristiky odhadnutého nového modelu B

Po srovnání všech tří modelů ((63), modelu A a B) dle indexů determinace, vyšel celkově nejlépe nový model A. Kromě akcie PHILIP MORRIS se u všech ostatních modelů akcií index determinace zvýšil. Oproti tomu, vyjma akcií TELEFÓNICA O2 a KOMERČNÍ BANKA, vychází nejlepší indexy determinace u třetího modelu B. Nabízí se zde tedy dvě alternativy: pro budoucí predikce použít model A pro všechny akcie a nebo model B bez predikcí cen akcií O2 a KB.

Nyní prozkoumám druhou možnost, tj. použití ZMNČ pro původní model (63). Odhad regresních parametrů ZMNČ se provádí dle (19) s novým parametrem  $\Omega$ . Obecně se ale aplikace ZMNČ při existenci multikolinearity v modelu nedoporučuje. Nyní už vím, proč je zde problém právě s určením matice  $\Omega$ . Zkusila jsem ji zastoupit varianční i korelační maticí regresorů (faktorů). Při výpočtu parametru  $\hat{\beta}$  vycházely ale vynásobené matice téměř singulární a výsledky potom byly nesmyslné. Dalším důvodem zamítnutí ZMNČ je také fakt, že chci konečný tvar regresního modelu použít pro predikce na další čtvrtletí. Musím tedy dávat pozor na to, zda jsou rozptyly jednotlivých faktorů neměnné na začátku i konci sledovaného období. Pokud by byly značně různé, predikce by potom nebyly kvalitní. Varianční matice pro prvních 15 hodnot a pro posledních 15 hodnot jednotlivých faktorů se od sebe výrazně lišily, žádané predikce by tedy mohly být dosti nepřesné.

Porovnejme všechny dosavadní možnosti. Odhad regresních parametrů pomocí ZMNČ není pro predikce vhodný. Rozhoduji se tedy mezi modelem A a B. U obou variant jsou nízké indexy determinace pro regresní modely akcií O2 a KB. Pokud bych tedy pro tuto chvíli přestala uvažovat zmíněné dvě akcie, indexy determinace u ostatních odhadnutých akcií jsou lepší v modelu B. Můžu říci, že jsem v rámci mých možností našla nejlepší regresní model s nejkvalitnějšími odhady regresních parametrů

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + \varepsilon_t, \quad t = 1, \dots, 23, \quad (64)$$

kde  $y_t$ ...cena akcie společnosti v čase  $t$ ,

$x_{t1}$ ...míra inflace v čase  $t$ ,

$x_{t2}$ ...meziroční růst HDP v čase  $t$ ,

$x_{t3}$ ...obecná míra nezaměstnanosti v čase  $t$ ,

$x_{t4}$ ...úroková sazba (2T reposazba) v čase  $t$ ,

$x_{t5}$ ...kurz měny CZK/EUR v čase  $t$ .

Nyní konečně přejdu k předpovědím. Nejprve mohu zkusit odhadnout ceny akcií k 1. 1. 2011 pro můj konečný regresní model (64) s hodnotami jednotlivých faktorů k 1. 10. 2010, predikce provedeme dle (57), viz příloha 6 [63]. Odhadnuté ceny akcií jsou uvedeny v tabulce 9.

Název akcie	odhadnutá cena akcie k 1. 1. 2011
ČEZ	914,5
TELEFÓNICA O2	<b>398,7</b>
UNIPETROL	<b>200,8</b>
PHILIP MORRIS	8177,2
ORCO	<b>84,3</b>
KOMERČNÍ BANKA	3856,8
CETV	<b>438,1</b>
ERSTE GROUP BANK	695,9

Tabulka 9: Odhadnuté ceny akcií k 1. 1. 2011

Pro srovnání kvality jednotlivých odhadů jsou v další tabulce uvedeny skutečné ceny akcií k 1. 1. 2011.

Název akcie	skutečná cena akcie k 1. 1. 2011
ČEZ	800,0
TELEFÓNICA O2	386,5
UNIPETROL	198,1
PHILIP MORRIS	10200
ORCO	182,9
KOMERČNÍ BANKA	4472
CETV	400,5
ERSTE GROUP BANK	899,4

Tabulka 10: Ceny akcií k 1. 1. 2011

Tučně zvýrazněné hodnoty správně predikovaly buď růst nebo pokles cen jednotlivých akcií.

V následující tabulce jsou uvedeny skutečné hodnoty faktorů k 1. 1. 2011.

Datum	inlace (%)	HDP (%)	nezam. (%)	CZK/EUR	úrok. sazba (%)
1. 1. 11	2,1	2,9	7,0	25,06	0,75

Tabulka 11: Hodnoty faktorů k 1. 1. 2011

Mohu se tedy pokusit o predikci dalšího vývoje kurzů akcií, tzn. odhadnout ceny akcií k 1. 4. 2011 pomocí nejnovějších hodnot faktorů z tabulky 11, které byly všechny zveřejněny do 12. 2. 2011 (výpočty predikcí v příloze 6 [63]). Dle (57) dostávám

Název akcie	Predikce k 1. 4. 2011
ČEZ	843,8
TELEFÓNICA O2	396,5
UNIPETROL	197,9
PHILIP MORRIS	8599,3
ORCO	68,9
KOMERČNÍ BANKA	3694,1
CETV	390,2
ERSTE GROUP BANK	645,6

Tabulka 12: Predikce vývoje cen akcií k 1. 4. 2011



Tyto predikce už ale bohužel nestihnou porovnat se skutečnými hodnotami cen akcií, ale mohly by mi pomoci při rozhodování, zda k tomuto období koupit či prodat určitou akcii podle jejich předpokládaného cenového poklesu či růstu. Na začátku března odpovídá skutečný vývoj cen všech akcií s výjimkou ORCO predikovanému trendu (poklesu či růstu).

## 8 Příklad simultánních rovnic (modelování prodejnosti bytů v developerských projektech)

Cílem příkladu bude modelovat poptávku po bytech prodávaných v developerských projektech na základě vhodných faktorů a v závislosti na prodejní ceně. Jako vhodné vysvětlované a vysvětlující proměnné jsem zvolila

- $y_{t1}$  průměrná cena 1 m<sup>2</sup> staršího bytu v lokalitě,
- $y_{t2}$  průměrná cena 1 m<sup>2</sup> bytové jednotky v developerském domě,
- $y_{t3}$  prodejnost developerských bytů v lokalitě,
- $x_{t1}$  průměrná mzda v regionu,
- $x_{t2}$  míra nezaměstnanosti v regionu,
- $x_{t3}$  poměr lidí ve věku 15 až 64 let k celkovému počtu lidí v regionu,
- $x_{t4}$  ohodnocení polohy nového bytu,
- $x_{t5}$  reprodukční cena 1 m<sup>2</sup> nového bytu v regionu,
- $x_{t6}$  cena 1 m<sup>2</sup> stavebního pozemku v lokalitě,
- $x_{t7}$  doba od zahájení prodeje nového bytu,
- $x_{t8}$  počet konkurenčních nových bytů v lokalitě,
- $x_{t9}$  modifikovaná reprodukční cena 1 m<sup>2</sup> nově stavěné bytové jednotky,
- $t$  index identifikující lokalitu developerského projektu,  $t = 1, \dots, 16$ .

Uvažujme následující soustavu tří simultánních rovnic

$$y_{t1} = \beta_{01} + \beta_{11}x_{t1} + \beta_{21}x_{t2} + \varepsilon_{t1} \quad (65)$$

$$y_{t2} = \gamma_{12}y_{t1} + \beta_{02} + \beta_{12}x_{t5} + \varepsilon_{t2} \quad (66)$$

$$y_{t3} = \gamma_{13}y_{t1} + \gamma_{23}y_{t2} + \beta_{03} + \beta_{13}x_{t1} + \beta_{23}x_{t3} + \beta_{33}x_{t4} + \\ + \beta_{43}x_{t7} + \beta_{53}x_{t8} + \beta_{63}x_{t9} + \varepsilon_{t3} \quad (67)$$

Mějme k dispozici následující data endogenních proměnných

$t$	lokalita	$y_{t1}$ [Kč/m <sup>2</sup> ]	$y_{t2}$ [Kč/m <sup>2</sup> ]	$y_{t3}$ [%]
1	Uničov, Nemocniční, A	13555	25120	57
2	Uničov, Nemocniční, B	13555	27730	71
3	Olomouc, Pražská	26819	32661	76
4	Olomouc, Horní Lán	25422	27618	84
5	Boskovice, Na Chmelnici	18643	26746	62
6	Boskovice, S. Čecha (Terasy na výsluní)	20675	26729	71
7	Brno, Obřany	29231	36483	69

8	Brno, Kamechy	26800	28361	53
9	Praha 9, Chlebovická (Barevné Letňany)	40572	41083	91
10	Praha 4, Na Pankráci (Vivus Pankrác)	45277	56138	30
11	Ostrava, 28. října (Nové byty Ostrava)	15789	35203	38
12	Ostrava, Masná (Viladomy Gemini)	16139	37864	33
13	Č. Budějovice, B. Smetany	24243	37763	39
14	Č. Budějovice, Lidická (Byty Rožnov)	26924	31753	83
15	Jihlava, Hruškové Dvory III-B	20317	26605	36
16	Liberec, Krejčího (Byt. dům Panorama)	17217	26902	27

Hodnoty všech predeterminovaných proměnných jsou znázorněny v následující tabulce

$t$	$x_{t1}$ [Kč]	$x_{t2}$ [%]	$x_{t3}$ [%]	$x_{t4}$	$x_{t5}$ [Kč]	$x_{t6}$ [Kč]	$x_{t7}$ [měs]	$x_{t8}$	$x_{t9}$ [Kč]
1	19788	14,1	70,8	2	14268	221	16	1	20789
2	19788	14,1	70,8	2	15751	221	16	1	20789
3	22369	10,6	70,5	3	18552	820	30	7	21272
4	22369	10,6	70,5	3	15687	800	18	7	21255
5	21889	10,7	69,7	3	15192	432	28	3	20959
6	21889	10,7	69,7	3	15182	432	26	3	20959
7	23834	8,5	69,6	4	20722	2590	10	42	22699
8	23834	8,5	69,6	4	16109	2210	16	42	22393
9	29423	4	71,5	3	23335	2660	12	266	22755
10	29423	4	71,5	1	31886	8400	16	266	27385
11	22694	11,8	70,6	3	19995	600	36	7	21094
12	22694	11,8	70,6	2	21507	730	22	7	21199
13	21951	5,9	71,4	1	21449	800	10	8	21255
14	21951	5,9	71,4	4	16396	800	22	8	21255
15	21884	8,8	71	4	15112	800	12	5	21255
16	22832	10,1	70,8	3	15280	800	6	3	21255

Nejdříve je zapotřebí podrobně popsat původ a výpočty všech výše uvedených dat. Pro jednotlivé endogenní proměnné platí

- $y_{t1}$  – vlastní výpočet ze zdroje [33] v podobě průměrných hodnot 1 m<sup>2</sup> nabízených starších bytů v lokalitě, nejlépe ve stejných či přilehlých ulicích, kde je realizován přísl. developerský projekt. Při výpočtu byl vytvořen datový soubor obsahující podlahové plochy prodávaných bytů a poptávanou kupní

cenu. Cena 1 m<sup>2</sup> se pak získala jako podíl kupní ceny a podlahové plochy (do plochy není započtena plocha balkonů, lodžii a teras).

- $y_{t2}$  – jde o hodnotu průměrné ceny bytové jednotky v domě postaveném developerem. Ceny bytů v domě se liší podle velikosti bytů, podlaží, orientace hlavních místností na světové strany, výhledu, atd. Vlastní datový soubor byl vytvořen ze zdrojů [34-48].
- $y_{t3}$  – hodnota v intervalu  $\langle 0, 1 \rangle$  udává podíl prodaných bytů v domě oproti celkovému počtu bytů v domě prodávaných developerem, informace viz zdroje [34-48].

Postupy pro zjišťování predeterminovaných proměnných jsou následující

- $x_{t1}$  – průměrná mzda v obci, jednotlivé hodnoty viz [49].
- $x_{t2}$  – míra nezaměstnanosti v regionu, informace dostupné z článku [50].
- $x_{t3}$  – poměr lidí ve věku od 15 do 64 let v populaci regionu, jednotlivé hodnoty viz [49].
- $x_{t4}$  – klasifikace polohy domu. Hodnoty byly voleny v rozmezí 1 – 5, přičemž číslo 1 charakterizuje polohu bytu v centru města a 5 je lokalita nejdále od centra.
- $x_{t5}$  – reprodukční cena 1 m<sup>2</sup> bytu (skutečná cena za kterou je možné stavbu realizovat, může být také určena pomocí kalkulace položek stavebního rozpočtu). Z nejmenovaných zdrojů bylo zjištěno, že přibližný zisk investorů, prodávajících nové byty, může činit až 43%. Hodnoty  $x_{t5}$  tedy byly získány jako rozdíly hodnot proměnných  $y_{t2}$  a uvedeného přibližného zisku.
- $x_{t6}$  – cena 1 m<sup>2</sup> stavebního pozemku v dané lokalitě. U velkých měst je dohledána z cenové mapy, u menších měst vypočtena z počtu obyvatel obce [49] podle vyhlášky Ministerstva financí ČR č. 364/2010 Sb., viz [55]. Cenová mapa pro Olomouc viz [51], Brno viz [52], Prahu viz [53] a Ostravu viz [54].

- $x_{t7}$  – doba od zahájení prodeje v měsících. Údaj byl dohledán na webových stránkách developerů, viz [34-48].
- $x_{t8}$  – proměnná, která má sledovat míru konkurence developerských projektů v regionu. Hodnota je dána celkovým počtem prodávaných bytů v lokalitě od všech developerů, údaje dohledány na [56].
- $x_{t9}$  – modifikovaná reprodukční cena 1 m<sup>2</sup> podlahové plochy prodávaného bytu. Tato cena byla odhadnuta jako součet známe částky reprodukční ceny 1 m<sup>2</sup> bytu u jednoho domu ve výši 20 780 Kč, která je upravena (navýšena) o podíl kupní ceny pozemku (uvažován vždy pozemek o výměře 1000 m<sup>2</sup>) přináležející k bytové jednotce. Spoluvlastnický podíl pozemku je určen dle zákona o vlastnictví bytů č. 72/1994 Sb. jako podíl podlahové plochy bytu a celkové podlahové plochy všech bytů, podrobněji viz příloha 7 [63].

Hlavním cílem tohoto příkladu je odhadnout všechny neznámé parametry soustavy rovnic (65), (66) a (67). Nutno dodat, že tvary těchto rovnic jsem hledala podobným způsobem jako v předchozím příkladu, až jsem dospěla k co nejvyšším indexům determinace. Vzhledem k rozsahu práce už zde ale tyto postupy znovu uvádět nebudu.

K tomu, abych mohla k odhadům parametrů použít M2NČ, musím nejprve ověřit podmínky identifikovatelnosti rovnic, viz (kap. 5.2). Protože ve všech třech rovnicích se najde matice, jejíž determinant je nenulový, hodnostní podmínka identifikace je splněna. Po dosazení do nerovnosti (48) dostávám pro první rovnici  $9 - 2 \geq 1 - 1$ , pro druhou  $9 - 1 \geq 2 - 1$  a pro třetí rovnici  $9 - 7 \geq 3 - 1$ . Řádová podmínka identifikace je tedy také splněna a všechny rovnice jsou identifikované.

Nyní už pro odhad parametrů strukturních rovnic (65), (66) a (67) můžu použít M2NČ. Datový soubor viz příloha 8 a výpočty v příloze 9 [63]. Tato metoda omezené informace je založena na opakované aplikaci klasické MNČ. Postupuji následovně

1. nejprve vypočítám vyrovnané hodnoty vysvětlované endogenní proměnné  $y_{t1}$  dle (52), přičemž matice  $\mathbf{X}$  je složena z absolutního členu a z hodnot vysvětlujících proměnných  $x_{t1}$  a  $x_{t2}$ ,
2. těmito vyrovnanými hodnotami  $\hat{y}_{t1}$  nahradím vysvětlující endogenní proměnnou  $y_{t1}$  v druhé rovnici (66) dle vztahu (53),
3. provedu odhad parametrů této upravené rovnice pomocí (55),
4. postup se opakuje: vyrovnanými vysvětlovanými endogenními proměnnými  $\hat{y}_{t1}$  a  $\hat{y}_{t2}$  z první a druhé rovnice nahradím proměnné  $y_{t1}$  a  $y_{t2}$ , vystupující v třetí rovnici (67) jako vysvětlující endogenní proměnné,
5. opět provedu odhad parametrů dle (55) a vypočítám vyrovnané hodnoty rovnice (67)  $\hat{y}_{t3}$  pomocí (56).

Dostávám odhady neznámých parametrů z rovnic (65), (66) a (67) jako

$$\hat{y}_{t1} = -89,7250 + 1,5449 x_{t1} - 1278,2487 x_{t2} \quad (68)$$

$$\hat{y}_{t2} = 0,0428 y_{t1} + 752,9062 + 1,6745 x_{t5} \quad (69)$$

$$\begin{aligned} \hat{y}_{t3} = & 0,0024 y_{t1} - 0,0012 y_{t2} + 363,4307 - 0,0178 x_{t1} - 3,8276 x_{t3} + \\ & + 0,7197 x_{t4} + 1,2934 x_{t7} + 0,4780 x_{t8} + 0,0146 x_{t9} \end{aligned} \quad (70)$$

Indexy determinace jednotlivých rovnic jsou uvedeny v následující tabulce

Rovnice	index determinace
(68)	0,9316
(69)	0,9908
(70)	0,7405

Tabulka 13: Indexy determinace odhadnutých modelů

Třetí index determinace není tak vysoký jako dva předcházející. Vzhledem k existenci spousty jiných veličin ovlivňujících prodejnost nových bytů ale nejsem schopna všechny tyto zakomponovat do modelu. Vyhodnotila jsem tedy i tento index jako uspokojivý.

Nyní se pokusím vžít do role investora, který váhá, zda v určité lokalitě investovat do prodeje bytů či nikoliv. Nejdůležitějším faktorem při jeho rozhodování bude úvaha nad předpokládanou úspěšností prodeje nových bytů v lokalitě. Řekněme, že investor by chtěl prodat poslední volné nové byty z developerského projektu Brno, Obřany. Nyní musí uvažovat, jaká by byla úspěšnost prodeje v závislosti na prodejní ceně, kterou určí. V této situaci může pomoci můj sestavený odhadnutý model (70).

Budu uvažovat tyto prodejní ceny

- a) 25 000 Kč/m<sup>2</sup>,
- b) 30 000 Kč/m<sup>2</sup>,
- c) 35 000 Kč/m<sup>2</sup>.

Po výpočtech v Matlabu, viz příloha 9 [63], dostávám následující hodnoty potenciační úspěšnosti prodeje při jednotlivých cenách

- a) 74,8 %,
- b) 68,9 %,
- c) 63,1 %.

Vidíme, že prodejnost klesá s rostoucí prodejní cenou, což je logické. Původními hodnotami pro Brno, Obřany byla prodejní cena 36 483 Kč a prodejnost 69 %, výsledné hodnoty odhadnuté úspěšnosti se tedy zdají být reálné. Teď už by záleželo pouze na mínění investora a na jeho finančních možnostech, pro kterou z variant by se rozhodl.

Investora by také mohly zajímat jiné potenciační developerské projekty v jiných městech než mnou zahrnutých. Protože jsem ještě neuvažovala např. Plzeňský kraj, budu se zajímat o potenciační úspěšnost prodeje nových bytů v Plzni 3. Postup je podobný jako v předcházejícím. Ze stejných zdrojů jako doposud a z webových stránek jiného již existujícího developerského projektu v Plzni 3 (viz [56]) dostávám následující hodnoty.

$y_{t1}$	$y_{t2}$	$y_{t3}$	$x_{t1}$	$x_{t2}$	$x_{t3}$	$x_{t4}$	$x_{t5}$	$x_{t6}$	$x_{t7}$	$x_{t8}$	$x_{t9}$
21477	34018	63,9	23189	6,6	70,1	2	19390	800	28	10	21255,4

Tabulka 14: Data pro nové byty v Plzni

Řekněme, že investor by chtěl tyto byty prodávat za určitou cenu a zajímala by ho prodejnost bytů po 12-ti měsících. Z výpočtů v příloze 9 [63] dostávám odhadnutou úspěšnost prodeje bytů v Plzni 3 ve výši

- 64,9 % při ceně 30 000 Kč,
- 59,0 % při ceně 35 000 Kč,
- 53,2 % při ceně 40 000 Kč.

Stejným způsobem bych mohla odhadovat úspěšnost prodeje jakýchkoliv jiných bytů v kterémkoliv jiném městě. Tento ukazatel mohu také zkoumat např. při různých prodejních cenách nebo při různé době trvání od prodeje těchto bytů.



## Závěr

V práci jsem se pokusila čtenáře seznámit s klasickým modelem lineární regrese, popsala MNČ pro odhad regresních parametrů, podmínky pro její aplikaci a vlastnosti všech odhadů. Pozornost byla také věnována zobecněnému lineárnímu regresnímu modelu, spolu se ZMNČ a postupům použitelným při výskytu heteroskedasticity, autokorelace a multikolinearity. Dále jsem se zabývala simultánními rovnicemi, tvary jejich maticového vyjádření a M2NČ pro odhad všech neznámých parametrů modelu. V šesté kapitole jsem uvedla postupy při ekonometrickém prognózování, jichž jsem později využila jako finálních cílů v praktické části publikace.

Cílem diplomové práce bylo především aplikovat teoretické poznatky z ekonometrické analýzy na příkladech z oblasti ekonomie a finančnictví. V prvním příkladu, zv. faktorový model, jsem se zabývala cenami akcií na burze a faktory, které cenu těchto akcií ovlivňují. Hlavním cílem bylo najít optimální tvar regresního modelu pro kvalitní odhady parametrů s co nejlepšími vlastnostmi. Zaměřila jsem se na použití i méně známých metod a postupů a podařilo se najít nejvhodnější model pro predikci cen akcií na budoucí období. Tyto závěry by mohly být nápomocny investorům na burze při rozhodování, zda akcii k určitému datu prodat či koupit podle jejího odhadnutého poklesu či růstu ceny.

V druhém příkladu využívajícím simultánních rovnic jsem se zabývala poptávkovou a nabídkovou funkcí po nově stavěných bytech v rámci uskutečněných developerských projektů v různých krajích České republiky. Při této studii dalo největší práci vyhledat, popř. vypočítat všechny hodnoty proměnných, potřebných k vytvoření jednotlivých simultánně závislých rovnic. Po několika alternativách se mi v rámci simultánního modelu podařilo sestavit tři rovnice s co nejvyššími indexy determinace. Pro odhady neznámých parametrů se využilo metody zobecněných nejmenších čtverců, spočívající v opakované aplikaci metody nejmenších čtverců, použité v prvním příkladu. Výsledky jsem využila nikoli pro predikci na následující období, ale pro odhad úspěšnosti prodeje bytů jiného potencionálního developerského projektu. Ukazatel prodejnosti jsem zkoumala také v závislosti

na různých prodejních cenách nových bytů. Tyto informace by mohly opět pomoci investorům při zvažování, zda realizovat určitý developerský projekt ať již ve městě samém nebo v jakémkoli jiném městě ČR.

Při psaní diplomové práce jsem získala spoustu cenných zkušeností. V první řadě to je zlepšení dovedností v rámci výpočtů v programu Matlab, dále také práce s velkým počtem zdrojů v literatuře. Nejvíce jsem vděčná za aplikovanou část práce a za praktické využití mých příkladů v dnešním reálném ekonomickém světě. Při sběru dat jsem narazila na spoustu veřejnosti dostupných informací, o kterých jsem neměla ani tušení a získala tak mnoho nových a užitečných vědomostí.

Existují i další ekonometrické úlohy, které by si zasloužily pozornost. Rozsah mé diplomové práce mi už ale neumožňuje jiné ekonometrické přístupy uvést. Přesto by tento text mohl být čtenářům nápomocen a třeba by se i odvážili dle mých závěrů investovat. Věřím, že úsilí věnované diplomové práci mi bude přínosem v dalším studiu popřípadě v budoucí praxi.

## Literatura

- [1] Anděl, J.: Statistické metody, UK, Praha 2007.
- [2] Anděl, J.: Základy matematické statistiky, UK, Praha 2007.
- [3] Cipra, T.: Finanční ekonometrie, 1. vydání, Praha 2008.
- [4] Čechura, L., Hálová, P., Kroupová, Z., Malý, M., Peterová, J., Šobrová, L.: Cvičení z ekonometrie, 2. vydání, ISBN, 2008.
- [5] Fiala, P.: Úvod do ekonometrie, ČVUT v Praze, Praha 2008.
- [6] Hebák, P., Svobodová, A.: Regrese II. část, VŠE v Praze, Praha 2001.
- [7] Hušek, R.: Aplikovaná ekonometrie, Praha 2001.
- [8] Hušek, R.: Ekonometrická analýza, 1. vydání, Praha 1999.
- [9] Hušek, R.: Ekonometrické metody, VŠE v Praze, Praha 1973.
- [10] Hušek, R.: Ekonometrické modely, VŠE v Praze, Praha 1983.
- [11] Hušek, R.: Základy ekonometrie, 2. vydání, VŠE v Praze, Praha 1992.
- [12] Hušek, R.: Základy ekonometrické analýzy II. (Speciální postupy a techniky), VŠE v Praze, Praha 1998.
- [13] Klímeck, P.: Ekonometrie, UTB ve Zlíně, Zlín 2001.
- [14] Klímeck, P.: Úvod do ekonometrie a hospodářské statistiky, ISBN, Zlín 2006.
- [15] Kožíšek, J.: Ekonomická statistika a ekonometrie, ČVUT v Praze, Praha 2005.
- [16] Kubáček, L., Kubáčková, L.: Statistika a metrologie, UP v Olomouci, Olomouc 2000.

- [17] Lagová, M.: Programové zabezpečení ekonometrických modelů, VŠE v Praze, Praha 2000.
- [18] Meloun, M.: Statistické zpracování dat [online], dostupné z: [http://meloun.upce.cz/index.php?option=com\\_content&view=article&id=4&Itemid=11](http://meloun.upce.cz/index.php?option=com_content&view=article&id=4&Itemid=11), [citováno 16.2.2011].
- [19] Mezník, I.: Ekonometrie pro magisterské studijní programy, CERM, Brno 2005.
- [20] Pojkarová, K.: Ekonometrie a prognostika v dopravě, DFJP Pardubice, 2006.
- [21] Tvrdík, J.: Analýza vícerozměrných dat [online], dostupné z: [http://prf.osu.cz/doktorske\\_studium/dokumenty/Multivariable\\_Data\\_Analysis.pdf](http://prf.osu.cz/doktorske_studium/dokumenty/Multivariable_Data_Analysis.pdf), [citováno 20.1.2011].
- [22] Tvrdoň, J., Peterová, J., Hálová, P.: Exercises in econometrics, ISBN, Praha 2001.
- [23] Tvrdoň, J.: Ekonometrie, 5. vydání, ISBN, 2005.
- [24] Víšek, J. A.: Ekonometrie I., UK, Praha 1997.
- [25] Zouhar, J., Ráčková, A., Lejnarová, Š.: Základy ekonometrie v příkladech, VŠE v Praze, Praha 2009.
- [26] Fundamentální analýza a základy makroekonomie [online], dostupné z: [http://www.xtb.cz/repository/cz/vzdelani/prezentace/Fundamenty\\_XTB.pdf](http://www.xtb.cz/repository/cz/vzdelani/prezentace/Fundamenty_XTB.pdf), [citováno 25. 1. 2011].
- [27] Data - inflace [online], dostupné z: *inflace* : [http://www.czso.cz/csu/redakce.nsf/i/mira\\_inflace](http://www.czso.cz/csu/redakce.nsf/i/mira_inflace), [citováno 21. 1. 2011].
- [28] Data - HDP [online], dostupné z: <http://www.czso.cz/csu/csu.nsf/kalendar/2010-hdp>, [citováno 23. 1. 2011].

- [29] Data - nezaměstnanost [online], dostupné z: <http://www.czso.cz/csu/csu.nsf/kalendar/2010-zam>, [citováno 16. 1. 2011].
- [30] Data - kurzy měn [online], dostupné z: <http://www.kurzy.cz/kurzy-men/historie/>, [citováno 23. 1. 2011].
- [31] Data - úroková sazba [online], dostupné z: <http://www.finance.cz/bankovnictvi/sazby-cnb/>, [citováno 23. 2. 2011].
- [32] Data - ceny akcií [online], dostupné z: <http://www.akcie.cz/>, [citováno 18. 12. 2010].
- [33] Data - cena  $m^2$  starého bytu v lokalitě [online], dostupné z: <http://www.sreality.cz/hledani/prodej/byty/>, [citováno 7. 12. 2010].
- [34] Data - cena  $m^2$  nového bytu Uničov [online], dostupné z: <http://www.byty-unicov.cz/cenik.do?dum=-1&typProstor=-1&velikost=-1&cena=-1&stav=-1&zobrazOblibene=false>, [citováno 29. 11. 2010].
- [35] Data - cena  $m^2$  nového bytu Olomouc, Pražská [online], dostupné z: <http://www.bytyprazska.cz/cenik.do?stranka=1>, [citováno 6. 12. 2010].
- [36] Data - cena  $m^2$  nového bytu Olomouc, Horní Lán [online], dostupné z: <http://www.hornilan.cz/cenik-bytu.do?dum=-1&typProstor=-1&velikost=-1&cena=-1&stav=-1&zobrazOblibene=>, [citováno 6. 12. 2010].
- [37] Data - cena  $m^2$  nového bytu Boskovice, Na Chmelnici [online], dostupné z: <http://web.realspektrum.cz/novostavby/byty/boskovice-na-chmelnici/dum-A1/>, [citováno 6. 12. 2010].
- [38] Data - cena  $m^2$  nového bytu Boskovice, S. Čecha [online], dostupné z: <http://www.terasynavyisluni.cz/default.asp?nav=cen&exe=2>, [citováno 6. 12. 2010].
- [39] Data - cena  $m^2$  nového bytu Brno, Obřany [online], dostupné z: <http://www.rekotrend.cz/vyhledavani-bytu.do>, [citováno 7. 12. 2010].

- [40] Data - cena  $m^2$  nového bytu Brno, Kamechy [online], dostupné z: [http : //www.byty – kamechy.cz/ceny – a – rezervace.do](http://www.byty-kamechy.cz/ceny-a-rezervace.do), [citováno 7. 12. 2010].
- [41] Data - cena  $m^2$  nového bytu Praha 9 [online], dostupné z: [http : //www.barevneletnany.cz/nove – byty/barevne – letnany – cenik.html](http://www.barevneletnany.cz/nove-byty/barevne-letnany-cenik.html), [citováno 7. 12. 2010].
- [42] Data - cena  $m^2$  nového bytu Praha 4 [online], dostupné z: [http : //www.vivus.cz/pankrac/hledej.php](http://www.vivus.cz/pankrac/hledej.php), [citováno 8. 12. 2010].
- [43] Data - cena  $m^2$  nového bytu Ostrava, Kubínova [online], dostupné z: [http : //www.novebytyostrava.cz/cenik.html](http://www.novebytyostrava.cz/cenik.html), [citováno 16. 2. 2011].
- [44] Data - cena  $m^2$  nového bytu Ostrava, Nová Bělá [online], dostupné z: [http : //www.viladomygemini.cz/index.php?action = flat\\_offers](http://www.viladomygemini.cz/index.php?action=flat_offers), [citováno 16. 2. 2011].
- [45] Data - cena  $m^2$  nového bytu České Budějovice, B. Smetany [online], dostupné z: [http : //hanzlikovareality.cz/developerske\\_projekty/projekt.html?pid = 64](http://hanzlikovareality.cz/developerske_projekty/projekt.html?pid=64), [citováno 17. 2. 2011].
- [46] Data - cena  $m^2$  nového bytu České Budějovice, Lidická [online], dostupné z: [http : //www.best – cb.cz/roznov/menu.html](http://www.best-cb.cz/roznov/menu.html), [citováno 17. 2. 2011].
- [47] Data - cena  $m^2$  nového bytu Jihlava, Hruškové Dvory [online], dostupné z: [http : //www.czstavhold.cz/projekty/jihlava\\_iii/apart.php](http://www.czstavhold.cz/projekty/jihlava_iii/apart.php), [citováno 17. 2. 2011].
- [48] Data - cena  $m^2$  nového bytu Liberec, Krejčího [online], dostupné z: [http : //www.bytypanorama.cz/byty.htm](http://www.bytypanorama.cz/byty.htm), [citováno 17. 2. 2011].
- [49] Data - mzdy; lidé [online], dostupné z: [http : //www.brno.czso.cz/xb/redakce.nsf/i/mesta\\_a\\_obce](http://www.brno.czso.cz/xb/redakce.nsf/i/mesta_a_obce), [citováno 8. 12. 2010].
- [50] Data - nezaměstnanost [online], článek dostupný z: [http : //tn.nova.cz/zpravy/ekonomika/priprava – nezamestnanost.html](http://tn.nova.cz/zpravy/ekonomika/priprava-nezamestnanost.html), [citováno 7. 12. 2010].

- [51] Data - cena  $m^2$  stavebního pozemku Olomouc [online], dostupné z: [http : //www.olomouc.eu : 85/](http://www.olomouc.eu:85/), [citováno 1. 3. 2011].
- [52] Data - cena  $m^2$  stavebního pozemku Brno [online], dostupné z: [http : //gis.brno.cz/tms/html/cenova\\_mapa/index.php?client\\_type = map\\_html](http://gis.brno.cz/tms/html/cenova_mapa/index.php?client_type=map_html), [citováno 1. 3. 2011].
- [53] Data - cena  $m^2$  stavebního pozemku Praha [online], dostupné z: [http : //wgp.praha – mesto.cz/tms/projects\\_h/cmp08/](http://wgp.praha-mesto.cz/tms/projects_h/cmp08/), [citováno 1. 3. 2011].
- [54] Data - cena  $m^2$  stavebního pozemku Ostrava [online], dostupné z: [http : //gisova.ostrava.cz/webmaps/mapacena/viewer.htm](http://gisova.ostrava.cz/webmaps/mapacena/viewer.htm), [citováno 2. 3. 2011].
- [55] Data - cena  $m^2$  stavebního pozemku [online], dostupné z: [http : //www.sbirka.cz/POSLATYD/NOVE/10 – 364.htm](http://www.sbirka.cz/POSLATYD/NOVE/10-364.htm), [citováno 2. 3. 2011].
- [56] Data - konkurence [online], dostupné z: [http : //www.byty – nemovitosti.cz/nove – byty – domy – novostavby – byty.html](http://www.byty-nemovitosti.cz/nove-byty-domy-novostavby-byty.html), [citováno 4. 3. 2011].
- [57] Durbin, J., Watson, G. S.: Testing for Serial Correlation in Least Squares Regression: I [online], článek dostupný z: [http : //www.jstor.org/stable/2332391](http://www.jstor.org/stable/2332391), [citováno 23. 2. 2011].
- [58] Goldfeld, S. M., Quandt, R. E.: Some Tests for Homoscedasticity [online], článek dostupný z: [http : //www.jstor.org/stable/2282689?&Search = yes &searchText = Goldfeld&searchText = Quandt&list = hide&searchUri = %2Faction%2FdoBasicSearch%3Facc%3Don%26Query%3DGoldfeld%2BQuandt%26gw%3Djtx%26acc%3Don%26prq%3DDurbin%2BWatson%26Search%3DSearch%26hp%3D25%26wc%3Don%26acc%3Don&prevSearch = &item = 8&tll = 916&returnArticleService = showFullText](http://www.jstor.org/stable/2282689?&Search=yes&searchText=Goldfeld&searchText=Quandt&list=hide&searchUri=%2Faction%2FdoBasicSearch%3Facc%3Don%26Query%3DGoldfeld%2BQuandt%26gw%3Djtx%26acc%3Don%26prq%3DDurbin%2BWatson%26Search%3DSearch%26hp%3D25%26wc%3Don%26acc%3Don&prevSearch=&item=8&tll=916&returnArticleService=showFullText), [citováno 23. 2. 2011].
- [59] Spearman, C.: The Proof and Measurement of Association between Two Things [online], článek dostupný z: [http : //www.jstor.org/stable/1412159 ?&Search = yes&searchText = Spearman&list = hide&searchUri = %2](http://www.jstor.org/stable/1412159?&Search=yes&searchText=Spearman&list=hide&searchUri=%2)

*Faction%2FdoBasicResults%3Fhp%3D25%26la%3D%26wc%3Don%26acc%3Don%26gw%3Djtx%26jpsi%3D1%26artsi%3D1%26Query%3DSpearman%26sbq%3DSpearman%26prq%3DSpearman%2Btest%26si%3D26%26jtxsi%3D26&prevSearch = &item = 49&tll = 18682&returnArticleService = showFullText*, [citováno 23. 2. 2011].

- [60] Breusch, T. S., Godfrey, L. G.: Data Transformation Tests [online], článek dostupný z:<http://www.jstor.org/stable/2232969?&Search=yes&searchText=Breusch&searchText=Godfrey&list=hide&searchUri=%2Faction%2FdoBasicResults%3Fhp%3D25%26la%3D%26wc%3Don%26acc%3Don%26gw%3Djtx%26jpsi%3D1%26artsi%3D1%26Query%3DBreusch%2BGodfrey%26prq%3DSpearman%26si%3D1%26jtxsi%3D1&prevSearch=&item=1&tll=278&returnArticleService=showFullText>, [citováno 23. 2. 2011].
- [61] Akaike, H: On the Likelihood of a Time Series Model [online], článek dostupný z: <http://www.jstor.org/stable/2988185?&Search=yes&searchText=Akaike&list=hide&searchUri=%2Faction%2FdoBasicResults%3Fhp%3D25%26la%3D%26wc%3Don%26acc%3Don%26gw%3Djtx%26jpsi%3D1%26artsi%3D1%26Query%3DAkaike%26sbq%3DAkaike%26prq%3DScott%26si%3D26%26jtxsi%3D26&prevSearch=&item=33&tll=6668&returnArticleService=showFullText>, [citováno 23. 2. 2011].
- [62] Mládek, J.: *Ekonomická analýza alkoholických nápojů v ČSSR*, Ekonomicko-matematický obzor (4), 1985.
- [63] Přílohy [online], dostupné z: <http://markova.hana2011.sweb.cz/main.html>, [citováno 30. 3. 2011].