

Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## DIPLOMOVÁ PRÁCE



Magdalena Zvejšková

# Odhady a testy v modelech panelových dat

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Prof. RNDr. Marie Hušková, DrSc.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická statistika  
a ekonometrie

Praha 2013

Na tomto místě bych chtěla poděkovat vedoucí své diplomové práce Prof. RNDr. Marii Huškové, DrSc., za mnoho cenných rad a trpělivé vedení. Děkuji také své rodině za morální a finanční podporu během celého studia.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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

V Praze dne 30. července 2013

Magdalena Zvejšková

**Název práce:** Odhady a testy v modelech panelových dat

**Autor:** Magdalena Zvejšková

**Katedra:** Katedra pravděpodobnosti a matematické statistiky

**Vedoucí diplomové práce:** Prof. RNDr. Marie Hušková, DrSc., Katedra pravděpodobnosti a matematické statistiky

**Abstrakt:** Tato práce se zaměřuje především na modelování panelových dat s nezávislými průřezy. V první části práce shrnujeme poznatky z oblasti poolových modelů a jednoduchých modelů komponentních chyb s pevnými a náhodnými efekty. Zaměřujeme se zejména na odhadování neznámých parametrů a testy významnosti efektů. Stručně je popsána i problematika oboustranných modelů komponentních chyb. Ve druhé části odvozujeme odhady parametrů v autoregresních modelech prvního řádu pro panelová data s pevnými i náhodnými koeficienty. Je ukázána nestrannost, konzistence a asymptotická normalita vybraných odhadů. Na základě těchto vlastností jsou navrženy testy hypotéz o příslušných parametrech. Použití modelů je demonstrováno na příkladech s reálnými a simulovanými daty.

**Klíčová slova:** panelová data, model komponentních chyb, autoregresní model prvního řádu, náhodné koeficienty

**Title:** Estimators and tests in panel data models

**Author:** Magdalena Zvejšková

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

**Supervisor:** Prof. RNDr. Marie Hušková, DrSc., Department of Probability and Mathematical Statistics

**Abstract:** This work investigates mainly panel data models in which cross-sections can be considered independent. In the first part, we summarize results in the field of pool models and one-way error component models with fixed and random effects. We focus especially on the ways of estimating unknown parameters and on effects significance tests. We also briefly describe two-way error component model issues. In the second part, estimators of first order autoregressive panel data model parameters are derived, for both fixed and random parameters case. The work proves unbiasedness, consistency and asymptotic normality of selected estimators. Using these features, hypothesis tests about corresponding parameters are derived. Application of models is illustrated using real data and simulated data examples.

**Keywords:** panel data, error component model, first order autoregressive model, random coefficient

# Obsah

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Úvod</b>   | <b>1</b>  |
| 1.1      | Historický přehled . . . . .                            | 2         |
| 1.2      | Aplikační přednosti a nedostatky . . . . .              | 5         |
| <b>2</b> | <b>Poolový model</b>                                    | <b>6</b>  |
| <b>3</b> | <b>Modely komponentních chyb</b>                        | <b>8</b>  |
| 3.1      | Jednoduchý model komponentních chyb . . . . .           | 9         |
| 3.1.1    | Jednoduchý model s pevnými efekty . . . . .             | 10        |
| 3.1.2    | Jednoduchý model s náhodnými efekty . . . . .           | 17        |
| 3.1.3    | Hausmanův specifikační test . . . . .                   | 23        |
| 3.2      | Oboustranný model komponentních chyb . . . . .          | 25        |
| 3.3      | Příklad . . . . .                                       | 27        |
| <b>4</b> | <b>AR(1) s pevným koeficientem</b>                      | <b>31</b> |
| 4.1      | Odhad pevného koeficientu a jeho vlastnosti . . . . .   | 33        |
| 4.2      | Odhad komponent rozptylu . . . . .                      | 37        |
| 4.3      | Test hodnoty koeficientu . . . . .                      | 40        |
| 4.4      | Aplikace na simulovaná data . . . . .                   | 44        |
| 4.5      | Příklad . . . . .                                       | 45        |
| <b>5</b> | <b>AR(1) s náhodným koeficientem</b>                    | <b>48</b> |
| 5.1      | Odhad náhodného koeficientu a jeho vlastnosti . . . . . | 50        |
| 5.2      | Odhad komponent rozptylu . . . . .                      | 54        |
| 5.3      | Test náhodnosti koeficientů . . . . .                   | 59        |
| 5.4      | Aplikace na simulovaná data . . . . .                   | 63        |
| 5.5      | Příklad . . . . .                                       | 65        |
| <b>6</b> | <b>Závěr</b>  | <b>67</b> |
| <b>A</b> | <b>Pomocná tvrzení</b>                                  | <b>69</b> |
| <b>B</b> | <b>Data</b>   | <b>73</b> |
| <b>C</b> | <b>Zdrojový kód v R</b>                                 | <b>76</b> |
|          | <b>Literatura</b>                                       | <b>80</b> |

# Kapitola 1

## Úvod

Panelová data jsou kombinací dat časových a průřezových. Průřezová data vznikají pozorováním vícero různých subjektů (firem, územních celků, domácností) v jednom časovém okamžiku a slouží tak především k porovnání těchto subjektů. Časová data oproti tomu zaznamenávají vývoj jednoho subjektu v čase (každý rok, měsíc, týden).

Panelová data jakožto kombinace těchto dvou typů tedy vznikají opakovaným pozorováním souboru subjektů a mají tak dvě dimenze, časovou a průřezovou. V případě časové dimenze jsou pozorování chronologicky uspořádána, asymptotické vlastnosti závisí na stacionaritě, svou roli zde často hraje dynamická povaha dat. V průřezové dimenzi typicky žádné přirozené uspořádání subjektů není a většinou je lze považovat za nezávislé.

V teorii panelových dat se obvykle uvažuje takový soubor, který se v průběhu sběru dat nemění a ve kterém jsou si subjekty (průřezy, panely) v jistém smyslu podobné natolik, že lze na všechny aplikovat model se stejnou strukturou. Takové typy souborů budeme uvažovat i v této práci. Zároveň se budeme zajímat jen o tzv. vyvážené soubory, tedy takové, ve kterých je v každém průřezu stejný počet pozorování, tj.  $T_i = T$ , pro každý průřez  $i = 1, \dots, N$ .

Příkladem panelových dat je třeba velikost HDP v jednotlivých státech USA za posledních 15 let, spotřeba benzínu v zemích Evropské unie sledovaná každý měsíc či například vývoj výše příjmu vybraných domácností ve Středočeském kraji v průběhu posledních 10 let.

Základní lineární regresní model, který se pro panelová data obvykle uvažuje, lze popsat následovně:

$$y_{it} = x_{1it}\beta_1 + x_{2it}\beta_2 + \dots + x_{Kit}\beta_K + u_{it},$$

kde  $t = 1, \dots, T$  označuje časový a  $i = 1, \dots, N$  průřezový rozměr pozorování.  $\beta_1, \dots, \beta_K$ ,  $K \in \mathbb{N}_0$ , označují koeficienty lineární regrese a  $u_{it}$ , označující reziduální (chybovou) složku regrese, jsou pro všechna  $i$  a  $t$  nezávislé stejně rozdělené náhodné veličiny se střední hodnotou 0 a rozptylem  $\sigma_u^2$ .  $x_{1it}, \dots, x_{Kit}$  potom značí vysvětlující proměnné v čase  $t$  pro skupinu  $i$ , které v této práci budeme považovat za náhodné. Předpokládá se, že reziduální složka není korelována s vysvětlujícími proměnnými. Tento model, někdy nazývaný poolový model, blíže prostudujeme ve druhé kapitole.

Poté se v této práci zaměříme převážně na studium takových panelových dat, jejichž průřezy lze považovat za nezávislé, ale pozorování v rámci průřezů (tedy přes časovou dimenzi) mohou být korelována. Soustředíme se přitom na dva z praktického hlediska významné typy modelů. Prvním z nich je tzv. model komponentních chyb, jemuž věnujeme třetí kapitolu, kde se zabýváme zejména jednoduchými modely s pevnými a náhodnými individuálními efekty.

Druhým typem jsou autoregresní modely prvního řádu pro panelová data, jejichž parametry a některé jejich vlastnosti odvozujeme ve čtvrté a páté kapitole. Zabýváme se přitom modely s pevným i modely s náhodným koeficientem. Na tyto kapitoly navazuje aplikace teoretických získaných poznatků na simulovaná i reálná data. Vybrané definice a věty používané v této práci jsou pak shrnuty v Příloze A.

Nyní však pro ucelení souvislostí nejprve popíšeme historický vývoj modelování panelových dat a uvedeme, jaké výhody či nevýhody modelování panelových dat přináší.

## 1.1 Historický přehled

Počátek modelování panelových dat lze vysledovat do druhé poloviny 19. století (viz Nerlove, 2000), kdy v roce 1861 George Biddell Airy aplikoval statistické metody na astronomická měření v podobě panelových dat. Konkrétně použil tzv. model s náhodnými efekty, stejně tak jako o dva roky později matematik a astronom William Chauvenet. Až na výjimky se však podrobnějšímu výzkumu začali matematici věnovat až ve druhé polovině 20. století. Širší využití potom panelová data kvůli početní náročnosti našla až s rozvojem výpočetní techniky.

Významným pokrokem byla v roce 1918 práce R. A. Fishera *The Correlation between Relatives on the Supposition of Mendelian Inheritance*, viz Fisher

(1918), jež se týkala kvantitativní genetiky a v níž zavedl pojmy „rozptyl“ a „analýza rozptylu“, které s panelovými daty též úzce souvisí, jak uvidíme později. V pozdějších pracích potom podrobně zpracoval koncept modelů s pevnými a náhodnými efekty, což jsou dva základní modely pro analýzu panelových dat. Fisher však tyto dva modely zřetelně nerozlišoval, jak poznamenává Nerlove (2000).

Až později Ellis Daniels v práci *The Estimation of Components of Variance* a Churchill Eisenhart v práci *The Assumptions Underlying the Analysis of Variance*, viz Daniels (1939) a Eisenhart (1947), navrhli, aby byl model s náhodnými efekty používán především pro neexperimentální (pozorovaná) data, např. v astronomické či ekonomické oblasti. Model s pevnými efekty je naopak vhodný pro experimentální data, kdy je navržen konkrétní (nenáhodný) postup pokusu, jako je tomu např. v zemědělství či průmyslu.

V roce 1950 v práci *Combining Cross Section Data and Time Series* Clifford Hildreth u reziduálního členu v modelu simultánních rovnic rozlišuje tři komponenty. Dvě z nich, časová a individuální složka, jsou pevné (a tedy jsou zároveň i parametry, které mají být odhadnuty) a zbytková rezidua jsou náhodná.

Autorem jedné z prvních ekonomických aplikací byl Irving Hoch, který v práci *Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data*, viz Hoch (1962), odhadoval Cobb-Douglasovu produkční funkci pro data shromážděná za šest let ze 63 farem v Minnesotě.

Poté se výzkum soustředil zejména na problematiku dynamických modelů (tj. modelů zahrnujících zpožděné vysvětlující nebo vysvětlované proměnné), jež jsou z hlediska uplatnění v praxi velmi důležité, neboť především modely ekonomického chování lze obvykle považovat za dynamické, ať už explicitně nebo implicitně.

Příkladem může být model vývoje poptávky po zemním plynu v jednotlivých státech USA využívající zpožděných vysvětlujících proměnných a pevných individuálních efektů, který navrhli Pietro Balestra a Marc Nerlove v článku *Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas* (viz Balestra a Nerlove, 1966).

V roce 1968 Institute for Social Research na Univerzity of Michigan začal provádět v rámci projektu Panel Study of Income Dynamics (PSID) kontinuální šetření domácností. Zahrnovalo reprezentativní vzorek (vzhledem ke všem obyvatelům USA) čítající přibližně 18 000 obyvatel žijících v 4 800 rodinách v USA. Tento průzkum probíhá dodnes. PSID mimo jiné sleduje údaje ekonomické (příjmy, výdaje), demografické (zdraví, úmrtnost) či sociolo-



gické. Sběr a následné zveřejnění získaných informací znamenalo významné zvýšení dostupnosti panelových datových souborů, což umožnilo širší studium této oblasti.

Později vznikaly podobné projekty i v jiných zemích, např. v roce 1984 German Socio-Economic Panel v Německu, v roce 1991 The British Household Panel Survey ve Spojeném království a v roce 1994 The European Community Household Panel sledující domácnosti v Evropské unii.

V roce 1977 se v Paříži konala první konference zabývající se panelovými daty, jež měla významný vliv na další vývoj výzkumu a na zvýšení zájmu o tuto problematiku. K tomu přispěla i lepší dostupnost datových souborů. Hlavní témata, kterými se výzkum po této konferenci zabíral, a to jak v teoretické, tak i v aplikované rovině, lze shrnout do následujících bodů:

- testy specifikace - Jerry A. Hausman v roce 1978 představil test, kterým lze vyhodnotit, zda model správně odpovídá datům - používá se pro rozlišení mezi modely s pevnými a náhodnými efekty (viz Hausman 1978);
- problém korelovaných efektů - studují se např. takové modely, kde jsou zbytková rezidua či časové efekty modelovány jako autoregresní procesy;
- alternativní odhady pro dynamické modely panelových dat - použití diferencí či zobecněné metody momentů (GMM);
- modelování diskretních, kvalitativních či cenzorovaných proměnných - zejména metodou maximální věrohodnosti nebo GMM;
- modelování rizika a durace;
- bayesovský přístup.

V červenci tohoto roku v Londýně proběhla již 19. mezinárodní konference na téma panelových dat, jež se zabývala mimo jiné nelineárními modely, stabilitou panelových modelů, závislostí mezi panely, předpověďmi, a také různými aplikacemi v energetice či ekonomii.

## 1.2 Aplikační přednosti a nedostatky

Možnosti aplikace modelů panelových dat jsou velmi rozsáhlé. Nacházejí své využití v ekonomii a v sociálních vědách, ale také např. v již zmíněné astronomii, zemědělství a průmyslu, medicíně či ekologii.

Modely panelových dat přináší několik významných výhod, jak uvádí např. Baltagi (2005). Zejména se tak získá větší množství pozorování a tím se dosáhne i větší různorodosti dat, než při použití čistě průřezových dat, či časových řad. Kromě toho se značně sníží nebezpečí multikolinearity mezi regresory a zvýší se počet stupňů volnosti. Větší množství informací umožňuje sledovat historii jednotlivých subjektů a studovat dynamiku změn. Panelová data navíc umožňují konstruovat takové modely, které dovolují testovat složitější hypotézy ohledně vzájemného chování.

V neposlední řadě potom modely panelových dat umožňují lépe analyzovat jak pozorované, tak i nepozorované individuální odlišnosti (heterogenitu). Například za použití modelů s pevnými efekty lze rozlišit, zda a do jaké míry se zbylá disturbance skládá z individuální složky, tedy časově invariantního efektu specifického pro daný průřez, či z časové složky, neboli průřezově invariantního efektu, který se mění v čase pro všechny jednotky stejně, viz Kapitola 3.

Naopak problémy nastávají při designu průzkumů a sběru dat - při šetření dochází v průběhu let k postupnému úbytku respondentů nebo ke změně složení zkoumaného souboru subjektů. Problematictější je i případná imputace chybějících dat. Kromě toho může být časový rozměr příliš krátký na to, aby bylo možné data smysluplně modelovat panelově. Přesto v mnoha případech pozitiva převažují a užití modelových dat je stále častější.

## Kapitola 2

### Poolový model

Než přistoupíme k modelům se složitější strukturou, popíšeme v této kapitole podrobněji základní model zmíněný v úvodní části, neboť na něj budeme v následující kapitole odkazovat při testování některých hypotéz. Připomeňme si podobu tohoto modelu:

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_K x_{itK} + u_{it}, \quad (2.1)$$

kde  $K$  značí počet regresorů a  $x_{it1}, \dots, x_{itK}$  jsou (náhodné) regresory,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ .

Tzv. poolový model (2.1) je ze zde uvažovaných modelů nejjednodušší. Předpokládá se, že celková rezidua nejsou strukturovaná - neuvažují se individuální ani časové efekty. Předpokládá se, že  $u_{it}$  jsou nezávislé stejně rozdělené náhodné veličiny se střední hodnotou 0 a rozptylem  $\sigma_u^2 > 0$  a že tato rezidua nejsou korelovaná s regresory.

V takovém případě se všechna pozorování zahrnou do jedné regrese tak, jako kdyby data neměla panelovou strukturu. Parametry se pak odhadují standardní metodou nejmenších čtverců, viz např. Anděl (2007, kap. 5). Přepíšeme nejprve model (2.1) do maticové podoby

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

kde  $\mathbf{Y} = (y_{11}, y_{12}, \dots, y_{1T}, \dots, y_{NT})'$ ,  $\mathbf{U} = (u_{11}, u_{12}, \dots, u_{1T}, \dots, u_{NT})'$ ,  $\mathbf{X} = (\mathbf{x}'_{11}, \mathbf{x}'_{12}, \dots, \mathbf{x}'_{1T}, \dots, \mathbf{x}'_{NT})'$ ,  $\mathbf{x}'_{it} = (1, x_{it1}, \dots, x_{itK})$  a  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$ . Při použití metody nejmenších čtverců se minimalizuje výraz

$$S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Z rovnice

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

dostaneme odhad parametrů  $\boldsymbol{\beta}$ , a to  $\mathbf{b} = (b_1, \dots, b_K)'$  =  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , za předpokladu, že použitá inverzní matice je regulární. V případě, že je rozptyl  $\sigma_u^2$  také neznámý, odhadne se obvyklým způsobem pomocí výběrového rozptylu

$$s_u^2 = \frac{1}{NT - K}(\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}).$$

Tento model je vhodný zejména pro tzv. poolovaná data, kdy je v každém období vyšetřován nový náhodný výběr. Vznikají tak nezávislé poolované průřezy, na které může být použita metoda nejmenších čtverců. Výhodou tohoto postupu je zvýšení velikosti výběru, což vede k přesnějším odhadům parametrů a k větší síle testů.

## Kapitola 3

# Modely komponentních chyb

Modely komponentních chyb (*error component models*) jsou díky strukturovaným celkovým reziduíům základním prostředkem pro modelování nepozorované heterogenity dat. Tyto modely nejsou tolik striktní ve svých předpokladech jako poolový model (2.1). Umožňují například, aby každému průřezu příslušel vlastní intercept, a tak mnohdy lépe modelují pozorovaná data.

U jednoduchého modelu komponentních chyb (*one-way error component model*) se předpokládá, že se celková rezidua skládají jednak z klasických zbytkových reziduí, tj. z nezávislých náhodných veličin s nulovou střední hodnotou a konstantním rozptylem, ke kterým se nyní navíc přidává buď průřezově specifická složka, nebo časově specifická složka. U oboustranného modelu komponentních chyb (*two-way error component model*) pak uvažujeme všechny tři tyto složky současně.

Tyto individuální či časové složky lze přitom podle situace považovat buď za nenáhodné parametry, které lze odhadnout (pak se jedná o model s pevnými efekty) nebo se považují za náhodné veličiny (pak jde o model s náhodnými efekty). V podstatě se jedná o kombinaci regresní analýzy a analýzy rozptylu, tj. o analýzu kovariance.

V této kapitole nejprve popíšeme jednoduchý model komponentních chyb pro nezávislé průřezy, a to jak s pevnými, tak s náhodnými individuálními efekty, včetně Hausmanova specifikačního testu a testů významnosti těchto efektů. Poté stručně popíšeme rozšíření na oboustranný model komponentních chyb.

### 3.1 Jednoduchý model komponentních chyb

V této části se tedy budeme zabývat jednoduchým modelem komponentních chyb s individuálními efekty  $\mu_i$ :

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}, \quad u_{it} = \mu_i + \varepsilon_{it},$$

kde  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$  jsou parametry regrese,  $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itK})'$  jsou (náhodné) regresory a  $\varepsilon_{it}$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ , značí nezávislá a stejně rozdělená zbytková rezidua, pro něž předpokládáme

$$\begin{aligned} E(\varepsilon_{it}) &= 0, \\ E(\varepsilon_{it}\varepsilon_{js}) &= \sigma_\varepsilon^2 > 0 && i = j, t = s, \\ &= 0 && \text{jinak.} \end{aligned} \tag{3.1}$$

Individuální efekty  $\mu_i$ , které lze podle situace považovat buď za náhodné, nebo za pevné, vlastně odrážejí nepozorovatelné nebo nepozorované vlivy invariantní v čase, které jsou specifické pro jednotlivé subjekty. Může jít také o vlivy těch regresorů, které jsou samostatně nevýznamné, avšak společně ovlivňují hodnoty závisle proměnné. Příkladem takových vlivů může být např. pohlaví, národnost, nepozorované individuální schopnosti nebo třeba management firmy či způsob výpočtu ekonomických ukazatelů v dané zemi. Složky  $\varepsilon_{it}$  jsou potom zbylé nevysvětlené efekty měnící se jak s časem  $t$ , tak s individuí  $i$  - tedy klasická rezidua.

Nyní popíšeme ony dva již zmíněné základní přístupy, jak lze k modelům s efekty přistupovat. Nejprve se zaměříme na model s pevnými efekty a odhad jeho parametrů, poté se soustředíme na model s náhodnými efekty. Současně uvedeme i testy významnosti efektů a Hausmanův specifikační test.

Dále se ovšem v kapitole o jednoduchých modelech budeme zabývat pouze modely s individuálními efekty. V případě modelu (3.2) s časovými efekty  $\lambda_t$ , reprezentujícími časově specifické vlivy ovlivňující všechny subjekty stejnou měrou, se pak postupuje zcela analogicky. Příkladem časových efektů může být výše cen nebo úrokových měr, celková nálada na trhu, politika vlády apod.

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \lambda_t + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{3.2}$$

### 3.1.1 Jednoduchý model s pevnými efekty

V modelu s pevnými<sup>1</sup> efekty (*fixed-effects model*)

$$\begin{aligned}y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}, \\u_{it} &= \mu_i + \varepsilon_{it}\end{aligned}\tag{3.3}$$

považujeme individuální efekty  $\mu_i$ ,  $i = 1, \dots, N$ , za nenáhodné parametry, které je možné odhadnout stejně jako koeficienty  $\boldsymbol{\beta}$ . Zároveň připouštíme, že členy  $u_{it}$  mohou být korelované s ostatními regresory, nikoli však se zbytkovou chybou.

Stále předpokládáme, že  $\varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2)$ <sup>2</sup>. Metod pro odhad parametrů tohoto modelu je více, my zde uvedeme dva typické přístupy - metodu nejmenších čtverců s tzv. dummy proměnnými a odhad pomocí vnitrotřídního modelu. Zaměříme se nejprve na druhý jmenovaný.

#### Vnitrotřídní model

Vnitrotřídní model (*within-groups model*) měří vztah mezi individuálními odchylkami proměnných od časových průměrů. Získá se snadno tak, že se nejprve od modelování výchozích proměnných v modelu

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mu_i + \varepsilon_{it}\tag{3.4}$$

přejde k modelování jejich časových průměrů:

$$\bar{y}_i = \bar{\mathbf{x}}'_i\boldsymbol{\beta} + \mu_i + \bar{\varepsilon}_i,\tag{3.5}$$

kde

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it} \quad \text{a} \quad \bar{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}$$

a podobně

$$\bar{\mathbf{x}}_i = \left( \frac{1}{T} \sum_{t=1}^T x_{it1}, \frac{1}{T} \sum_{t=1}^T x_{it2}, \dots, \frac{1}{T} \sum_{t=1}^T x_{itK} \right)', \quad i = 1, \dots, N.$$

<sup>1</sup>Někdy se také v češtině používá výraz „fixní efekty“.

<sup>2</sup>Označení „iid“ je zkratkou anglického *independent and identically distributed*, tzn. nezávislé a stejně rozdělené.

Následně se od modelu (3.4) odečte model (3.5), čímž se využije variace dat v čase. Přecházíme tak k vnitrotřídnímu modelu

$$y_{it} - \bar{y}_i = \beta(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i), \quad (3.6)$$

který již neobsahuje individuální efekty. Je dobré si uvědomit, že touto úpravou v modelu (3.6) přicházíme o nezávislost reziduí  $(\varepsilon_{it} - \bar{\varepsilon}_i)$ . Přesto se obvykle postupuje dál tak, jako kdyby rezidua nezávislá byla. Model (3.6) zapišme maticově

$$\mathbf{Y}_W = \mathbf{X}_W \beta + \boldsymbol{\varepsilon}_W,$$

kde  $\mathbf{Y}_W = (y_{11} - \bar{y}_1, y_{12} - \bar{y}_1, \dots, y_{NT} - \bar{y}_N)'$ ,  $\boldsymbol{\varepsilon}_W = (\varepsilon_{11} - \bar{\varepsilon}_1, \varepsilon_{12} - \bar{\varepsilon}_1, \dots, \varepsilon_{NT} - \bar{\varepsilon}_N)'$  a

$$\mathbf{X}_W = \begin{pmatrix} (\mathbf{x}_{11} - \bar{\mathbf{x}}_1)' \\ (\mathbf{x}_{12} - \bar{\mathbf{x}}_1)' \\ \dots \\ (\mathbf{x}_{NT} - \bar{\mathbf{x}}_N)' \end{pmatrix}.$$

Parametry tohoto modelu lze odhadnout metodou nejmenších čtverců, tj. minimalizací výrazu

$$S(\beta) = (\mathbf{Y}_W - \mathbf{X}_W \beta)' (\mathbf{Y}_W - \mathbf{X}_W \beta).$$

Získáme tak vnitrotřídní odhad koeficientů

$$\hat{\beta}_W = (\mathbf{X}'_W \mathbf{X}_W)^{-1} \mathbf{X}'_W \mathbf{Y}_W, \quad (3.7)$$

existuje-li uvedená inverzní matice.

Z modelu (3.5) za použití odhadu  $\hat{\beta}_W$  potom snadno dopočteme odhady individuálních efektů:

$$\hat{\mu}_{i,W} = \bar{y}_i - \bar{\mathbf{x}}'_i \hat{\beta}_W. \quad (3.8)$$

Všimněme si ještě, že v modelu (3.6) se průměrováním a následným odečtením vytrácí případné další časově invariantní proměnné, takže koeficient takového regresoru nelze tímto způsobem odhadnout. V případě, že by model obsahoval intercept, označme ho pro tento případ  $\alpha$ , lze k jeho odhadu využít model celkových průměrů:

$$\bar{y} = \alpha + \beta \bar{\mathbf{x}} + \bar{\varepsilon}, \quad (3.9)$$



kde  $\bar{\mathbf{x}} = \left( \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T x_{it1}, \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T x_{it2}, \dots, \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T x_{itK} \right)'$ ,  
 $\bar{y} = \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T y_{it}$ ,  $\bar{\varepsilon} = \frac{1}{NT} \sum_{n=1}^N \sum_{t=1}^T \varepsilon_{it}$  a kde využíváme (zvolené) re-  
 strikce

$$\sum_{i=1}^N \mu_i = 0.$$

Jinak je v případech s časově invariantními proměnnými výhodnější použít alternativní způsob odhadu, např. metodu nejmenších čtverců s dummy proměnnými.

### Metoda nejmenších čtverců s dummy proměnnými

Poněkud jednodušší a přímější metodu nejmenších čtverců s dummy, tj. pomocnými, proměnnými (*least squares dummy variables, LSDV*) pro odhad  $\beta$  a  $\mu_i$  je vhodné použít, zejména pokud  $N$  není příliš vysoké.

Zaveďme do modelu (3.3) pomocné proměnné  $d_{hit}$ , pro které

$$\begin{aligned} d_{hit} &= 1 & i &= h, \\ &= 0 & \text{jinak,} \\ & & i, h &= 1, \dots, N, t = 1, \dots, T. \end{aligned}$$

Přejdeme tak k alternativnímu zápisu modelu (3.3):

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it} \beta + \mu_1 d_{1it} + \mu_2 d_{2it} + \dots + \mu_N d_{Nit} + \varepsilon_{it}, \\ i &= 1, \dots, N, t = 1, \dots, T. \end{aligned} \quad (3.10)$$

Maticově tento model pak zapíšeme jako

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{D}_\mu \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \quad (3.11)$$

kde  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)'$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \dots, \varepsilon_{NT})'$  a  $\mathbf{D}_\mu = \mathbf{I}_N \otimes \mathbf{1}_T$  označuje matici pomocných proměnných o rozměrech  $NT \times N$  ( $\mathbf{I}_N$  je jednotková matice dimenze  $N$ ,  $\mathbf{1}_T$  jednotkový vektor dimenze  $T$  a  $\otimes$  značí Kroneckerův součin, viz Definice A.1). Parametry tohoto modelu lze taktéž odhadnout metodou nejmenších čtverců, kdy se minimalizuje výraz

$$S(\beta, \boldsymbol{\mu}) = (\mathbf{Y} - \mathbf{X}\beta - \mathbf{D}_\mu \boldsymbol{\mu})' (\mathbf{Y} - \mathbf{X}\beta - \mathbf{D}_\mu \boldsymbol{\mu}).$$

Po zderivování podle příslušných proměnných sestavíme rovnice

$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\mu})}{\partial \boldsymbol{\beta}} = \mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}_\mu\boldsymbol{\mu}) = \mathbf{0}, \quad (3.12)$$

$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \mathbf{D}'_\mu(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{D}_\mu\boldsymbol{\mu}) = \mathbf{0}. \quad (3.13)$$

Z tvaru matice pomocných proměnných  $\mathbf{D}_\mu$  je zřejmé, že matice  $(\mathbf{D}'_\mu\mathbf{D}_\mu)^{-1}$  existuje. Z (3.13) tak dostaneme

$$\hat{\boldsymbol{\mu}}_{LSDV} = (\mathbf{D}'_\mu\mathbf{D}_\mu)^{-1}\mathbf{D}'_\mu(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LSDV}), \quad (3.14)$$

což dosadíme do rovnice z (3.12)

$$\begin{aligned} \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{LSDV} &= \mathbf{X}'(\mathbf{Y} - \mathbf{D}_\mu\hat{\boldsymbol{\mu}}_{LSDV}) \\ &= \mathbf{X}'\left[\mathbf{Y} - \mathbf{D}_\mu(\mathbf{D}'_\mu\mathbf{D}_\mu)^{-1}\mathbf{D}'_\mu\mathbf{Y} + \mathbf{D}_\mu(\mathbf{D}'_\mu\mathbf{D}_\mu)^{-1}\mathbf{D}'_\mu\mathbf{X}\hat{\boldsymbol{\beta}}_{LSDV}\right]. \end{aligned}$$

Odtud při označení tzv. projekční matice  $\mathbf{Q}_D = \mathbf{I}_{NT} - \mathbf{D}_\mu(\mathbf{D}'_\mu\mathbf{D}_\mu)^{-1}\mathbf{D}'_\mu$  dostáváme odhad

$$\hat{\boldsymbol{\beta}}_{LSDV} = (\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_D\mathbf{Y}, \quad (3.15)$$

Symetricky se odvodí i odhad individuálních efektů:

$$\hat{\boldsymbol{\mu}}_{LSDV} = (\mathbf{D}'_\mu\mathbf{Q}_X\mathbf{D}_\mu)^{-1}\mathbf{D}'_\mu\mathbf{Q}_X\mathbf{Y},$$

kde  $\mathbf{Q}_X = \mathbf{I}_{NT} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , přičemž předpokládáme existenci použitých inverzních matic.

Tento postup je výhodnější spíše při nižších počtech průřezů. Jestliže je totiž počet individuálních efektů  $N$  příliš vysoký, může být problémem invertování matice dimenze  $N + K$ . V takovém případě je vhodnější použít spíše postup s využitím vnitrotřídního odhadu.

Oba tyto postupy, jak LSDV, tak vnitrotřídní regrese, přitom dávají ekvivalentní odhady  $\boldsymbol{\beta}$ . Upravme nejprve vyjádření pro vnitrotřídní odhad  $\hat{\boldsymbol{\beta}}_W = (\mathbf{X}'_W\mathbf{X}_W)^{-1}\mathbf{X}'_W\mathbf{Y}_W$  z (3.7). Platí totiž

$$\mathbf{X}_W = \mathbf{Q}_W\mathbf{X} \text{ a } \mathbf{Y}_W = \mathbf{Q}_W\mathbf{Y},$$

kde  $\mathbf{Q}_W = \mathbf{I}_T - \mathbf{1}_T(\mathbf{1}'_T \mathbf{1}_T)^{-1} \mathbf{1}'_T$  je symetrická a idempotentní matice (viz Anděl (2007, str. 81)). Díky tomu dostáváme

$$\begin{aligned}\hat{\beta}_W &= [(\mathbf{Q}_W \mathbf{X})'(\mathbf{Q}_W \mathbf{X})]^{-1} (\mathbf{Q}_W \mathbf{X})' \mathbf{Q}_W \mathbf{Y} \\ &= \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{Q}_W \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{Q}_W \mathbf{Y}_i,\end{aligned}\quad (3.16)$$

kde  $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$ ,  $i = 1, \dots, N$ , a  $\mathbf{X}_i = (\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \dots, \mathbf{x}'_{iT})'$ .

Odhad  $\hat{\beta}_{LSDV} = (\mathbf{X}' \mathbf{Q}_D \mathbf{X})^{-1} \mathbf{X}' \mathbf{Q}_D \mathbf{Y}$  z (3.15) také můžeme dále upravit. Platí

$$\begin{aligned}\mathbf{Q}_D &= \mathbf{I}_{NT} - \mathbf{D}_\mu (\mathbf{D}'_\mu \mathbf{D}_\mu)^{-1} \mathbf{D}'_\mu \\ &= \mathbf{I}_{NT} - (\mathbf{I}_N \otimes \mathbf{1}_T) [(\mathbf{I}_N \otimes \mathbf{1}_T)' (\mathbf{I}_N \otimes \mathbf{1}_T)]^{-1} (\mathbf{I}_N \otimes \mathbf{1}_T)' \\ &= \mathbf{I}_{NT} - (\mathbf{I}_N \otimes \mathbf{1}_T) (\mathbf{I}_N \otimes \mathbf{1}'_T \mathbf{1}_T)^{-1} (\mathbf{I}_N \otimes \mathbf{1}_T)' \\ &= \mathbf{I}_{NT} - (\mathbf{I}_N \otimes \mathbf{1}_T) [\mathbf{I}_N \otimes (\mathbf{1}'_T \mathbf{1}_T)^{-1}] (\mathbf{I}_N \otimes \mathbf{1}_T)' \\ &= \mathbf{I}_{NT} \otimes (\mathbf{I}_T - \mathbf{1}_T (\mathbf{1}'_T \mathbf{1}_T)^{-1} \mathbf{1}'_T) = \mathbf{I}_{NT} \otimes \mathbf{Q}_W.\end{aligned}$$

Odtud dostáváme

$$\begin{aligned}\hat{\beta}_{LSDV} &= [\mathbf{X}' (\mathbf{I}_{NT} \otimes \mathbf{Q}_W) \mathbf{X}]^{-1} \mathbf{X}' (\mathbf{I}_{NT} \otimes \mathbf{Q}_W) \mathbf{Y} \\ &= \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{Q}_W \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{Q}_W \mathbf{Y}_i,\end{aligned}\quad (3.17)$$

což je stejný výsledek jako (3.16). Oba odhady jsou tedy skutečně ekvivalentní. Proto je dále budeme označovat jednotně  $\hat{\beta}_{FE}$ .

Z ekvivalence těchto odhadů  $\beta$  však také přímo plyne ekvivalence odhadů efektů  $\mu$ . Stačí si uvědomit, že po rozepsání (3.14) po složkách dostaneme přímo (3.8). Oba odhady pevných efektů tak také můžeme označit jednotně  $\hat{\mu}_{FE}$ .

Počet stupňů volnosti je v obou případech  $NT - N - K$ . U vnitrotřídního odhadu je od standardních  $NT - K$  nutno odečíst dalších  $N$  pro zohlednění pevných efektů. Pak lze jako odhad  $\sigma_\varepsilon^2$  použít výběrový rozptyl:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{NT - N - K} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{x}'_{it} \hat{\beta}_{FE} - \hat{\mu}_{i,LSDV})^2.$$

Jelikož odhad  $\hat{\beta}_{FE}$  má tvar Aitkenova odhadu, viz Zvára (2008, str. 23), můžeme vyjádřit rozptyl tohoto odhadu jako

$$\text{var}(\hat{\beta}_{FE}|\mathbf{X}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1}, \quad (3.18)$$

v případě neznámého rozptylu pak jako

$$\text{var}(\hat{\beta}_{FE}|\mathbf{X}) = \hat{\sigma}_\varepsilon^2(\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1}.$$

Doplňme ještě, že odhad  $\hat{\beta}_{FE}$  je konzistentním odhadem  $\beta$  při  $N, T \rightarrow \infty$ , jestliže platí následující předpoklady:

- (i)  $(NT)^{-1}(\mathbf{X}'\mathbf{Q}_D\mathbf{X})$  konverguje při  $N, T \rightarrow \infty$  podle pravděpodobnosti k pozitivně definitní konečné matici a
- (ii)  $(NT)^{-1}(\mathbf{X}'\mathbf{Q}_D\varepsilon) \xrightarrow{p} 0, N, T \rightarrow \infty$ .

Pak totiž

$$\text{plim}_{N,T \rightarrow \infty} \hat{\beta}_{FE} = \text{plim}_{N,T \rightarrow \infty} \left[ \beta + \left( \frac{(\mathbf{X}'\mathbf{Q}_D\mathbf{X})}{NT} \right)^{-1} \frac{\mathbf{X}'\mathbf{Q}_D\varepsilon}{NT} \right] = \beta. \quad (3.19)$$

Obdobný postup lze zopakovat i pro důkaz konzistence odhadu individuálních efektů, avšak pouze pro  $T \rightarrow \infty$  a pevné  $N$ . Při  $N \rightarrow \infty$  odhad parametru  $\mu$  není konzistentní - jedná se o tzv. incidentální parametr, což znamená, že jeho dimenze se zvyšuje s velikostí výběru, viz Neyman a Scott (1948).

Dosavadní poznatky o odhadech parametrů v jednoduchých modelech s pevnými efekty můžeme shrnout do následující věty:

**Věta 3.1.** *Odhady parametrů zobecněnou metodou nejmenších čtverců v modelu (3.3) s pevnými efekty s předpoklady (3.1) jsou*

$$\begin{aligned} \hat{\beta}_{FE} &= (\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_D\mathbf{Y} \\ \hat{\mu}_{LSDV} &= (\mathbf{D}'_\mu\mathbf{Q}_X\mathbf{D}_\mu)^{-1}\mathbf{D}'_\mu\mathbf{Q}_X\mathbf{Y} \end{aligned}$$

za předpokladu, že uvedené inverzní matice existují. Rozptyly těchto odhadů jsou pak rovny

$$\text{var}(\hat{\beta}_{FE}|\mathbf{X}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{Q}_D\mathbf{X})^{-1} \quad \text{a} \quad \text{var}(\hat{\mu}_{FE}|\mathbf{X}) = \sigma_\varepsilon^2(\mathbf{D}'_\mu\mathbf{Q}_X\mathbf{D}_\mu)^{-1}.$$

Předpokládáme-li navíc, že

(i)  $(NT)^{-1}(\mathbf{X}'\mathbf{Q}_D\mathbf{X})$  konverguje při  $N, T \rightarrow \infty$  podle pravděpodobnosti k pozitivně definitní konečné matici  $\mathbf{a}$

(ii)  $(NT)^{-1}(\mathbf{X}'\mathbf{Q}_D\boldsymbol{\varepsilon}) \xrightarrow{p} 0, N, T \rightarrow \infty,$

pak  $\hat{\boldsymbol{\beta}}_{FE}$  je konzistentním odhadem  $\boldsymbol{\beta}$  pro  $N, T \rightarrow \infty$ .

Podobně za předpokladů

(i)  $(T)^{-1}(\mathbf{D}'_{\mu}\mathbf{Q}_X\mathbf{D}_{\mu})$  konverguje  $T \rightarrow \infty$  podle pravděpodobnosti k pozitivně definitní konečné matici  $\mathbf{a}$

(ii)  $(T)^{-1}(\mathbf{D}'_{\mu}\mathbf{Q}_X\boldsymbol{\varepsilon}) \xrightarrow{p} 0, T \rightarrow \infty$

je  $\hat{\boldsymbol{\mu}}_{LSDV}$  konzistentním odhadem  $\boldsymbol{\mu}$  pro  $T \rightarrow \infty$ .

### Testování významnosti pevných efektů

Sdruženou významnost efektů v modelu s pevnými efekty (3.3) lze za dodatečného předpokladu  $\varepsilon_{it} \sim N(0, \sigma_{\varepsilon}^2), i = 1, \dots, N, t = 1, \dots, T$ , testovat F-testem, který je analogický F-testu pro analýzu rozptylu jednoduchého třídění<sup>3</sup>, viz např. Anděl (2007, str. 210). Testuje se hypotéza

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_N = 0.$$

Označme

$$RSS_0 = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

reziduální součet čtverců v poolovém modelu (2.1) odhadnutém metodou nejmenších čtverců, a

$$RSS = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{FE})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{FE})$$

reziduální součet čtverců v modelu s pevnými efekty. Pak má testová statistika

$$F_0 = \frac{NT - N - K}{N} \frac{(RSS_0 - RSS)}{RSS} \quad (3.20)$$

<sup>3</sup>V případě, že bychom v modelu (3.3) nepoužili žádné regresory, jednalo by se o model jednoduchého třídění analýzy rozptylu.

za platnosti nulové hypotézy F rozdělení s  $N$  a  $NT - N - K$  stupni volnosti (viz Baltagi (2005)). Nulovou hypotézu tak zamítneme na hladině  $\alpha$ , jestliže

$$F_0 \geq F_{1-\alpha, N, NT-N-K},$$

kde  $F_{1-\alpha, N, NT-N-K}$  je příslušný kvantil F rozdělení. V takovém případě je vhodnější použít model s pevnými efekty namísto poolového modelu.

### 3.1.2 Jednoduchý model s náhodnými efekty

Jestliže je počet průřezových jednotek v datovém souboru velmi vysoký, je při použití pevných efektů vysoký i počet parametrů, které je třeba odhadnout (koeficienty regresorů a pevný efekt pro každý průřez). Tím ovšem ztrácíme stupně volnosti. Proto je někdy výhodnější pohlížet na individuální efekty jako na nezávislé stejně rozdělené náhodné veličiny.

Uvažujme tedy model

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}, \\ u_{it} &= \mu_i + \varepsilon_{it}, i = 1, \dots, N, t = 1, \dots, T, \end{aligned} \quad (3.21)$$

ve kterém platí

$$\begin{aligned} \varepsilon_{it} &\sim \text{iid}(0, \sigma_\varepsilon^2 > 0), \\ \mu_i &\sim \text{iid}(0, \sigma_\mu^2 > 0), \\ E(\mu_i \mu_j) &= 0, i \neq j, \\ E(\varepsilon_{it}, \mu_j) &= 0, i, j = 1, \dots, N, t = 1, \dots, T. \end{aligned} \quad (3.22)$$

Kromě toho budeme předpokládat, že pro všechna  $i$  a  $t$  jsou  $\mu_i$  a  $\varepsilon_{it}$  nezávislé náhodné veličiny a že regresory  $x_{it}$  jsou nekorelované s  $\mu_i$  a  $\varepsilon_{it}$ .

Vzhledem k nezávislosti průřezů, jsou i celkové chyby příslušné různým subjektům nezávislé. Přítomnost další náhodné složky v celkové disturbanci však způsobuje její časovou korelovanost:

$$\begin{aligned} \text{cov}(u_{it}, u_{is}) &= \text{cov}(\mu_i + \varepsilon_{it}, \mu_i + \varepsilon_{is}) \\ &= \text{var}(\mu_i) + \text{cov}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_\mu^2 \neq 0, t \neq s, \end{aligned}$$

což znamená

$$\begin{aligned} \text{corr}(u_{it}, u_{js}) &= \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} && i = j, t \neq s, \\ &= 1 && i = j, t = s, \\ &= 0 && \text{jinak.} \end{aligned}$$

Varianční matice disturbancí tak má za použití značení zavedeném u (3.11) tvar

$$\Sigma = E(\mathbf{u}\mathbf{u}') = \mathbf{D}_\mu E(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{D}_\mu' + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\varepsilon^2(\mathbf{I}_N \otimes \mathbf{I}_T), \quad (3.23)$$

kde  $\mathbf{J}_T$  je čtvercová matice jedniček dimenze  $T$ . Je tak porušen základní předpoklad pro použití standardní metody nejmenších čtverců, totiž nekorelovanost reziduí, což v důsledku může vést například ke zkreslení výsledků některých testů, jak uvádí např. Cipra (2008, kap. 4.3). Proto se v takovém případě použije zobecněná metoda nejmenších čtverců (*generalized least squares*, GLS), neboli Aitkenův odhad, viz např. Zvára (2008, str. 23)).

Předpokládejme nejprve, že jsou rozptyly  $\sigma_\varepsilon^2$  a  $\sigma_\mu^2$  známé. Pak přímo dostaneme Aitkenův odhad

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}, \quad (3.24)$$

existují-li uvedené inverzní matice. Tento odhad je také konzistentním odhadem  $\boldsymbol{\beta}$  při  $N, T \rightarrow \infty$ , jestliže

$$\begin{aligned} \text{plim}_{N,T \rightarrow \infty} \frac{\mathbf{X}'\Sigma^{-1}\mathbf{u}}{NT} &= \mathbf{0}, \\ \text{plim}_{N,T \rightarrow \infty} \frac{\mathbf{X}'\Sigma^{-1}\mathbf{X}}{NT} &= \mathbf{W}, \end{aligned} \quad (3.25)$$

kde  $\mathbf{W}$  je pozitivně definitní konečná matice. Pro důkaz stačí zopakovat stejný postup jako ve (3.19). Rozptyl  $\hat{\boldsymbol{\beta}}_{GLS}$  se potom odhadne jako

$$\text{var}(\hat{\boldsymbol{\beta}}_{GLS}|\mathbf{X}) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}. \quad (3.26)$$

$\Sigma$ , matice dimenze  $NT$ , přitom typicky bývá velká. Proto je při invertování této matice výhodnější využít spektrální rozklad

$$\Sigma = (T\sigma_\mu^2 + \sigma_\varepsilon^2)\mathbf{H} + \sigma_\varepsilon^2\mathbf{M},$$

kde  $\mathbf{H} = \mathbf{D}_\mu(\mathbf{D}_\mu'\mathbf{D}_\mu)^{-1}\mathbf{D}_\mu'$  a  $\mathbf{M} = \mathbf{I}_{NT} - \mathbf{H}$ . Pro podrobné odvození viz např. Baltagi (2005, str. 15). Odtud pak snadno dostaneme inverzi matice  $\Sigma$ :

$$\Sigma^{-1} = \frac{1}{(T\sigma_\mu^2 + \sigma_\varepsilon^2)}\mathbf{H} + \frac{1}{\sigma_\varepsilon^2}\mathbf{M}.$$

V případě, že je varianční matice  $\Sigma$  také neznámá, je třeba nalézt její konzistentní odhad  $\hat{\Sigma}$ , aby mohl být parametr  $\beta$  odhadnut pomocí tzv. přípustného Aitkenova odhadu (*feasible generalized least squares, FGLS*), viz Greene (2002, kap. 10.5).

Díky tvaru  $\Sigma$  vyjádřeném v (3.23) stačí, když konzistentně odhadneme rozptyly  $\sigma_\mu^2$  a  $\sigma_\varepsilon^2$ . To lze např. využitím vnitrotřídní a mezitřídní regrese. Nejprve se zaměříme na vnitrotřídní regresní model (viz (3.6)):

$$y_{it} - \bar{y}_i = \beta(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i), \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

V tomto modelu jsme již dříve odhadli  $\beta$  jako  $\hat{\beta}_W = (\mathbf{X}'_W \mathbf{X}_W)^{-1} \mathbf{X}'_W \mathbf{Y}_W$ , kde

$$\mathbf{X}_W = \begin{pmatrix} (\mathbf{x}_{11} - \bar{\mathbf{x}}_1)' \\ (\mathbf{x}_{12} - \bar{\mathbf{x}}_1)' \\ \dots \\ (\mathbf{x}_{NT} - \bar{\mathbf{x}}_N)' \end{pmatrix}.$$

Protože  $\text{var}(\varepsilon_{it} - \bar{\varepsilon}_i) = \text{E}(\varepsilon_{it} - \bar{\varepsilon}_i)^2 = \sigma_\varepsilon^2$ , odhadneme tento rozptyl pomocí výběrového rozptylu

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_{within}^2 = \frac{1}{NT - N - K} \sum_{i=1}^N \sum_{t=1}^T \left[ y_{it} - \bar{y}_i - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \hat{\beta}_W \right]^2. \quad (3.27)$$

Odhad  $\sigma_\mu^2$  potom dostaneme z modelu mezitřídní regrese (3.5)

$$\bar{y}_i = \bar{\mathbf{x}}_i' \beta + \mu_i + \bar{\varepsilon}_i, \quad (3.28)$$

v němž pro celková rezidua vzhledem k předpokladům (3.22) platí

$$\text{var}(\mu_i + \bar{\varepsilon}_i) = \sigma_{between}^2 = \sigma_\mu^2 + \frac{\sigma_\varepsilon^2}{T}. \quad (3.29)$$

Rozptyl  $\sigma_{between}^2$  přitom opět můžeme odhadnout výběrovým rozptylem

$$\hat{\sigma}_{between}^2 = \frac{1}{(N - K)} \sum_{i=1}^N (\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}_B)^2,$$

kde  $\hat{\beta}_B = (\bar{\mathbf{X}}' \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}' \bar{\mathbf{Y}}$ ,  $\bar{\mathbf{Y}} = (\bar{y}_1, \dots, \bar{y}_N)'$  a  $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1', \dots, \bar{\mathbf{x}}_N')'$ . Z (3.29) potom můžeme vyjádřit

$$\hat{\sigma}_\mu^2 = \hat{\sigma}_{between}^2 - \frac{\hat{\sigma}_\varepsilon^2}{T}, \quad (3.30)$$



za předpokladu, že je tento rozdíl nezáporný. V opačném případě definujeme  $\hat{\sigma}_\mu^2 = 0$ .

Výběrové rozptyly  $\hat{\sigma}_\varepsilon^2$ ,  $\hat{\sigma}_{between}^2$  i odhad  $\hat{\sigma}_\mu^2$  jsou za rutinních předpokladů konzistentními odhady  $\sigma_\varepsilon^2$ ,  $\sigma_{between}^2$  a  $\sigma_\mu^2$ . Zde ukážeme konzistenci  $\hat{\sigma}_\varepsilon^2$ , a to za předpokladu, že  $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E\varepsilon_{it}^4 \rightarrow 0$  a že

$$\begin{aligned} & \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^2, \\ & \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \end{aligned}$$

konvergují při  $N, T \rightarrow \infty$  podle pravděpodobnosti ke konečné hodnotě. Upravme

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 = & \frac{NT}{NT - N - K} \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 + \right. \\ & \left. + \frac{2}{NT} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_W) \sum_{i=1}^N \sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \frac{1}{NT} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_W)^2 \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^2 \right]. \end{aligned}$$

Vzhledem k tomu, že  $\hat{\boldsymbol{\beta}}_W$  je konzistentním odhadem  $\boldsymbol{\beta}$ , viz (3.19), stačí už jen dokázat, že

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 \xrightarrow{p} \sigma_\varepsilon^2$$

při  $N, T \rightarrow \infty$ . Po úpravě dostaneme

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}^2 - \left( \frac{1}{T} \sum_{t=1}^T \varepsilon_{it} \right)^2 \right].$$

Protože  $\varepsilon_{it}$ ,  $t = 1, \dots, T = T_N$ , jsou nezávislé stejně rozdělené náhodné veličiny se střední hodnotou 0 a rozptylem  $\sigma_\varepsilon^2$ , podle slabého zákona velkých čísel A.10 platí

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it} \xrightarrow{p} 0, \quad T \rightarrow \infty, \quad i = 1, \dots, N.$$

Podobně protože  $\varepsilon_{it}^2$  jsou podle (3.22) nezávislé stejně rozdělené náhodné veličiny se střední hodnotou  $\sigma_\varepsilon^2$ , platí za našich předpokladů

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it}^2 \xrightarrow{p} \sigma_\varepsilon^2, \quad T \rightarrow \infty, \quad i = 1, \dots, N.$$

Z těchto výsledků je již vidět, že skutečně  $\hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2, N, T \rightarrow \infty$ . Konzistence  $\sigma_{between}^2$  lze dokázat analogicky, z čehož pak již plyne i konzistence  $\sigma_\mu^2$ .

Dosadíme-li tedy do vyjádření varianční matice (3.23) odhady (3.27) a (3.30), dostaneme konzistentní odhad  $\Sigma$ :

$$\hat{\Sigma} = \hat{\sigma}_\mu^2 (\mathbf{I}_N \otimes \mathbf{J}_T) + \hat{\sigma}_\varepsilon^2 (\mathbf{I}_N \otimes \mathbf{I}_T).$$

Použitím tohoto odhadu ve (3.23) potom za předpokladu existence inverzních matic získáme odhad parametru  $\beta$

$$\hat{\beta}_{FGLS} = (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}^{-1} \mathbf{Y}, \quad (3.31)$$

který je za analogických předpokladů jako ve (3.25) také konzistentní. Rozptyl tohoto odhadu odhadneme jako

$$\text{var}(\hat{\beta}_{FGLS} | \mathbf{X}) = (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1}. \quad (3.32)$$

Výsledky o jednoduchých modelech s náhodnými individuálními efekty shrneme v následující větě:

**Věta 3.2.** *Odhady parametrů zobecněnou metodou nejmenších čtverců v modelu (3.21) s náhodnými efekty s předpoklady (3.22) jsou*

$$\hat{\beta}_{GLS} = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{Y}$$

*a mají rozptyl  $\text{var}(\hat{\beta}_{GLS} | \mathbf{X}) = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}$ , existuje-li uvedená inverzní matice.*

*Jsou-li splněny předpoklady*

(i)  $\text{plim}_{N,T \rightarrow \infty} \frac{\mathbf{X}' \Sigma^{-1} \mathbf{u}}{NT} = \mathbf{0}$  a

(ii)  $\text{plim}_{N,T \rightarrow \infty} \frac{\mathbf{X}' \Sigma^{-1} \mathbf{X}}{NT} = \mathbf{W}$ , kde  $\mathbf{W}$  je pozitivně definitní konečná matice,

pak  $\hat{\beta}_{GLS}$  je konzistentním odhadem  $\beta$ .

Předpokládáme-li navíc, že  $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E\varepsilon_{it}^4 \rightarrow 0$ , pak jsou odhady

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{NT - N - K} \sum_{i=1}^N \sum_{t=1}^T \left[ y_{it} - \bar{y}_i - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \hat{\beta}_W \right]^2 \text{ a}$$

$$\hat{\sigma}_\mu^2 = \frac{1}{(N - K)} \sum_{i=1}^N (\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}_B)^2 - \frac{1}{T} \hat{\sigma}_\varepsilon^2$$

konzistentními odhady  $\sigma_\varepsilon^2$  a  $\sigma_\mu^2$ .

### Breuschův-Paganův LM test

U modelu s náhodnými efekty můžeme provést jistou analogii testování významnosti pevných efektů - v případě náhodných efektů se v modelu (3.21) testuje

$$H_0 : \sigma_\mu^2 = 0,$$

$$H_1 : \sigma_\mu^2 > 0.$$

Tuto hypotézu lze za dodatečného předpokladu normality reziduí, tj.  $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ , testovat Breuschovým-Paganovým LM testem. K tomu slouží statistika

$$LM_\mu = \frac{NT}{2(T-1)} \left[ \frac{\sum_{i=1}^N (\sum_{t=1}^T \hat{u}_{it})^2}{\sum_{i=1}^N \sum_{t=1}^T (\hat{u}_{it})^2} - 1 \right]^2, \quad (3.33)$$

kde  $\hat{u}_{it} = y_{it} - b_1 x_{it1} - \dots - b_K x_{itK}$  značí rezidua odhadnutá na základě poolové regrese, viz (2.1), tj. za platnosti nulové hypotézy.

Testová statistika (3.33) má za platnosti  $H_0$  a při  $T, N \rightarrow \infty$  asymptoticky  $\chi_1^2$  rozdělení, viz Breusch a Pagan (1980) či Baltagi (2005, kap. 4.2). Nulovou hypotézu tak zamítneme na hladině  $\alpha$ , jestliže

$$LM_\mu \geq \chi_{1-\alpha,1}^2$$

kde  $\chi_{1-\alpha,1}^2$  je  $(1 - \alpha)$ -kvantil příslušný  $\chi_1^2$  rozdělení. Je-li nulová hypotéza zamítnuta, pak lze usoudit, že model s náhodnými efekty je vhodným pro naše data.

### 3.1.3 Hausmanův specifikační test

Popsali jsme, jak se modely s pevnými a náhodnými efekty odhadují, avšak mnohdy nemusí být zřejmé, jaký model je v dané situaci vhodnější. Pak je možné řídit se následujícími doporučeními, viz také např. Baltagi (2005, kap. 2.2 a 2.3):

Pevné efekty jsou vhodné zejména v případech, kdy bude platnost našich závěrů podmíněna těmi konkrétními subjekty, které jsou sledovány. Například vyšetřujeme-li  $N$  konkrétních evropských států nebo třeba  $N$  konkrétních českých pojišťoven, budou naše závěry platit pouze pro těchto  $N$  konkrétních států, resp. pojišťoven. Model s pevnými efekty navíc dovoluje, aby složky  $u_{it} = \mu_i + \varepsilon_{it}$  byly korelované s regresory. Nevýhodou modelu s pevnými efekty je, že nelze předpovídat individuální efekty pro nové subjekty, resp. časové efekty pro nové časové okamžiky.

Náhodné efekty je dobré uvažovat v případě, že subjekty byly náhodně vybrány a na základě údajů o těchto subjektech chceme činit závěry o celé populaci, ze které tyto subjekty pocházejí. Příkladem mohou být již v úvodu zmíněné panelové studie pro šetření domácností USA, kdy se získávají data pouze od reprezentativního vzorku domácností, ale závěry, které se na základě těchto dat činí, jsou považovány za platné pro celou populaci. V takových případech obvykle bývá  $N$  vysoké. Výhodou je, že tento model netrpí tak velkou ztrátou stupňů volnosti jako třeba metoda nejmenších čtverců s dummy proměnnými. Protože však předpokládáme, že efekty pocházejí ze stejného rozdělení, není v případě modelu s náhodnými efekty žádoucí silná heterogenita mezi subjekty. U modelu s náhodnými efekty je však také třeba, aby tyto efekty nebyly korelovány s regresory, protože v opačném případě není splněna podmínka pro konzistenci odhadů parametrů (3.25). V takovém případě je pak vhodnější považovat efekty za pevné.

Kromě těchto doporučení lze použít také Hausmanův specifikační test (viz Hausman (1978) či např. Baltagi (2005, kap. 4.3)), kde se testuje právě hypotéza nekorelovanosti reziduální složky s regresory:

$$H_0 : E(u_{it}X_{it}) = 0.$$

Nechť  $\hat{\beta}_{FE}$  je i nadále odhad parametru  $\beta$  v modelu s pevnými efekty a  $\hat{\beta}_{RE}$  bude značit odhad parametru  $\beta$  v modelu s náhodnými efekty (tj.  $\hat{\beta}_{GLS}$  nebo  $\hat{\beta}_{FGLS}$ ).  $\hat{\beta}_{FE}$ , jak bylo ukázáno v (3.19), je konzistentním odhadem  $\beta$  jak při  $H_0$ , tak při alternativní hypotéze. Odhad  $\hat{\beta}_{RE}$  je konzistentním odhadem  $\beta$

pouze při  $H_0$ , viz podmínka (3.25). Těchto vlastností se využije při konstrukci testu.

Označme  $\mathbf{q} = \hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$  a  $V(\mathbf{q}) = \text{var}(\hat{\boldsymbol{\beta}}_{FE}) - \text{var}(\hat{\boldsymbol{\beta}}_{RE})$ . Potom za platnosti nulové hypotézy platí  $\text{var}(\mathbf{q}) = V(\mathbf{q})$ , neboť

$$\begin{aligned}\text{var}(\mathbf{q}) &= \text{var}(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \\ &= \text{var}(\hat{\boldsymbol{\beta}}_{FE}) + \text{var}(\hat{\boldsymbol{\beta}}_{RE}) - 2\text{cov}(\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{RE}).\end{aligned}\quad (3.34)$$

Při nekorelovanosti, tj. za platnosti  $H_0$ , platí

$$\begin{aligned}\text{cov}(\hat{\boldsymbol{\beta}}_{RE}, \mathbf{q}) &= \text{cov}(\hat{\boldsymbol{\beta}}_{RE}, \hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \\ &= \text{var}(\hat{\boldsymbol{\beta}}_{RE}) - \text{cov}(\hat{\boldsymbol{\beta}}_{RE}, \hat{\boldsymbol{\beta}}_{FE}) = 0.\end{aligned}\quad (3.35)$$

Tento nezbytný fakt je dokázán v Lemmatu 2.1 v (Hausman, 1978). Stručně řečeno, pokud není kovariance  $\text{cov}(\hat{\boldsymbol{\beta}}_{RE}, \mathbf{q})$  nulová, pak lze sestrojít odhad s nižším rozptylem, než má  $\hat{\boldsymbol{\beta}}_{RE}$ , což je ovšem ve sporu s tím, že  $\hat{\boldsymbol{\beta}}_{RE}$  je již eficientní odhad.

Z (3.35) potom okamžitě dostáváme

$$\text{cov}(\hat{\boldsymbol{\beta}}_{RE}, \hat{\boldsymbol{\beta}}_{FE}) = \text{var}(\hat{\boldsymbol{\beta}}_{RE}),$$

z čehož po dosazení do (3.34) vyplývá, že  $\text{var}(\mathbf{q}) = V(\mathbf{q})$ . Pro vyhodnocení testu se za předpokladu existence  $[V(\mathbf{q})]^{-1}$  použije testová statistika

$$m = \mathbf{q}'[V(\mathbf{q})]^{-1}\mathbf{q},$$

která má za platnosti  $H_0$  asymptoticky  $\chi_K^2$  rozdělení,  $N, T \rightarrow \infty$ . Nulovou hypotézu tak zamítneme na asymptotické hladině  $\alpha$ , jestliže  $m \geq \chi_{1-\alpha, K}^2$ , kde  $\chi_{1-\alpha, K}^2$  je kvantil příslušný  $\chi_K^2$  rozdělení. V takovém případě je pak vhodné použít model s pevnými efekty.

Následující Tabulka 3.1 shrnuje tři popsané testy pro jednoduché modely komponentních chyb - F-test, LM test a Hausmanův test. Zároveň jsou v ní uvedeny modely, které je vhodné použít, je-li daná nulová hypotéza zamítnuta, či nikoli - konkrétně poolový model, tj. model bez efektů, jednoduchý model s pevnými efekty a jednoduchý model s náhodnými efekty.

| test           | H <sub>0</sub> | non H <sub>0</sub> |
|----------------|----------------|--------------------|
| F-test         | poolový model  | pevné efekty       |
| LM test        | poolový model  | náhodné efekty     |
| Hausmanův test | náhodné efekty | pevné efekty       |

Tabulka 3.1: Testování v jednoduchých modelech komponentních chyb

## 3.2 Oboustranný model komponentních chyb

Přirozeným rozšířením jednoduchého modelu komponentních chyb, ve kterém uvažujeme buď pouze individuální efekty, nebo pouze časové efekty, je oboustranný model komponentních chyb, který zahrnuje oba druhy efektů současně. Zde si tento model popíšeme pouze velmi stručně, neboť se odhadování jeho parametrů v principu neliší od jednoduchého modelu. V případě zájmu lze nalézt podrobnější popis vhodných metod a vlastností odhadů parametrů jimi získaných např. v Baltagi (2005, kap. 3).

Matematicky se tento oboustranný model popíše následovně

$$\begin{aligned}
 y_{it} &= x'_{it}\beta + u_{it}, \\
 u_{it} &= \mu_i + \lambda_t + \varepsilon_{it}, i = 1, \dots, N, t = 1, \dots, T, \\
 \varepsilon_{it} &\sim \text{iid}(0, \sigma_\varepsilon^2 > 0),
 \end{aligned}
 \tag{3.36}$$

kde předpokládáme, že reziduální složka  $\varepsilon_{it}$  je nekorelovaná s regresory. Opět můžeme efekty  $\mu_i$  a  $\lambda_t$ , reprezentující nepozorovanou individuální a časovou heterogenitu, v závislosti na povaze dat považovat za pevné nebo za náhodné.

### Oboustranný model s pevnými efekty

V případě, že se rozhodneme efekty považovat za pevné, je vhodné pro odhad parametrů díky nižší náročnosti na stupně volnosti použít vnitrotřídní transformaci. Uvažujme pro  $i = 1, \dots, N, t = 1, \dots, T$  model

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\mathbf{x}})' \boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon}),
 \tag{3.37}$$

případně totéž maticově

$$\mathbf{Y}_{2W} = \mathbf{X}_{2W} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{2W},$$

kde se držíme značení zavedeného v sekci 3.1.1. Navíc je zde  $\bar{y}_t$  průměrem  $y_{it}$  přes průřezy, přičemž analogický vztah platí i pro  $\bar{\mathbf{x}}_t$  a  $\bar{\varepsilon}_t$ .

Odhad regresních koeficientů získáme aplikací metody nejmenších čtverců na model (3.37):  $\beta_{2W} = (\mathbf{X}'_{2W}\mathbf{X}_{2W})^{-1}\mathbf{X}'_{2W}\mathbf{Y}_{2W}$ , existuje-li inverzní matice. Efekty lze odhadnout za použití zvolených restrikcí  $\sum_{i=1}^N \mu_i = 0$  a  $\sum_{t=1}^T \lambda_t = 0$  ze

$$\begin{aligned}\tilde{\mu}_i &= (\bar{y}_i - \bar{y}) - \beta_{2W}(\bar{x}_i - \bar{x}), \quad i = 1, \dots, N, \\ \tilde{\lambda}_t &= (\bar{y}_t - \bar{y}) - \beta_{2W}(\bar{x}_t - \bar{x}), \quad i = 1, \dots, T.\end{aligned}$$

Odhady parametrů  $\beta$  jsou v tomto případě konzistentní, ať už pro  $N \rightarrow \infty$  nebo  $T \rightarrow \infty$  (zdůvodnění je analogické (3.19)).

Závěry plynoucí z tohoto modelu jsou i zde podmíněné konkrétními subjekty a nadto ještě konkrétními časovými obdobími, po které byly tyto subjekty pozorovány.

### Oboustranný model s náhodnými efekty

Považujme nyní naopak efekty  $\mu_i, i = 1, \dots, N$ , a  $\lambda_t, t = 1, \dots, T$ , v modelu (3.36) za náhodné a předpokládejme, že  $\mu_i \sim \text{iid}(0, \sigma_\mu^2 > 0)$ ,  $\lambda_t \sim \text{iid}(0, \sigma_\lambda^2 > 0)$  a  $\varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2 > 0)$ . Dále předpokládejme, že jsou všechny tři složky celkové reziduální chyby vzájemně nezávislé a nekorelované s vysvětlujícími proměnnými. Pak platí

$$\begin{aligned}\text{cov}(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2 & i = j, t = s, \\ &= \sigma_\mu^2 & i = j, t \neq s, \\ &= \sigma_\lambda^2 & i \neq j, t = s, \\ &= 0 & \text{jinak.}\end{aligned}$$

Za těchto předpokladů můžeme pro odhad parametru  $\beta$  opět použít Aitkenův odhad:

$$\tilde{\beta}_{GLS} = (\mathbf{X}'\mathbf{\Gamma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}^{-1}\mathbf{Y}, \quad (3.38)$$

kde

$$\mathbf{\Gamma} = \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{I}_T) + \sigma_\lambda^2(\mathbf{J}_N \otimes \mathbf{I}_T) + \sigma_\varepsilon^2(\mathbf{I}_N \otimes \mathbf{I}_T).$$

Je-li rozptylová matice neznámá, lze namísto ní použít některý její konzistentní odhad  $\hat{\mathbf{\Gamma}}$  a pro odhad  $\beta$  přípustný Aitkenův odhad.

Podobně jako u jednoduchého modelu, lze pak závěry z takové analýzy zobecnit na celou populaci, ze které byl vzorek subjektů náhodně vybrán.

Na závěr této kapitoly uveďme, že podobně jako pro jednoduché modely, tak i pro oboustranné modely komponentních chyb lze odvodit test významnosti pevných efektů, test nulovosti rozptylů náhodných efektů a Hausmanův specifikační test, a to analogickým způsobem, jako je uvedeno v 2.2. Podrobný popis těchto testů, stejně tak jako další poznatky z oblasti modelů komponentních chyb lze nalézt např. v knize Baltagi (2005).

### 3.3 Příklad

Pro ilustraci uvedené metody aplikujeme na reálná data o meziročním procentním růstu hrubého domácího produktu (*HDP*), průmyslové výroby (*PROD*), objemu investic (*INV*) a spotřeby domácností (*CONS*) v zemích Organizace pro hospodářskou spolupráci a rozvoj (OECD) v období mezi lety 1992 a 2011. Při modelování nejsou z důvodu chybějících údajů použita data Estonska, Islandu, Slovenska a Slovinska. Údaje o HDP ve zbylých 30 zemích uvádíme v Příloze B. Data pocházejí z <http://stats.oecd.org/>, viz OECD (2013).

Uvažujme model s individuálními efekty

$$HDP_{it} = \mu_i + \beta_1 PROD_{it} + \beta_2 INV_{it} + \beta_3 CONS_{it} + \varepsilon_{it},$$
$$i = 1, \dots, 30, t = 1992, \dots, 2011,$$

a považujme nejprve efekty  $\mu_i$  za pevné. Pak použitím vzorců z Věty 3.1 dostaneme<sup>4</sup> odhady

$$\hat{\beta}_1 = 0,23, \hat{\beta}_2 = 0,10, \hat{\beta}_3 = 0,34,$$

které jsou všechny na základě t-testu významné (p-hodnoty jsou nulové). Dostaneme také odhady pevných efektů, které jsou uvedeny v Tabulce 3.2. Sdruženou významnost efektů můžeme testovat podle kapitoly 3.1.1. Testová statistika má hodnotu 3,71 a p-hodnota příslušná tomuto testu má nulovou hodnotu. Zamítáme tak (na hladině 0,05) nulovou hypotézu o nevýznamnosti pevných efektů.

---

<sup>4</sup>K výpočtům bylo použito prostředí R, konkrétně `plm` package, viz Croissant a Millo (2008).



| Země            | Efekt | Země        | Efekt | Země           | Efekt |
|-----------------|-------|-------------|-------|----------------|-------|
| Austrálie       | 1,17  | Japonsko    | 0,57  | Polsko         | 0,82  |
| Belgie          | 0,49  | Kanada      | 0,82  | Portugalsko    | 0,84  |
| Česká republika | 0,64  | Korea       | 1,53  | Rakousko       | 0,63  |
| Dánsko          | 0,55  | Lucembursko | 2,01  | Řecko          | 0,97  |
| Finsko          | 0,64  | Maďarsko    | -0,11 | Španělsko      | 1,32  |
| Francie         | 0,83  | Mexiko      | 0,71  | Švédsko        | 0,92  |
| Chile           | 1,41  | Německo     | 0,57  | Švýcarsko      | 0,47  |
| Irsko           | 1,58  | Nizozemsko  | 1,10  | Turecko        | 1,21  |
| Itálie          | 1,52  | Norsko      | 0,81  | USA            | 0,80  |
| Izrael          | 1,40  | Nový Zéland | 1,00  | Velká Británie | 1,12  |

Tabulka 3.2: Odhady pevných individuálních efektů  $\mu_i$  v jednoduchém modelu.

Zkusme nyní uvažovat náhodné efekty v modelu

$$HDP_{it} = \alpha + \mu_i + \beta_1 PROD_{it} + \beta_2 INV_{it} + \beta_3 CONS_{it} + \varepsilon_{it},$$

$$i = 1, \dots, 30, t = 1992, \dots, 2011.$$

Pak podle (3.31) dostaneme odhady

$$\hat{\alpha} = 0,87, \hat{\beta}_1 = 0,22, \hat{\beta}_2 = 0,10, \hat{\beta}_3 = 0,36,$$

které jsou opět všechny na základě výsledků t-testů významné (p-hodnoty jsou nulové). Odhad rozptylu  $\mu_i$  je podle (3.32) roven 0,08.

Náhodné efekty jsou také významné - testová statistika Breuschova-Paganova testu má hodnotu 62,12, p-hodnota tohoto testu je nulová. Zamítáme tak na hladině 0,05 hypotézu nulového rozptylu náhodných efektů, viz str. 22.

Zda se přiklonit k pevným, nebo náhodným efektům, můžeme rozhodnout na základě Hausmanova testu, viz str. 23. Jelikož p-hodnota tohoto testu je přibližně rovna 0,001 (testová statistika má hodnotu 16,74), zamítáme na hladině 0,05 nulovou hypotézu Hausmanova testu. Tudíž bychom se měli přiklonit k použití pevných efektů.

Zaměříme se ještě na oboustranný model komponentních chyb s pevnými efekty:

$$HDP_{it} = \mu_i + \lambda_t + \beta_1 PROD_{it} + \beta_2 INV_{it} + \beta_3 CONS_{it} + \varepsilon_{it},$$

$$i = 1, \dots, 30, t = 1992, \dots, 2011.$$

Pak podle kapitoly 3.2 odhadneme

$$\hat{\beta}_1 = 0,21, \hat{\beta}_2 = 0,10, \hat{\beta}_3 = 0,32,$$

což jsou i zde významné parametry (p-hodnoty t-testů jsou nulové). Odhady individuálních efektů jsou uvedeny v Tabulce 3.3, odhady časových efektů v Tabulce 3.4.

| Země            | Efekt | Země        | Efekt | Země           | Efekt |
|-----------------|-------|-------------|-------|----------------|-------|
| Austrálie       | 0,27  | Japonsko    | 0,58  | Polsko         | 0,99  |
| Belgie          | 0,56  | Kanada      | 0,90  | Portugalsko    | 0,87  |
| Česká republika | 0,73  | Korea       | 1,72  | Rakousko       | 0,71  |
| Dánsko          | 0,60  | Lucembursko | 2,07  | Řecko          | 1,00  |
| Finsko          | 0,74  | Maďarsko    | 0,00  | Španělsko      | 1,37  |
| Francie         | 0,86  | Mexiko      | 0,80  | Švédsko        | 0,99  |
| Chile           | 1,58  | Německo     | 0,61  | Švýcarsko      | 0,53  |
| Irsko           | 1,77  | Nizozemsko  | 1,15  | Turecko        | 1,35  |
| Itálie          | 0,54  | Norsko      | 0,87  | USA            | 0,88  |
| Izrael          | 1,55  | Nový Zéland | 1,07  | Velká Británie | 1,17  |

Tabulka 3.3: Odhady pevných individuálních efektů  $\mu_i$  v oboustranném modelu.

| Rok  | Efekt | Rok  | Efekt | Rok  | Efekt | Rok  | Efekt |
|------|-------|------|-------|------|-------|------|-------|
| 1992 | 0,96  | 1997 | 1,16  | 2002 | 0,95  | 2007 | 1,03  |
| 1993 | 0,91  | 1998 | 0,95  | 2003 | 0,72  | 2008 | 0,62  |
| 1994 | 0,77  | 1999 | 1,26  | 2004 | 1,21  | 2009 | 0,49  |
| 1995 | 1,26  | 2000 | 1,62  | 2005 | 1,18  | 2010 | 0,50  |
| 1996 | 0,93  | 2001 | 1,14  | 2006 | 1,35  | 2011 | 0,87  |

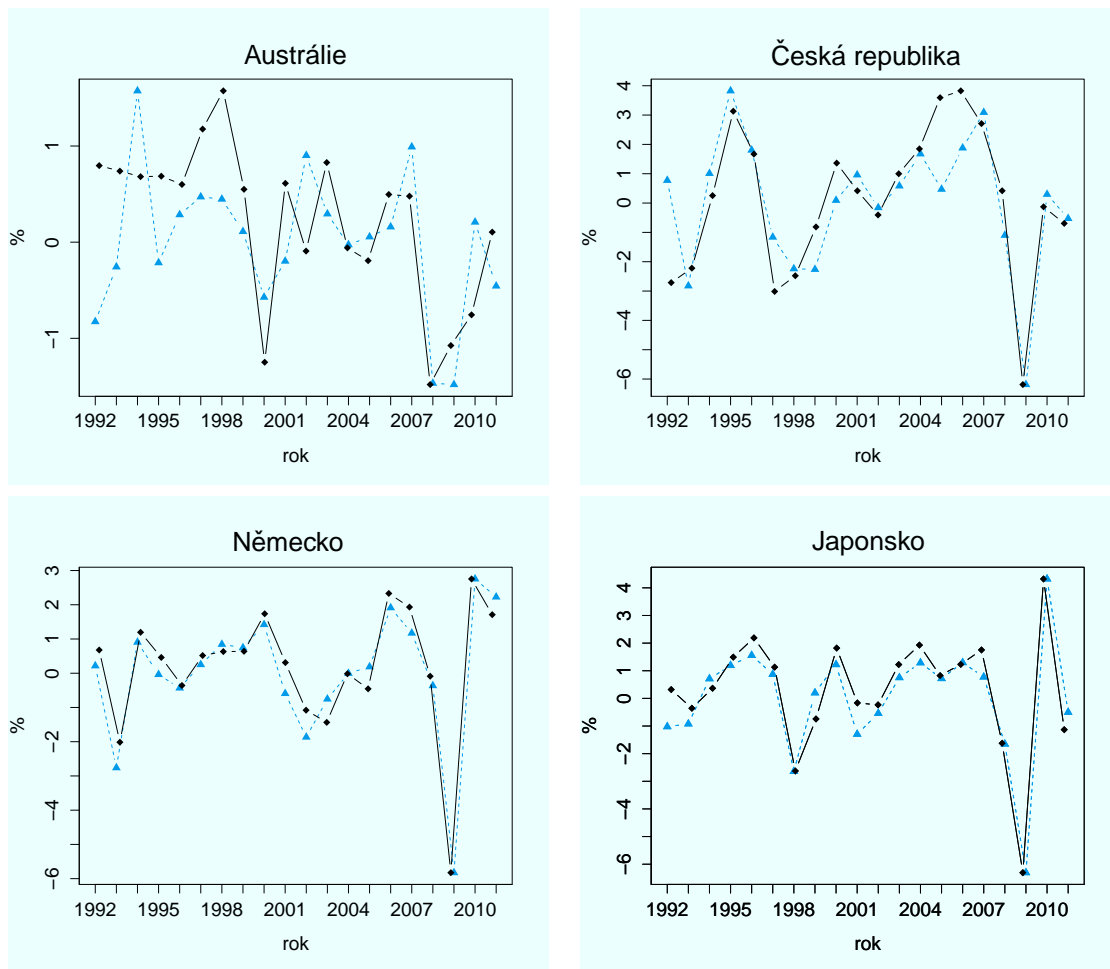
Tabulka 3.4: Odhady pevných časových efektů  $\lambda_t$  v oboustranném modelu.

K testování sdružené významnosti časových efektů můžeme použít F-test. P-hodnota tohoto testu je rovna 0,0005 (testová statistika má hodnotu 2,47), tedy zamítáme nulovou hypotézu, že časové efekty jsou nevýznamné (bez ohledu na individuální efekty).

Avšak při zahrnutí časových efektů došlo oproti jednoduchému modelu ke snížení korigovaného koeficientu determinace (počítaném stejným způsobem jako u klasické lineární regrese), a to z 0,82 na 0,73. Navíc z praktického hlediska může být výhodnější uvažovat pouze individuální efekty, což umožňuje předpovídat vysvětlovanou veličinu na další časová období.

V tomto případě bychom se tedy zřejmě přiklonili k jednoduchému modelu komponentních chyb s pevnými individuálními efekty, ačkoliv pro konečné závěry by bylo ještě třeba model korigovat s ohledem na různou velikost rozptylů v jednotlivých průřezích.

Odhadnutý vývoj HDP na základě tohoto v porovnání se skutečnými hodnoty je pro vybrané země graficky znázorněn na Obr. 3.1.



Obrázek 3.1: Porovnání skutečného vývoje (černě) a odhadnutého vývoje (modře) růstu HDP ve vybraných zemích mezi roky 1992-2011. Vývoj je odhadnut na základě modelu s pevnými individuálními efekty.

# Kapitola 4

## AR(1) s pevným koeficientem

V předchozích kapitolách jsme věnovali pozornost statickým modelům, tedy takovým, jimiž lze modelovat pouze současné vztahy mezi vysvětlujícími a vysvětlovanými proměnnými. Avšak již v úvodu bylo zmíněno, že možnost modelování dynamiky dat, tj. vlivu předchozích hodnot na současné, je nejen z hlediska ekonomické praxe velmi důležitá.

Jedním z možných přístupů k modelování stacionárních řad je využití autoregresních modelů. Na takové modely, konkrétně na autoregresní modely prvního řádu (tj. AR(1) modely) se zaměříme nyní. V této kapitole odvodíme odhady parametrů takového autoregresního modelu za předpokladu nenáhodnosti jeho koeficientu, v následující kapitole potom budeme u téhož modelu uvažovat náhodný koeficient.

Mějme tedy takové reálné náhodné procesy  $\{X_{it}, t \in \mathbb{N}_0\}$ , které jsou autoregresními posloupnostmi prvního řádu a které jsou přes  $i$  nezávislé. Navíc předpokládejme, že všechny tyto posloupnosti mají totožný koeficient  $\varphi$ , v této kapitole nenáhodný. Matematicky zapsáno

$$X_{it} = \varphi X_{i,t-1} + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (4.1)$$

kde  $N$  i zde značí počet průřezů a  $T$  počet časových pozorování. Není-li řečeno jinak, budeme  $N$  i  $T$  považovat za konečné. Zároveň předpokládejme, že platí následující:

- (A1)  $\{e_{it}, t \in \mathbb{N}\} \sim N(0, \sigma_i^2)$  jsou pro každé  $i = 1, \dots, N$  nezávislé stejně rozdělené náhodné veličiny s nenulovým a konečným rozptylem;
- (A2) procesy  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , jsou slabě stacionární, tzn. mají konečné druhé momenty, konstantní střední hodnotu a  $\text{cov}(X_{it}, X_{it-k})$  jsou funkcí pouze  $k$ ;

(A3) rezidua  $\{e_{it}, t \in \mathbb{N}\}$  a modelované řady  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , jsou nekorelované.

Z teorie náhodných procesů víme, viz např. Prášková (2007, kap. 5), že předpoklad (A2) je splněn, platí-li podmínka

$$|\varphi| < 1. \quad (4.2)$$

Jelikož z předpokladu stacionarity vyplývá  $\text{var } X_{it} = \text{var } X_{i,t-1}$  pro každé  $i$  a  $t$  a protože

$$\text{var } X_{it} = \text{var}(\varphi X_{i,t-1} + e_{it}) = \varphi^2 \text{var } X_{i,t-1} + \sigma_i^2,$$

je snadno vidět, že v rámci průřezu mají kovariance tvar

$$\begin{aligned} \text{cov}(X_{it}, X_{it-k}) &= \text{cov}(\varphi^k X_{i,t-k}, X_{i,t-k}) = \frac{\sigma_i^2}{1 - \varphi^2}, \quad k = 0, \\ &= \varphi^k \frac{\sigma_i^2}{1 - \varphi^2}, \quad k = 1, \dots, t. \end{aligned} \quad (4.3)$$

Díky předpokladu (A2), resp. podmínce (4.2) je pro libovolné  $t \in \mathbb{N}_0$  splněno

$$\sum_{k=0}^{\infty} |\text{cov}(X_{it}, X_{it-k})| = \frac{\sigma_i^2}{1 - \varphi^2} \sum_{k=0}^t |\varphi^k| = \frac{\sigma_i^2}{1 - \varphi^2} \frac{1 - |\varphi|^{t+1}}{1 - |\varphi|} < \infty,$$

což je podle A.7 postačující podmínka pro to, aby procesy  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , byly ergodické<sup>1</sup>, což znamená, že pro každé  $i$  platí

$$\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T X_{it} = \text{E}(X_{i0}).$$

Za předpokladů (A1)-(A3) a při uvážení jejich důsledků bude nyní naším cílem odhadnout parametr  $\varphi$  a ukázat asymptotickou nestrannost a konzistenci tohoto odhadu. Následně pro případ, kdy nejsou rozptyly známé, odvodíme i jejich odhady. Poté zkonstruujeme test pro tento parametr, pomocí kterého bude možné ověřit, zda je splněna podmínka stacionarity.

<sup>1</sup>Formální definice ergodicity je velmi technická, lze ji nalézt např. v práci Vaněček (2007, kap. 6).

## 4.1 Odhad pevného koeficientu a jeho vlastnosti

Nechť jsou tedy splněny předpoklady (A1), (A2) a (A3). Na jejich základě v této části textu odvodíme maximálně věrohodný odhad parametru  $\varphi$  a jeho základní vlastnosti. Předpokládejme nejprve, že rozptyly  $\sigma_i^2$  jsou známé.

Označme filtraci  $\mathcal{F}_t = \sigma\{X_{is}, s \leq t, i = 1, \dots, N\}$  a sestavme podmíněnou věrohodnostní funkci  $L(\varphi)$ , jež je definována jako podmíněná hustota vektorů  $(X_{i1}, \dots, X_{iT})$ ,  $i = 1, \dots, N$ , za podmínky  $\mathcal{F}_0$ :

$$L(\varphi) = \prod_{i=1}^N \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{(X_{it} - \varphi X_{i,t-1})^2}{2\sigma_i^2} \right\}.$$

Maximum této funkce je totožné s maximem jejího logaritmu:

$$l(\varphi) = -\frac{1}{2} \left[ NT \log(2\pi) + \sum_{i=1}^N \sum_{t=1}^T \log(\sigma_i^2) + \sum_{i=1}^N \sum_{t=1}^T \frac{(X_{it} - \varphi X_{i,t-1})^2}{\sigma_i^2} \right].$$

Po zderivování této logaritmické věrohodnostní funkce podle  $\varphi$  sestavíme věrohodnostní rovnici

$$\frac{\partial l(\varphi)}{\partial \varphi} = \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} X_{it} - \varphi X_{i,t-1}^2}{\sigma_i^2} = 0,$$

z níž odvodíme maximálně věrohodný odhad parametru  $\varphi$ :

$$\hat{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} X_{it}}{\sigma_i^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}. \quad (4.4)$$

### Asymptotická nestrannost

Tento odhad je asymptoticky nestranným odhadem  $\varphi$ . Vzhledem ke vztahu (4.1) totiž můžeme výraz (4.4) dále upravit na

$$\hat{\varphi} - \varphi = \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} e_{it}}{\sigma_i^2}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}. \quad (4.5)$$

Pro čítelel zlomku na pravé straně (4.5) díky nekorelovanosti reziduí a modelované řady platí

$$\mathbb{E} \left( \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} e_{it}}{\sigma_i^2} \right) = 0.$$

Co se týče jmenovatele téhož zlomku, ukážeme, že konverguje podle pravděpodobnosti k nenulové hodnotě. Již víme, že řady  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , jsou ergodické a stacionární, z čehož dále podle věty o ergodicitě funkcí A.9 vyplývá, že také řady  $\{X_{i,t-1}^2/\sigma_i^2, t \in \mathbb{N}\}$  jsou ergodické a stacionární. Proto platí

$$\begin{aligned} \text{var} \left( \frac{X_{it}^2}{\sigma_i^2} \right) &= \frac{1}{\sigma_i^4} [\varphi^4 \text{var} X_{i,t-1}^2 + 4\varphi^4 \text{var} X_{i,t-1} \text{var} e_{it} + \text{var} e_{it}^2] \\ &= \varphi^4 \text{var} \left( \frac{X_{i,t-1}^2}{\sigma_i^2} \right) + \frac{4\varphi^4}{1-\varphi^2} + 2, \end{aligned}$$

kde jsme využili vlastnosti čtvrtých momentů normálního rozdělení

$$\mathbb{E}(e_{it} - \mathbb{E}e_{it})^4 = 3(\text{var } e_{it})^2.$$

Využijeme-li nyní stacionarity, dostaneme

$$\text{var} \left( \frac{X_{it}^2}{\sigma_i^2} \right) = \text{var} \left( \frac{X_{i,t-1}^2}{\sigma_i^2} \right) = \frac{4\varphi^4 + 2(1-\varphi^2)}{(1-\varphi^2)(1-\varphi^4)}.$$

Odtud

$$\begin{aligned} &\mathbb{E} \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \frac{X_{i,t-1}^2}{\sigma_i^2} - \frac{1}{1-\varphi^2} \right) \right]^2 \\ &= \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{t=1}^T \text{var} \left( \frac{X_{i,t-1}^2}{\sigma_i^2} \right) + \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^{t-1} \text{cov} \left( \frac{X_{i,t-1}^2}{\sigma_i^2}, \frac{X_{i,t-k-1}^2}{\sigma_i^2} \right) \\ &= \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{t=1}^T \left[ \text{var} \left( \frac{X_{i,t-1}^2}{\sigma_i^2} \right) + \sum_{k=1}^{t-1} \varphi^{2k} \text{var} \left( \frac{X_{i,t-1}^2}{\sigma_i^2} \right) \right] \\ &= \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{t=1}^T \text{var} \left( \frac{X_{i,t-1}^2}{\sigma_i^2} \right) \frac{1-\varphi^{2t}}{1-\varphi^2} \xrightarrow{N, T \rightarrow \infty} 0. \end{aligned}$$

Protože

$$E\left(\frac{X_{it}^2}{\sigma_i^2}\right) = E\left(\frac{X_{i,t-1}^2}{\sigma_i^2}\right) = \frac{1}{1-\varphi^2},$$

platí podle slabého zákona velkých čísel pro trojúhelníkové schéma A.10

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \frac{X_{i,t-1}^2}{\sigma_i^2} - \frac{1}{1-\varphi^2} \right) \xrightarrow{p} 0, \quad N, T \rightarrow \infty. \quad (4.6)$$

přičemž víme, že  $1/(1-\varphi^2) \neq 0$ . Jmenovatel pravé strany (4.5) tedy konverguje podle pravděpodobnosti k nenulové hodnotě, zatímco střední hodnota čitatele je nulová. Použitím Sluckého věty A.11 tak dostáváme asymptotickou nestrannost odhadu  $\hat{\varphi}$ .

### Konzistence

Ukážeme ještě, že odhad  $\hat{\varphi}$  je konzistentním odhadem  $\varphi$ . Vzhledem k právě dokázané nestrannosti zbývá podle A.5 dokázat, že rozptyl odhadu  $\hat{\varphi}$  konverguje s rostoucím počtem pozorování k nule.

Protože jsme již ukázali, že jmenovatel zlomku na pravé straně (4.5) konverguje podle pravděpodobnosti k nenulové hodnotě, stačí díky Sluckého větě ukázat, že rozptyl čitatele téhož zlomku konverguje k nule. Označme

$$Y_{i,T} = \frac{1}{T} \sum_{t=1}^T \frac{e_{it} X_{i,t-1}}{\sigma_i^2}.$$

Výraz

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} e_{it}}{\sigma_i^2} = \frac{1}{N} \sum_{i=1}^N Y_{i,T}$$

je pak vzhledem k uvažované nezávislosti průřezů součtem nezávislých náhodných veličin. Použitím (4.3) dostáváme

$$\text{var } Y_{i,T} = \frac{1}{T^2} \frac{1}{\sigma_i^4} \sum_{t=1}^T \text{var } e_{it} \text{ var } X_{i,t-1} = \frac{1}{T} \frac{1}{1-\varphi^2}.$$



Konečně vzhledem k nezávislosti veličin  $Y_{i,T}$  dostáváme

$$\begin{aligned} \text{var} \left( \frac{1}{N} \sum_{i=1}^N Y_{i,T} \right) &= \frac{1}{N^2} \sum_{i=1}^N \text{var} Y_{i,T} \\ &= \frac{1}{N} \frac{1}{T} \frac{1}{1 - \varphi^2} \rightarrow 0, \end{aligned}$$

při  $N, T \rightarrow \infty$ .

Aplikujeme-li nyní opět Sluckého větu, dostaneme konvergenci rozptylu  $\hat{\varphi}$  k nule, a tím i konzistenci odhadu  $\hat{\varphi}$  podle věty A.5.

Z dosavadních výsledků navíc vyplývá, že tato konzistence je dokonce typu  $\sqrt{NT}(\hat{\varphi} - \varphi) = O_P(1)$  při  $N, T \rightarrow \infty$ , tedy že  $\hat{\varphi}$  je  $\sqrt{NT}$ -konzistentním odhadem  $\varphi$ .

Pro přehlednost uvádíme tyto výsledky v následující větě:

**Věta 4.1.** *Za předpokladů (A1)-(A3) má odhad metodou maximální věrohodnosti parametru  $\varphi$  v modelu (4.1) tvar*

$$\hat{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} X_{it}}{\sigma_i^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}.$$

*Tento odhad je při  $N, T \rightarrow \infty$  asymptoticky nestranným a  $\sqrt{NT}$ -konzistentním odhadem parametru  $\varphi$ .*

Pokud se znovu podíváme, jakým způsobem jsme výše uvedené vlastnosti odhadu  $\hat{\varphi}$  odvodili, zjistíme, že není nutné předpokládat normalitu reziduí. Závěry Věty 4.1 tak zůstanou v platnosti i v případě, že předpoklad (A1) nahradíme předpokladem

(A1\*)  $\{e_{it}, t \in \mathbb{N}\}$  jsou pro každé  $i = 1, \dots, N$  nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou, s nenulovým a konečným rozptylem  $\sigma_i^2$  a s  $Ee_{it}^4 < \infty$ .

Větu 4.1 tak můžeme uvést také v následující podobě:

**Věta 4.2.** *Za předpokladů (A1\*), (A2), (A3) a  $EX_{it}^4 < \infty$  je odhad*

$$\hat{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} X_{it}}{\sigma_i^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}$$

parametru  $\varphi$  v modelu (4.1) asymptoticky nestranným a  $\sqrt{NT}$ -konzistentním odhadem parametru  $\varphi$ .

## 4.2 Odhad komponent rozptylu

Dosud jsme předpokládali, že rozptyly  $\sigma_i^2$ ,  $i = 1, \dots, N$ , jsou známy. V praxi tomu však obvykle není a je také třeba je odhadnout. K odhadu těchto rozptylů použijeme opět metodu maximální věrohodnosti, tentokrát však pro každý průřez zvlášť. Předpokládejme, že platí (A1)-(A3). Pak pro každé  $i = 1, \dots, N$  má nyní logaritmická věrohodnostní funkce podmíněná  $\mathcal{F}_0$  tvar:

$$l(\varphi, \sigma_i^2) = -\frac{1}{2} \left[ T \log(2\pi) + T \log(\sigma_i^2) + \sum_{t=1}^T \frac{(X_{it} - \varphi X_{i,t-1})^2}{\sigma_i^2} \right].$$

Po zderivování podle  $\sigma_i^2$  z věrohodnostních rovnic

$$\frac{\partial l(\varphi, \sigma_i^2)}{\partial \sigma_i^2} = -\frac{1}{2} \left[ \frac{T}{\sigma_i^2} - \sum_{t=1}^T \frac{(X_{it} - \varphi X_{i,t-1})^2}{\sigma_i^4} \right] = 0$$

dostaneme při známém  $\varphi$

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T (X_{it} - \varphi X_{i,t-1})^2, \quad i = 1, \dots, N. \quad (4.7)$$

Při neznámém koeficientu  $\varphi$  dosadíme místo něj jeho odhad, který získáme metodou nejmenších čtverců na základě všech pozorování (takový odhad tak bude korespondovat s naším modelem v tom smyslu, že odhad parametru  $\varphi$  bude identický pro všechny průřezy, stejně jako samotný parametr  $\varphi$ ). Získáme jej minimalizací výrazu

$$S(\varphi) = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \varphi X_{i,t-1})^2.$$

Po jeho zderivování a vyřešení normální rovnice dostáváme odhad

$$\hat{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{it}^2}. \quad (4.8)$$

Analogickým postupem jako u  $\hat{\varphi}$  lze i u odhadu  $\hat{\varphi}_{LS}$  zdůvodnit asymptotickou nestrannost a  $\sqrt{T}$ -konzistenci při pevném  $N$ , jestliže uvážíme rovnost

$$\hat{\varphi}_{LS} = \varphi + \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it} X_{i,t-1}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2},$$

kde

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2 \xrightarrow{p} \frac{1}{N} \sum_{i=1}^N \frac{\sigma_i^2}{1 - \varphi^2} \quad (4.9)$$

a

$$\text{var} \left( \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it} X_{i,t-1} \right) = \frac{1}{N^2 T} \sum_{i=1}^N \frac{\sigma_i^4}{1 - \varphi^2} \longrightarrow 0, \quad (4.10)$$

kde  $T \rightarrow \infty$  a  $N$  je pevné. Této vlastnosti  $\sqrt{T}$ -konzistence za okamžik využijeme.

Dosazením  $\hat{\varphi}_{LS}$  do (4.7) dostaneme maximálně věrohodné odhady rozptylů  $\sigma_i^2$  tvaru

$$\hat{\sigma}_{i,ML}^2 = \frac{1}{T} \sum_{t=1}^T (X_{it} - \hat{\varphi}_{LS} X_{i,t-1})^2, \quad i = 1, \dots, N. \quad (4.11)$$

Ukážeme, že tyto odhady rozptylů jsou také  $\sqrt{T}$ -konzistentní (při pevném  $N$ ). Vyjádření (4.11) můžeme upravit na následující tvar:

$$\begin{aligned} \hat{\sigma}_{i,ML}^2 &= \frac{1}{T} \sum_{t=1}^T [X_{it} - \varphi X_{i,t-1} + (\varphi - \hat{\varphi}_{LS}) X_{i,t-1}]^2 \\ &= \frac{1}{T} \sum_{t=1}^T e_{it}^2 + (\varphi - \hat{\varphi}_{LS}) \frac{2}{T} \sum_{t=1}^T e_{it} X_{i,t-1} + (\varphi - \hat{\varphi}_{LS})^2 \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2. \end{aligned} \quad (4.12)$$

Protože pro každé  $i = 1, \dots, N$  jsou  $\{e_{it}^2, t \in \mathbb{N}\}$  nezávislé stejně rozdělené náhodné veličiny se střední hodnotou  $\sigma_i^2 < \infty$ , podle slabého zákona velkých čísel, resp. podle ergodické věty platí, že pro každé  $i$

$$\frac{1}{T} \sum_{t=1}^T e_{it}^2 \xrightarrow{p} \sigma_i^2, \quad T \rightarrow \infty.$$

Zároveň již z předchozích úvah víme, že  $\hat{\varphi}_{LS}$  je  $\sqrt{T}$ -konzistentním odhadem  $\varphi$ , tzn.  $\sqrt{T}(\varphi - \hat{\varphi}_{LS}) \xrightarrow{p} 0$  a  $T(\varphi - \hat{\varphi}_{LS})^2 \xrightarrow{p} 0, T \rightarrow \infty$ .

Vzhledem k předpokladům ergodicity a stacionarity  $X_{it}$  jsou opět podle Věty A.9 ergodické a stacionární i procesy  $X_{i,t-1}^2$  a  $e_{it}X_{i,t-1} = (X_{it} - \varphi X_{i,t-1})X_{i,t-1}$ . Tudíž z ergodické věty A.8 plyne, že při  $T \rightarrow \infty$

$$\frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2 \xrightarrow{p} \frac{\sigma_i^2}{1 - \varphi^2} < \infty,$$

$$\frac{1}{T} \sum_{t=1}^T e_{it}X_{i,t-1} \xrightarrow{p} 0.$$

Srovnáme-li tyto výsledky s tvarem (4.12), je již zřejmé, že pro každé  $i = 1, \dots, N$  je  $\hat{\sigma}_{i,ML}^2$   $\sqrt{T}$ -konzistentním odhadem  $\sigma_i^2$ .

Odhady  $\hat{\sigma}_{i,ML}^2$  použijeme při odhadu  $\varphi$  metodou maximální věrohodnosti při neznámém rozptylu - dosazením do (4.4) dostaneme

$$\hat{\varphi}_{ML} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}X_{it}}{\hat{\sigma}_{i,ML}^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\hat{\sigma}_{i,ML}^2}}, \quad (4.13)$$

což je díky Sluckého větě opět  $\sqrt{NT}$ -konzistentní odhad  $\varphi$ .

V případě potřeby lze dále postupovat iteračně - po vypočtení  $\hat{\varphi}_{ML}$  se tento odhad dosadí zpět do (4.7), čímž se získají nové odhady rozptylů, jež se použijí pro nový výpočet  $\hat{\varphi}_{ML}$  atd.

Právě uvedené výsledky můžeme shrnout ve větě:

**Věta 4.3.** *Za předpokladů (A1)-(A3) mají odhady rozptylů  $\sigma_i^2, i = 1, \dots, N$ , v modelu (4.1) (modifikovanou) metodou maximální věrohodnosti tvar*

$$\hat{\sigma}_{i,ML}^2 = \frac{1}{T} \sum_{t=1}^T (X_{it} - \hat{\varphi}_{LS}X_{i,t-1})^2,$$

kde

$$\hat{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it}X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{it}^2}.$$

$\hat{\varphi}_{LS}$  je  $\sqrt{T}$ -konzistentní odhad parametru  $\varphi$  a odhady  $\hat{\sigma}_{i,ML}^2$ ,  $i = 1, \dots, N$ , jsou (po řadě)  $\sqrt{T}$ -konzistentními odhady parametrů  $\sigma_i^2$ ,  $i = 1, \dots, N$ , když  $T \rightarrow \infty$  a  $N$  je pevné.

Tak jako v případě Věty 4.1 i zde si povšimněme toho, že pro konzistenci odhadu komponent rozptylu není nutné předpokládat normální rozdělení reziduální složky. Větu 4.3 tedy také můžeme dále upravit do následovné podoby:

**Věta 4.4.** Za předpokladů (A1\*), (A2) a (A3) jsou odhady komponent rozptylu  $\sigma_i^2$ ,  $i = 1, \dots, N$ , v modelu (4.1) tvaru

$$\hat{\sigma}_{i,ML}^2 = \frac{1}{T} \sum_{t=1}^T (X_{it} - \hat{\varphi}_{LS} X_{i,t-1})^2,$$

kde

$$\hat{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{it}^2}$$

je  $\sqrt{T}$ -konzistentní odhad parametru  $\varphi$ , (po řadě)  $\sqrt{T}$ -konzistentními odhady parametrů  $\sigma_i^2$ ,  $i = 1, \dots, N$ .

### 4.3 Test hodnoty koeficientu

Dosud jsme pracovali za předpokladu, že jsou řady  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , stacionární, tj. že  $|\varphi| < 1$ . V této části odvodíme test, pomocí kterého bude možné tento předpoklad ověřit. Za předpokladů (A1)-(A3) zkonstruujeme test hypotézy

$$H_0 : \varphi = \varphi_0, \tag{4.14}$$

$$H_1 : \varphi < \varphi_0, \tag{4.15}$$

kde  $|\varphi_0| < 1$ .

Při konstrukci vhodné testové statistiky využijeme toho, že  $\sqrt{NT}(\hat{\varphi} - \varphi)$  má asymptoticky normální rozdělení, což dokážeme pomocí centrální limitní věty pro martingalové diference A.13 a využitím Sluckého věty A.11. Podívejme se tedy na tuto záležitost podrobněji.

Již víme, že jmenovatel zlomku na pravé straně

$$\hat{\varphi} - \varphi = \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} e_{it}}{\sigma_i^2}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}$$

konverguje podle pravděpodobnosti k  $1/(1 - \varphi^2)$ , viz (4.6). Stačí tedy dokázat, že  $(1/\sqrt{NT})(\sum_{i=1}^N \sum_{t=1}^T e_{it} X_{i,t-1} \sigma_i^{-2})$  má asymptoticky normální rozdělení. K tomu použijeme zmíněnou centrální limitní větu pro martingalové diference.

Označme nejprve

$$Z_{Nt} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{e_{it} X_{i,t-1}}{\sigma_i^2}.$$

Je zřejmé, že  $Z_{Nt}$ ,  $t = 1, \dots, T = T_N$ , je podle Definice A.12 posloupnost martingalových diferencí. Přitom vzhledem k předpokladům (A1)-(A3) máme

$$\mathbb{E} Z_{Nt}^2 = \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left( \frac{e_{it}^2 X_{i,t-1}^2}{\sigma_i^4} \right) = \frac{1}{1 - \varphi^2} > 0.$$

Ověřme nyní předpoklady Věty A.13.

(1) První předpoklad je zřejmě splněn, neboť

$$\frac{1}{T} \sum_{t=1}^T \frac{1}{1 - \varphi^2} = \frac{1}{1 - \varphi^2}, \quad T = 1, 2, \dots$$

(2) Díky ergodicitě máme ihned podle A.9

$$\frac{1}{T} \sum_{t=1}^T Z_{Nt}^2 \xrightarrow{p} \mathbb{E} Z_{Nt}^2 = \frac{1}{1 - \varphi^2}, \quad T \rightarrow \infty.$$

(3) Počítejme  $E |Z_{Nt}|^{2+\delta}$  pro  $\delta = 2$ :

$$\begin{aligned} EZ_{Nt}^4 &= \frac{1}{N^2} E \left( \sum_{i=1}^N \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \right)^4 \\ &= \frac{1}{N^2} E \left[ \sum_{i=1}^N \left( \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \right)^2 + 2 \sum_{i=1}^N \sum_{j=1}^{i-1} \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \frac{e_{jt} X_{j,t-1}}{\sigma_j^2} \right]^2 \leq \\ &\leq \frac{1}{N^2} \left[ \sqrt{E \left( \sum_{i=1}^N \frac{e_{it}^2 X_{i,t-1}^2}{\sigma_i^4} \right)^2} + 2 \sqrt{E \left( \sum_{i=1}^N \sum_{j=1}^{i-1} \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \frac{e_{jt} X_{j,t-1}}{\sigma_j^2} \right)^2} \right]^2, \end{aligned}$$

kde jsme využili Minkowského nerovnost (viz např. Lachout (1998, str. 19)).  
Dále díky (A1)-(A3) platí

$$E \left( \sum_{i=1}^N \frac{e_{it}^2 X_{i,t-1}^2}{\sigma_i^4} \right)^2 = \sum_{i=1}^N E \frac{e_{it}^4 X_{i,t-1}^4}{\sigma_i^8} + 2 \cdot \sum_{i=1}^N \sum_{j=1}^{i-1} E \frac{e_{it}^2 X_{i,t-1}^2}{\sigma_i^4} \frac{e_{jt}^2 X_{j,t-1}^2}{\sigma_j^4} < \infty,$$

neboť všechny momenty normálně rozdělených náhodných veličiny jsou konečné. Podobně za využití nezávislosti panelů platí také

$$E \left( \sum_{i=1}^N \sum_{j=1}^{i-1} \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \frac{e_{jt} X_{j,t-1}}{\sigma_j^2} \right)^2 = \sum_{i=1}^N \sum_{j=1}^{i-1} E \left( \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \frac{e_{jt} X_{j,t-1}}{\sigma_j^2} \right)^2 < \infty.$$

Spojením těchto výsledků tak dostaneme požadovanou vlastnost  $EZ_{Nt}^4 < \infty$ .

Všechny předpoklady Věty A.13 jsou tedy splněny, tudíž

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{e_{it} X_{i,t-1}}{\sigma_i^2} \xrightarrow{D} N \left( 0, \frac{1}{1 - \varphi^2} \right), \quad N, T \rightarrow \infty. \quad (4.16)$$

Čitatel pravé strany v (4.5) tak konverguje v distribuci k normálnímu rozdělení a jmenovatel konverguje podle pravděpodobnosti k nenulové hodnotě  $1/(1 - \varphi^2)$ . Podle Sluckého věty tedy platí

$$\sqrt{NT}(\hat{\varphi} - \varphi) = \frac{\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \frac{e_{it} X_{i,t-1}}{\sigma_i^2}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}} \xrightarrow{D} N(0, 1 - \varphi^2), \quad N, T \rightarrow \infty,$$

respektive

$$\sqrt{NT} \frac{(\hat{\varphi} - \varphi)}{\sqrt{1 - \varphi^2}} \xrightarrow{D} N(0, 1), \quad N, T \rightarrow \infty.$$

Díky těmto výsledkům a díky konzistenci odhadů rozptylů (4.12) můžeme pro test hypotézy (4.15) použít testovou statistiku

$$\sqrt{NT} \frac{(\hat{\varphi}_{ML} - \varphi_0)}{\sqrt{1 - \hat{\varphi}_{ML}^2}}, \quad (4.17)$$

která má také za platnosti  $H_0$  asymptoticky rozdělení  $N(0, 1)$ .  $H_0$  tak zamítneme na asymptotické hladině  $\alpha \in (0, 1)$ , jestliže

$$\sqrt{NT} \frac{(\hat{\varphi}_{ML} - \varphi_0)}{\sqrt{1 - \hat{\varphi}_{ML}^2}} \leq u_\alpha,$$

kde  $u_\alpha$  je  $\alpha$ -kvantil normovaného normálního rozdělení.

Speciálně můžeme uvažovat  $\varphi_0$  blízké 1 zleva nebo  $\varphi_0 = 0$ . Ve druhém jmenovaném případě má testová statistika tvar

$$\sqrt{NT} \frac{\hat{\varphi}_{ML}}{\sqrt{1 - \hat{\varphi}_{ML}^2}} \leq u_\alpha.$$

Obecně v případě, že skutečný parametr  $\varphi$  není roven  $\varphi_0$ , pak při pevném  $(\varphi - \varphi_0)$

$$\sqrt{NT} \frac{|\hat{\varphi} - \varphi_0|}{\sqrt{1 - \varphi^2}} = \sqrt{NT} \frac{|\hat{\varphi} - \varphi + \varphi - \varphi_0|}{\sqrt{1 - \varphi^2}} \xrightarrow{p} \infty, \quad N, T \rightarrow \infty,$$

jinými slovy při alternativě  $H_1 : \varphi \neq \varphi_0$  konverguje testová statistika podle pravděpodobnosti do nekonečna.

Shrňme nové poznatky do následující věty:

**Věta 4.5.** Předpokládejme, že platí (A1)-(A3). Pak při  $N, T \rightarrow \infty$  pro odhad parametru  $\varphi$  v modelu (4.1) tvaru

$$\hat{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1} X_{it}}{\sigma_i^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}$$

platí

$$\sqrt{NT}(\hat{\varphi} - \varphi) \xrightarrow{D} N(0, 1 - \varphi^2).$$



Pokud bychom i zde chtěli vypustit předpoklad normality, musíme jej nahradit podmínkou na konečnost momentů tak, aby byl splněn poslední předpoklad centrální limitní věty pro martingalové diference:

**Věta 4.6.** *Za předpokladů (A1\*), (A2) a (A3) a za předpokladu, že  $Ee_{it}^4 < \infty$  a  $EX_{it}^4 < \infty, i = 1, \dots, N, t = 1, \dots, T$ , pro odhad*

$$\hat{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}X_{it}}{\sigma_i^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2}}$$

parametru  $\varphi$  v modelu (4.1) při  $N, T \rightarrow \infty$  platí

$$\sqrt{NT}(\hat{\varphi} - \varphi) \xrightarrow{D} N(0, 1 - \varphi^2).$$

## 4.4 Aplikace na simulovaná data

Pro ilustraci zkusme výše uvedené poznatky použít k odhadu parametrů simulovaných AR(1) procesů s pevným koeficientem. Pro tyto účely jsme (v prostředí R) vygenerovali soubor panelových dat o 10 průřezích a 500 časových pozorování. V každém průřezu byla data generována jako AR(1) proces s koeficientem  $\varphi = 0,5$  s normálně rozdělenými rezidui. Zvolené rozptyly reziduálních složek pro jednotlivé průřezy jsou uvedeny v Tabulce 4.1.

Podle (4.4), (4.8) a (4.13) jsme získali tyto odhady koeficientu  $\varphi$ :

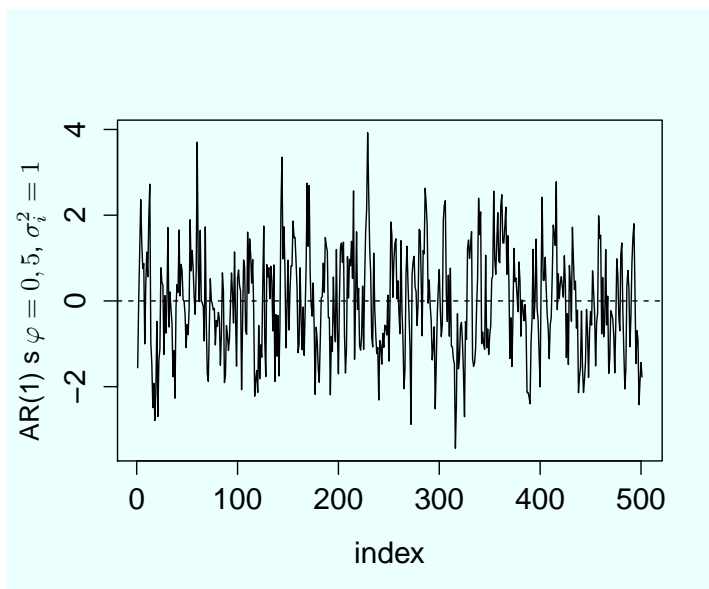
$$\hat{\varphi} = 0,4932, \hat{\varphi}_{LS} = 0,4860, \hat{\varphi}_{ML} = 0,4925.$$

Přitom  $\hat{\varphi}$  jsme vypočetli za použití skutečných rozptylů, zatímco při výpočtu  $\hat{\varphi}_{ML}$  byly použity odhady těchto rozptylů podle (4.11) - ty jsou taktéž uvedeny v Tabulce 4.1.

|                  | $\sigma_1^2$ | $\sigma_2^2$ | $\sigma_3^2$ | $\sigma_4^2$ | $\sigma_5^2$ | $\sigma_6^2$ | $\sigma_7^2$ | $\sigma_8^2$ | $\sigma_9^2$ | $\sigma_{10}^2$ |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|
| skutečné hodnoty | 1,00         | 2,00         | 1,50         | 0,50         | 0,90         | 1,70         | 3,00         | 1,10         | 0,20         | 0,80            |
| odhady           | 1,06         | 2,02         | 1,55         | 0,47         | 0,94         | 1,72         | 2,96         | 1,22         | 0,20         | 0,81            |

Tabulka 4.1: Skutečné hodnoty rozptylů  $\sigma_i^2$  a jejich odhady v AR(1) s pevným koeficientem.

Vidíme, že jak odhady  $\varphi$ , tak odhady  $\sigma_i^2$  jsou poměrně dosti blízké skutečným hodnotám, přičemž odhady využívající informaci o hodnotách rozptylů, tj.  $\hat{\varphi}$  a  $\hat{\varphi}_{ML}$ , jsou o něco přesnější než  $\hat{\varphi}_{LS}$ .



Obrázek 4.1: AR(1) proces s pevným koeficientem,  $\varphi = 0,5, \sigma_{\varepsilon}^2 = 1$ .

## 4.5 Příklad

Odvozené výsledky můžeme ilustrovat i na reálných datech. Opět použijeme soubor obsahující údaje o meziročním procentním růstu hrubého domácího produktu v zemích Organizace pro hospodářskou spolupráci a rozvoj (OECD) v letech 1991-2012. Uvažujeme ty země, které jsou členem organizace OECD k roku 2013, vyjma Estonska, Maďarska a Slovenska, které jsme ze souboru vyňali kvůli chybějícím pozorováním. Údaje o HDP ve zbylých 31 zemích uvádíme v Příloze B a pocházejí z internetové databáze <http://stats.oecd.org/>, viz OECD (2013).

Protože uvažujeme model bez interceptu, je třeba data pro každý stát zvlášť upravit odečtením jejich průměru (přes všechny roky):

$$HDP_{it}^* = HDP_{it} - \frac{1}{22} \sum_{t=1}^{22} HDP_{it}, \quad i = 1, \dots, 31. \quad (4.18)$$

Pro takto upravená data jsme podle vyjádření (4.13) odhadli<sup>2</sup> hodnotu parametru  $\varphi$  v modelu

$$HDP_{it}^* = \varphi HDP_{i,t-1}^* + e_{it}, \quad i = 1, \dots, 31, \quad t = 2, \dots, 22,$$

tak, že  $\hat{\varphi}_{ML} = 0,33$ .

Protože rozptyly jsou neznámé, bylo třeba je také odhadnout. Jejich odhady  $\hat{\sigma}_i^2$ ,  $i = 1, \dots, 31$ , jsou uvedeny v Tabulce 4.2.

| Země            | Rozptyl | Země        | Rozptyl | Země           | Rozptyl |
|-----------------|---------|-------------|---------|----------------|---------|
| Austrálie       | 2,77    | Japonsko    | 4,72    | Rakousko       | 2,54    |
| Belgie          | 2,66    | Kanada      | 2,62    | Řecko          | 4,24    |
| Česká republika | 7,30    | Korea       | 13,10   | Slovinsko      | 9,72    |
| Dánsko          | 6,81    | Lucembursko | 7,45    | Španělsko      | 3,39    |
| Finsko          | 4,26    | Mexiko      | 12,29   | Švédsko        | 6,64    |
| Francie         | 8,96    | Německo     | 2,01    | Švýcarsko      | 1,98    |
| Chile           | 8,13    | Nizozemsko  | 3,00    | Turecko        | 26,91   |
| Irsko           | 11,21   | Norsko      | 2,00    | USA            | 2,43    |
| Island          | 9,84    | Nový Zéland | 3,07    | Velká Británie | 2,86    |
| Itálie          | 3,51    | Polsko      | 2,70    |                |         |
| Izrael          | 5,00    | Portugalsko | 4,33    |                |         |

Tabulka 4.2: Odhady rozptylů  $\sigma_i^2$  v modelu s pevným koeficientem.

Odhad  $\hat{\varphi}_{ML}$  dále můžeme použít k testování stacionarity podle kapitoly 4.3, zejména pak podle výsledků na str. 43. V kapitole 4.3 jsme však odvodili test pouze za podmínky, že  $|\varphi_0| < 1$ . Proto v tomto případě budeme testovat

$$H_0 : \varphi = 0,99,$$

$$H_1 : \varphi < 0,99.$$

Testová statistika

$$\sqrt{NT} \frac{(\hat{\varphi}_{ML} - 0,99)}{\sqrt{1 - \hat{\varphi}_{ML}^2}}$$

má hodnotu  $-18,25$ , což je menší, než příslušný kvantil normovaného normálního rozdělení  $u_{0,05} = -1,64$ . Hypotézu  $H_0$  tedy na (asymptotické) hladině  $0,05$  zamítáme, což v tomto případě není překvapující.

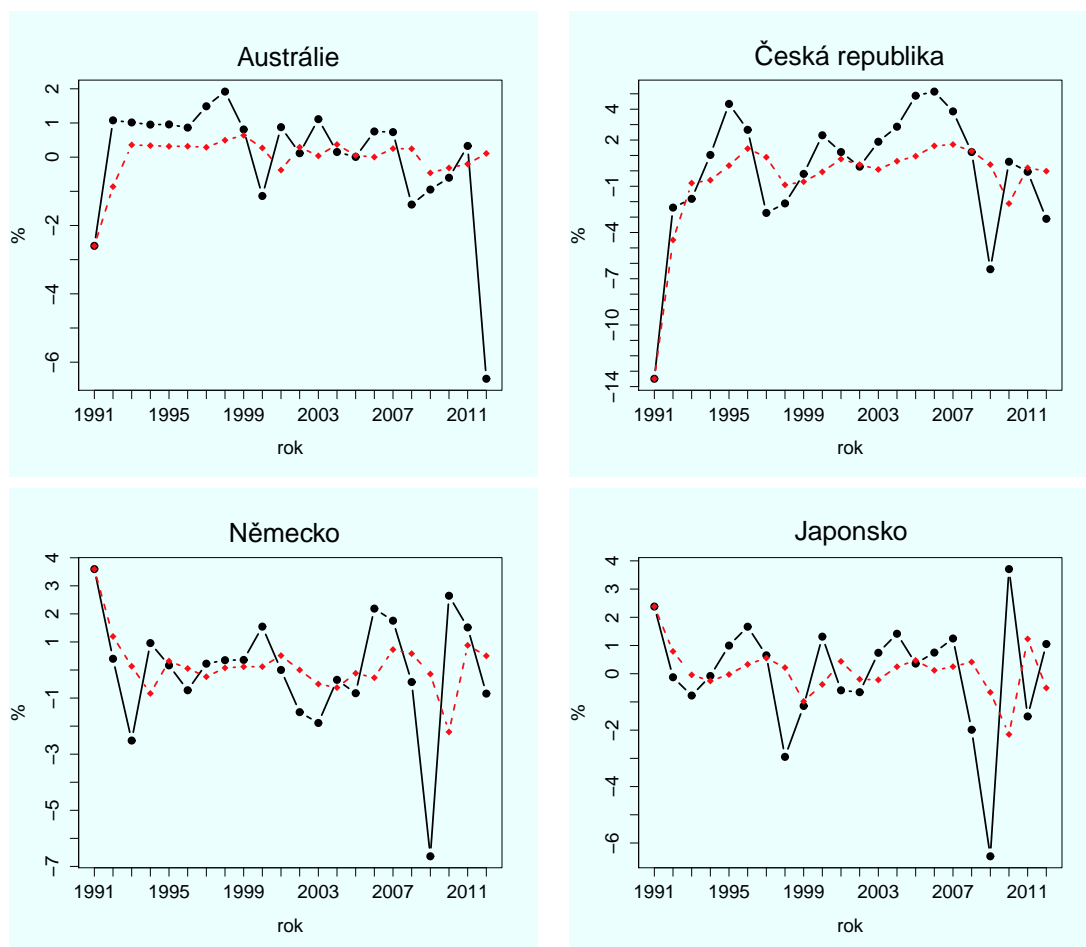
Na závěr kapitoly přikládáme Obr. 4.2 s grafickým přehledem vývoje růstu HDP pro vybrané země<sup>3</sup> v porovnání s jeho odhady získané pomocí  $\hat{\varphi}_{ML}$  následovným způsobem:

$$\widehat{HDP}_{i1}^* = HDP_{i1}^*,$$

$$\widehat{HDP}_{it}^* = \hat{\varphi}_{ML} HDP_{i,t-1}^*, \quad t = 2, \dots, 22.$$

<sup>2</sup>K výpočtům bylo použito výpočetní prostředí R.

<sup>3</sup>Za povšimnutí stojí v případě Japonska a Německa výkyv v roce 2009, což lze zřejmě považovat za důsledek světové finanční krize v roce 2008



Obrázek 4.2: Porovnání skutečného vývoje (černě) a odhadnutého vývoje (červeně) růstu HDP (po odečtení průměru) ve vybraných zemích mezi roky 1991-2012. Vývoj je odhadnut na základě modelu AR(1) s pevným koeficientem.

## Kapitola 5

### AR(1) s náhodným koeficientem

V předchozí kapitole jsme předpokládali, že řady  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , lze modelovat jako autoregresní posloupnosti prvního řádu s pevným koeficientem. Nyní budeme uvažovat velmi podobnou situaci, avšak s tím rozdílem, že v této kapitole půjde o AR(1) s náhodným koeficientem  $\varphi_{it}$ . Jde tedy o obecnější přístup, neboť nyní je možné, aby se koeficienty v různých průřezech do jisté míry lišily. Konkrétně budeme uvažovat model

$$X_{it} = \varphi_{it}X_{i,t-1} + e_{it} = (\varphi + b_{it})X_{i,t-1} + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (5.1)$$

s následujícími předpoklady:

- (B1)  $\{e_{it}, t \in \mathbb{N}\} \sim N(0, \sigma_i^2)$  jsou pro každé  $i = 1, \dots, N$  nezávislé stejně rozdělené náhodné veličiny s nenulovým a konečným rozptylem;
- (B2)  $\{b_{it}, t \in \mathbb{N}\} \sim N(0, \omega^2)$  jsou pro každé  $i = 1, \dots, N$  nezávislé stejně rozdělené náhodné veličiny s nenulovým a konečným rozptylem;
- (B3)  $\{e_{it}, t \in \mathbb{N}\}$  a  $\{b_{it}, t \in \mathbb{N}\}$  jsou pro každé  $i = 1, \dots, N$  vzájemně nezávislé;
- (B4) procesy  $\{X_{it}, t \in \mathbb{N}_0\}$ ,  $i = 1, \dots, N$ , jsou slabě stacionární;
- (B5) procesy  $\{X_{it}, t \in \mathbb{N}_0\}$  jsou pro každé  $i = 1, \dots, N$  nekorelované s procesy  $\{e_{it}, t \in \mathbb{N}\}$  a  $\{b_{it}, t \in \mathbb{N}\}$ .

I nadále uvažujeme nezávislé průřezy, a také  $N, T$  konečné, není-li řečeno jinak.

Za předpokladu (B4) platí pro každé  $i$  rovnost  $\text{var } X_{it} = \text{var } X_{i,t-1}$  a vzhledem ke tvaru modelu (5.1) také

$$\text{var } X_{it} = \text{var}((\varphi + b_{it})X_{it} + u_{it}) = \varphi^2 \text{var } X_{i,t-1} + \omega^2 \text{var } X_{i,t-1} + \sigma_i^2.$$

Kovariance tohoto procesu tak mají obecný tvar

$$\begin{aligned} \text{cov}(X_{it}, X_{it-k}) &= \frac{\sigma_i^2}{1 - \varphi^2 - \omega^2}, \quad k = 0, \\ &= \varphi^k \frac{\sigma_i^2}{1 - \varphi^2 - \omega^2}, \quad k = 1, \dots, t. \end{aligned}$$

Aby však proces vůbec mohl být slabě stacionární, musí mít podle definice konečné druhé momenty. Proto jestliže předpokládáme (B4), předpokládáme tak současně, že platí podmínka

$$\varphi^2 + \omega^2 < 1. \quad (5.2)$$

Za platnosti (5.2) jsou pak stacionární řady  $\{X_{it}, t \in \mathbb{N}_0\}$  dokonce ergodické (viz A.7), neboť pro každé  $t \in \mathbb{N}_0$  platí

$$\begin{aligned} \sum_{k=0}^{\infty} |\text{cov}(X_{it}, X_{it-k})| &= \frac{\sigma_i^2}{1 - \varphi^2 - \omega^2} \sum_{k=0}^t |\varphi^k| \\ &= \frac{\sigma_i^2}{1 - \varphi^2 - \omega^2} \frac{1 - |\varphi|^{t+1}}{1 - |\varphi|} < \infty, \end{aligned} \quad (5.3)$$

jelikož  $1 > \varphi^2 + \omega^2 > \varphi^2$ , a tudíž i  $|\varphi| < 1$ .

Než začneme odhadovat koeficient  $\varphi$ , definujme ještě filtraci  $\mathcal{G}_t = \sigma\{b_{is}, X_{is}, s \leq t, i = 1, \dots, N\}$  a poznamenejme, že model (5.1) lze ekvivalentně zapsat jako

$$\begin{aligned} X_{it} &= \varphi X_{i,t-1} + u_{it}, \\ u_{it} &= b_{it} X_{i,t-1} + e_{it}. \end{aligned} \quad (5.4)$$

Potom lze snadno vidět, že

$$\text{var}[u_{it} | \mathcal{G}_{t-1}] = \omega^2 X_{i,t-1}^2 + \sigma_i^2. \quad (5.5)$$

V případě, že by  $\omega$  bylo rovné nule, jednalo by se o případ AR(1) s pevným koeficientem z minulé kapitoly.

## 5.1 Odhad náhodného koeficientu a jeho vlastnosti

Pro odhad parametru  $\varphi$  opět použijeme metodu maximální věrohodnosti. Předpokládejme, že platí (B1)-(B5) a že rozptyly  $\sigma_i^2$ ,  $i = 1, \dots, N$ , a  $\omega^2$  jsou známé. Logaritmická věrohodnostní funkce podmíněná  $\mathcal{G}_0$  potom bude mít tvar

$$l(\varphi) = -\frac{1}{2} \left[ NT \log(2\pi) + \sum_{i=1}^N \sum_{t=1}^T \log(\sigma_i^2 + \omega^2 X_{i,t-1}^2) + \sum_{i=1}^N \sum_{t=1}^T \frac{(X_{it} - \varphi X_{i,t-1})^2}{\sigma_i^2 + \omega^2 X_{i,t-1}^2} \right].$$

Maximalizací této funkce dostaneme maximálně věrohodný odhad

$$\tilde{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{it} X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}} = \varphi + \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}. \quad (5.6)$$

### Asymptotická nestrannost

Nyní ukážeme, že tento odhad je asymptoticky nestranným odhadem parametru  $\varphi$  při  $T \rightarrow \infty$  a pevném  $N$ . Označme  $w(X_{it}) = \sigma_i^2 + \omega^2 X_{it}^2$  a označme také čitatele a jmenovatele zlomku na pravé straně předchozího výrazu (5.6):

$$W_1^{(NT)} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} X_{i,t-1}}{w(X_{i,t-1})},$$

$$W_2^{(NT)} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{w(X_{i,t-1})}.$$

Stejně jako v Kapitole 4 budeme při dokazování nestrannosti postupovat tak, že ukážeme, že jmenovatel  $W_2^{(NT)}$  konverguje podle pravděpodobnosti k nenulové hodnotě a že střední hodnota čitatele  $W_1^{(NT)}$  je rovna nule. Následně použijeme Sluckého větu A.11.

Již víme, že pro každé  $i$  je řada  $\{X_{it}, t \in \mathbb{N}_0\}$  stacionární a ergodická. Díky větě o ergodicitě funkcí A.9 proto platí:

$$\frac{1}{T} \sum_{t=1}^T \frac{X_{i,t-1}^2}{w(X_{i,t-1})} \xrightarrow{p} \mathbb{E} \left( \frac{X_{i0}^2}{w(X_{i0})} \right), \quad T \rightarrow \infty, \quad (5.7)$$

tudíž pro  $T \rightarrow \infty$  a  $N$  pevné máme

$$W_2^{(NT)} \xrightarrow{p} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left( \frac{X_{i0}^2}{w(X_{i0})} \right) > 0. \quad (5.8)$$

Pro střední hodnotu  $W_1^{(NT)}$  platí

$$\begin{aligned} \mathbb{E}W_1^{(NT)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left( \mathbb{E} \left[ \frac{X_{i,t-1}}{w(X_{i,t-1})} (b_{it}X_{i,t-1} + e_{it}) \mid \mathcal{G}_{t-1} \right] \right) \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left( \frac{X_{i,t-1}^2}{w(X_{i,t-1})} \cdot \mathbb{E}[b_{it} | \mathcal{G}_{t-1}] + \frac{X_{i,t-1}}{w(X_{i,t-1})} \cdot \mathbb{E}[e_{it} | \mathcal{G}_{t-1}] \right) = 0, \end{aligned} \quad (5.9)$$

což ve spojení s výsledkem (5.8) podle Sluckého věty znamená, že odhad  $\tilde{\varphi}$  je skutečně asymptoticky nestranným odhadem  $\varphi$ .

### Konzistence

Podobně za použití A.5 dokážeme konzistenci odhadu  $\tilde{\varphi}$  při  $T \rightarrow \infty$  a pevném  $N$ . Ukážeme, že rozptyl  $W_1^{(NT)}$  konverguje k nule při  $T \rightarrow \infty$ . Potom již opět stačí aplikovat Sluckého větu.

Rozptyl  $W_1^{(NT)}$  za využití nezávislosti průřezů vyjádříme jako

$$\begin{aligned} \text{var } W_1^{(NT)} &= \mathbb{E}(W_1^{(NT)})^2 = \frac{1}{N^2T^2} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left( \frac{X_{i,t-1}^2}{w(X_{i,t-1})^2} \cdot \mathbb{E}[u_{it}^2 | \mathcal{G}_{t-1}] \right) \\ &= \frac{1}{N^2T^2} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} \left( \frac{X_{i,t-1}^2}{w(X_{i,t-1})^2} \cdot w(X_{i,t-1}) \right) \\ &= \frac{1}{N^2T} \cdot \sum_{i=1}^N \mathbb{E} \left( \frac{X_{i0}^2}{w(X_{i0})} \right) \xrightarrow{T \rightarrow \infty} 0, \end{aligned} \quad (5.10)$$

z čehož opět ve spojení s výsledkem (5.8) podle výše uvedených vět plyne konzistence odhadu  $\tilde{\varphi}$ . Podobně jako v Kapitole 4 se dokonce jedná o konzistenci typu  $\sqrt{T}(\tilde{\varphi} - \varphi) = O_P(1)$ .



### Asymptotická normalita

Ukážeme ještě, že odhad  $\tilde{\varphi}$  má při pevném  $N$  a  $T \rightarrow \infty$  asymptoticky normální rozdělení, přičemž stačí tuto vlastnost dokázat pouze pro čítec zlomku na pravé straně

$$\tilde{\varphi} - \varphi = \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}.$$

K tomu použijeme centrální limitní větu pro ergodické a stacionární martingalové diference A.14. Pro celý zlomek pak dostaneme asymptotickou normalitu opět na základě použití Sluckého věty.

Označme

$$U_{it} = \frac{u_{it} X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2} = \frac{(X_{it} - \varphi X_{i,t-1}) X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}.$$

$U_{it}$ ,  $t = 1, \dots, T$ , je zřejmě pro každé  $i = 1, \dots, N$  posloupnost martingalových diferencí, které jsou stacionární a ergodické. Tyto posloupnosti jsou přes  $i = 1, \dots, N$  nezávislé a  $N$  je pevné a konečné. Dále platí

$$\mathbb{E} U_{it}^2 = \mathbb{E} \left[ \frac{X_{i,t-1}^2}{w(X_{i,t-1})^2} \mathbb{E}(u_{it}^2 | \mathcal{G}_{t-1}) \right] = \mathbb{E} \left( \frac{X_{i,t-1}^2}{w(X_{i,t-1})} \right).$$

Z nezávislosti konečného počtu průřezů pak podle Věty A.14 pro pevné  $N$  máme

$$\sqrt{T} \sum_{i=1}^N \sum_{t=1}^T U_{it} \xrightarrow{D} \mathbb{N} \left( 0, \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left( \frac{X_{i0}^2}{w(X_{i0})} \right) \right), \quad T \rightarrow \infty,$$

a aplikací Sluckého věty za použití (5.8) opět pro pevné  $N$  dostaneme

$$\sqrt{T}(\tilde{\varphi} - \varphi) \xrightarrow{D} \mathbb{N} \left( 0, \frac{1}{\sum_{i=1}^N \mathbb{E} \left( \frac{X_{i0}^2}{w(X_{i0})} \right)} \right), \quad T \rightarrow \infty.$$

Shrňme tyto výsledky v následující větě:

**Věta 5.1.** *Za předpokladů (B1)-(B5) má odhad metodou maximální věrohodnosti parametru  $\varphi$  v modelu (5.1) tvar*

$$\tilde{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{it}X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}.$$

*Tento odhad je asymptoticky nestranným a  $\sqrt{T}$ -konzistentním odhadem parametru  $\varphi$  při  $T \rightarrow \infty$  a pevném  $N$ . Navíc platí*

$$\sqrt{T}(\tilde{\varphi} - \varphi) \xrightarrow{D} N \left( 0, \frac{1}{\sum_{i=1}^N E \left( \frac{X_{i0}^2}{w(X_{i0})} \right)} \right), T \rightarrow \infty.$$

Kdybychom si znovu prošli způsob odvození těchto vlastností, podobně jako v Kapitole 4 zjistíme, že ani zde není třeba předpoklad normality. Uvažujme tedy alternativní předpoklady:

(B1\*)  $\{e_{it}, t \in \mathbb{N}\}$  jsou pro každé  $i = 1, \dots, N$  nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou a s nenulovým a konečným rozptylem  $\sigma_i^2$  a s  $Ee_{it}^4 < \infty$ ;

(B2\*)  $\{b_{it}, t \in \mathbb{N}\}$  jsou pro každé  $i = 1, \dots, N$  nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou s nenulovým a konečným rozptylem  $\omega^2$  a s  $Eb_{it}^4 < \infty$ .

Pak můžeme přeformulovat větu 5.1 následovným způsobem:

**Věta 5.2.** *Za předpokladů (B1\*), (B2\*), (B3)-(B5) je odhad*

$$\tilde{\varphi} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{it}X_{i,t-1}}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\sigma_i^2 + \omega^2 X_{i,t-1}^2}}$$

*parametru  $\varphi$  v modelu (5.1) asymptoticky nestranným a  $\sqrt{T}$ -konzistentním odhadem tohoto parametru, jestliže  $T \rightarrow \infty$  a  $N$  je pevné. Platí také*

$$\sqrt{T}(\tilde{\varphi} - \varphi) \xrightarrow{D} N \left( 0, \frac{1}{\sum_{i=1}^N E \left( \frac{X_{i0}^2}{w(X_{i0})} \right)} \right), T \rightarrow \infty.$$

## 5.2 Odhad komponent rozptylu

Při konstrukci podmíněného maximálně věrohodného odhadu  $\tilde{\varphi}$  jsme předpokládali, že známe rozptyly  $\sigma_i^2, i = 1, \dots, N$ , a  $\omega^2$ . Pokud tomu tak není, lze namísto nich použít jejich odhady. Při jejich odvození nejprve použijeme metodu nejmenších čtverců, jejíž pomocí nalezneme alternativní odhad  $\varphi$ . To vše za předpokladů (B1)-(B5). Tento alternativní odhad pak dále použijeme pro vlastní odhady rozptylů.

### Odhad $\varphi$ metodou nejmenších čtverců a jeho vlastnosti

Hledáme tedy nejdříve takový odhad parametru  $\varphi$ , který minimalizuje

$$S(\varphi) = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \varphi X_{i,t-1})^2,$$

což je

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2} = \varphi + \frac{\sum_{i=1}^N \sum_{t=1}^T u_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2}. \quad (5.11)$$

Tento odhad má podobné vlastnosti jako maximálně věrohodný odhad, což shrnuje následující věta:

**Věta 5.3.** *Za předpokladů (B1)-(B5) je odhad*

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2}$$

*parametru  $\varphi$  v modelu (5.1) asymptoticky nestranným a  $\sqrt{T}$ -konzistentním odhadem tohoto parametru při  $T \rightarrow \infty$  a pevném  $N$ . Dále platí*

$$\sqrt{T}(\tilde{\varphi} - \varphi) \xrightarrow{D} N \left( 0, \frac{\sum_{i=1}^N \mathbb{E}(X_{i0}^2 w(X_{i0}))}{\left(\sum_{i=1}^N \mathbb{E}X_{i0}^2\right)^2} \right), T \rightarrow \infty.$$

*Důkaz.* Pro důkaz stačí zopakovat týž postup jako u  $\tilde{\varphi}$  (viz (5.7) - (5.10)) s tím, že se místo  $X_{i,t-1}/w(X_{i,t-1})$  počítá pouze s  $X_{i,t-1}$ . Nastíníme alespoň hlavní body:

Jmenovatel pravé strany výrazu

$$\tilde{\varphi}_{LS} - \varphi = \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T u_{it} X_{i,t-1}}{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2}.$$

konverguje při  $T \rightarrow \infty$  díky ergodicitě podle pravděpodobnosti ke

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} X_{i0}^2 = \frac{1}{N} \sum_{i=1}^N \frac{\sigma_i^2}{1 - \varphi^2 - \omega^2} > 0.$$

Pro čítelel téhož zlomku platí

$$\mathbb{E} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T u_{it} X_{i,t-1} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E} (X_{i,t-1} \mathbb{E} [u_{it} | \mathcal{G}_{t-1}]) = 0$$

a

$$\begin{aligned} \text{var} \left( \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T u_{it} X_{i,t-1} \right) &= \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{t=1}^T (\mathbb{E} X_{i,t-1}^2 \mathbb{E} [u_{it}^2 | \mathcal{G}_{t-1}]) \\ &= \frac{1}{N^2 T} \sum_{i=1}^N \mathbb{E} (X_{i0} w(X_{i0})) \xrightarrow{T \rightarrow \infty} 0, \end{aligned}$$

z čehož při využití Sluckého věty vyplývá  $\sqrt{T}$ -konzistence a asymptotická nestrannost.

Při důkazu asymptotické normality lze pak postupovat stejně jako na str. 52.  $\square$

Z důkazu také plyne, že tytéž vlastnosti můžeme odvodit i bez předpokladu normality:

**Věta 5.4.** *Za předpokladů (B1\*), (B2\*), (B3)-(B5) je odhad*

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2}$$

parametru  $\varphi$  v modelu (5.1) asymptoticky nestranným a  $\sqrt{T}$ -konzistentním odhadem tohoto parametru při  $T \rightarrow \infty$  a pevném  $N$ . Platí také

$$\sqrt{T}(\tilde{\varphi} - \varphi) \xrightarrow{D} \mathbb{N} \left( 0, \frac{\sum_{i=1}^N \mathbb{E} (X_{i0}^2 w(X_{i0}))}{\left( \sum_{i=1}^N \mathbb{E} X_{i0}^2 \right)^2} \right).$$

### Odhad rozptylů $\omega^2$ a $\sigma_i^2$

Odhad  $\tilde{\varphi}_{LS}$  nyní použijeme k odhadu reziduí:

$$\tilde{u}_{it} = X_{it} - \tilde{\varphi}_{LS} X_{i,t-1}.$$

Dále budeme vycházet z postupu v práci Vaněček (2007, str. 15) a využijeme vlastnosti (5.5), tj.  $E[u_{it}^2 | \mathcal{G}_{t-1}] = \sigma_i^2 + \omega^2 X_{i,t-1}^2$ . Za předpokladu, že  $\{X_{it}^2, t \in \mathbb{N}_0\}, i = 1, \dots, N$ , nejsou konstantní řady, můžeme neznámé rozptyly odhadnout z lineární regrese

$$\tilde{u}_{it}^2 = \sigma_i^2 + \omega^2 X_{i,t-1}^2 + \nu_{it}, \quad (5.12)$$

kde  $\nu_{it}, i = 1, \dots, N, t = 1, \dots, T$ , je (přibližně) bílý šum. Metodou nejmenších čtverců nyní snadno získáme odhady všech  $N + 1$  koeficientů v (5.12):

$$\tilde{\omega}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (\tilde{u}_{it}^2 - \tilde{\sigma}_i^2) X_{i,t-1}^2}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^4}, \quad (5.13)$$

$$\tilde{\sigma}_i^2 = \frac{\sum_{t=1}^T \tilde{u}_{it}^2}{T} - \tilde{\omega}^2 \frac{\sum_{t=1}^T X_{i,t-1}^2}{T}, \quad i = 1, \dots, N. \quad (5.14)$$

Ty můžeme dále upravit dosazením (5.14) do (5.13). Pak je

$$\begin{aligned} \tilde{\omega}^2 &= \frac{\sum_{i=1}^N \sum_{t=1}^T \tilde{u}_{it}^2 X_{i,t-1}^2 - \sum_{i=1}^N \left( \sum_{t=1}^T \tilde{u}_{it}^2 \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2 \right)}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^4 - \sum_{i=1}^N \left( \sum_{t=1}^T X_{i,t-1}^2 \right)^2} \\ &= \frac{\sum_{i=1}^N \sum_{t=1}^T \tilde{u}_{it}^2 \left( X_{i,t-1}^2 - \overline{X_i^2} \right)}{\sum_{i=1}^N \sum_{t=1}^T \left( X_{i,t-1}^2 - \overline{X_i^2} \right)^2} \\ &= \frac{\sum_{i=1}^N \sum_{t=1}^T \left( X_{it} - \tilde{\varphi}_{LS} X_{i,t-1} \right)^2 \left( X_{i,t-1}^2 - \overline{X_i^2} \right)}{\sum_{i=1}^N \sum_{t=1}^T \left( X_{i,t-1}^2 - \overline{X_i^2} \right)^2}, \end{aligned} \quad (5.15)$$

kde

$$\overline{X_i^2} = \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2, \quad i = 1, \dots, N.$$

Podobně můžeme zapsat

$$\tilde{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \left( X_{it} - \tilde{\varphi}_{LS} X_{i,t-1} \right)^2 - \tilde{\omega}^2 \overline{X_i^2}, \quad i = 1, \dots, N. \quad (5.16)$$

### Konzistence odhadů rozptylů

Tyto odhady rozptylů  $\omega^2$  a  $\sigma_i^2$ ,  $i = 1, \dots, N$ , jsou  $\sqrt{T}$ -konzistentní (při pevném  $N$ ), což lze ukázat postupem navrženým v práci Schick (1996, Věta 3), přizpůsobeným naší situaci panelových modelů. Nejprve ukážeme nenulovost následujícího rozptylu:

$$\begin{aligned} \text{var}(X_{it}^2) &= \text{var}[(\varphi + b_{it})X_{i,t-1} + e_{it}]^2 \\ &= \text{var}[(\varphi + b_{it})^2 X_{i,t-1}^2 + 2(\varphi + b_{it})X_{i,t-1}e_{it} + e_{it}^2] \\ &= (\varphi^4 + 4\varphi^2\omega^2 + 3\omega^4)\text{var}X_{i,t-1}^2 + \frac{4\sigma_i^4(\varphi^2 + \omega^2)}{1 - \varphi^2 - \omega^2} + 3\sigma_i^4 > 0, \end{aligned}$$

kde využíváme vlastnosti čtvrtých momentů normálního rozdělení:

$$\begin{aligned} \frac{\text{E}(b_{it} - \text{E}b_{it})^4}{(\text{var } b_{it})^2} - 3 &= 0 \\ \text{tj. } \text{E}b_{it}^4 &= 3(\text{var } b_{it})^2 \end{aligned} \quad (5.17)$$

a analogicky pro  $e_{it}$ . Označme

$$\gamma_i = \text{var}(X_{i0}^2),$$

pak platí

$$\text{cov}(X_{i0}^2, w(X_{i0})) = \text{cov}(X_{i0}^2, \sigma_i^2 + \omega^2 X_{i0}^2) = \omega^2 \gamma_i,$$

takže za použití Vět A.8 a A.9 dostáváme pro pevné  $N$  a  $T \rightarrow \infty$

$$\begin{aligned} A_{NT} &= \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \varphi X_{i,t-1})^2 (X_{i,t-1}^2 - \overline{X_i^2}) \xrightarrow{p} \sum_{i=1}^N (\text{E}u_{i1}^2 X_{i0}^2 - \text{E}u_{i1}^2 \text{E}X_{i0}^2) = \\ &= \sum_{i=1}^N (\text{E}[X_{i0}^2 \cdot \text{E}(u_{i1}^2 | \mathcal{G}_0)] - \text{E}X_{i0}^2 \cdot \text{E}[\text{E}(u_{i1}^2 | \mathcal{G}_0)]) \\ &= \sum_{i=1}^N \text{cov}(X_{i0}^2, w(X_{i0})) = \omega^2 \sum_{i=1}^N \gamma_i, \\ B_{NT} &= \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T X_{i,t-1} (X_{it} - \varphi X_{i,t-1}) (X_{i,t-1}^2 - \overline{X_i^2}) \\ &\xrightarrow{p} \sum_{i=1}^N \{ \text{E}[X_{i0}^3 \cdot \text{E}(u_{i1} | \mathcal{G}_0)] - \text{E}X_{i0}^2 \cdot \text{E}[X_{i0} \cdot \text{E}(u_{i1} | \mathcal{G}_0)] \} = 0, \end{aligned}$$

$$C_{NT} = \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 - \overline{X_i^2})^2 \xrightarrow{p} \sum_{i=1}^N \text{var}(X_{i0}^2) = \sum_{i=1}^N \gamma_i.$$

Toho využijeme k důkazu konzistence:

$$\begin{aligned} \tilde{\omega}^2 &= \frac{\frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T [(X_{it} - \varphi X_{i,t-1}) - (\tilde{\varphi}_{LS} - \varphi) X_{i,t-1}]^2 (X_{i,t-1}^2 - \overline{X_i^2})}{\frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 - \overline{X_i^2})^2} \\ &= \frac{A_{NT} - 2(\tilde{\varphi}_{LS} - \varphi) B_{NT} + (\tilde{\varphi}_{LS} - \varphi)^2 C_{NT}}{C_{NT}} \xrightarrow{p} \omega^2, \quad T \rightarrow \infty, \end{aligned}$$

neboť  $\tilde{\varphi}_{LS}$  je  $\sqrt{T}$ -konzistentním odhadem  $\varphi$ . Z ergodické věty potom také máme

$$\tilde{\sigma}_i^2 \xrightarrow{p} \sigma_i^2 + \omega^2 \cdot \text{E}(X_{i0}^2) - \omega^2 \cdot \text{E}(X_{i0}) = \sigma_i^2, \quad T \rightarrow \infty.$$

Odhady  $\omega$  a  $\sigma_i^2$ ,  $i = 1, \dots, N$ , jsou tedy skutečně  $\sqrt{T}$ -konzistentní. Po jejich dosazení do (5.6) dostáváme odhad

$$\tilde{\varphi}_{ML} = \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{it} X_{i,t-1}}{\tilde{\sigma}_i^2 + \tilde{\omega}^2 X_{i,t-1}^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\tilde{\sigma}_i^2 + \tilde{\omega}^2 X_{i,t-1}^2}} = \varphi + \frac{\sum_{i=1}^N \sum_{t=1}^T \frac{u_{it} X_{i,t-1}}{\tilde{\sigma}_i^2 + \tilde{\omega}^2 X_{i,t-1}^2}}{\sum_{i=1}^N \sum_{t=1}^T \frac{X_{i,t-1}^2}{\tilde{\sigma}_i^2 + \tilde{\omega}^2 X_{i,t-1}^2}}. \quad (5.18)$$

Právě odvozené výsledky můžeme opět shrnout do následující věty:

**Věta 5.5.** *Za předpokladů (B1)-(B5), při pevném  $N$  a  $T \rightarrow \infty$  jsou odhady*

$$\begin{aligned} \tilde{\omega}^2 &= \frac{\sum_{i=1}^N \sum_{t=1}^T (X_{it} - \tilde{\varphi}_{LS} X_{i,t-1})^2 (X_{i,t-1}^2 - \overline{X_i^2})}{\sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 - \overline{X_i^2})^2}, \\ \tilde{\sigma}_i^2 &= \frac{1}{T} \sum_{t=1}^T (X_{it} - \tilde{\varphi}_{LS} X_{i,t-1})^2 - \tilde{\omega}^2 \overline{X_i^2}, \quad i = 1, \dots, N, \end{aligned}$$

kde

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2},$$

$\sqrt{T}$ -konzistentními odhady parametrů  $\omega^2$  a  $\sigma_i^2$ ,  $i = 1, \dots, N$ .

Je dobré si uvědomit, že i zde k obdobným závěrům dojdeme i bez předpokladu normality:

**Věta 5.6.** *Předpokládejme, že platí (B1\*), (B2\*), (B3)-(B5) a nechť  $EX_{it}^4 < \infty$ ,  $i = 1, \dots, N, t = 1, \dots, T$ . Pak jsou odhady*

$$\tilde{\omega}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (X_{it} - \tilde{\varphi}_{LS} X_{i,t-1})^2 (X_{i,t-1}^2 - \overline{X_i^2})}{\sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 - \overline{X_i^2})^2},$$

$$\tilde{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T (X_{it} - \tilde{\varphi}_{LS} X_{i,t-1})^2 - \tilde{\omega}^2 \overline{X_i^2}, \quad i = 1, \dots, N,$$

kde

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2},$$

$\sqrt{T}$ -konzistentními odhady parametrů  $\omega^2$  a  $\sigma_i^2$ ,  $i = 1, \dots, N$ .

### 5.3 Test náhodnosti koeficientů

Přirozenou otázkou je, zda-li je v daném případě skutečně třeba uvažovat náhodné koeficienty, nebo zda je dostatečně vhodný model s pevným koeficientem z Kapitoly 4. Tato otázka vede přímo k testování hypotézy

$$H_0 : \omega^2 = 0,$$

$$H_1 : \omega^2 > 0.$$

Při zamítnutí  $H_0$  bychom se přiklonili k použití náhodných koeficientů, při nezamítnutí potom k použití pevného koeficientu.

Pro konstrukci testu této hypotézy předpokládejme (B1)-(B5) a uvažujme  $N$  pevné. Dále použijeme rozklad  $\tilde{\omega}$  ze str. 57:

$$\tilde{\omega}^2 = \frac{A_{NT} - 2(\tilde{\varphi}_{LS} - \varphi)B_{NT} + (\tilde{\varphi}_{LS} - \varphi)^2 C_{NT}}{C_{NT}}.$$



Upravme

$$\begin{aligned}
A_{NT} &= \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (b_{it}X_{i,t-1} + e_{it})^2 \left( X_{i,t-1}^2 - \overline{X_i^2} \right) \\
&= \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (b_{it}^2 - \mathbb{E}b_{it}^2) X_{i,t-1}^2 \left( X_{i,t-1}^2 - \overline{X_i^2} \right) \\
&\quad + \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (e_{it}^2 - \mathbb{E}e_{it}^2) \left( X_{i,t-1}^2 - \overline{X_i^2} \right) \\
&\quad + \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 \mathbb{E}b_{it}^2 + \mathbb{E}e_{it}^2) \left( X_{i,t-1}^2 - \overline{X_i^2} \right) \\
&\quad + 2 \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (b_{it}X_{i,t-1}e_{it}) \left( X_{i,t-1}^2 - \overline{X_i^2} \right).
\end{aligned}$$

Nyní se zaměříme na každý člen pravé strany zvlášť:

$$I_1^{(iT)} = \frac{1}{T} \sum_{t=1}^T (b_{it}^2 - \mathbb{E}b_{it}^2) X_{i,t-1}^2 \left( X_{i,t-1}^2 - \overline{X_i^2} \right)$$

je průměrem stacionární a ergodické posloupnosti martingalových diferencí, přičemž za využití (5.17)

$$\mathbb{E} \left[ (b_{it}^2 - \omega^2) X_{i,t-1}^2 \left( X_{i,t-1}^2 - \overline{X_i^2} \right) \right]^2 = 2\omega^4 \mathbb{E} \left[ X_{i0}^4 (X_{i0}^2 - \mathbb{E}X_{i0}^2)^2 \right].$$

Můžeme proto použít větu A.14, podle které

$$\sqrt{T}I_1^{(iT)} \xrightarrow{D} \mathbb{N} \left( 0, 2\omega^4 \mathbb{E} \left[ X_{i0}^4 (X_{i0}^2 - \mathbb{E}X_{i0}^2)^2 \right] \right), \quad T \rightarrow \infty.$$

Totéž můžeme provést pro

$$I_2^{(iT)} = \frac{1}{T} \sum_{t=1}^T (e_{it}^2 - \sigma_i^2) \left( X_{i,t-1}^2 - \overline{X_i^2} \right)$$

a dostaneme

$$\sqrt{T}I_2^{(iT)} \xrightarrow{D} \mathbb{N} \left( 0, 2\sigma_i^4 \text{var} X_{i0}^2 \right), \quad T \rightarrow \infty.$$

A konečně platí také

$$\begin{aligned}\sqrt{T}I_3^{(iT)} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (b_{it}X_{i,t-1}e_{it}) \left( X_{i,t-1}^2 - \bar{X}_i^2 \right) \xrightarrow{D} \\ &\xrightarrow{D} N \left( 0, \omega^2 \sigma_i^2 E \left[ X_{i0}^2 (X_{i0}^2 - EX_{i0}^2)^2 \right] \right), T \rightarrow \infty,\end{aligned}$$

podle téže věty.

Pro zbývající člen na pravé straně rozkladu  $A_{NT}$  pak platí

$$\begin{aligned}I_4^{iT} &= \frac{1}{T} \sum_{t=1}^T (\omega^2 X_{i,t-1}^2 + \sigma_i^2) \left( X_{i,t-1}^2 - \bar{X}_i^2 \right) \\ &= \omega^2 \frac{1}{T} \sum_{t=1}^T \left( X_{i,t-1}^2 - \bar{X}_i^2 \right)^2.\end{aligned}$$

Pokud spojíme tyto výsledky s těmi na str. 57 a použijeme Sluckého větu, pak vzhledem k nezávislosti průřezů dostaneme při pevném  $N$

$$\sqrt{T}(\tilde{\omega}^2 - \omega^2) \xrightarrow{D} N(0, V_N), T \rightarrow \infty,$$

kde

$$\begin{aligned}V_N &= \frac{2\omega^4 \sum_{i=1}^N E \left[ X_{i0}^4 (X_{i0}^2 - EX_{i0}^2)^2 \right] + 2 \sum_{i=1}^N \sigma_i^4 \text{var} X_{i0}^2}{\left( \sum_{i=1}^N \text{var} X_{i0}^2 \right)^2} \\ &\quad + \frac{4 \sum_{i=1}^N \omega^2 \sigma_i^2 E \left[ X_{i0}^2 (X_{i0}^2 - EX_{i0}^2)^2 \right]}{\left( \sum_{i=1}^N \text{var} X_{i0}^2 \right)^2}.\end{aligned}\tag{5.19}$$

Nyní můžeme odvodit rozdělení  $\tilde{\omega}^2$  za platnosti  $H_0 : \omega^2 = 0$ :

$$\sqrt{T}\tilde{\omega}^2 \underset{asympt., H_0}{\sim} N \left( 0, \frac{2 \sum_{i=1}^N \sigma_i^4 \text{var} X_{i0}^2}{\left( \sum_{i=1}^N \text{var} X_{i0}^2 \right)^2} \right).$$

Neznámé střední hodnoty a rozptyly pak můžeme odhadnout příslušnými průměry, které, jak jsme již ukázali dříve, jsou díky ergodicitě konzistentními odhady těchto středních hodnot. Podobně v případě neznámých rozptylů

nahradíme  $\sigma_i^2$  konzistentním odhadem  $\tilde{\sigma}_i^2$ . Díky konzistenci těchto odhadů zůstává podle Sluckého věty zachována asymptotická normalita:

$$\frac{\sqrt{T}\tilde{\omega}^2}{\sqrt{\frac{2\sum_{i=1}^N \tilde{\sigma}_i^4 \frac{1}{T} \sum_{t=1}^T (X_{i,t-1}^2 - \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2)^2}{(\sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T (X_{i,t-1}^2 - \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2)^2)^2}}} \stackrel{asympt., H_0}{\sim} N(0, 1).$$

Pomocí této statistiky můžeme nulovou hypotézu  $H_0 : \omega^2 = 0$  testovat - zamítneme ji na asymptotické hladině  $\alpha$ , jestliže

$$\frac{\sqrt{T}\tilde{\omega}^2}{\sqrt{\frac{2\sum_{i=1}^N \tilde{\sigma}_i^4 \frac{1}{T} \sum_{t=1}^T (X_{i,t-1}^2 - \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2)^2}{(\sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T (X_{i,t-1}^2 - \frac{1}{T} \sum_{t=1}^T X_{i,t-1}^2)^2)^2}}} \geq u_{1-\alpha}, \quad (5.20)$$

kde  $u_{1-\alpha}$  je  $(1 - \alpha)$ -kvantil normovaného normálního rozdělení.

Následující věta shrnuje odvozené poznatky o rozdělení  $\tilde{\omega}^2$ :

**Věta 5.7.** *Za předpokladů (B1)-(B5) a pevného  $N$  pro odhad rozptylu  $\omega^2$  tvaru*

$$\tilde{\omega}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (X_{it} - \tilde{\varphi}_{LS} X_{i,t-1})^2 (X_{i,t-1}^2 - \overline{X_i^2})}{\sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 - \overline{X_i^2})^2},$$

kde

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2},$$

platí

$$\sqrt{T}(\tilde{\omega}^2 - \omega^2) \xrightarrow{D} N(0, V_N), \quad T \rightarrow \infty,$$

kde  $V_N$  je tvaru (5.19).

Větu i zde můžeme zformulovat bez předpokladů normality, pokud ošetříme konečnost vyšších momentů  $X_{it}$ .

**Věta 5.8.** Předpokládejme, že platí (B1\*), (B2\*), (B3)-(B5) a nechť platí  $EX_{it}^8 < \infty$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ . Pak pro odhad

$$\tilde{\omega}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T (X_{it} - \tilde{\varphi}_{LS} X_{i,t-1})^2 (X_{i,t-1}^2 - \overline{X_i^2})}{\sum_{i=1}^N \sum_{t=1}^T (X_{i,t-1}^2 - \overline{X_i^2})^2},$$

kde

$$\tilde{\varphi}_{LS} = \frac{\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T X_{i,t-1}^2},$$

při pevném  $N$  platí

$$\sqrt{T}(\tilde{\omega}^2 - \omega^2) \xrightarrow{D} N(0, V_N^*), \quad T \rightarrow \infty,$$

kde

$$V_N^* = \frac{\sum_{i=1}^N \text{var } b_{i0}^2 \text{E} \left[ X_{i0}^4 (X_{i0}^2 - \text{E}X_{i0}^2)^2 \right] + \sum_{i=1}^N \text{var } e_{i0}^2 \text{var } X_{i0}^2}{\left( \sum_{i=1}^N \text{var } X_{i0}^2 \right)^2} + \frac{4 \sum_{i=1}^N \omega^2 \sigma_i^2 \text{E} \left[ X_{i0}^2 (X_{i0}^2 - \text{E}X_{i0}^2)^2 \right]}{\left( \sum_{i=1}^N \text{var } X_{i0}^2 \right)^2}.$$

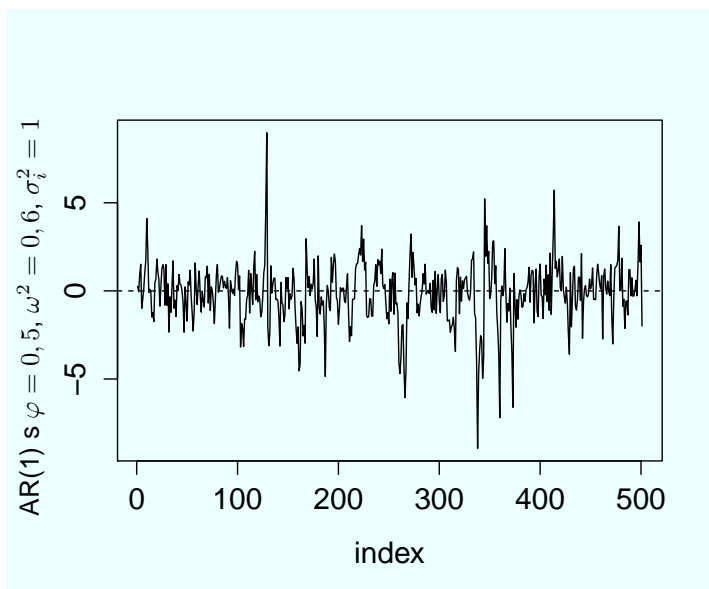
## 5.4 Aplikace na simulovaná data

Pro ilustraci výše uvedených poznatků použijeme tyto výsledky k odhadu parametrů simulovaných AR(1) procesů s náhodným koeficientem. Pro tyto účely jsme (v prostředí R) vygenerovali soubor panelových dat o 10 průřezích a 500 časových pozorování tak, že v každém průřezu byla data generována jako AR(1) proces s náhodným koeficientem, který má normální rozdělení se střední hodnotou  $\varphi = 0,5$  a rozptylem  $\omega^2 = 0,6$ . Je tedy splněna podmínka stacionarity (5.2):  $\varphi^2 + \omega^2 = 0,25 + 0,6 = 0,85 < 1$ .

Zvolené rozptyly reziduálních složek pro jednotlivé průřezy, které jsou stejné jako v Kapitole 4, jsou uvedeny v Tabulce 5.1. Pro  $\sigma_1^2$  je proces znázorněn na Obr. 5.1. Pro rezidua také uvažujeme normální rozdělení.

Podle (5.6), (5.11), (5.18) a (5.15) odhadneme

$$\tilde{\varphi} = 0,4888, \quad \tilde{\varphi}_{LS} = 0,4565, \quad \tilde{\varphi}_{ML} = 0,4794, \quad \tilde{\omega}^2 = 0,3120.$$



Obrázek 5.1: AR(1) proces s náhodným koeficientem,  $\varphi = 0,5$ ,  $\omega^2 = 0,6$ ,  $\sigma_i^2 = 1$ .

Podobně jako v Kapitole 4 jsme  $\hat{\varphi}$  mohli vypočíst díky daným rozptylům, zatímco při výpočtu  $\hat{\varphi}_{ML}$  používáme jejich odhady podle (5.16), které jsou také uvedeny v Tabulce 5.1.

|                  | $\sigma_1^2$ | $\sigma_2^2$ | $\sigma_3^2$ | $\sigma_4^2$ | $\sigma_5^2$ | $\sigma_6^2$ | $\sigma_7^2$ | $\sigma_8^2$ | $\sigma_9^2$ | $\sigma_{10}^2$ |
|------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|
| skutečné hodnoty | 1,00         | 2,00         | 1,50         | 0,50         | 0,90         | 1,70         | 3,00         | 1,10         | 0,20         | 0,80            |
| odhady           | 1,38         | 3,12         | 2,63         | 1,11         | 2,13         | 2,75         | 6,00         | 2,45         | 0,76         | 1,40            |

Tabulka 5.1: Skutečné hodnoty rozptylů  $\sigma_i^2$  a jejich odhady v AR(1) s náhodným koeficientem.

Z výsledků je vidět, že pokud využijeme alespoň přibližné znalosti rozptylů, obvykle dostaneme přesnější odhad skutečného koeficientu. Odhady rozptylů jsou však v tomto případě poměrně nepřesné, jak můžeme porovnat v Tabulce 5.1.

Pro srovnání ještě uveďme, že kdybychom zanedbali náhodnost koeficientu (tj. uvažovali bychom pevný koeficient a metody pro jeho odhad, pokud je ve skutečnosti koeficient náhodný), dostali bychom méně přesné odhady

$$\hat{\varphi} = 0,4487, \hat{\varphi}_{LS} = \tilde{\varphi}_{LS} = 0,4565, \hat{\varphi}_{ML} = 0,4441.$$

## 5.5 Příklad

Navažme na příklad z kapitoly 4.5, kde jsme odhadovali pevný koeficient autoregresního panelového modelu pro soubor dat sestávající z hodnot meziročních procentních přírůstků HDP z let 1991-2012 zemí OECD. Nyní pro stejný soubor dat (po téže úpravě jako ve (4.18)) budeme uvažovat model s náhodným koeficientem, tedy

$$HDP_{it}^* = (\varphi + b_{it})HDP_{i,t-1}^* + e_{it}, i = 1, \dots, 31, t = 1, \dots, 22.$$

V tomto případě chceme odhadnout nejen parametr  $\varphi$  a rozptyly reziduí  $\sigma_i^2$ , ale i rozptyl  $\omega^2$  náhodné složky  $b_{it}$ . Použitím vzorců odvozených v této kapitole, konkrétně (5.6), (5.15) a (5.16), dostaneme odhady  $\tilde{\varphi} = 0,40$  a  $\tilde{\omega}^2 = 0,10$ . Odhady rozptylů uvádíme v Tabulce 5.2.

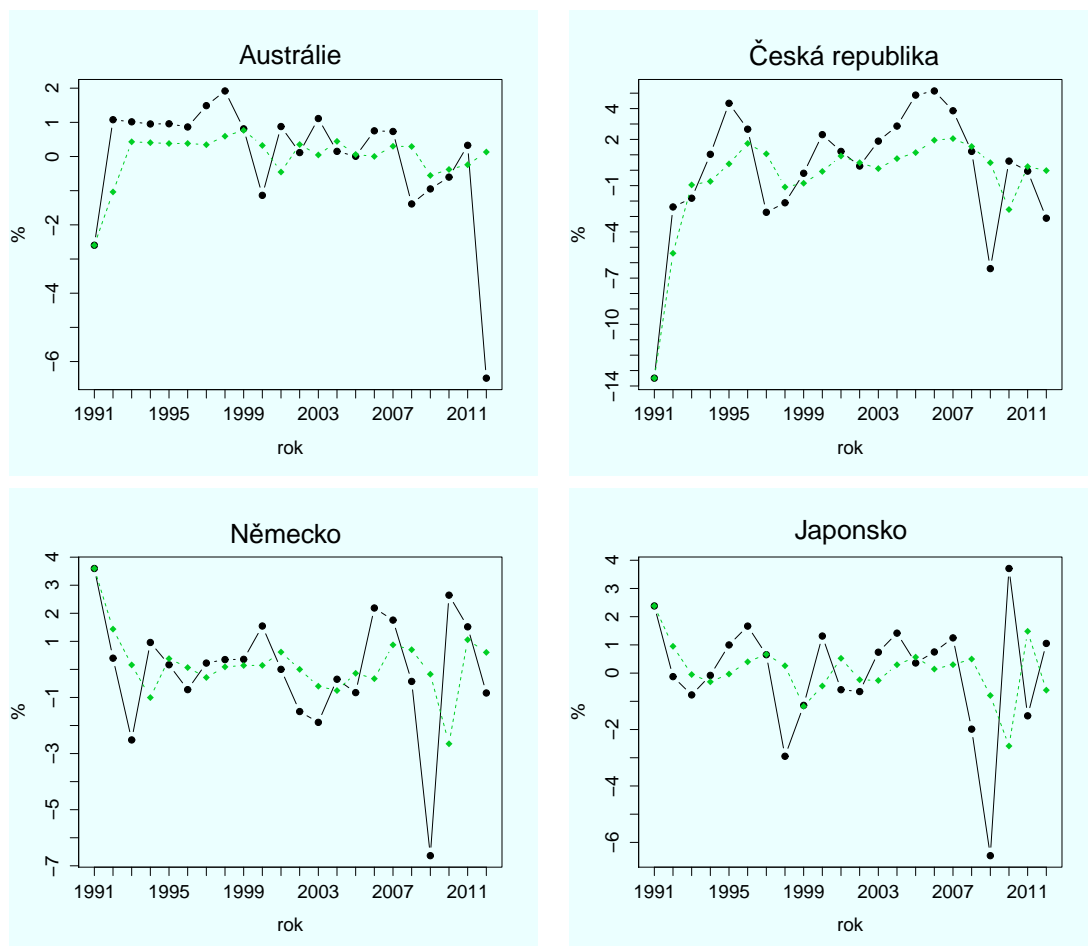
| Země            | Rozptyl | Země        | Rozptyl | Země           | Rozptyl |
|-----------------|---------|-------------|---------|----------------|---------|
| Austrálie       | 2,77    | Japonsko    | 4,50    | Rakousko       | 2,38    |
| Belgie          | 2,54    | Kanada      | 2,32    | Řecko          | 3,99    |
| Česká republika | 6,66    | Korea       | 12,42   | Slovensko      | 8,32    |
| Dánsko          | 5,37    | Lucembursko | 6,82    | Španělsko      | 3,09    |
| Finsko          | 3,98    | Mexiko      | 11,70   | Švédsko        | 6,19    |
| Francie         | 7,94    | Německo     | 1,88    | Švýcarsko      | 1,82    |
| Chile           | 7,35    | Nizozemsko  | 2,80    | Turecko        | 25,67   |
| Irsko           | 9,80    | Norsko      | 1,80    | USA            | 2,18    |
| Island          | 8,90    | Nový Zéland | 2,78    | Velká Británie | 2,54    |
| Itálie          | 3,34    | Polsko      | 1,90    |                |         |
| Izrael          | 4,63    | Portugalsko | 4,00    |                |         |

Tabulka 5.2: Odhady rozptylů  $\sigma_i^2$  v modelu s náhodným koeficientem.

Na Obr. 5.2 je znázorněn grafický přehled vývoje růstu HDP pro vybrané země v porovnání s jeho odhadu získané pomocí  $\tilde{\varphi}_{ML}$  následovným způsobem:

$$\begin{aligned} \widehat{HDP}_{i1}^* &= HDP_{i1}^*, \\ \widehat{HDP}_{it}^* &= \tilde{\varphi}_{ML} HDP_{i,t-1}^*, t = 2, \dots, 22. \end{aligned}$$

Nyní je přirozenou otázkou, zda-li je nutné uvažovat náhodný koeficient. Otestujme tedy podle výsledků kapitoly 5.3 odhad  $\tilde{\omega}^2$ . Hodnota testové statistiky (5.20) je 0,14, přičemž 95%-kvantil normovaného normálního rozdělení má hodnotu 1,64. Na (asymptotické) hladině  $\alpha = 0,05$  tak nezamítáme nulovou hypotézu o nulovém rozptylu  $\omega^2$ .



Obrázek 5.2: Porovnání skutečného vývoje (černě) a odhadnutého vývoje (zeleň) růstu HDP (po odečtení průměru) ve vybraných zemích mezi roky 1991-2012. Vývoj je odhadnut na základě modelu AR(1) s náhodným koeficientem.

Podle těchto výsledků můžeme usoudit, že náhodná složka koeficientu  $\varphi_{it}$  není významná. Proto bychom se v tomto případě měli přiklonit spíše k (jednoduššímu) modelu s nenáhodným koeficientem. Všimněme si, že i v grafických záznamech na Obr. 4.2 a 5.2 dostáváme velice podobné odhady růstu HDP. Pokud však porovnáme tyto obrázky s Obr. 3.1 je zřejmé a jistě ne překvapující, že v tomto případě nejkvalitnější odhady dostaneme za použití vhodných regresorů.

# Kapitola 6

## Závěr

V této práci jsme se zabývali základními koncepty modelování panelových dat s nezávislými průřezy. V první části práce jsme shrnuli již známé postupy při odhadování parametrů modelů komponentních chyb. Zejména jsme se soustředili na jednoduchý model komponentních chyb s individuálními efekty.

Nejprve jsme tyto efekty považovali za pevné parametry a popsali jsme vybrané postupy při jejich odhadování společně s ostatními koeficienty regrese či rozptylem reziduální složky. Následně jsme uvažovali efekty náhodné a za tohoto předpokladu jsme uvedli odhady ostatních parametrů. V práci je ukázána konzistence vybraných odhadů. Tuto problematiku jsme doplnili popisem hlavní myšlenky Hausmanova specifikačního testu pro rozlišení mezi náhodnými a pevnými efekty a stručným nástinem metod pro oboustranné modely komponentních chyb.

Ve druhé části práce jsme přesunuli pozornost od statických modelů k dynamickým autoregresním modelům prvního řádu pro panelová data. V Kapitole 4 jsme se zabývali AR(1) modelem s pevným koeficientem. Odvodili jsme za předpokladu normálního rozdělení reziduí maximálně věrohodný odhad a odhad metodou nejmenších čtverců tohoto parametru, a také maximálně věrohodné odhady rozptylů reziduálních složek. Zároveň jsme studovali vlastnosti těchto odhadů, zejména jejich asymptotickou ne-strannost, konzistenci a asymptotickou normalitu. Zjistili jsme také, že tyto vlastnosti zůstávají v platnosti i při vypuštění předpokladu normality. Použití AR(1) modelu s pevným koeficientem jsme ilustrovali na simulovaných i reálných datech.

V Kapitole 5 jsme představili autoregresní model prvního řádu



s náhodným koeficientem. Za předpokladu normálního rozdělení tohoto koeficientu i reziduálních složek jsme opět odvodili maximálně věrohodný odhad a odhad metodou nejmenších čtverců koeficientu. Při odhadu komponent rozptylů pomocí lineární regrese jsme navázali na práci Vaněčka (2007), kdy jsme tento koncept rozšířili na panelová data. Dále jsme podrobně studovali chování vybraných odhadů a ukázali jsme jejich konzistenci a asymptotickou normalitu, a to i při vynechání předpokladu o normálním rozdělení náhodného koeficientu a reziduí. Na základě těchto vlastností jsme navrhli test náhodnosti koeficientu autoregresního modelu. V závěru kapitoly jsme ilustrovali použití AR(1) modelu s náhodným koeficientem na simulovaných i reálných datech.

Z aplikací na simulovaná data se dá usoudit, že maximálně věrohodné odhady koeficientu AR(1) modelu, založené na předpokladu normality reziduí, resp. náhodného koeficientu, obvykle dávají lepší výsledky než odhady metodou nejmenších čtverců. Jsme si však vědomi, že v tomto ohledu je zde stále prostor pro podrobnější simulační studii.

Jako námět pro další výzkum také může sloužit rozšíření autoregresních modelů pro panelová data i na vyšší řády, jak s pevnými, tak s náhodnými koeficienty, či zpracování problematiky dynamických modelů za současného použití nezávisle proměnných.

# Příloha A

## Pomocná tvrzení

**Definice A.1** (Kroneckerův součin). Nechť  $A = (a_{ij})_{i=1,j=1}^{M,N}$  je matice  $m \times n$  a  $B$  je matice  $p \times q$ . Pak jejich Kroneckerův součin  $A \otimes B$  je definován jako matice o velikosti  $mp \times nq$ :

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \dots & \dots & \dots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}.$$

**Definice A.2** (Konvergence náhodných veličin). Nechť  $X_n, n \in \mathbb{N}_0$ , je posloupnost reálných náhodných veličin s distribuční funkcí  $F_n$  a  $X$  je reálná náhodná veličina s distribuční funkcí  $F$ . Pak

- (i) řekneme, že  $X_n$  konverguje k  $X$  podle pravděpodobnosti, jestliže  $\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$  pro každé  $\varepsilon > 0$ . Značí se  $\text{plim}_{n \rightarrow \infty} X_n = X$  nebo také  $X_n \xrightarrow{p} X, n \rightarrow \infty$ ;
- (ii) řekneme, že  $X_n$  konverguje k  $X$  skoro jistě, jestliže  $P(\lim_{n \rightarrow \infty} X_n = X) = 1$ . Značí se  $X_n \xrightarrow{s.j.} X, n \rightarrow \infty$ .
- (iii) řekneme, že  $X_n$  konverguje k  $X$  v distribuci, jestliže  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ . Používá se značení  $X_n \xrightarrow{D} X, n \rightarrow \infty$  nebo také, je-li např.  $X \sim N(\mu, \sigma^2)$ ,  $X_n \xrightarrow{D} N(\mu, \sigma^2), n \rightarrow \infty$ .

**Definice A.3.** Nechť  $X_n, n \in \mathbb{N}_0$ , je posloupnost reálných náhodných veličin a  $a_n > 0, n \in \mathbb{N}_0$  je posloupnost čísel. Řekneme, že  $X_n = O_P(a_n)$ , jestliže  $\lim_{C \rightarrow \infty} \sup_{n \in \mathbb{N}_0} P(|X_n| > Ca_n) = 0$ .

**Definice A.4** (Konzistence). Řekneme, že odhad  $T_n = g_n(X_1, \dots, X_n)$  parametru  $\tau$  je jeho konzistentním odhadem, jestliže  $T_n \xrightarrow{p} \tau$ ,  $n \rightarrow \infty$ .

**Věta A.5.** Nechť  $ET_n^2 < \infty$  pro každé  $n \in \mathbb{N}_0$ . Jestliže  $ET_n \rightarrow \tau$  a  $\text{var } T_n \rightarrow 0$ ,  $n \rightarrow \infty$ , pak je  $T_n$  konzistentním odhadem parametru  $\tau$ .

Důkaz. Viz Anděl (2007, kap. 7.2). □

**Definice A.6** (Slabá stacionarita). Řekneme, že náhodný proces  $X_t$ ,  $t \in \mathbb{N}_0$ , s konečnými druhými momenty je (slabě) stacionární, jestliže má konstantní střední hodnotu a funkce  $\text{cov}(X_s, X_t)$  je funkcí pouze  $s - t$ .

**Věta A.7** (O ergodicitě). Jestliže stacionární náhodný proces  $X_t$ ,  $t = 1, 2, \dots$ , se střední hodnotou  $EX_t = \mu$  splňuje

$$\sum_{j=0}^{\infty} |\text{cov}(X_t, X_{t-j})| < \infty,$$

pak je tento proces ergodický.

Důkaz. Viz např. Hamilton (1994, kap. 3.1 a kap. 7.2). □

**Věta A.8** (Ergodická věta). Nechť  $X_t$ ,  $t = 1, \dots, T$ , je stacionární a ergodický náhodný proces takový, že  $E|X_t| < \infty$  a  $EX_t = \mu$ . Pak

$$\bar{X}_t = \frac{1}{T} \sum_{t=1}^T X_t \xrightarrow{p} \mu, \quad t \rightarrow \infty.$$

Důkaz. Viz White (1984, kap. 3). □

**Věta A.9** (Ergodicitá funkcí). Nechť  $X_t$ ,  $t = 1, \dots, T$ , je stacionární a ergodický náhodný proces, pak  $f(X_t)$ ,  $t = 1, \dots, T$ , je také stacionární a ergodický proces pro každou  $f(\cdot)$  měřitelnou funkci.

Důkaz. Viz White (1984, kap. 3). □

**Věta A.10** (Slabý zákon velkých čísel pro trojúhelníkové schéma). *Nechť  $\{X_{nt}\}_{t=1}^{T_n}$  jsou pro každé  $n \in \mathbb{N}$  nezávislé náhodné veličiny. Jestliže  $T_n^{-1}EX_{nt}^2 \rightarrow 0$ ,  $n, T_n \rightarrow \infty$ , pak*

$$\frac{1}{T_n} \sum_{t=1}^{T_n} (X_{nt} - EX_{nt}) \xrightarrow{p} 0, \quad n, T_n \rightarrow \infty.$$

*Důkaz.* Větu lze snadno dokázat pomocí Čebyševovy nerovnosti - pro každé  $\varepsilon > 0$  platí:

$$P \left( \left| \frac{1}{T_n} \sum_{t=1}^{T_n} (X_{nt} - EX_{nt}) \right| > \varepsilon \right) \leq \frac{E \left( \sum_{t=1}^{T_n} (X_{nt} - EX_{nt}) \right)^2}{T_n \varepsilon^2} \rightarrow 0, \quad n, T_n \rightarrow \infty.$$

□

**Věta A.11** (Sluckého věta). *Nechť  $X_n, X$  a  $Y_n$  jsou náhodné veličiny. Jestliže  $X_n \xrightarrow{D} X$  a  $X_n \xrightarrow{p} c, n \rightarrow \infty$ , pak*

$$\begin{aligned} X_n + Y_n &\xrightarrow{D} X + c, \\ Y_n X_n &\xrightarrow{D} cX, \\ Y_n^{-1} X_n &\xrightarrow{D} c^{-1}X, \text{ je-li } c \neq 0. \end{aligned}$$

*Důkaz.* Viz Rao (1973, str. 122).

□

**Definice A.12** (Posloupnost martingalových diferencí). *Náhodné veličiny  $\{Y_t\}_{t=1}^{\infty}$  s  $EY_t = 0, t = 1, 2, \dots$ , nazveme posloupností martingalových diferencí, jestliže  $E(Y_t | \Omega_{t-1}) = 0$ , kde  $\Omega_{t-1} = \sigma \{Y_{t-1}, Y_{t-2}, \dots, Y_1\}$ .*

**Věta A.13** (Centrální limitní věta pro posloupnosti martingalových diferencí). *Nechť je  $\{Y_{nt}\}_{t=1}^n$  posloupnost martingalových diferencí (vzhledem k  $\Omega_{nt} = \sigma \{Y_{nt-1}, \dots, Y_{n1}\}$ ) takových, že  $EY_{nt}^2 = \sigma_{nt}^2 > 0$ . Nechť dále platí*

1.  $\frac{1}{n} \sum_{t=1}^n \sigma_{nt}^2 \rightarrow \sigma^2 > 0, n \rightarrow \infty$ ,
2.  $\frac{1}{n} \sum_{t=1}^n Y_{nt}^2 \xrightarrow{p} \sigma^2, n \rightarrow \infty$ , a
3.  $E|Y_{nt}|^{2+\delta} < \Delta < \infty$  pro nějaké  $\delta > 0$  a všechna  $t = 1, \dots, n$ .

Pak při  $n \rightarrow \infty$  platí  $\frac{1}{\sqrt{n}} \sum_{t=1}^n Y_{nt} \stackrel{asympt.}{\sim} N(0, \sigma^2)$

Důkaz. Viz např. White (1984, str. 130). □

**Věta A.14** (Centrální limitní věta pro ergodické a stacionární posloupnosti martingalových diferencí). *Nechť je  $\{Y_t\}_{t=1}^T$  stacionární a ergodická posloupnost martingalových diferencí (vzhledem k  $\Omega_t = \sigma\{Y_{t-1}, \dots, Y_1\}$ ) takových, že  $EY_t^2 = \sigma^2 > 0$ . Pak při  $T \rightarrow \infty$  platí  $\frac{1}{\sqrt{T}} \sum_{t=1}^T Y_t \stackrel{asympt.}{\sim} N(0, \sigma^2)$*

Důkaz. Viz Billingsley (1961). □

# Příloha B

## Data

| Země            | 1991   | 1992  | 1993  | 1994  | 1995  | 1996 | 1997  | 1998  | 1999  |
|-----------------|--------|-------|-------|-------|-------|------|-------|-------|-------|
| Austrálie       | 0,44   | 4,11  | 4,05  | 3,99  | 3,99  | 3,90 | 4,52  | 4,95  | 3,85  |
| Belgie          | 1,83   | 1,53  | -0,96 | 3,23  | 2,38  | 1,42 | 3,74  | 1,93  | 3,54  |
| Česká republika | -11,61 | -0,51 | 0,06  | 2,91  | 6,22  | 4,54 | -0,85 | -0,24 | 1,68  |
| Dánsko          | 1,30   | 1,98  | -0,09 | 5,53  | 3,07  | 2,83 | 3,20  | 2,16  | 2,56  |
| Finsko          | -6,00  | -3,48 | -0,81 | 3,65  | 3,96  | 3,57 | 6,21  | 5,03  | 3,91  |
| Francie         | 1,04   | 1,48  | -0,67 | 2,25  | 2,05  | 1,07 | 2,18  | 3,38  | 3,29  |
| Chile           | 7,97   | 12,28 | 6,99  | 5,71  | 10,63 | 7,37 | 6,60  | 3,26  | -0,73 |
| Irsko           | 1,93   | 3,34  | 2,69  | 5,76  | 9,63  | 9,32 | 11,50 | 8,79  | 11,05 |
| Island          | -0,22  | -3,37 | 1,31  | 3,61  | 0,12  | 4,79 | 4,91  | 6,32  | 4,09  |
| Itálie          | 1,54   | 0,83  | -0,85 | 2,15  | 2,89  | 1,13 | 1,87  | 1,45  | 1,45  |
| Izrael          | 7,73   | 7,76  | 4,12  | 7,43  | 6,60  | 5,43 | 3,18  | 4,25  | 3,29  |
| Japonsko        | 3,32   | 0,82  | 0,17  | 0,86  | 1,94  | 2,61 | 1,60  | -2,00 | -0,20 |
| Kanada          | -2,09  | 0,88  | 2,34  | 4,80  | 2,81  | 1,62 | 4,23  | 4,10  | 5,53  |
| Korea           | 9,71   | 5,77  | 6,33  | 8,77  | 8,93  | 7,19 | 5,77  | -5,71 | 10,73 |
| Lucembursko     | 8,64   | 1,82  | 4,20  | 3,82  | 1,43  | 1,52 | 5,94  | 6,49  | 8,42  |
| Maďarsko        | -      | -3,06 | -0,58 | 2,95  | 1,49  | 0,16 | 3,13  | 4,07  | 3,20  |
| Mexiko          | 4,22   | 3,63  | 1,95  | 4,42  | -6,17 | 5,15 | 6,77  | 5,03  | 3,76  |
| Německo         | 5,11   | 1,91  | -1,00 | 2,47  | 1,68  | 0,79 | 1,74  | 1,86  | 1,87  |
| Nizozemsko      | 2,44   | 1,71  | 1,26  | 2,96  | 3,12  | 3,41 | 4,28  | 3,92  | 4,68  |
| Norsko          | 3,11   | 3,52  | 2,79  | 5,05  | 4,19  | 5,10 | 5,39  | 2,68  | 2,03  |
| Nový Zéland     | -1,10  | 1,11  | 6,37  | 5,01  | 4,18  | 3,25 | 2,72  | 1,31  | 5,12  |
| Polsko          | -7,02  | 2,51  | 3,74  | 5,29  | 6,95  | 6,24 | 7,09  | 4,98  | 4,52  |
| Portugalsko     | 4,37   | 1,09  | -2,04 | 0,96  | 4,28  | 3,69 | 4,41  | 5,14  | 4,07  |
| Rakousko        | 3,44   | 2,09  | 0,53  | 2,40  | 2,67  | 2,47 | 2,31  | 3,79  | 3,54  |
| Řecko           | 3,10   | 0,70  | -1,60 | 2,00  | 2,10  | 2,36 | 3,64  | 3,36  | 3,42  |
| Slovinsko       | -8,90  | -5,46 | 2,84  | 5,33  | 4,11  | 3,65 | 4,96  | 3,51  | 5,33  |
| Španělsko       | 2,55   | 0,93  | -1,03 | 2,38  | 2,76  | 2,42 | 3,87  | 4,47  | 4,75  |
| Švédsko         | -1,12  | -1,20 | -2,06 | 4,01  | 3,94  | 1,61 | 2,71  | 4,20  | 4,66  |
| Švýcarsko       | -0,92  | -0,04 | -0,13 | 1,27  | 0,48  | 0,48 | 2,04  | 2,73  | 1,40  |
| Turecko         | 0,93   | 5,98  | 8,04  | -5,46 | 7,19  | 7,01 | 7,53  | 3,09  | -3,37 |
| USA             | -0,26  | 3,40  | 2,87  | 4,11  | 2,55  | 3,79 | 4,51  | 4,40  | 4,87  |
| Velká Británie  | -1,78  | 0,86  | 3,09  | 4,58  | 3,18  | 3,11 | 3,86  | 3,51  | 3,17  |

Tabulka B.1: HDP v zemích OECD, 1991-1999.

| Země            | 2000  | 2001  | 2002  | 2003  | 2004 | 2005 | 2006 | 2007 |
|-----------------|-------|-------|-------|-------|------|------|------|------|
| Austrálie       | 1,90  | 3,91  | 3,15  | 4,15  | 3,19 | 3,04 | 3,79 | 3,77 |
| Belgie          | 3,67  | 0,81  | 1,36  | 0,81  | 3,27 | 1,75 | 2,67 | 2,88 |
| Česká republika | 4,19  | 3,10  | 2,15  | 3,77  | 4,74 | 6,75 | 7,02 | 5,74 |
| Dánsko          | 3,53  | 0,70  | 0,47  | 0,38  | 2,30 | 2,45 | 3,39 | 1,58 |
| Finsko          | 5,32  | 2,28  | 1,83  | 2,01  | 4,12 | 2,92 | 4,41 | 5,34 |
| Francie         | 3,68  | 1,84  | 0,93  | 0,90  | 2,54 | 1,83 | 2,47 | 2,29 |
| Chile           | 4,46  | 3,35  | 2,17  | 3,96  | 7,02 | 6,18 | 5,69 | 5,16 |
| Irsko           | 10,74 | 5,30  | 5,64  | 3,88  | 4,36 | 5,88 | 5,40 | 5,45 |
| Island          | 4,32  | 3,92  | 0,14  | 2,43  | 7,84 | 7,23 | 4,71 | 5,99 |
| Itálie          | 3,65  | 1,86  | 0,45  | -0,05 | 1,73 | 0,93 | 2,20 | 1,68 |
| Izrael          | 8,66  | -0,18 | -0,07 | 1,48  | 4,90 | 4,92 | 5,80 | 5,90 |
| Japonsko        | 2,26  | 0,36  | 0,29  | 1,69  | 2,36 | 1,30 | 1,69 | 2,19 |
| Kanada          | 5,23  | 1,78  | 2,92  | 1,88  | 3,12 | 3,02 | 2,82 | 2,20 |
| Korea           | 8,80  | 3,97  | 7,15  | 2,80  | 4,62 | 3,96 | 5,18 | 5,11 |
| Lucembursko     | 8,44  | 2,52  | 4,09  | 1,67  | 4,37 | 5,25 | 4,94 | 6,59 |
| Maďarsko        | 4,23  | 3,71  | 4,51  | 3,85  | 4,80 | 3,96 | 3,89 | 0,11 |
| Mexiko          | 6,59  | -0,03 | 0,77  | 1,39  | 4,07 | 3,28 | 5,06 | 3,36 |
| Německo         | 3,06  | 1,51  | 0,01  | -0,38 | 1,16 | 0,68 | 3,70 | 3,27 |
| Nizozemsko      | 3,94  | 1,93  | 0,08  | 0,34  | 2,24 | 2,05 | 3,39 | 3,92 |
| Norsko          | 3,25  | 1,99  | 1,50  | 0,98  | 3,96 | 2,59 | 2,30 | 2,65 |
| Nový Zéland     | 2,26  | 3,58  | 4,88  | 3,89  | 3,51 | 3,22 | 1,69 | 3,56 |
| Polsko          | 4,26  | 1,21  | 1,44  | 3,87  | 5,34 | 3,62 | 6,23 | 6,79 |
| Portugalsko     | 3,92  | 1,97  | 0,76  | -0,91 | 1,56 | 0,78 | 1,45 | 2,37 |
| Rakousko        | 3,67  | 0,86  | 1,69  | 0,87  | 2,59 | 2,40 | 3,67 | 3,71 |
| Řecko           | 4,48  | 4,20  | 3,44  | 5,94  | 4,37 | 2,28 | 5,51 | 3,54 |
| Slovinsko       | 4,27  | 2,94  | 3,83  | 2,93  | 4,40 | 4,01 | 5,85 | 6,96 |
| Španělsko       | 5,05  | 3,67  | 2,71  | 3,09  | 3,26 | 3,58 | 4,08 | 3,48 |
| Švédsko         | 4,45  | 1,26  | 2,48  | 2,34  | 4,23 | 3,16 | 4,30 | 3,31 |
| Švýcarsko       | 3,67  | 1,24  | 0,19  | 0,02  | 2,42 | 2,69 | 3,75 | 3,85 |
| Turecko         | 6,77  | -5,70 | 6,16  | 5,27  | 9,36 | 8,40 | 6,89 | 4,67 |
| USA             | 4,17  | 1,09  | 1,83  | 2,55  | 3,48 | 3,08 | 2,66 | 1,91 |
| Velká Británie  | 4,24  | 2,89  | 2,43  | 3,81  | 2,91 | 2,77 | 2,60 | 3,63 |

Tabulka B.2: HDP v zemích OECD, 2000-2007.

| Země            | 2008  | 2009  | 2010  | 2011  | 2012  | Průměr |
|-----------------|-------|-------|-------|-------|-------|--------|
| Austrálie       | 1,65  | 2,09  | 2,43  | 3,36  | -3,45 | 3,04   |
| Belgie          | 0,99  | -2,79 | 2,42  | 1,84  | -0,28 | 2,09   |
| Česká republika | 3,10  | -4,51 | 2,47  | 1,82  | -1,24 | 1,73   |
| Dánsko          | -0,78 | -5,67 | 1,58  | 1,10  | -0,47 | 2,40   |
| Finsko          | 0,29  | -8,54 | 3,32  | 2,78  | -0,21 | 5,36   |
| Francie         | -0,08 | -3,15 | 1,72  | 2,03  | 0,01  | 1,88   |
| Chile           | 3,29  | -1,04 | 6,10  | 5,99  | 5,56  | 1,50   |
| Irsko           | -2,11 | -5,46 | -0,77 | 1,43  | 0,94  | 1,91   |
| Island          | 1,19  | -6,57 | -4,10 | 2,89  | 1,64  | 1,50   |
| Itálie          | -1,16 | -5,49 | 1,72  | 0,37  | -2,37 | 1,51   |
| Izrael          | 4,11  | 1,11  | 4,98  | 4,60  | 3,18  | 1,41   |
| Japonsko        | -1,04 | -5,53 | 4,65  | -0,57 | 2,00  | 2,42   |
| Kanada          | 0,69  | -2,77 | 3,21  | 2,57  | 1,84  | 4,76   |
| Korea           | 2,30  | 0,32  | 6,32  | 3,63  | 2,04  | 4,51   |
| Lucembursko     | -0,73 | -4,08 | 2,91  | 1,66  | 0,31  | 0,82   |
| Maďarsko        | 0,89  | -6,77 | 1,32  | 1,65  | -1,73 | 1,70   |
| Mexiko          | 1,22  | -6,00 | 5,32  | 3,92  | 3,92  | 0,94   |
| Německo         | 1,08  | -5,13 | 4,16  | 3,03  | 0,67  | 5,17   |
| Nizozemsko      | 1,80  | -3,67 | 1,63  | 0,99  | -0,96 | 3,65   |
| Norsko          | 0,07  | -1,63 | 0,48  | 1,22  | 3,09  | 2,80   |
| Nový Zéland     | -1,85 | 0,94  | 0,24  | 1,08  | 2,97  | 2,07   |
| Polsko          | 5,13  | 1,63  | 3,87  | 4,52  | 1,86  | 2,63   |
| Portugalsko     | -0,01 | -2,91 | 1,94  | -1,55 | -3,25 | 2,56   |
| Rakousko        | 1,44  | -3,78 | 2,05  | 2,70  | 0,85  | 3,82   |
| Řecko           | -0,22 | -3,14 | -4,94 | -7,11 | -6,37 | 1,46   |
| Slovinsko       | 3,38  | -7,84 | 1,24  | 0,60  | -2,34 | 2,07   |
| Španělsko       | 0,89  | -3,74 | -0,32 | 0,42  | -1,42 | 2,17   |
| Švédsko         | -0,61 | -5,03 | 6,56  | 3,71  | 0,74  | 2,17   |
| Švýcarsko       | 2,16  | -1,94 | 3,03  | 1,93  | 0,98  | 1,42   |
| Turecko         | 0,66  | -4,83 | 9,16  | 8,77  | 2,24  | 4,04   |
| USA             | -0,36 | -3,11 | 2,38  | 1,80  | 2,21  | 2,14   |
| Velká Británie  | -0,97 | -3,97 | 1,80  | 0,99  | 0,27  | 2,45   |

Tabulka B.3: HDP v zemích OECD, 2007-2012. Průměr HDP pro 1991-2012.



# Příloha C

## Zdrojový kód v R

```
#####
gen<-function(phi,omega2,sigmaEps,t){
#####
# Funkce: Generuje panelový AR(1) model s pevným nebo náhodným koeficientem.
#
# Předpoklady: Reziduální složka a náhodný koeficient mají normální rozdělení.
#               Počet průřezů je alespoň 2.
#
# Vstup:
#   phi - koeficient AR(1),
#   omega2 - rozptyl náhodné složky koeficientu (jestliže je roven 0, pak
#           generujeme AR(1) s pevným koeficientem),
#   sigmaEps - vektor rozptylů reziduálních složek průřezů, jeho délka udává
#             počet průřezů,
#   t - časová délka simulovaných řad.
#
# Výstup:
#   sim.data - soubor dat obsahující ve sloupcích length(sigmaEps) řad o délce t+1.
#
# Příklad použití: gen(0.5,0.6,c(1,2,0.8),500)
#####

N=length(sigmaEps)
T=2*t

cells<-rnorm(T, 0, sqrt(sigmaEps[1]))
for(i in 2:N) {
  k<-rnorm(T, 0, sqrt(sigmaEps[i]))
  cells = rbind(cells,k)
}
eps<-t(cells)

x<-matrix(rep(0,N*T), nrow=T, ncol=N, byrow=TRUE)

for(j in 1:N) {
  for(i in 2:T) {
    bij<-0
    bij<-rnorm(1,0, sqrt(omega2))
    x[i,j]= phi*x[i-1,j]+bij*x[i-1,j] + eps[i,j]
  }
}
```

```

    }
  }

sim.data<-data.frame(x[t:T, 1:N])
return(sim.data)
}

#####
fixed<-function(data){
#####
# Funkce: Odhaduje panelový AR(1) model s pevným koeficientem.
#
# Předpoklady: Počet průřezů je alespoň 2.
#
# Vstup:
#   data - soubor panelových dat se sloupci reprezentujícími průřezy.
#
# Výstup:
#   phiLS - odhad koeficientu odvozený na základě metody nejmenších čtverců,
#   phiML - odhad koeficientu odvozený na základě metody maximální věrohodnosti,
#   sigma2i - vektor odhadů rozptylů reziduálních složek jednotlivých průřezů.
#####

N<-length(data[,1])
time<-length(data[,1])

LSSoucet<-0
LSSoucet2<-0
for(j in 2:time) {
  for(i in 1:N) {
    LSSoucet<-LSSoucet+data[j,i]*data[j-1,i]
    LSSoucet2<-LSSoucet2+data[j-1,i]*data[j-1,i]
  }
}
phiLS<-LSSoucet/LSSoucet2
names(phiLS)<-"phi.LS"

sigma2i<-rep(0,N)
for (i in 1:N) {
  sigma2i[i] = sum( (data[2:time,i] - phiLS*data[1:(time-1),i])^2 )/time
}

MLSoucet<-0
MLSoucet2<-0
for(j in 2:time) {
  for(i in 1:N) {
    MLSoucet<-MLSoucet+data[j,i]*data[j-1,i]/(sigma2i[i])
    MLSoucet2<-MLSoucet2+data[j-1,i]*data[j-1,i]/(sigma2i[i])
  }
}
phiMLE<-MLSoucet/MLSoucet2
names(phiMLE)<-"phi.MLE"

return(c(phiLS,phiMLE, sigma2i))
}

```

```

#####
random<-function(data){
#####
# Funkce: Odhaduje panelový AR(1) model s náhodným koeficientem.
#
# Předpoklady: Počet průřezů je alespoň 2.
#
# Požadavek: Reshape2 Package.
#
# Vstup:
#   data - soubor panelových dat se sloupci reprezentujícími průřezy.
#
# Výstup:
#   phiLS - odhad koeficientu odvozený na základě metody nejmenších čtverců,
#   phiML - odhad koeficientu odvozený na základě metody maximální věrohodnosti,
#   omega2 - odhad rozptylu náhodného koeficientu,
#   sigma2 - vektor odhadů rozptylů reziduálních složek jednotlivých průřezů.
#####

N<-length(data[,1])
time<-length(data[,1])

LScoucet<-0
LScoucet2<-0
for(j in 2:time) {
  for(i in 1:N) {
    LScoucet<-LScoucet+data[j,i]*data[j-1,i]
    LScoucet2<-LScoucet2+data[j-1,i]*data[j-1,i]
  }
}
phiLS<-LScoucet/LScoucet2
names(phiLS)<-"phi.LS"

u2 <-(data[2:time,] - phiLS*data[1:(time-1),])^2
data.T<-data[-time,]
pom.data<-data.frame(data.T,u2)

require(reshape2)
col.X<-melt(pom.data[,1:N])
col.U<-melt(pom.data[, (N+1):(2*N)])
panel.f<-factor(col.U$variable)
regrese<-lm(col.U$value~panel.f+I(col.X$value^2))

sigma2<-rep(0,N)
for (i in 2:N) {
  sigma2[i]<-regrese$coef[1]+regrese$coef[i]
}
sigma2[1]<-regrese$coef[1]
for (i in 1:N) { if (sigma2[i]<0) sigma2[i]<-0 }
names(sigma2)<-"komponenty rozptylu"

if (regrese$coef[N+1]>0) omega2<-regrese$coef[N+1]
else omega2<-0
names(omega2)<-"omega2"

MLscoucet<-0
MLscoucet2<-0
for(j in 2:time) {

```

```

for(i in 1:N) {
  MLsoucet<-(MLsoucet+
    data[j,i]*data[j-1,i]/(sigma2[i]+omega2*data[j-1,i]*data[j-1,i]))
  MLsoucet2<-(MLsoucet2+
    data[j-1,i]*data[j-1,i]/(sigma2[i]+omega2*data[j-1,i]*data[j-1,i]))
}
}
phiML<-MLsoucet/MLsoucet2
names(phiMLE)<-"phi.ML"

return(c(phiLS, phiML, omega2, sigma2))
}

```

# Literatura

- Anděl, J. (2007). *Základy matematické statistiky*, Matfyzpress.
- Balestra, P. a Nerlove, M. (1966). Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas, *Econometrica* **34**: 585–612.
- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd.
- Billingsley, P. (1961). The Lindeberg-Levy Theorem for Martingales, **12**: 788–792.
- Breusch, T. S. a Pagan, A. R. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics, *Review of Economic Studies* **47**(1): 239–253.
- Cipra, T. (2008). *Finanční ekonometrie*, Ekopress.
- Croissant, Y. a Millo, G. (2008). Panel Data Econometrics in R: The plm Package, *Journal of Statistical Software* **27**(2): 1–43.
- Daniels, H. E. (1939). The Estimation of Components of Variance, *Supplement to the Journal of the Royal Statistical Society* **6**(2): 186–197.
- Eisenhart, C. (1947). The Assumptions Underlying the Analysis of Variance, *Biometrics* **3**(1): 1–21.
- Fisher, R. A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance, *Transactions of the Royal Society of Edinburgh* **52**: 399–433.
- Greene, W. H. (2002). *Econometric Analysis*, Prentice Hall.

- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press.
- Hausman, J. A. (1978). Specification Tests in Econometrics, *Econometrica* **46**(6): 1251–1271.
- Hoch, I. (1962). Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data, *Econometrica* **30**(1): 34–53.
- Lachout, P. (1998). *Teorie pravděpodobnosti*, Karolinum.
- Nerlove, M. (2000). An Essay on the History of Panel Data Econometrics, 9<sup>th</sup> International Conference on Panel Data, červen 2000, Geneva, Švýcarsko.
- Neyman, J. a Scott, E. L. (1948). Consistent estimation from partially consistent observations, *Econometrica* **16**(1): 1–32.
- OECD (2013). OECD.StatExtracts, <http://stats.oecd.org/>. [Online, naposledy navštíveno 30. 6. 2013].
- Prášková, Z. (2007). *Základy náhodných procesů II*, Karolinum.
- Rao, C. R. (1973). *Linear statistical inference and its applications*, John Wiley & Sons, Inc.
- Schick, A. (1996). Root-N Consistent Estimation in a Random Coefficient Autoregressive Model, *The Australian Journal of Statistics* **38**(2): 155–160.
- Vaněček, P. (2007). *Estimators of Random Coefficient Autoregressive models*, PhD thesis, Charles University in Prague.
- White, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press.
- Zvára, K. (2008). *Regrese*, Matfyzpress.