

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Magdalena Komzáková

Zdánlivá regrese ekonomických ukazatelů

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Doc. RNDr. Petr Lachout, CSc.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická statistika
a ekonometrie

Studijní plán: Ekonometrie

Praha 2006

Poděkování

Poděkování patří především vedoucímu mé diplomové práce Doc. RNDr. Petru Lachoutovi, CSc. za cenné připomínky. Dále bych ráda poděkovala své rodině a přátelům za potřebnou podporu v průběhu celého studia.

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

V Praze dne 17.4.2006

Magdalena Komzáková

Obsah

Úvod	i
1 Základní definice	1
1.1 Náhodný proces	1
1.1.1 Stacionarita	2
1.1.2 Ergodicita	3
1.2 Bílý šum	4
1.3 Procesy klouzavých součtů	4
1.3.1 $MA(1)$ proces	4
1.3.2 $MA(q)$ proces	5
1.3.3 $MA(\infty)$ proces	6
1.4 Autoregresní procesy	7
1.4.1 $AR(1)$ proces	7
1.4.2 $AR(p)$ proces	7
1.5 Smíšené autoregresní procesy klouzavých součtů	8
1.5.1 $ARMA(p, q)$ proces	8
2 Nestacionární procesy	10
2.1 Procesy s deterministickým trendem	11
2.2 Procesy se stochastickým trendem	
Jednotkové kořeny	13
2.2.1 Jednoduché příklady	13
2.2.2 Integrované procesy	15
2.2.3 $ARIMA(p, d, q)$ proces	15
2.2.4 $SARIMA(p, d, q) \times (P, D, Q)_s$ proces	16
3 Zdánlivá regrese	18
3.1 Vlastnosti $I(0)$ a $I(1)$ procesů	18
3.2 Ekvilibrium	19
3.3 Durbin-Watsonův test	19
3.4 Zdánlivá regrese: Granger & Newbold	21
3.4.1 Příklad	22
3.5 Zdánlivá regrese obecně	23
3.6 Léky zdánlivé regrese	25

<i>OBSAH</i>	3
4 Rozpoznání zdánlivé regrese	27
4.1 Kointegrace	27
4.1.1 Příklad	27
4.1.2 Definice	28
4.2 Testy jednotkového kořene a stacionarity	30
4.2.1 Testy jednotkového kořene	30
4.2.2 Testy stacionarity	36
4.3 Testy kointegrace	37
5 Příklad	39
6 Závěr	42
Literatura	43

Název práce: Zdánlivá regrese ekonomických ukazatelů

Autor: Magdalena Komzáková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Doc. RNDr. Petr Lachout, CSc.

e-mail vedoucího: Petr.Lachout@mff.cuni.cz

Abstrakt: V analýze časových řad je častým jevem, že se dvě nebo více časových řad navzájem ovlivňuje. Pokud ovšem generující náhodné procesy těchto řad nemají stacionární strukturu, ale jsou stochasticky nestacionární, tj. jejich charakteristický polynom obsahuje jednotkový kořen, stává se nezdánlivá, že i regrese modelující vzájemnou závislost několika naprosto nesouvisejících řad dává statisticky významné odhady parametrů a statistiky běžné k posuzování výstižnosti použitého modelu nenaznačují nic o jeho nevhodnosti. Tento problém se nazývá zdánlivou regresí a je řešen teorií kointegrace, neboť můžeme prohlásit, že pokud jsou řady kointegrované, ukazuje jejich model skutečnou závislost, nikoliv jen zdánlivou. Proto jsou testy kointegrace současně využívány i k testování přítomnosti zdánlivé regrese. Tyto testy vycházejí z testů přítomnosti jednotkového kořene v generujícím procesu (nebo z testů stacionarity řady), protože časové řady mohou být kointegrované pouze v případě, kdy jejich lineární kombinace je sama stacionární řadou.

Klíčová slova: zdánlivá regrese, kointegrace, procesy s jednotkovým kořenem

Title: Seeming regression of economic indices

Author: Magdalena Komzáková

Department: Department of Probability and Mathematical Statistics

Supervisor: Doc. RNDr. Petr Lachout, CSc.

Supervisor's e-mail address: Petr.Lachout@mff.cuni.cz

Abstract: In the time series analysis it often appears that two or more time series influence each other. When the generating stochastic processes of these series do not have stationary structure but they are stochastically non-stationary, i.e. the characteristic polynomial has a unit root, it happens that the regression modelling the dependence of some absolutely independent series gives statistically significant parameter estimations and statistics used to judge the model fitting do not indicate anything about its impropriety. This phenomenon is called seeming regression (spurious regression) and is solved with the theory of cointegration. We can say that when the series are cointegrated, their model shows their real dependence, not only the seeming one. Due to this fact, cointegration tests are also used for testing for the presence of seeming regression. These tests are based on unit root tests in generating process (or on stationarity tests), because time series can be cointegrated only if their linear combination is a stationary series.

Keywords: seeming regression, spurious regression, cointegration, unit root processes

Úvod

Pro modelování vztahů mezi ekonomickými veličinami jsou v ekonometrii často používané jednorovnicové regresní modely. Při tomto postupu je jedna ekonomická veličina určena jako závislá proměnná (vysvětlovaná, endogenous) a druhá jako nezávislá (vysvětlující, exogenous), s jejíž pomocí se závislá proměnná modeluje. Obě proměnné mohou být prostorově i časově uspořádané, a právě díky specifickým vlastnostem časového uspořádání může při tomto modelování docházet k určitým problémům. Jedním z nich je i fenomén zvaný "Zdánlivá regrese".

Lze ji ilustrovat situací, kdy máme k dispozici dvě časové řady, které jsou nezávislé. Pokud ale provedeme regresi jedné na druhou, tj. budeme jednu řadu považovat za vysvětlovanou a druhou za vysvětlující, může se stát, že metodou nejmenších čtverců získáme statisticky významné odhady parametrů dané regresní funkce. Z logiky věci (z nezávislosti zkoumaných řad) však vyplývá, že je tento závěr chybný. Tato skutečnost pak v praxi často vede k mylným závěrům o vztazích ekonomických veličin.

Problém zdánlivé regrese je znám již 80 let, přesto je stále aktuální a ne mnoho praktikujících ekonomů s ním umí zacházet. Na toto téma již bylo sepsáno hodně článků, ne tak však v české literatuře. Proto je cílem této práce shrnout základní známé informace o zdánlivé regresi ze zahraniční literatury, které by mohly nadále posloužit zájemcům jako startovní čára pro hlubší studium.

Předpokladem pro porozumění následujícího textu je alespoň základní znalost matematické statistiky a teorie náhodných procesů.

Historie

Zdánlivá regrese, neboli "nesmyslná korelace", jak byla původně nazývána, má ve statistice dlouhou minulost. Poprvé byl tento problém zaznamenán v práci G.U.Yulea v roce 1926. V ekonometrické a statistické literatuře je uvedeno mnoho zajímavých, často však spíše humorných příkladů tohoto zvláštního - a na první pohled nepochopitelného - úkazu.

Jedním z nich je například vysoká korelace mezi počtem ministrů a mírou alkoholismu ve Velké Británii v 19. století. Další vyslovil sám Yule, když zjistil korelaci

0,95 mezi poměrem sňatků v anglikánské církvi k celkovému počtu sňatků (ve Velké Británii) a mírou úmrtnosti v období let 1866 až 1911. A do třetice ekonometrický příklad, který zmínil Hendry (1980) a který prokazuje vysokou spojitost mezi cenovou hladinou a kumulativními srážkami ve Spojeném království. Tento "vztah" pak byl svým autorem žertovně nazván "novou teorií inflace". Díky těmto příkladům se tedy zdánlivá regrese dostala do povědomí mnoha lidí.

V roce 1974 vyšel v časopise *Journal of Econometrics* článek od autorů Grangera a Newbolda, který se stal velice známým. Autoři zde na základě Monte Carlo simulací prokázali, že se zdánlivá regrese projeví, pokud testujeme závislost mezi dvěma nezávislými náhodnými procházkami. Regresní koeficient ve většině případů vyšel statisticky signifikantní a hodnota koeficientu determinace tohoto modelu velice blízká jedničce. Tento článek zahájil moderní diskusi o problému zdánlivé regrese a vedl k vytvoření konceptu kointegrace.

Uveďme na tomto místě pro zajímavost ještě jeden kuriózní příklad: za vysvětlovanou proměnnou považujeme míru úmrtnosti novorozeňat v Egyptě v letech 1971 až 1990 (roční údaje), za vysvětlující pak příjem amerických farmářů a celkovou zásobu peněz Hondurasu. Pak klíčové statistiky vycházejí následovně: $R^2 = 0,918$, $F = 95,17$, což ukazuje na významnou vzájemnou závislost použitých ekonomických ukazatelů.

Kapitola 1

Základní definice

V této úvodní kapitole si připomeneme některé základní definice a pojmy z teorie časových řad, jako *náhodný proces*, *stacionarita*, *ergodicita*, a zavedeme *MA*, *AR*, *ARMA* procesy.

1.1 Náhodný proces

Definice 1.1: Náhodný proces je rodina náhodných veličin $\{Y_t, t \in T\}$, $T \subset \mathbf{R}$, definovaných na pravděpodobnostním prostoru (Ω, A, P) .

Pokud platí, že $T = \mathbf{Z}$ nebo $T \subset \mathbf{Z}$, jedná se o *náhodný proces s diskrétním časem* neboli o *časovou řadu*. V případě, že $T = [a, b]$, kde $-\infty \leq a < b \leq \infty$, mluvíme o *náhodném procesu se spojitým časem*.

V naší analýze se budeme zabývat prvním z výše zmíněných, tedy *časovou řadou*. Přitom nebudeme rozlišovat mezi samotným náhodným procesem a jeho realizací.

Je dobré si také připomenout, že rozdělení náhodného procesu $\{Y_t, t \in T\}$ je jednoznačně určeno, a to rozdělením všech konečněrozměrných náhodných vektorů $(Y_{t_1}, \dots, Y_{t_n})$, $n \in \mathbf{N}$, což vyplývá z Daniellovy-Kolmogrovovy věty. Její plné znění viz např. Prášková, Lachout (1998).

Pro náhodný proces $\{Y_t, t \in T\}$ definujeme následující:

- **střední hodnota:** $\mu_t = EY_t$, pokud pro všechna $t \in T$ existuje.

Pokud $\mu_t = 0$ pro $\forall t \in T$, nazveme $\{Y_t, t \in T\}$ *centrovaným procesem*.

- **autokovarianční funkce:** $\gamma(s, t) = cov(Y_s, Y_t)$, pokud pro $\forall t \in T$: $E|Y_t|^2 < \infty$.

Rozptyl procesu $\{Y_t, t \in T\}$ je speciálním případem autokovarianční funkce, když $s = t$. Tedy $Var(Y_t) = \gamma(t, t)$ je rozptyl procesu v čase t .

- autokorelační funkce:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}},$$

pokud $\gamma(t, t) > 0$ pro $\forall t \in T$.

Distribuce náhodného procesu je určena prvními a druhými momenty (střední hodnotou a autokorelační funkcí). Pokud je proces *gaussovský*, tj. všechna jeho konečněrozměrná rozdělení jsou normální, je jeho distribuce těmito momenty určena jednoznačně.

1.1.1 Stacionarita

Definice 1.2: Náhodný proces $\{Y_t, t \in \mathbf{Z}\}$ nazveme *striktně stacionárním*, pokud sdružené distribuční funkce vektorů $(Y_{t_1}, \dots, Y_{t_n})$ a $(Y_{t_1+h}, \dots, Y_{t_n+h})$ jsou stejné po všechna $n \in \mathbf{N}$ a všechna $t_1, \dots, t_n, h \in \mathbf{Z}$.

Pokud je náhodný proces striktně stacionární, nemění se distribuční vlastnosti (střední hodnota, rozptyl,...) náhodných veličin Y_t v čase. Tato definice je poměrně přísná, proto uvedeme slabší verzi stacionarity.

Definice 1.3: Náhodný proces $\{Y_t, t \in \mathbf{Z}\}$ nazveme *slabě (kovariančně) stacionárním*, pokud splňuje následující podmínky:

- $E|Y_t|^2 < \infty$
- $E(Y_t) = \mu, \quad \forall t \in \mathbf{Z}$
- $\gamma(s, t) = \gamma(s + r, t + r), \quad \forall t, s, r \in \mathbf{Z}$

Pokud je proces slabě stacionární, lze jeho autokovarianční funkci vyjádřit jako funkci jedné proměnné, protože závisí pouze na časovém intervalu mezi dvěma pozorováními:

$$\gamma(s, t) = \gamma(t - s, 0)$$

Proto predefinujeme autokovarianční funkci pro slabě stacionární procesy:

$$\gamma(h) = \gamma(h, 0) = \text{cov}(Y_t, Y_{t-h}) \quad \forall h, t \in \mathbf{Z}.$$

Dále platí, že autokovarianční funkce je symetrická podle nuly:

$$\gamma(h) = \gamma(-h) \quad \forall h \in \mathbf{Z}.$$

Z výše uvedeného je vidět, že pro slabě stacionární proces také platí, že jeho rozptyl je konstantní:

$$\text{Var}(Y_t) = \gamma(0)$$

a autokorelační funkce vypadá následovně:

$$\rho(t) = \frac{\gamma(t)}{\gamma(0)}.$$

Věta 1.1: Každý striktně stacionární náhodný proces $\{Y_t, t \in \mathbf{Z}\}$ s konečnými druhými momenty je i slabě stacionární.

Důkaz je velmi snadný, viz. např. Prášková (2004).

Opačná implikace bohužel neplatí obecně. Platí ovšem, pokud je náhodný proces gaussovský.

Pro analýzu časových řad se málokdy používají striktně stacionární řady, protože se velmi zřídka objevují v praxi. Proto v následujícím textu budeme pokaždé výrazem "stacionární" myslet slabě stacionární procesy.

1.1.2 Ergodicita

Jelikož máme pro každou časovou řadu pouze jednu realizaci (nemůžeme se vrátit v čase zpět a nechat ji "běžet" znova), zavádíme pojem ergodicity, který nám zajistí jistou dávku stability procesu.

Definice 1.4: Řekneme, že stacionární proces $\{Y_t, t \in \mathbf{Z}\}$ je *ergodický podle (kvadratického) středu*, pokud platí:

$$\bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t \xrightarrow{P} EY_t = \mu$$

pro $T \rightarrow \infty$.

Stacionární proces je ergodický podle středu, pokud je jeho kovarianční funkce absolutně sčítatelná, tj. splňuje následující podmínku:

$$\sum_{h=0}^{\infty} |\gamma(h)| < \infty.$$

Tato podmínka nám zajišťuje, že s rostoucím časem jde kovarianční funkce dostatečně rychle k nule. Důkaz uvedl např. Hamilton (1994).

Definice 1.5: Řekneme, že stacionární proces $\{Y_t, t \in \mathbf{Z}\}$ je *ergodický do druhých momentů*, pokud platí:

$$\hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=h+1}^T (Y_t - \mu)(Y_{t-h} - \mu) \xrightarrow{P} \gamma(h)$$

pro $T \rightarrow \infty$ a $\forall h \in \mathbf{Z}$.

V mnoha případech se podmínky stacionarity a ergodicity slučují ve stejné požadavky. Není tomu tak ale vždy, proto nelze ergodicitu zanedbávat.

1.2 Bílý šum

Bílý šum je základním stavebním kamenem pro většinu dalších procesů, proto je jeho znalost nezbytně nutná.

Definice 1.6: Náhodný proces $\{\epsilon_t, t \in \mathbf{Z}\}$ nazýváme *procesem bílého šumu* právě tehdy, když splňuje následující podmínky:

- $E(\epsilon_t) = 0$ pro $\forall t \in \mathbf{Z}$
- $E(\epsilon_t \epsilon_s) = \begin{cases} \sigma^2, & t = s \\ 0, & \text{jinak} \end{cases}$

Z definice je zřejmé, že proces bílého šumu je vždy stacionární.

Někdy se proces bílého šumu značí *WN*, z anglického "white noise". V tomto případě se můžeme setkat se zápisem: $\{\epsilon_t, t \in \mathbf{Z}\} \sim WN(0, \sigma^2)$.

Pokud jsou navíc ϵ_t a ϵ_s nezávislá pro $t \neq s$, jedná se o *nezávislý bílý šum* a značí se $\{\epsilon_t\} \sim IWN(0, \sigma^2)$ (*I* pochází z anglického "independent").

Pokud ještě kromě toho pro všechna t platí, že $\epsilon_t \sim N(0, \sigma^2)$, pak nazýváme bílý šum *gaussovským* a značíme ho $\{\epsilon_t\} \sim GWN(0, \sigma^2)$.

1.3 Procesy klouzavých součtů

1.3.1 MA(1) proces

Definice 1.7: Nechť je $\{\epsilon_t\}$ bílý šum. Uvažujme proces

$$Y_t = \mu + \epsilon_t + \theta \epsilon_{t-1},$$

kde μ a θ jsou libovolné konstanty. Tato časová řada se nazývá *proces klouzavých součtů prvního řádu* a značí se *MA(1)*.

Základní vlastnosti $MA(1)$ procesu jsou následující:

- $E(Y_t) = \mu$ pro $\forall t$
- $\gamma(h) = \begin{cases} \sigma^2(1 + \theta^2), & h = 0 \\ \theta\sigma^2, & h = 1 \\ 0, & h > 1 \end{cases}$

Z toho vyplývá, že $MA(1)$ proces je vždy stacionární a ergodický podle středu.

1.3.2 $MA(q)$ proces

Definice 1.8: Nechť je $\{\epsilon_t\}$ bílý šum. Uvažujme proces

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q},$$

kde μ a $(\theta_1, \theta_2, \dots, \theta_q)$ jsou libovolné konstanty. Tato časová řada se nazývá *proces klouzavých součtů q -tého řádu* a značí se $MA(q)$.

$MA(q)$ proces se dá také zapsat pomocí sumy:

$$Y_t = \mu + \sum_{j=0}^q \theta_j \epsilon_{t-j},$$

kde $\theta_0 = 1$.

Další častý způsob zápisu je pomocí operátoru zpětného posunutí L (z anglického "lag"), který je definován tak, že $L^j(Y_t) = Y_{t-j}$:

$$Y_t = \mu + \theta(L)\epsilon_t,$$

kde $\theta(L)$ značí polynom

$$\theta(L) = \sum_{j=0}^q \theta_j L^j$$

a $\theta_0 = 1$.

Řekneme, že $MA(q)$ proces je invertibilní, pokud je splněna tzv. *podmínka invertibility*: všechny kořeny charakteristického polynomu $\theta(z) = \sum_{j=0}^q \theta_j z^j$ (pro $z \in \mathbf{C}$) leží vně jednotkového kruhu, tj. pro $\forall z \in \mathbf{C}$ taková, že $\theta(z) = 0$, platí: $|z| > 1$. Potom existuje *AR* reprezentace (viz dále) $MA(q)$ procesu.

Základní vlastnosti $MA(q)$ procesu jsou následující:

- $E(Y_t) = \mu$ pro $\forall t$
- $\gamma(h) = \begin{cases} \sigma^2(\theta_h + \theta_1\theta_{h+1} + \cdots + \theta_{q-h}\theta_q), & h \leq q \\ 0, & h > q \end{cases}$

Z toho vyplývá, že $MA(q)$ proces je vždy stacionární a ergodický podle středu.

1.3.3 $MA(\infty)$ proces

Definice 1.9: Necht' je $\{\epsilon_t\}$ bílý šum. Uvažujme proces

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots,$$

kde μ a (ψ_1, ψ_2, \dots) jsou libovolné konstanty, $\psi_0 = 1$. Tato časová řada se nazývá *proces klouzavých součtů ∞ -ného řádu* a značí se $MA(\infty)$.

$MA(\infty)$ proces lze opět zapsat pomocí operátoru zpětného posunutí L :

$$Y_t = \mu + \psi(L)\epsilon_t,$$

kde $\psi(L)$ značí polynom

$$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$$

a $\psi_0 = 1$.

Řekneme, že $MA(\infty)$ proces je dobře definován, pokud koeficienty ψ splňují následující podmínku kvadratické sčítatelnosti:

$$\sum_{j=0}^{\infty} |\psi_j|^2 < \infty.$$

Často bývá ovšem zvykem pracovat s poněkud silnější podmínkou absolutní sčítatelnosti:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Základní vlastnosti $MA(\infty)$ procesu jsou následující, jedná se o limitní vlastnosti procesu $MA(q)$ pro $q \rightarrow \infty$:

- $E(Y_t) = \mu$ pro $\forall t$
- $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$ pro $h = 0, 1, 2, \dots$

Z toho tedy vyplývá, že $MA(\infty)$ proces, který je dobře definován (tj. splňuje podmínku kvadratické sčítatelnosti koeficientů), je vždy stacionární a ergodický podle středu.

1.4 Autoregresní procesy

1.4.1 $AR(1)$ proces

Definice 1.10: Necht' je $\{\epsilon_t\}$ bílý šum. Uvažujme proces

$$Y_t = c + \phi Y_{t-1} + \epsilon_t,$$

kde c a ϕ jsou libovolné konstanty. Tato časová řada se nazývá *autoregresní proces prvního řádu* a značí se $AR(1)$.

$AR(1)$ proces lze také pomocí rekurze zapsat následujícím způsobem:

$$Y_t = \sum_{j=0}^{\infty} \phi^j (c + \epsilon_{t-j})$$

.

Pokud je navíc $|\phi| < 1$, existuje jednoznačně určený náhodný proces, který splňuje danou $AR(1)$ rovnici. Tento proces je stacionární a navíc kauzální, tj. existuje jeho $MA(\infty)$ reprezentace.

Základní vlastnosti stacionárního $AR(1)$ procesu jsou následující:

- $E(Y_t) = \mu = \frac{c}{1-\phi}$ pro $\forall t$
- $\gamma(h) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}$ pro $\forall h$

Stacionární $AR(1)$ proces je vždy také ergodický podle středu.

1.4.2 $AR(p)$ proces

Definice 1.11: Necht' je $\{\epsilon_t\}$ bílý šum. Uvažujme proces

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t,$$

kde c a (ϕ_1, \dots, ϕ_p) jsou libovolné konstanty. Tato časová řada se nazývá *autoregresní proces p -tého řádu* a značí se $AR(p)$.

$AR(p)$ proces se často zapisuje pomocí operátoru zpětného posunutí L :

$$\phi(L)Y_t - c = \epsilon_t,$$

kde $\phi(L)$ značí polynom

$$\phi(L) = 1 - \sum_{j=1}^p \phi_j L^j.$$

Proces $AR(p)$ je stacionární, pokud splňuje tzv. *podmínku stacionarity*: Pokud všechny kořeny charakteristického polynomu $\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$ (pro $z \in \mathbf{C}$) leží vně jednotkového kruhu, tj. pokud pro $\forall z \in \mathbf{C}$ taková, že $\phi(z) = 0$, platí: $|z| > 1$, pak existuje jednoznačně určený náhodný proces, který splňuje danou $AR(p)$ rovnici. Tento proces je stacionární a navíc kauzální, tj. existuje jeho $MA(\infty)$ reprezentace:

$$Y_t = \mu + \psi(L)\epsilon_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

kde $\psi(L) = \phi(L)^{-1}$.

Můžeme si všimnout, že požadavek kladený na koeficient ϕ $AR(1)$ procesu je totožný s podmínkou stacionarity.

Základní vlastnosti stacionárního $AR(p)$ procesu jsou následující:

- $E(Y_t) = \mu = \frac{c}{1 - \sum_{j=1}^p \phi_j} = \frac{c}{\phi(1)}$ pro $\forall t$
- $\gamma(h) = \begin{cases} \sum_{j=1}^p \phi_j \gamma(h-j), & h > 0 \\ \sum_{j=1}^p \phi_j \gamma(j) + \sigma^2, & h = 0 \end{cases}$

Stacionární $AR(p)$ proces je vždy také ergodický podle středu.

1.5 Smíšené autoregresní procesy klouzavých součtů

1.5.1 $ARMA(p, q)$ proces

Definice 1.12: Necht' je $\{\epsilon_t\}$ bílý šum. Uvažujme proces

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q},$$

kde c , (ϕ_1, \dots, ϕ_p) a $(\theta_1, \dots, \theta_q)$ jsou libovolné konstanty. Tato časová řada se nazývá *autoregresní proces klouzavých součtů řádu p, q* a značí se $ARMA(p, q)$.

Z definice je vidět, že se $ARMA(p, q)$ proces skládá z autoregresní části a z části klouzavých součtů. Pomocí operátoru zpětného posunutí má následující tvar:

$$\phi(L)Y_t = c + \theta(L)\epsilon_t,$$

kde

$$\phi(L) = 1 - \sum_{j=1}^p \phi_j L^j,$$

$$\theta(L) = \sum_{j=0}^q \theta_j L^j$$

a $\theta_0 = 1$.

Platí:

$$\begin{aligned} ARMA(p, 0) &= AR(p), & p \geq 1, \\ ARMA(0, q) &= MA(q), & q \geq 0. \end{aligned}$$

$ARMA(p, q)$ proces je stacionární, když je splněna podmínka stacionarity pro autoregresní část. Pokud tedy platí, že všechny kořeny charakteristického polynomu

$$\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j = 0$$

leží vně jednotkového kruhu, existuje jednoznačně určený náhodný proces, který splňuje danou $ARMA(p, q)$ rovnici. Tento proces je stacionární a navíc kauzální, tj. existuje jeho $MA(\infty)$ reprezentace.

Základní vlastnosti stacionárního $ARMA(p, q)$ procesu jsou následující:

- $E(Y_t) = \mu = \frac{c}{1 - \sum_{j=1}^p \phi_j} = \frac{c}{\phi(1)}$ pro $\forall t$
- $\gamma(h) = \begin{cases} \sum_{j=1}^p \phi_j \gamma(h-j), & h > q \\ \sum_{j=1}^p \phi_j \gamma(h-j) + \sigma^2 \sum_{k=h}^q \theta_k \psi_{k-h}, & 0 \leq h \leq q, \end{cases}$

kde ψ_j jsou koeficienty z $MA(\infty)$ reprezentace $ARMA(p, q)$ procesu.

Kapitola 2

Nestacionární procesy

V minulé kapitole jsme uvažovali procesy stacionární a ergodické podle středu, tj. náš požadavek byl, aby

1. střední hodnota procesu $Y_t, t \in \mathbf{Z}$ byla nezávislá na čase:

$$E(Y_t) = \mu$$

2. předpověď procesu konvergovala ke střední hodnotě procesu:

$$\lim_{s \rightarrow \infty} \hat{Y}_{t+s|t} = \lim_{s \rightarrow \infty} \hat{E}(Y_{t+s} | Y_t, Y_{t-1}, \dots) = \mu.$$

Tyto požadavky jsou ale pro mnoho ekonomických a finančních časových řad nerealistické. Existují dva důvody pro nestacionaritu řad:

- nestálá střední hodnota kvůli deterministickému trendu, sezónním komponentám či zlomům v deterministických trendech
 - tyto řady se pak nazývají *deterministicky nestacionární*;
- jednotkový kořen v *AR*-polynomu
 - pak se jedná o *stochasticky nestacionární* řady.

V této kapitole se budeme věnovat právě těmto nestacionárním procesům a doplníme výčet používaných procesů o *ARIMA* a sezónní *ARIMA* (*SARIMA*) procesy.

2.1 Procesy s deterministickým trendem

Příkladem procesu s deterministickým trendem je následující proces vycházející z $MA(\infty)$ reprezentace:

$$Y_t = c + \beta t + \psi(L)\epsilon_t,$$

kde $\{\epsilon_t\}$ je bílý šum a $\psi(L)$ je definován v předchozí kapitole.

Někdy se takovým procesům říká *trendově stacionární*, protože po odečtení trendu βt získáme klasický $MA(\infty)$ proces, který je vždy stacionární. Samozřejmě musí i zde být splněna podmínka: $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$.

Nejjednodušším případem je proces, kde $\psi(L) = 1$:

$$Y_t = c + \beta t + \epsilon_t,$$

$$\epsilon_t \sim WN(0, \sigma^2), t = 1, \dots, T.$$

Takto dostáváme vlastně klasický lineární regresní model s maticí regresorů

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{bmatrix}$$

a s neznámými parametry c a β , které lze odhadnout pomocí metody nejmenších čtverců.

Dalším příkladem je stacionární $ARMA(p, q)$ proces *kolem deterministického trendu*:

$$Y_t = c + \beta t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \dots + \theta_q \epsilon_{t-q},$$

kde $\epsilon_t \sim WN(0, \sigma^2)$.

Pomocí operátoru zpětného posunutí pak vypadá takto:

$$\phi(L)Y_t = c + \beta t + \theta(L)\epsilon_t.$$

Pro stacionaritu musí platit:

$$\phi(z) \neq 0 \quad \text{pro} \quad \forall z : |z| \leq 1.$$

Tento proces se po několika úpravách dá také zapsat jako součet deterministického lineárního trendu a $ARMA(p, q)$ procesu X_t s nulovou střední hodnotou:

$$Y_t = \mu_0 + \mu_1 t + X_t,$$

kde

$$\mu_0 = \frac{c}{\phi(1)} - \kappa,$$

$$\mu_1 = \frac{\beta}{\phi(1)},$$

$$\kappa = \frac{\beta \sum_{j=1}^p j \phi_j}{\phi(1)^2}$$

a

$$X_t = \phi^{-1}(L)\theta(L)\epsilon_t$$

je stacionární proces s $E(X_t) = 0$.

S takto upraveným procesem je nadále výhodnější pracovat, neboť platí:

$$E(Y_t) = \mu_0 + \mu_1 t.$$

Uveďme ještě několik příkladů, jak je možné deterministickou složku trendu rozšířit. Uvažujme proces

$$Y_t = d_t + X_t,$$

kde d_t znázorňuje deterministickou složku a X_t stacionární $ARMA(p, q)$ proces s nulovou střední hodnotou. Platí:

$$d_t = \mu_0 + \mu_1 t + b_t,$$

kde b_t značí rozšiřující zlomovou funkci ("break function").

Následují příklady takové zlomové funkce, se kterou je možné se v praxi setkat:

- $b_t = \begin{cases} \omega & t = m \\ 0 & \text{jinak} \end{cases}$

– odlehlé pozorování v bodě $t = m$

- $b_t = \begin{cases} \omega & t \geq m \\ 0 & \text{jinak} \end{cases}$

– posunutí úrovně

- $b_t = \begin{cases} \omega(t - m) & t \geq m \\ 0 & \text{jinak} \end{cases}$

– změna v trendu

2.2 Procesy se stochastickým trendem

Jednotkové kořeny

Definice 2.1: Necht' je $\{Y_t, t \in \mathbf{Z}\}$ $AR(p)$ proces:

$$Y_t = c + \sum_{j=1}^p \phi_j Y_{t-j} + \epsilon_t,$$

$$\phi(L)Y_t = c + \epsilon_t,$$

kde $\epsilon_t \sim WN(0, \sigma^2)$. Proces Y_t nazveme *procesem s jednotkovým kořenem*, pokud platí:

$$\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j = 0$$

pro $z_1 = 1$ a $|z_i| > 1, i = 2, \dots, p$. Kořen z_1 se nazývá *jednotkovým kořenem* polynomu $\phi(z)$.

Pak také platí:

$$\phi(L) = (1 - L)\phi^*(L),$$

kde $\phi^*(L)$ je $AR(p - 1)$ - polynom.

2.2.1 Jednoduché příklady

1. Nejjednodušším příkladem procesu s jednotkovým kořenem je proces, kde $p = 1$ a $\phi_1 = 1$:

$$Y_t = Y_{t-1} + \epsilon_t.$$

Tento proces se nazývá *náhodná procházka*, anglicky "random walk". Pokud použijeme rekurzi, dostaneme zápis ve tvaru

$$Y_t = Y_0 + \sum_{j=1}^t \epsilon_j.$$

Pokud proces začal v čase $t = 0$ s počáteční hodnotou $Y_0 = 0$, vychází nám

$$Y_t = \sum_{j=1}^t \epsilon_j.$$

Náhodná procházka je nestacionární proces, neboť obsahuje *stochastický trend* $\sum_{j=1}^t \epsilon_j$. Z tohoto zápisu je vidět, že každá *inovace* ϵ_t má trvalý vliv na hodnotu procesu. Tato vlastnost je typická pro procesy se stochastickým trendem.

Základní statistické vlastnosti náhodné procházky jsou následující:

- *střední hodnota*: $E(Y_t) = 0$
- *rozptyl*: $Var(Y_t) = t\sigma^2$
 - tj. závisí na čase
- *autokovarianční funkce*: $\gamma(t, t-h) = (t-h)\sigma^2$
 - také závisí na čase a platí, že

$$\gamma(t, t-h) \rightarrow \infty,$$

když $(t-h) \rightarrow \infty$ a $h/t \rightarrow c < 1$.

Z těchto vlastností je vidět, že Y_t a Y_{t-h} jsou silně korelované, i když je h vysoké (tj. časový rozdíl mezi oběma veličinami je značný).

2. Dalším často používaným příkladem je tzv. *modifikovaná náhodná procházka*. Je založená na tom, že k původní náhodné procházce přičteme konstantu:

$$Y_t = \mu + Y_{t-1} + \epsilon_t.$$

Stále platí, že $\epsilon_t \sim WN(0, \sigma^2)$. Pokud i nadále předpokládáme, že proces začal v čase $t = 0$ s počáteční hodnotou $Y_0 = 0$, lze ho vyjádřit jako

$$Y_t = \mu t + \sum_{j=1}^t \epsilon_j.$$

Kromě stochastického trendu tedy tento proces obsahuje i deterministický trend μt .

Základní statistické vlastnosti modifikované náhodné procházky jsou následující:

- *střední hodnota*: $E(Y_t) = \mu t$
 - oproti minulému příkladu tedy nepodmíněná střední hodnota závisí na čase
- *rozptyl a autokovarianční funkce*: zůstávají stejné

I když modifikovaná náhodná procházka obsahuje deterministický trend, nemůžeme s ní zacházet jako s "procesem s deterministickým trendem", protože odečtením tohoto trendu nedostaneme stacionární proces. Proto je potřeba pro tento druh procesů vyvinout jinou koncepci jejich stacionarizace.

2.2.2 Integrované procesy

Definice 2.2: Proces s jedním jednotkovým kořenem v AR -polynomu se nazývá *integrováný proces řádu jedna* a značí se

$$Y_t \sim I(1).$$

Proces bez jednotkového kořene v AR -polynomu (tj. stacionární proces) se značí

$$Y_t \sim I(0).$$

Náhodná procházka i modifikovaná náhodná procházka jsou tedy integrované procesy řádu jedna. Takové procesy lze stacionarizovat jejich první diferencí. Platí:

$$Y_t \sim I(1) \quad \Rightarrow \quad \Delta Y_t \sim I(0),$$

kde $\Delta Y_t = Y_t - Y_{t-1}$. Definujme obecně k -tou diferencí procesu jako

$$\Delta^k Y_t = (1 - L)^k Y_t.$$

Platí, že $\Delta^k Y_t = \Delta^{k-1}(\Delta Y_t)$.

Je zřejmé, že $\Delta Y_t = Y_t - Y_{t-1}$ a $\Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - L)^2 Y_t$.

Definice 2.3: Proces $\{Y_t, t \in \mathbf{Z}\}$ se nazývá *integrováným procesem d -tého řádu*, pokud

$$\Delta^d Y_t \sim I(0) \quad a \quad \Delta^{d-1} Y_t \sim I(1),$$

tj. AR -polynom má d jednotkových kořenů a po d -té diferencí se z něj stane stacionární proces.

Platí, že pokud je $Y_t \sim I(0)$, pak je i $\Delta Y_t \sim I(0)$.

Časové řady generované procesy $I(d)$ se označují jako řady typu $I(d)$.

2.2.3 $ARIMA(p, d, q)$ proces

Zkratka $ARIMA$ značí *autoregresní integrováný proces klouzavých součtů*.

Definice 2.4: Nechtě $p, d, q \in \mathbf{N}$. Proces $\{Y_t, t \in \mathbf{Z}\}$ nazveme $ARIMA(p, d, q)$ procesem, pokud

$$X_t = \Delta^d Y_t = (1 - L)^d Y_t$$

je stacionární $ARMA(p, q)$ proces, kde d nazýváme *integračním řádem* procesu.

Pokud je Y_t $ARIMA(p, d, q)$ proces, pak platí, že

$$Y_t \sim I(d)$$

a Y_t splňuje:

$$\phi(L)Y_t = \phi^*(L)(1-L)^d Y_t = \phi^*(L)X_t = c + \theta(L)\epsilon_t,$$

kde

- $\epsilon_t \sim WN(0, \sigma^2)$,
- $\phi^*(z)$ je polynom řádu p ,
- $\theta(z)$ je polynom řádu q ,
- $\phi^*(z) \neq 0$ pro $\forall z : |z| \leq 1$,
- $\phi(z)$ má kořen 1 řádu d .

Proces $\{Y_t\}$ je stacionární právě tehdy, když $d = 0$. Platí:

$$ARIMA(p, 0, q) = ARMA(p, q).$$

Poznamenejme, že pokud $d \geq 1$, můžeme k $ARIMA(p, d, q)$ procesu přidat polynomiální trend až do řádu $(d-1)$ bez ohrožení stacionarity $ARMA$ procesu po d -té diferenci.

Je zřejmé, že náhodná procházka je $ARIMA(0, 1, 0)$ proces.

Nyní rozšířme tuto koncepci integrovaných procesů o sezónní složku.

2.2.4 $SARIMA(p, d, q) \times (P, D, Q)_s$ proces

Nejprve si definujeme sezónní $ARMA$ proces.

Definice 2.5: Nechtě $p, q, P, Q, s \in \mathbf{N}$. Nechtě s označuje délku jedné periody (sezóny). Proces $\{Y_t, t \in \mathbf{Z}\}$ nazveme $SARMA(p, q) \times (P, Q)_s$ procesem, pokud

$$\phi(L)\Phi(L^s)Y_t = c + \theta(L)\Theta(L^s)\epsilon_t,$$

kde $\epsilon_t \sim WN(0, \sigma^2)$ a

$$\begin{aligned}\theta(z) &= \sum_{j=0}^q \theta_j z^j, \\ \phi(z) &= 1 - \sum_{j=1}^p \phi_j z^j, \\ \Theta(z) &= \sum_{j=0}^Q \Theta_j z^j,\end{aligned}$$

$$\Phi(z) = 1 - \sum_{j=1}^P \Phi_j z^j,$$

s $\theta_0 = 1$ a $\Theta_0 = 1$.

Měli bychom také pro jednoznačnost zápisu předpokládat, že $\phi(z)$, $\Phi(z)$, $\theta(z)$ a $\Theta(z)$ nemají společné kořeny. Ty by se totiž pak mohly "vykrátit" a tím by se snížily hodnoty p , P , q a Q .

Aby byl $SARMA(p, q) \times (P, Q)_s$ proces stacionární, musí platit, že všechny kořeny polynomů $\phi(z)$ a $\Phi(z)$ leží vně jednotkového kruhu.

Dá se snadno dokázat, že každý $SARMA(p, q) \times (P, Q)_s$ proces může být zapsaný jako $ARMA(p + Ps, q + Qs)$ proces s určitými omezujícími podmínkami na parametry.

Nyní proces $SARMA$ zevšeobecníme a dovolíme polynomům $\phi(z)$ a $\Phi(z)$ mít jednotkové kořeny.

Definice 2.6: Necht' $p, d, q, P, D, Q, s \in \mathbf{N}$. Necht' s označuje délku jedné periody (sezóny). Proces $\{Y_t, t \in \mathbf{Z}\}$ nazveme $SARIMA(p, d, q) \times (P, D, Q)_s$ procesem, pokud

$$X_t = \Delta^d \Delta_s^D Y_t = (1 - L)^d (1 - L^s)^D Y_t$$

je stacionární $SARMA(p, q) \times (P, Q)_s$ proces. Platí:

$$\phi^*(L)\Phi^*(L^s)X_t = c + \theta(L)\Theta(L^s)\epsilon_t,$$

kde $\epsilon_t \sim WN(0, \sigma^2)$ a

$$\theta(z) = \sum_{j=0}^q \theta_j z^j,$$

$$\phi^*(z) = 1 - \sum_{j=1}^p \phi_j^* z^j,$$

$$\Theta(z) = \sum_{j=0}^Q \Theta_j z^j,$$

$$\Phi^*(z) = 1 - \sum_{j=1}^P \Phi_j^* z^j,$$

s $\theta_0 = 1$ a $\Theta_0 = 1$.

Kapitola 3

Zdánlivá regrese

V této kapitole bude naší snahou vysvětlit problémy, které mohou nastat při modelování vztahů mezi více časovými řadami. Jelikož se v praxi téměř nesetkáme s $I(0)$ řadami, zaměříme se na častější skupinu $I(1)$ řad.

V první části se budeme věnovat některým dalším vlastnostem $I(0)$ a $I(1)$ procesů, a to právě v souvislosti s modelováním jejich vztahů. Dále si vysvětlíme pojem ekvilibria, které chceme modelováním vystihnout. Hlavní částí této kapitoly však bude *zdánlivá regrese*, jak ji poprvé vystihli Granger a Newbold (1974).

3.1 Vlastnosti $I(0)$ a $I(1)$ procesů

Při modelování vztahů více časových řad musíme brát ohled na řády jejich integrace, protože platí následující jednoduchá pravidla. Nechť $a, b \in \mathbf{R}$:

1. jestliže $\{y_t\} \sim I(0)$, potom $\{a + by_t\} \sim I(0)$,
2. jestliže $\{y_t\} \sim I(1)$, potom $\{a + by_t\} \sim I(1)$,
3. jestliže $\{y_t\} \sim I(0)$ a $\{z_t\} \sim I(0)$, potom $\{ay_t + bz_t\} \sim I(0)$,
4. jestliže $\{y_t\} \sim I(1)$ a $\{z_t\} \sim I(0)$, potom $\{ay_t + bz_t\} \sim I(1)$,
5. obecně platí, že jestliže $\{y_t\} \sim I(1)$ a $\{z_t\} \sim I(1)$, potom $\{ay_t + bz_t\} \sim I(1)$. Pokud existuje takové $a, b \in \mathbf{R}$, že $\{y_t\} \sim I(1)$, $\{z_t\} \sim I(1)$ a současně $\{ay_t + bz_t\} \sim I(0)$, pak jsou řady $\{y_t\}$ a $\{z_t\}$ "kointegrované".

O kointegraci časových řad se blíže zmíníme ve 4. kapitole této práce.

3.2 Ekvilibrium

Při modelování vícerozměrných ekonomických časových řad je potřeba rozlišovat mezi *krátkodobými* ("short-run") a *dlouhodobými vztahy* ("long-run relationships"). Krátkodobé vztahy většinou vyvstanou po nějakém šoku, kterému je ekonomika vystavena, a zpravidla po určité době odeznívají. Oproti tomu dlouhodobé vztahy s časem nemizí, vyjadřují jistou "stabilitu" ekonomického systému.

Toto velmi úzce souvisí s pojmem *ekvilibría*. Můžeme si ho představit jako určitý stav, ke kterému je systém stále přitahován. Nadále budeme předpokládat, že se jedná o *stabilní ekvilibríum*, tj. nemění se v čase. Obecně ho lze vyjádřit jako funkci n proměnných

$$f(x_1, \dots, x_n) = 0.$$

Jelikož je však každý ekonomický systém stále vystaven šokům, nemůže nikdy v ekvilibríu permanentně setrvat. Proto budeme uvažovat *dlouhodobé ekvilibríum*, tj. ekvilibríum, ke kterému systém konverguje v čase.

3.3 Durbin-Watsonův test

V dalším textu bude potřeba znát Durbin-Watsonův test autokorelovanosti reziduí. Uvažujme lineární model

$$Y_t = \mathbf{X}'_t \beta + \epsilon_t, \quad (3.1)$$

kde pro chybu ϵ platí:

$$\epsilon_t = \rho \epsilon_{t-1} + v_t. \quad (3.2)$$

Chyby v_t jsou nekorelované, platí $\{v_t\} \sim WN(0, \sigma_v^2)$.

Je zřejmé, že chybový vektor $\{\epsilon_t\}$ je uvažován jako řada typu $AR(1)$. Aby byl stacionární, musí pro koeficient autokorelace ρ platit, že $|\rho| < 1$.

Durbin-Watsonův test testuje hypotézu autokorelace reziduí prvního řádu, kde

$$H_0 : \rho = 0,$$

a jako testovou statistiku uvažujeme statistiku DW :

$$DW = \frac{\sum_{t=2}^T (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\epsilon}_t^2}, \quad (3.3)$$

kde $\hat{\epsilon}_t$ značí reziduum modelu (3.1). Statistiku DW lze také pomocí vektorů zapsat ve tvaru

$$DW = \frac{\hat{\epsilon}' A \hat{\epsilon}}{\hat{\epsilon}' \hat{\epsilon}},$$

kde $\hat{\epsilon}$ značí vektor všech reziduí $\hat{\epsilon}_t$ a

$$A = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

Pro Durbin-Watsonovu statistiku platí:

- $0 < DW < 4$,
- $DW \approx 2 \Rightarrow$ nekorelovanost reziduí,
- $DW > 2 \Rightarrow$ negativní korelovanost reziduí ($\rho < 0$),
- $DW < 2 \Rightarrow$ pozitivní korelovanost reziduí ($\rho > 0$).

DW statistiku můžeme po několika matematických úpravách zapsat pomocí odhadu metodou nejmenších čtverců $\hat{\rho}$ z rovnice (3.2) jako

$$DW = 2 - \gamma_1 - 2\gamma_2\hat{\rho},$$

kde

$$\begin{aligned} \gamma_1 &= \frac{\hat{\epsilon}_1^2 + \hat{\epsilon}_T^2}{\sum_{t=1}^T \hat{\epsilon}_t^2}, \\ \gamma_2 &= \frac{\sum_{t=2}^T \hat{\epsilon}_{t-1}^2}{\sum_{t=1}^T \hat{\epsilon}_t^2}, \\ \hat{\rho} &= \frac{\sum_{t=2}^T \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=2}^T \hat{\epsilon}_{t-1}^2}. \end{aligned}$$

Pro vysokou hodnotu T platí: $\gamma_1 \approx 0$,
 $\gamma_2 \approx 1$,

a tudíž $DW \approx 2 - 2\hat{\rho}$,
 $\hat{\rho} \approx 1 - \frac{1}{2}DW$.

Jelikož ale DW stále závisí na matici vysvětlujících proměnných \mathbf{X} , Durbin a Watson získali pomocí metody Monte Carlo pravděpodobnostní rozdělení jiných dvou statistik DW_L a DW_U , které nejsou závislé na \mathbf{X} a které ohraničují DW :

$$DW_L < DW < DW_U.$$

Pro tyto statistiky jsou kritické hodnoty DW_L^* a DW_U^* tabelovány.

3.4 Zdánlivá regrese: Granger & Newbold

Při modelování a odhadování rovnovážného stavu (ekvilibria) se v praxi často uplatňuje klasická regresní analýza. Základním předpokladem pro asymptotickou teorii odhadů metodou nejmenších čtverců je stacionarita a vzájemná nezávislost vysvětlující i vysvětlované proměnné. Někdy však statistici možnost závislosti či nestacionarity neprověří a z analýzy pak získávají nesprávné výsledky.

Na tento již dříve známý problém upozornili svým článkem Granger a Newbold (1974), kteří ukázali nejjednodušší příklad zdánlivé regrese u odhadování závislosti mezi dvěma nezávislými náhodnými procházkami. Tento článek se stal velice slavným a často citovaným.

Uvažujme dvě časové řady:

$$y_t = y_{t-1} + \epsilon_{1t}, \quad (3.4)$$

$$x_t = x_{t-1} + \epsilon_{2t}, \quad (3.5)$$

kde $\{\epsilon_{1t}\} \sim IWN(0, \sigma_1^2)$ a $\{\epsilon_{2t}\} \sim IWN(0, \sigma_2^2)$, obě řady $\{\epsilon_{1t}\}$ a $\{\epsilon_{2t}\}$ jsou "nezávisle stejně rozdělené" (*i.i.d.* = identically independent distributed). Z definice bílého šumu je zřejmé, že jsou i obě řady (3.4) a (3.5) navzájem nezávislé.

Granger a Newbold se pokusili vyjádřit vztah mezi náhodnými procházkami jako regresi typu

$$y_t = \beta_0 + \beta_1 x_t + u_t. \quad (3.6)$$

U tohoto typu regresního modelu se parametry běžně odhadují metodou nejmenších čtverců, jak to učinili i výše zmínění autoři. Pro použití této metody je základním předpokladem nezávislost a nekorelovanost procesu $\{u_t\}$, tj. $\{u_t\}$ musí být *i.i.d.*, navíc nezávislého na procesu $\{x_t\}$. Při testování signifikantnosti takto odhadnutých parametrů pomocí *t*-testu často vychází, že odhady jsou signifikantní na velmi vysoké hladině spolehlivosti, tj. často je zamítna hypotéza o nulové hodnotě parametrů β_0 a především β_1 . Také vysoká hodnota koeficientu determinace (statistika R^2) ukazuje, že model je odhadnutými parametry dobře vystižen.

Jediný problém tedy nastal u *Durbin-Watsonovy* statistiky, která testuje autokorelovanost reziduí odhadnutého modelu. Zde, zatím bez ohledu na známost nezávislosti vysvětlující a vysvětlované proměnné, by bylo logické předpokládat, že pokud je vztah mezi y_t a x_t odhadnutým modelem dobře vystižen, neměla by se autokorelovanost reziduí prokázat, což ale neodpovídá výsledkům testů. Hodnota *DW* statistiky totiž autorům vyšla jako velice nízká (blízká 0), což poukazuje na velký problém s předpokladem nezávislosti reziduí.

Jelikož je však od začátku znám fakt nezávislosti vysvětlované a vysvětlující proměnné, je namísto předpokládat, že celý regresní model (3.6) nemůže být smysluplný a že pomocí klasického *t*-testu bude potvrzena hypotéza nulovosti

koeficientů β_0 a β_1 . Toto je ale v rozporu s dosaženými výsledky. Z tvrzení v kapitole 3.1 také víme, že když jsou obě řady $\{y_t\}$ a $\{x_t\}$ náhodné procházky, tedy $I(1)$ řady, měla by být jejich lineární kombinace $\{u_t\}$ ve většině případů také $I(1)$. Z DW statistiky plyne, že je nesplněn základní požadavek na použití metody nejmenších čtverců, tedy stacionarita reziduí modelu, což při odhadování parametru β_1 vede ke *zdánlivé regresi*.

Granger a Newbold ve svém článku navrhli, aby nerovnost

$$R^2 > DW \quad (3.7)$$

byla považována za znamení zdánlivé regrese modelu, neboť tento vztah může znamenat, že rezidua mají nestacionární charakter.

3.4.1 Příklad

V této části práce se pokusíme po vzoru Grangera a Newbolda nasimulovat dvě nezávislé náhodné procházky a ukážeme, jak se zdánlivá regrese projevuje v praxi. Pro simulaci použijeme statistický program R, výsledky uvedené v textu najdeme v tabulce 3.1.

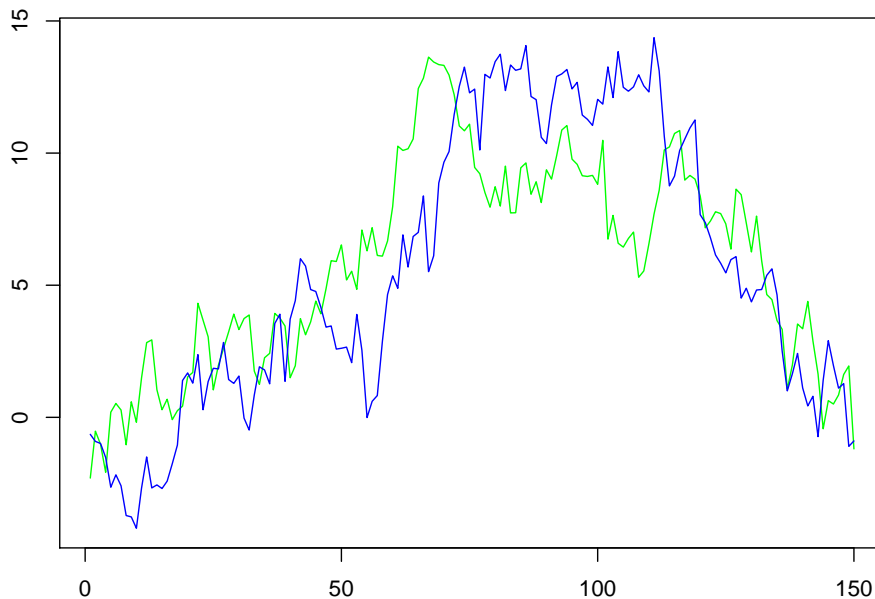
Na základě procesů (3.4) a (3.5) generujeme časové řady y_t a x_t , chybové vektory $\{\epsilon_{1t}\}$ a $\{\epsilon_{2t}\}$ bereme z normovaného normálního rozdělení, tedy $\sigma_1^2 = \sigma_2^2 = 1$ a $\epsilon_{1t} \sim N(0, 1)$, $\epsilon_{2t} \sim N(0, 1)$. Časové řady y_t a x_t mají délku 150 hodnot a jsou zachycené na obrázku 3.1.

Protože jsou časové řady generovány podle procesů (3.4) a (3.5), jsou navzájem nezávislé a nestacionární. Již z pohledu na obrázek je zřejmé, že obě vykazují trend a že v jejich průběhu můžeme vysledovat jisté podobnosti. Pomocí metody nejmenších čtverců odhadneme model jejich závislosti jako

$$y_t = 2,56909 + 0,57520x_t. \quad (3.8)$$

Na základě t -testů jednotlivých parametrů zjistíme, že oba parametry jsou statisticky významné (dokonce na 1% hladině významnosti). Pomocí F -testu také zjistíme, že celkový model je vhodný k modelování závislosti mezi našimi dvěma řadami (opět na 1% hladině významnosti). Koeficient determinace R^2 je 0,6159. Durbin-Watsonův test vykazuje silnou míru autokorelace reziduí modelu: $DW = 0,3032$.

Protože platí, že koeficient determinace je vyšší než DW statistika ($R^2 > DW$), můžeme konstatovat, že mezi řadami je vztah zvaný zdánlivá regrese. To je vidět i z obrázku 3.2, na kterém jsou vyznačena rezidua modelu (3.8). Je zřejmé, že tato časová řada není stacionární.



Obrázek 3.1: Vygenerované náhodné procházky

3.5 Zdánlivá regrese obecně

Uvažujme nyní regresi ve tvaru

$$y_t = \mathbf{x}_t' \beta + u_t, \quad (3.9)$$

kde veličiny y_t a \mathbf{x}_t mohou být nestacionární. Pokud neexistuje žádná hodnota parametru β , pro kterou by rezidua $u_t = \mathbf{x}_t' \beta - y_t$ byla $I(0)$, je velice pravděpodobné, že metodou nejmenších čtverců budou vycházet "podezřelé" výsledky.

Tyto výsledky analyticky vysvětlil ve svém článku Phillips (1986). Odvodil asymptotickou teorii regrese, která je aplikovatelná na obecně integrované procesy, což objasňuje i poznatky Grangera a Newbolda. Nyní shrneme základní fakta, která z jeho studie vyplývají.

Předpokládejme tedy, že máme regresi typu (3.6), kde ale na chybové vektory ϵ_{1t} a ϵ_{2t} ze (3.4) a (3.5) klademe daleko slabší požadavky než v kapitole 3.4. Tyto požadavky (matematicky vyjádřené v článku Phillips (1986) jako Předpoklad 1) dovolují chybovým vektorům být obecně integrovanými procesy (řádu jedna), jejichž diference mohou být závislé a heterogenně rozdělené. Na základě těchto předpokladů Phillips odvodil následující:

1. Obvyklé t -statistiky odhadů parametrů β_0 a β_1 (t_{β_0} a t_{β_1}), které jsou používány k posouzení signifikantnosti koeficientů v regresní analýze, nemají

```

Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7266 -1.6080 -0.1509  1.3921  7.8913

Coefficients:
            Estimate Std. Error t value Pr(> |t|)
(Intercept)  2.56909    0.28615    8.978  1.16e-15 ***
x            0.57520    0.03734   15.404 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.414 on 148 degrees of freedom
Multiple R-Squared:  0.6159,    Adjusted R-squared:  0.6133
F-statistic: 237.3 on 1 and 148 DF, p-value: < 2.2e-16

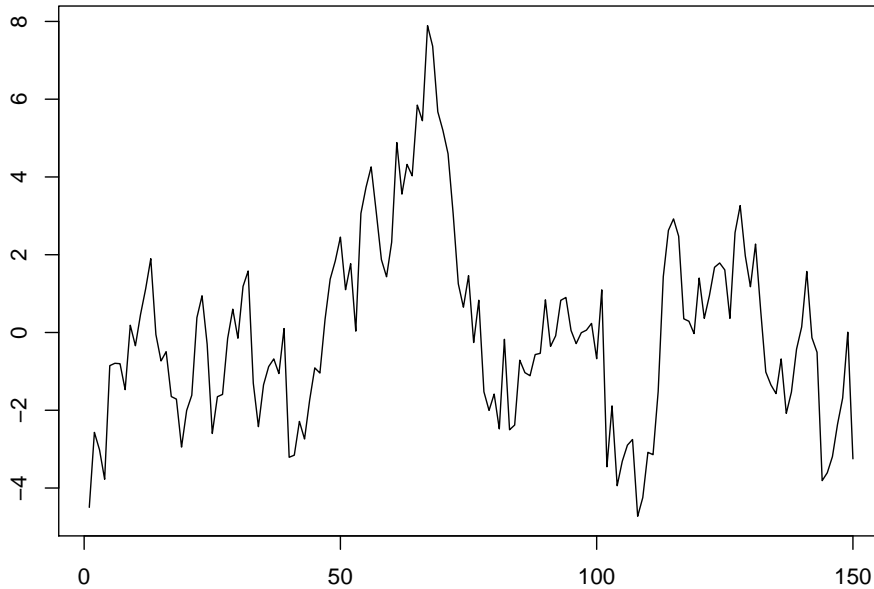
```

Tabulka 3.1: Model pro závislost dvou náhodných procházek - výstup z R

v tomto případě limitní rozdělení. Navíc rozdělení t_{β_0} a t_{β_1} divergují s $T \rightarrow \infty$, takže neexistují asymptoticky správné kritické hodnoty pro tyto obvyklé testy signifikance. Je zřejmé, že pro jakoukoliv kritickou hodnotu poměr zamítnutí nulové hypotézy $\beta_1 = 0$ roste s rostoucím počtem pozorování.

2. Na rozdíl od běžných výsledků regresní teorie koeficienty $\hat{\beta}_0$ a $\hat{\beta}_1$ nekonvergují v pravděpodobnosti s rostoucím T ke konstantám. Konkrétně $\hat{\beta}_1$ má nedegenerované limitní rozdělení při $T \rightarrow \infty$ a rozdělení $\hat{\beta}_0$ diverguje při $T \rightarrow \infty$. Tento rozdíl oproti běžné regresní teorii vyvstane ještě zřetelněji, pokud vezmeme za x_t a y_t dva nezávislé stacionární autoregresní procesy. Pak oba parametry $\hat{\beta}_0$ i $\hat{\beta}_1$ konvergují dle očekávání v pravděpodobnosti k nule.
3. Durbin-Watsonova statistika konverguje v pravděpodobnosti k nule, zatímco pokud spolu řady x_t a y_t skutečně souvisí, konverguje k nenulové hodnotě. Tím je vysvětleno chování DW , které odhalili Granger a Newbold, a je teorií ospravedlněné, že tuto statistiku můžeme použít k rozpoznání zdánlivé regrese.
4. Koeficient determinace R^2 má nedegenerované limitní rozdělení při $T \rightarrow \infty$.

Všechny tyto výsledky se odlišují od konvenční regresní teorie stacionárních procesů. Důvodem k těmto charakteristickým vlastnostem je fakt, že procesy x_t a y_t nejsou ergodické. Ve skutečnosti jejich výběrové momenty a sdružené výběrové momenty nekonvergují ke konstantám, jak tomu je u ergodických procesů. Základní důkaz konzistence odhadů získaných metodou nejmenších čtverců vychází z předpokladu, že $\text{plim}(1/T)(X'X) = Q$, kde X je matice dat a Q je pevná matice. Tento předpoklad však v našem případě splněn není. Navíc víme, že problém zdánlivé regrese není možné vyřešit větším počtem pozorování. S rostoucím T se vše naopak zhoršuje.



Obrázek 3.2: Rezidua modelu (3.8)

Když budeme chtít výše uvedené výsledky rozšířit na vícerozměrný model, tj. budeme jako základní uvažovat model typu (3.9), dojdeme ke stejným výsledkům jako při hledání souvislosti mezi dvěma nezávislými náhodnými procházkami. Je potřeba ještě dodat, že i F -test pro testování signifikance modelu (3.9) diverguje s rostoucím T . Navíc pokud jsou y_t a \mathbf{x}_t obecně korelované časové řady, došel Phillips ke stejným výsledkům, což znamená, že se tedy nemusíme v této teorii zabývat pouze nezávislými časovými řadami.

3.6 Léky zdánlivé regrese

Z výše uvedeného je zřejmé, že zdrojem existence zdánlivé regrese je přítomnost stochastických trendů v generujících procesech časových řad obsažených v regresním modelu. Existují dva jednoduché způsoby, kterými můžeme problémům spojeným se zdánlivou regresí zabránit.

Prvním přístupem je přidat do modelu časově zpožděné hodnoty závislé i nezávislé proměnné. Uvažujme například následující model jako alternativu k modelu (3.6):

$$y_t = \beta_0 + \alpha y_{t-1} + \beta_1 x_t + \gamma x_{t-1} + u_t. \quad (3.10)$$

Lze dokázat, že metoda nejmenších čtverců poskytuje konzistentní odhady všech parametrů tohoto modelu (3.10). Každý z koeficientů $\hat{\beta}_1$ a $\hat{\gamma}$ konverguje ke gaussov-

skému rozdělení a t -testy hypotéz, že $\beta_1 = 0$ a $\gamma = 0$, jsou asymptoticky normální. Přesto F -test sdružené nulové hypotézy, že oba koeficienty β_1 i γ jsou nulové, má nestandardní limitní rozdělení. Proto zahrnutí zpožděných proměnných do modelu vyřeší mnoho problémů, avšak ne všechny.

Druhou metodou je všechna data před odhadováním vztahu diferencovat. Pak získáme místo modelu (3.6) model následující:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + u_t. \quad (3.11)$$

Protože je zřejmé, že v tomto případě jsou vysvětlující i vysvětlovaná proměnná stacionární, tedy $I(0)$, je i chybový vektor u_t stacionární časovou řadou. Potom se jedná o běžný typ regrese, odhady $\hat{\beta}_0$ a $\hat{\beta}_1$ jsou konzistentní a konvergují ke gaussovským proměnným. Všechny t a F testy založené na modelu (3.11) mají obvyklá limitní rozdělení.

Jelikož se model (3.11) vyhýbá problémům se zdánlivou regresí stejně tak jako nestandardním rozdělením určitých hypotéz spojených s úrovnovým modelem (3.6), mnoho analytiků doporučuje rutinně diferencovat zjevně nestacionární řady ještě před odhadováním vzájemné závislosti. Přestože se tento způsob zdá být ideálním, existují dvě rozdílné situace, ve kterých by rutinní diferencování bylo nevhodné.

Nejprve uvažujme situaci, kdy hodnota α v modelu (3.10) je rovna spíše například 0,9 než 1, tj. data jsou skutečně stacionární. Potom by diferenciací jako v modelu (3.11) vedla ke špatně specifikované regresii. Ve druhé situaci uvažujeme takovou regresii, kde i když obě řady y_t a x_t jsou skutečně $I(1)$ procesy, chybový vektor u_t je stacionární $I(0)$ řada. Této zajímavé třídě řad se říká "kointegrované řady" a budeme se jim částečně věnovat v následující kapitole.

Kapitola 4

Rozpoznání zdánlivé regrese

Jak už bylo řečeno v minulé kapitole, zdrojem zdánlivé regrese je přítomnost stochastických trendů v generujících procesech časových řad obsažených v regresním modelu. To znamená, že tento problém nastává, pokud jsou procesy integrované minimálně řádem jedna. Avšak i u těchto procesů dochází k situaci, kdy regrese mezi nimi nemusí být nutně zdánlivá, a to tehdy, pokud chybový vektor u_t , tedy lineární kombinace nestacionárních procesů, je sám stacionární. V takovém případě nazýváme procesy kointegrovanými.

Můžeme tedy prohlásit, že ke zdánlivé regresi nedochází, pokud jsou integrované procesy v regresním modelu kointegrované. Z tohoto důvodu si v této kapitole povíme něco o kointegrovaných procesech a o testování kointegrace, která nám zajišťuje nepřítomnost zdánlivé regrese.

4.1 Kointegrace

4.1.1 Příklad

Uveďme nejprve příklad kointegrovaných časových řad převzatý z knihy Hamilton (1994). Uvažujme jednoduchý systém dvou časových řad

$$y_t = \gamma x_t + u_{1t} \quad (4.1)$$

$$x_t = x_{t-1} + u_{2t}, \quad (4.2)$$

kde u_{1t} a u_{2t} jsou nekorelované procesy bílého šumu. Z reprezentace x_t vidíme, že se jedná o náhodnou procházku

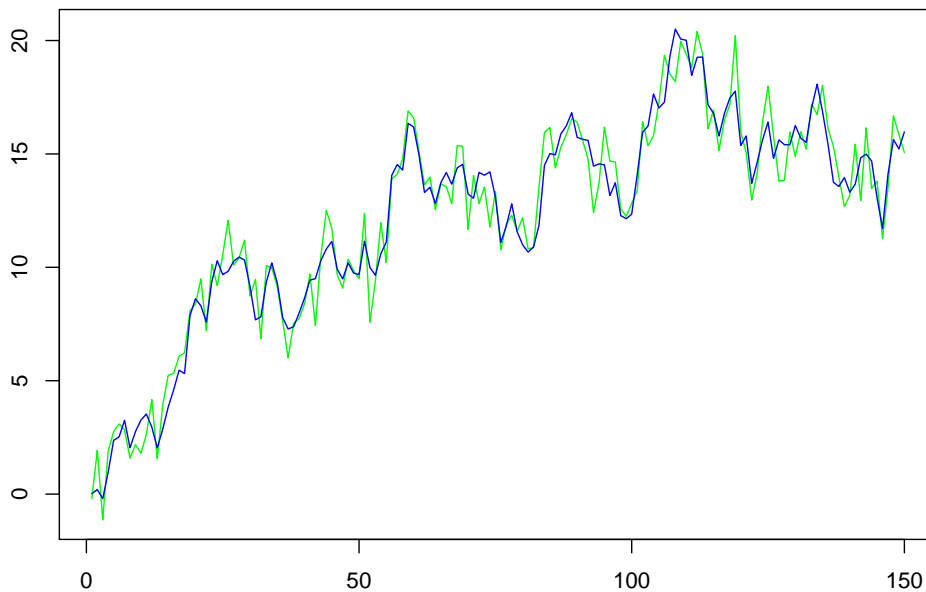
$$\Delta x_t = u_{2t}, \quad (4.3)$$

a z rovnice (4.1) diferencováním získáme

$$\Delta y_t = \gamma \Delta x_t + \Delta u_{1t} = \gamma u_{2t} + u_{1t} - u_{1,t-1}. \quad (4.4)$$

Vidíme, že na pravé straně rovnice (4.4) se nachází lineární kombinace $I(0)$ procesů, a jak již bylo uvedeno v kapitole 3.1, taková kombinace je také sama $I(0)$ procesem. Neboť x_t je $I(1)$ proces, platí, že y_t je lineární kombinací $I(1)$ a $I(0)$ procesů; navíc z rovnice (4.4) víme, že se po první diferenci stává sám $I(0)$ procesem. Můžeme tedy konstatovat, že y_t je také $I(1)$ proces. Z rovnice (4.1) je vidět, že lineární kombinace obou procesů $(y_t - \gamma x_t)$ je definovaná jako stacionární časová řada, tedy řada typu $I(0)$. O tomto případě jsme již dříve hovořili jako o příkladu kointegrace. Řekneme tedy, že řady y_t a x_t jsou *kointegrované*, a vektor $(1, -\gamma)'$ nazveme *kointegračním vektorem*.

Na obrázku 4.1 jsou zakresleny vygenerované časové řady y_t a x_t podle modelu (4.1) a (4.2), kde $\gamma = 1$ a u_{1t} a u_{2t} jsou dvě nezávislé $N(0, 1)$ proměnné. Poznamenejme, že je z obrázku vidět, že se obě řady y_t a x_t mohou libovolně daleko vzdálit od počáteční hodnoty, avšak od sebe navzájem se vzdálí vždy maximálně jen do pevně dané vzdálenosti. Tomuto vztahu pak říkáme také "dlouhodobé ekvilibrium" (viz kapitola 3.2).



Obrázek 4.1: Model kointegrace

4.1.2 Definice

Uveďme nyní obecnou definici, jak tento vztah kointegrace ve své práci uveřejnili Engle a Granger (1987). Pro dva procesy ji lze vyjádřit následujícím způsobem:

Definice 4.1: Procesy $\{x_t\}$ a $\{y_t\}$ se nazývají *kointegrované řádu d , b* a označují se $\{x_t\}, \{y_t\} \sim CI(d, b)$, pokud:

- a) oba jsou typu $I(d)$,
- b) existuje lineární kombinace $\{ax_t + cy_t\} \sim I(d - b)$, kde $b > 0$.

Vektor $(a, c)'$ se nazývá *kointegrační vektor*.

Definici lze zobecnit na množinu N proměnných:

Definice 4.2: Komponenty vektoru \mathbf{y}_t se nazývají *kointegrované řádu d , b* ($\mathbf{y}_t \sim CI(d, b)$), pokud:

- a) všechny komponenty jsou typu $I(d)$,
- b) existuje vektor α ($\alpha \neq 0$) takový, že $z_t = \alpha' \mathbf{y}_t \sim I(d - b)$, $b > 0$.

Vektor α se nazývá *kointegrační vektor*.

V empirické ekonometrii je nejzajímavější případ, kdy $d = b$, tj. kdy kointegrační vektor vede ke stacionární lineární kombinaci. Vztah

$$\alpha' \mathbf{y}_t = 0 \quad (4.5)$$

popisuje dlouhodobé ekvilibrium a výraz

$$\alpha' \mathbf{y}_t = z_t \quad (4.6)$$

představuje odchylku od dlouhodobého ekvilibria, a proto se nazývá *chyba ekvilibria*. Všeobecně mezi N procesy může existovat maximálně $N - 1$ kointegračních vektorů. V případě dvou procesů může existovat pouze jeden kointegrační vektor, tedy jen jedna lineární kombinace, která je stacionární. V případě dvou $I(1)$ procesů lze dlouhodobé ekvilibrium zapsat ve tvaru

$$y_t = \beta_0 + \beta_1 x_t \quad (4.7)$$

a kointegrační vektor se tudíž rovná $(1, -\beta_1)$. Je vidět, že vyjádření kointegračního vektoru není jednoznačné, protože po vynásobení obou stran rovnice (4.6) stejným nenulovým číslem nám rovnost zůstane zachována.

4.2 Testy jednotkového kořene a stacionarity

Před tím, než se začneme věnovat samotným testům kointegrace, je třeba zjistit, jakého typu jsou analyzované časové řady. Jelikož se zde zabýváme integrovanými řadami, potřebujeme vědět, jakého řádu jejich integrace je, tj. kolik má řada jednotkových kořenů.

Prvním krokem v analýze časové řady bývá prozkoumání jejího grafu a subjektivní posouzení jeho tvaru. Většinou tímto postupem můžeme zjistit, zda je daná časová řada stacionární, či zda ji k její stacionarizaci potřebujeme jednou nebo vícekrát diferencovat. Dalším krokem pak je posouzení tvaru autokorelační funkce analyzované řady. Je-li první hodnota této funkce blízká jedné a ostatní hodnoty se zmenšují jen velmi pomalu, lze očekávat, že daná řada nebude stacionární. Toto je ovšem také jen subjektivní metoda.

I když výše zmíněné subjektivní metody mohou často pomoci ke konkrétní představě o tvaru analyzované řady, většinou nám poradí pouze v tom, zda je daná řada stacionární či nikoliv. Proto musíme na tomto místě použít přesné statistické metody ke zjištění, zda naše časová řada obsahuje jednotkový kořen (tedy i stochastický trend).

Existují dva druhy statistických testů: jedna skupina je založena na testování nulové hypotézy o existenci jednotkového kořene, zatímco druhá testuje nulovou hypotézu stacionarity generujícího procesu. Podrobněji se budeme zbývat první skupinou testů.

4.2.1 Testy jednotkového kořene

Dickey-Fullerův test

Nejnámější test se jmenuje podle svých autorů *Dickey-Fullerův test* a zkráceně se často označuje jako *DF test*. Tento test se používá pro rozlišení, zda je časová řada typu $I(1)$ nebo $I(0)$. Uvažujme jednoduchý $AR(1)$ proces bez deterministického trendu

$$y_t = \phi y_{t-1} + \epsilon_t, \quad (4.8)$$

kde $y_0 = 0$, $t = 1 \dots T$ a $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. Při testování hypotézy $H_0 : \phi = \phi_0$ můžeme použít klasický t -test pouze v případě, že $|\phi_0| < 0$, tedy pokud je y_t stacionární časová řada. Pak testové kritérium vypadá následovně:

$$t = \frac{\hat{\phi} - \phi_0}{S_{\hat{\phi}}}, \quad (4.9)$$

kde $S_{\hat{\phi}}$ je odhad směrodatné odchylky parametru ϕ definovaný jako

$$S_{\hat{\phi}} = \sqrt{\frac{s^2}{\sum_{t=1}^T y_{t-1}^2}},$$

$$s^2 = \frac{RSS}{n-1}$$

a $\hat{\phi}$ je odhad parametru pomocí metody nejmenších čtverců. Za platnosti nulové hypotézy má t statistika v malých výběrech přibližně rozdělení t s $(T-1)$ stupni volnosti.

Ovšem v případě testování jednotkového kořene $H_0 : \phi = 1$ výše uvedené výsledky neplatí.

Model (4.8) lze odečtením y_{t-1} od obou stran upravit následně:

$$\begin{aligned} y_t - y_{t-1} &= (\phi - 1)y_{t-1} + \epsilon_t \\ \Delta y_t &= \rho y_{t-1} + \epsilon_t, \end{aligned} \quad (4.10)$$

kde $\rho = \phi - 1$. Potom testování toho, že $\phi = 1$, je totéž jako testování $\rho = 0$. Naopak pokud je řada stacionární a $\phi < 1$, platí také $\rho < 0$. Proto definujeme nulovou a alternativní hypotézu jako

$$H_0 : \rho = 0 \quad vs. \quad H_1 : \rho < 0.$$

Dickey a Fuller pak uvažovali t statistiku pro testování platnosti nulové hypotézy:

$$t_\rho^{DF} = \frac{\hat{\rho}}{S_\rho}, \quad (4.11)$$

kde $\hat{\rho}$ je odhad parametru ρ pomocí metody nejmenších čtverců v modelu (4.10) a S_ρ je odhad směrodatné odchylky parametru ρ v modelu (4.10).

Statistika t_ρ^{DF} má za platnosti nulové hypotézy takzvané *Dickey-Fullerovo (DF) rozdělení*, jehož kritické hodnoty byly tabelovány právě Dickeyem a Fullerem v roce 1976.

Dosud jsme uvažovali hypotézu o generujícím procesu jako náhodné procházce proti alternativní hypotéze $AR(1)$ procesu s nulovou střední hodnotou. V časových řadách je však zajímavější zabývat se takovými procesy, ve kterých jsou navíc obsaženy další parametry jako konstanta a lineární trend. Dickey a Fuller tedy dále zkoumali procesy tvaru

$$\Delta y_t = \alpha + \rho y_{t-1} + \epsilon_t \quad (4.12)$$

$$\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \epsilon_t. \quad (4.13)$$

Pro oba procesy (4.12) a (4.13) je nulová hypotéza o existenci jednotkového kořene totožná: $H_0 : \rho = 1$. Jelikož se ale rozdělení testové statistiky pro tyto modely oproti modelu (4.10) změní, je nutné zavést rozlišující označení. Testová statistika modelu (4.12) se značí $t_\rho^{DF,\mu}$, pro model (4.13) pak $t_\rho^{DF,t}$. Kritické hodnoty opět

Dickey a Fuller tabelovali a najdeme je společně s kritickými hodnotami statistiky t_ρ^{DF} například v Banerjee a kol. (1993).

Je důležité si uvědomit, že jsme dosud uvažovali jako skutečný generující proces model (4.8), což ovšem v praxi bývá jen málokdy. Na základě modelů (4.10), (4.12), (4.13) jsme pak sestavovali testovou statistiku. Víme ale, že rozdělení testových statistik závisí jak na zkoumaném modelu, tak i na skutečném generujícím modelu. Proto je třeba zkonstruovat testy pro jiné generující procesy, než je (4.8), které by nezávisely na tzv. přebytech ("nuisance") parametrech.

Tyto další testy se nazývají *podobné testy*, protože jsou postaveny na základě modelů (4.10), (4.12), (4.13). V tabulce 4.1 jsou shrnuty výsledky zkoumání pro některé další generující procesy. V praxi to pak znamená, že testové statistiky a jejich kritické hodnoty jsou totožné se zkonstruovanými na základě modelů pro konstrukci podobných testů (t_ρ^{DF} , $t_\rho^{DF,\mu}$, $t_\rho^{DF,t}$).

Generující proces	Modely pro konstrukci podobných testů
(i) $y_t = \phi y_{t-1} + \epsilon_t, y_0 = 0$	(4.10), (4.12), (4.13)
(ii) $y_t = \phi y_{t-1} + \epsilon_t, y_0$ libovolné	(4.12), (4.13)
(iii) $y_t = \alpha + \phi y_{t-1} + \epsilon_t, y_0$ libovolné	(4.13)
(iv) $y_t = \alpha + \beta t + \phi y_{t-1} + \epsilon_t, y_0$ libovolné	nutné rozšíření modelu (4.13)

Tabulka 4.1: Modely pro konstrukci podobných testů ($\alpha \neq 0, \beta \neq 0$)

Obecně platí, že model, který poskytuje podobný test, musí obsahovat více parametrů než skutečný generující proces. Proto je také v případě (iv) nutné do modelu (4.13) navíc zahrnout kvadrát časové proměnné.

Rozšířený Dickey-Fullerův test

V praxi může být proces (4.8) rozšířen nejen o konstantu a lineární trend, ale také jeho reziduální složka může mít bohatší autokorelační strukturu. Test jednotkového kořene pro tento typ modelů vyvinuli současně s DF testem autoři Dickey a Fuller a nazývá se *rozšířený Dickey-Fullerův test*, neboli *ADF test* z anglického "augmented".

Uvažujme $AR(p)$ proces ($p > 1$)

$$y_t = \sum_{j=1}^p \phi_j y_{t-j} + \epsilon_t, \quad (4.14)$$

neboli

$$\phi(L)y_t = \epsilon_t,$$

kde

$$\phi(L) = 1 - \sum_{j=1}^p \phi_j L^j.$$

Pokud platí, že

$$\phi(L) = (1 - L)\phi^*(L),$$

kde všechny kořeny polynomu $\phi^*(L)$ leží vně jednotkového kruhu, je y_t proces typu $I(1)$. Pro testování jednotkového kořene v $AR(p)$ procesu je potřeba nejprve rovnici (4.14) upravit tak, že do regrese zahrneme $p - 1$ posunutí diference Δy_t . Dostaneme tak rovnici

$$\Delta y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \epsilon_t, \quad (4.15)$$

kde

$$\rho = \sum_{j=1}^p \phi_j - 1 = -\phi(1)$$

a

$$\phi_j^* = -(\phi_{j+1} + \phi_{j+2} + \dots + \phi_p)$$

pro $j = 1, \dots, p - 1$. Hypotéza přítomnosti jednotkového kořene je stejná jako v Dickey-Fullerově testu, testujeme tedy opět

$$H_0 : \rho = 0 \quad vs. \quad H_1 : \rho < 0.$$

Dickey a Fuller uvažovali tutéž testovou statistiku pro testování platnosti nulové hypotézy, a to

$$t_\rho^{ADF} = \frac{\hat{\rho}}{S_\rho}, \quad (4.16)$$

kde jsou ale odhady $\hat{\rho}$ a S_ρ založeny na rozšířeném modelu (4.15). Tato statistika má stejné rozdělení jako statistika t_ρ^{DF} , tedy DF rozdělení, proto můžeme použít stejné kritické hodnoty.

Podobně jako u testování $AR(1)$ modelů na přítomnost jednotkového kořene můžeme v případě $AR(p)$ procesů do jejich struktury přidat konstantu nebo lineární trend. Dostaneme pak procesy tvaru

$$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \epsilon_t \quad (4.17)$$

$$\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \epsilon_t \quad (4.18)$$

a t -statistiky pro testování přítomnosti jednotkového kořene založené na modelech (4.17) a (4.18) ($t_\rho^{ADF,\mu}$, $t_\rho^{ADF,t}$) mají popořadě stejné rozdělení jako $t_\rho^{DF,\mu}$ a $t_\rho^{DF,t}$.

Pokud není v modelu (4.15) znám přesný počet autoregresních koeficientů p , je vždy bezpečnější volit vyšší číslo. Zahrneme-li totiž do (4.15) příliš velký počet posunutí, odhady přebytečných parametrů budou ležet blízko nuly, i když za cenu jisté ztráty eficientnosti. Naopak nedostatek posunutí nám zajistí určitou autokorelaci v chybovém vektoru rovnice, a proto není možná aplikace předchozích asymptotických výsledků založených na předpokladu jeho nezávislosti.

Said-Dickeyův test

Said a Dickey v roce 1984 zobecnili Dickey-Fullerovy testy na třídu modelů, jejichž reziduální složky jsou stacionárními a invertibilními $ARMA(p, q)$ procesy. Výhodou těchto testů je, že je můžeme aplikovat i na procesy, u nichž řád MA a AR procesů není znám. Metoda je založená na aproximaci procesu pomocí AR polynomu, neboť víme, že každý invertibilní a stacionární $ARMA$ proces má $AR(\infty)$ reprezentaci.

Nejprve tedy předpokládejme proces

$$y_t = \phi y_{t-1} + \epsilon_t, \quad (4.19)$$

kde reziduální složka má následující tvar

$$\epsilon_t + \sum_{i=1}^p \alpha_i \epsilon_{t-i} = u_t + \sum_{j=1}^q \theta_j u_{t-j}$$

a $\{u_t\} \sim WN(0, \sigma_u^2)$. Předpokládejme dále, že ϵ_t má tvar $ARMA(p, q)$ procesu, který je invertibilní a stacionární. Generující proces pak můžeme přepsat do tvaru

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^k \alpha_i \Delta y_{t-i} + v_t, \quad (4.20)$$

kde k je dostatečně velké, aby byl model (4.19) dobře aproximován modelem (4.20) tak, že $\{v_t\}$ je přibližně bílý šum. I v tomto případě je nulová hypotéza přítomnosti jednotkového kořene stejná: $H_0 : \rho = 0$. Said a Dickey ukázali, že je test platný, pokud k roste s rozsahem výběru tak, že existují čísla c a r ($c > 0$, $r > 0$), pro která $ck > T^{1/r}$ a $T^{-1/3}k \rightarrow 0$. Jelikož odhad parametru ρ metodou nejmenších čtverců je konzistentní, test může být založen na testové statistice

$$t_\rho^{SD} = \frac{\hat{\rho}}{S_{\hat{\rho}}}, \quad (4.21)$$

která má totéž rozdělení jako příslušná testová statistika ADF testu.

Testy mnohonásobného jednotkového kořene

Uvažujme nyní problém přítomnosti $d > 1$ jednotkových kořenů v časové řadě. Nebylo by správné testovat tuto možnost způsobem, že jako první nulovou hypotézu zvolíme přítomnost jednoho jednotkového kořene proti alternativní hypotéze stacionárního procesu, a pokud nezamítneme nulovou hypotézu, testujeme dále novou hypotézu o přítomnosti jednotkového kořene v jednu diferencovanou řadě, atd. Tato procedura neposkytuje statisticky platnou testovací posloupnost.

Dickey a Pantula v roce 1987 ve svém článku navrhli takovou sekvenční testovou proceduru, kde jako první nulovou hypotézu uvažujeme největší možný počet jednotkových kořenů, a dále snižujeme řád difference pokaždé, když aktuální nulovou hypotézu zamítneme. Uveďme příklad tohoto postupu na $AR(2)$ procesu.

Uvažujme $AR(2)$ model

$$(1 - \phi_1 L)(1 - \phi_2 L)y_t = \epsilon_t.$$

Tento model můžeme reparametrizovat do tvaru

$$\Delta^2 y_t = \beta_1 \Delta y_{t-1} + \beta_2 y_{t-1} + \epsilon_t,$$

kde $\beta_1 = (\phi_1 \phi_2 - 1)$ a $\beta_2 = -(1 - \phi_1)(1 - \phi_2)$.

Testová procedura sestává z následujících kroků:

1. Testujme nulovou hypotézu dvou jednotkových kořenů proti alternativní hypotéze jednoho jednotkového kořenu. Za platnosti nulové hypotézy musí být $\beta_1 = \beta_2 = 0$ a můžeme použít F -test pro testování této sdružené hypotézy. F -test však není jednostranné povaze alternativní hypotézy uzpůsoben. Daleko lépe vystihne takovou hypotézu t -test, který můžeme použít, pokud uvažujeme následovně. Při platnosti H_0 i H_1 platí, že $\beta_2 = 0$. Naopak při platnosti H_0 je $\beta_1 = 0$, ale při platnosti H_1 je $\beta_1 < 0$. Proto je účinnější testovat odhad β_1 na základě regrese $\Delta^2 y_t$ na Δy_{t-1} . Takto sestavená t statistika má opět DF rozdělení.
2. Pokud na základě t -testu zamítneme výše zmíněnou nulovou hypotézu, pokračujeme dále s testováním nové nulové hypotézy přítomnosti jednoho jednotkového kořene proti alternativní hypotéze stacionárního procesu. V tomto případě za platnosti H_0 je $\beta_1 < 0$, $\beta_2 = 0$ a za platnosti H_1 je $\beta_1 < 0$, $\beta_2 < 0$. Proto sestojíme t statistiku odhadu β_2 a porovnáme ji s kritickými hodnotami DF rozdělení.

Tuto testovou proceduru lze samozřejmě zobecnit na testování tří i více jednotkových kořenů. Všeobecně platí, že t -test má větší sílu než F -test.

4.2.2 Testy stacionarity

Pokud chceme testovat stacionaritu časových řad nebo jejich lineární kombinace, bylo by zajímavé testovat tuto hypotézu přímo. Jelikož víme, že klasická metodologie testování hypotéz nám zajišťuje přijetí nulové hypotézy, dokud proti ní není velmi silná statistická evidence, není překvapivé, že některé empirické práce ukazují, že standardní testy jednotkového kořene nezamítají nulovou hypotézu, přestože je časová řada ve skutečnosti stacionární. Proto je užitečné použít k testování oba druhy testů, jak s nulovou hypotézou stacionarity procesů, tak i s nulovou hypotézou přítomnosti jednotkového kořene.

Kwiatkowski, Phillips, Schmidt a Shin vytvořili v roce 1992 test nulové hypotézy stacionarity procesu proti alternativě přítomnosti jednotkového kořene, tzv. **KPSS test**. Uvažujme následující generující proces

$$y_t = \beta t + \alpha_t + \epsilon_t. \quad (4.22)$$

Výraz α_t nechť je náhodná procházka

$$\alpha_t - \alpha_{t-1} = u_t,$$

kde počáteční hodnota α_0 je pevně daná a hraje roli konstanty, reziduální složka $\{u_t\} \sim WN(0, \sigma_u^2)$ a $\{\epsilon_t\}$ je stacionární proces nezávislý na $\{u_t\}$. Pokud by ovšem neplatil předpoklad a α_t by nebyla náhodná procházka, byla by pak časová řada y_t trendově stacionární. Proto je nulová hypotéza stacionarity vyjádřena jako $H_0 : \sigma_u^2 = 0$.

Nechť $\hat{\epsilon}_t$, $t = 1, \dots, T$, jsou odhadnutá rezidua metodou nejmenších čtverců z pomocné regrese

$$y_t = \alpha_0 + \beta t + \epsilon_t$$

a definujme částečný součet reziduí jako

$$S_t = \sum_{i=1}^t \hat{\epsilon}_i.$$

Idea testu je taková, že za platnosti H_0 by S_t měl být relativně malý, ale za platnosti H_1 by měl být významný. Testová statistika KPSS testu tedy vypadá následovně:

$$\eta = \frac{\sum_{t=1}^T S_t^2}{T^2 \hat{\sigma}^2}, \quad (4.23)$$

kd $\hat{\sigma}^2$ je konzistentní odhad rozptylu reziduální složky ϵ_t .

Autoři tabelovali kritické hodnoty rozdělení testové statistiky η za platnosti silnějších předpokladů, že u_t je normální a $\epsilon_t \sim GWN(0, \sigma^2)$. Jelikož η nabývá pouze kladných hodnot, nulová hypotéza trendově stacionárního procesu je zamítnuta, pokud η překročí příslušnou kritickou hodnotu.

4.3 Testy kointegrace

Jelikož z teorie kointegrace víme, že komponenty vektoru \mathbf{y}_t jsou kointegrované, pokud existuje jejich stacionární lineární kombinace, nepřekvapí nás, že Engle a Granger navrhli jednoduchý test kointegrace vektoru \mathbf{y}_t jako testování existence jednotkového kořenu v reziduích statického modelu (model bez časově zpožděných proměnných). V praxi se nejčastěji setkáváme s $I(1)$ řadami, proto je tedy hlavní otázkou, zda rezidua tvaru

$$\alpha' \mathbf{y}_t = u_t \quad (4.24)$$

jsou typu $I(1)$ nebo $I(0)$. Pokud by byla typu $I(0)$, byly by komponenty vektoru \mathbf{y}_t kointegrované, tedy $CI(1, 1)$.

K testování hypotézy jednotkového kořene v reziduální složce však nestačí pouze použít testy popsané v předchozí části práce. Musíme si uvědomit, že je nutná modifikace těchto testů, a to proto, že v tomto případě nemůžeme test založit na pozorovaných hodnotách, ale pouze na jejich odhadu. Abychom mohli získat odhad reziduí \hat{u}_t , musíme nejprve odhadnout kointegrační vektor.

Předpokládejme, že všechny komponenty vektoru \mathbf{y}_t jsou $I(1)$ řady, a že jsou kointegrované s kointegračním vektorem α . Tento vektor lze odhadnout metodou nejmenších čtverců tak, že přeformulujeme rovnici (4.24). Předpokládejme, že vektor \mathbf{y}_t má K komponent. Pak rovnice vypadá následovně:

$$y_{t1} = \alpha_2 y_{t2} + \alpha_3 y_{t3} + \dots + \alpha_K y_{tK} + u_t. \quad (4.25)$$

Regrese tohoto typu se nazývá *kointegrační regrese* a $\alpha = (1, -\alpha_2, -\alpha_3, \dots, -\alpha_K)'$. Odhad kointegračního vektoru $\hat{\alpha}$ je velmi dobrou aproximací skutečného vektoru α , neboť, jak ukázal Stock v roce 1987, konverguje ke skutečné hodnotě rychlostí T^{-1} . Jelikož běžný odhad parametru metodou nejmenších čtverců konverguje "jen" rychlostí $T^{-1/2}$, nazývá se $\hat{\alpha}$ superkonzistentním odhadem.

Odhadneme tedy rezidua modelu (4.24) jako

$$\hat{u}_t = \hat{\alpha}' \mathbf{y}_t \quad (4.26)$$

a budeme testovat nulovou hypotézu přítomnosti jednotkového kořene v \hat{u}_t proti alternativní hypotéze stacionarity odhadnutých reziduí. K testování těchto hypotéz se používá převážně následujících pět testů:

1. Durbin-Watsonův test kointegrační regrese

$$CRDW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2}, \quad (4.27)$$

kde \hat{u}_t je odhad reziduí z kointegrační regrese. Nulová hypotéza je v tomto případě existence jednoho jednotkového kořene, tedy za platnosti nulové

hypotézy je u_t náhodná procházka. Upozorníme na to, že v klasickém DW testu je nulovou hypotézou neautokorelovanost reziduí.

Opět platí, že jelikož CRDW statistika závisí na počtu regresorů v kointegrační rovnici, neexistují tabelované kritické hodnoty jejího rozdělení, jen jejich horní a dolní limitní rozdělení.

2. **Dickey-Fullerův test** (odhad parametru ρ)

$$DF(\rho) = T\hat{\rho}, \quad (4.28)$$

kde $\hat{\rho}$ získáme z regrese $\Delta\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + \hat{v}_t$.

3. **Dickey-Fullerův test** (statistika t)

$$DF(t) = t_{\rho=0}, \quad (4.29)$$

kde statistika t vychází z regrese $\Delta\hat{u}_t = \rho\hat{u}_{t-1} + \hat{v}_t$.

4. **Rozšířený Dickey-Fullerův test** (odhad parametru ρ)

$$ADF(\rho) = T\hat{\rho}, \quad (4.30)$$

kde $\hat{\rho}$ získáme z regrese $\Delta\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + \sum_{j=1}^p \phi_j^* \Delta\hat{u}_{t-j} + \hat{v}_t$.

5. **Rozšířený Dickey-Fullerův test** (statistika t)

$$ADF(t) = t_{\rho=0}, \quad (4.31)$$

kde statistika t vychází z regrese $\Delta\hat{u}_t = \rho\hat{u}_{t-1} + \sum_{j=1}^p \phi_j^* \Delta\hat{u}_{t-j} + \hat{v}_t$.

Všechny výše zmíněné DF a ADF testy vychází z odhadu reziduí v kointegrační regresi (4.24). V DF testech předpokládáme tvar reziduální složky jako $AR(1)$ proces a v ADF testech pak $AR(p)$ proces. Stejně jako u testů jednotkového kořene je možné rozšíření na $ARMA$ tvar reziduí, a to aproximací AR procesu s dostatečně velkým p .

Kritické hodnoty těchto testů nejsou shodné s kritickými hodnotami testů přítomnosti jednotkového kořenu, protože hodnota u_t není známá a musí být nejprve odhadnuta. Kritické hodnoty jsou však i pro tento případ tabelovány, najdeme je například v Banerjee a kol. (1993) či v Engle, Granger (1991).

Kapitola 5

Příklad

V této kapitole nyní ukážeme praktické použití výše popsaných metod. U konkrétních dvou časových řad se pokusíme dokázat jejich vzájemnou závislost. Pokud se prokáže, že jsou skutečně závislé, tedy kointegrované, najdeme jejich kointegrační vektor. V opačném případě pomocí statistických testů dokážeme, že se jedná pouze o zdánlivou regresi.

Máme k dispozici dvě časové řady, index dovozních cen (*IDC*) a index vývozních cen (*IVC*) České republiky za všechny oblasti importu i exportu (zdroj dat: ČSÚ). Obě jsou zaznamenány měsíčně v období od ledna 2001 do ledna 2006, máme tedy 61 pozorování. Jejich průběh je zachycen na obrázku 5.1, řada *IDC* je vyznačena tmavší, řada *IVC* světlejší barvou.

Z pohledu na obrázek můžeme konstatovat, že se řady za celé pozorovací období chovají velice podobně. Proto se dá očekávat, že spolu budou nějak souviset, tedy že budou kointegrované. Jelikož chceme odhadovat model vzájemné závislosti těchto řad, musíme vědět, zda jsou řady stacionární, či zda obsahují jednotkový kořen. K testování těchto hypotéz využijeme Dickey-Fullerovy a KPSS testy. Vychází nám, že u ADF testu nemůžeme zavrhnout hypotézu existence jednotkového kořenu ani pro jednu z obou studovaných řad, a na základě KPSS testu zavrhuje hypotézu o stacionaritě procesů. Proto můžeme konstatovat, že obě řady jsou $I(1)$.

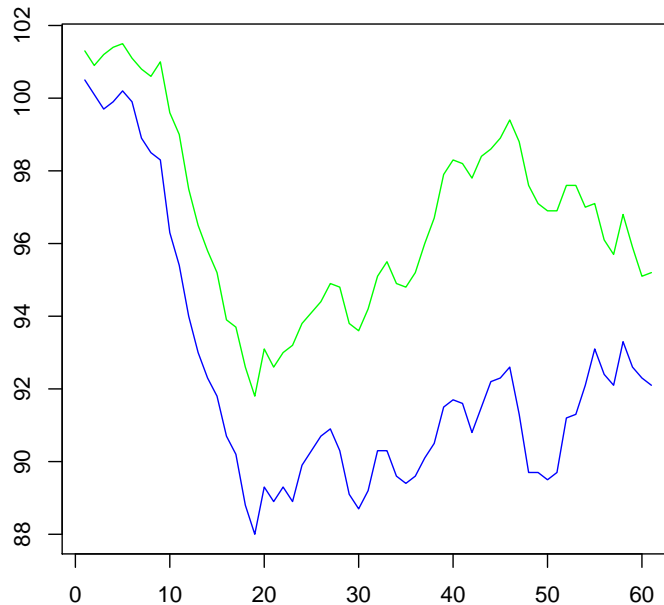
Z obrázku je také zřejmé, že ani jedna z řad neobsahuje lineární trend, takže ho nemusíme zahrnovat ani do celkové regrese. Metodou nejmenších čtverců získáme odhady parametrů β_0 a β_1 v modelu

$$IDC_t = \beta_0 + \beta_1 IVC_t + \epsilon_t \quad (5.1)$$

a zjistíme, zda model dobře vystihuje vzájemnou závislost řad. Vychází nám, že

$$IDC_t = -16,419 + 1,124ICND_t. \quad (5.2)$$

U obou parametrů musíme na základě t -testu zamítnout nulovou hypotézu o jejich nulovosti a na základě F -testu konstatujeme celkovou signifikantnost modelu (na

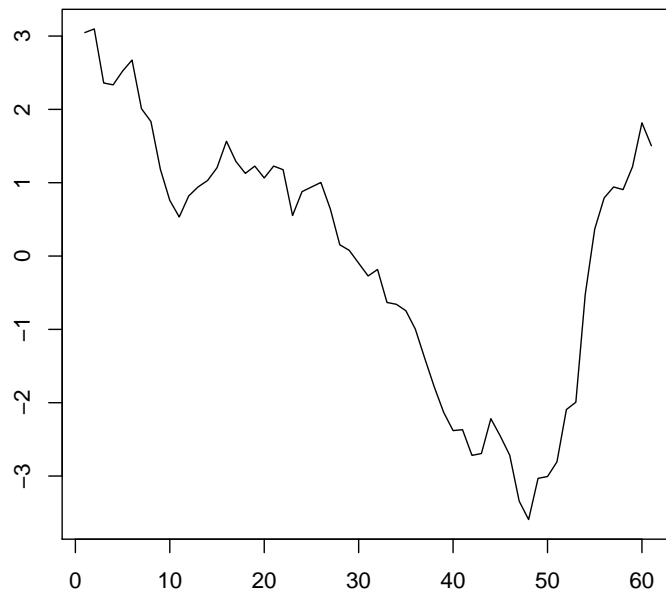
Obrázek 5.1: Časové řady *IDC* a *IVC*

5% hladině významnosti). Koeficient determinace je vysoký: $R^2 = 0,725$. Může se tedy zdát, že model (5.2) dobře vystihuje závislost našich řad.

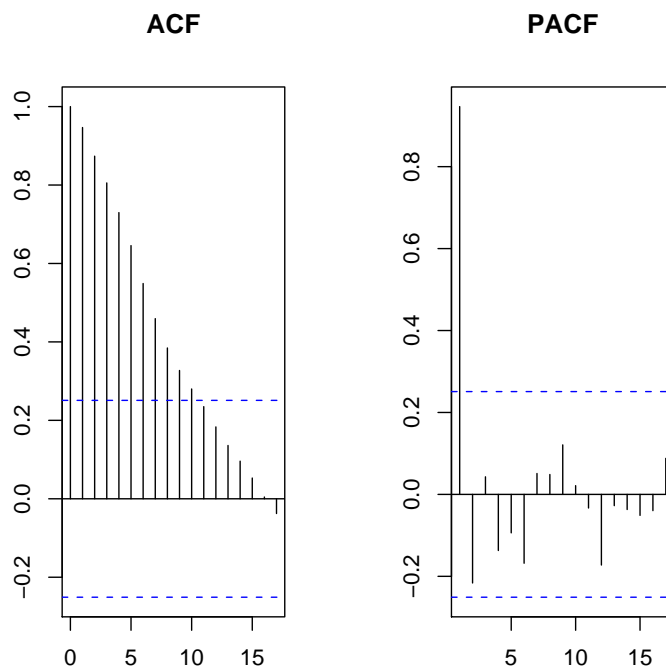
Je ovšem ještě potřeba zkontrolovat průběh reziduí našeho modelu. Ta najdeme graficky znázorněná na obrázku 5.2. Z tohoto obrázku je zřejmé, že rezidua obsahují systematický pohyb, proto můžeme předpokládat, že jsou autokorelovaná, což nám potvrzuje i velmi nízká hodnota Durbin-Watsonovy statistiky ($DW = 0,048$).

Z obrázku 5.2 tedy očekáváme přítomnost zdánlivé regrese a zároveň platí pomocné pravidlo pro její rozpoznávání (dle Grangera a Newbolda): $R^2 > DW$. Navíc z obrázku 5.3, který zachycuje autokorelační (ACF) a parciální autokorelační (PACF) funkce vektoru reziduí, můžeme předpokládat přítomnost jednotkového kořene, protože první hodnoty obou grafů jsou velice blízké jedné. Zbývá nám tedy tuto domněnku podpořit statistickým testem. To učiníme pomocí ADF testu kointegrace modelu. Statistika t ADF testu vychází rovna $-1,015$, což je hodnota daleko vyšší než kritické hodnoty uvedené v Banerjee a kol. (1993). Proto nemůžeme zavrhnout hypotézu o přítomnosti jednotkového kořene ve vektoru reziduí našeho modelu.

Můžeme tedy konstatovat, že časové řady *IDC* a *IVC* nejsou kointegrované, jejich podobný průběh není způsoben vzájemnou závislostí a při snaze modelovat tuto závislost končíme u problému zdánlivé regrese.



Obrázek 5.2: Rezidua modelu (5.2)



Obrázek 5.3: ACF a PACF reziduí modelu (5.2)

Kapitola 6

Závěr

Záměrem této diplomové práce bylo objasnit pojem zdánlivé regrese. Na základě prostudování zahraniční literatury bylo potřeba shrnout všechny důležité vlastnosti tohoto fenoménu a nastínit možnosti jeho odhalení a řešení.

V první části jsme si připomněli základní pojmy z teorie stochastických procesů a časových řad, jako jsou pojmy stacionarity či ergodicity, a prezentovali jsme seznam základních a převážně používaných generujících procesů časových řad. Ve druhé části jsme se zabývali nestacionárními procesy a jejich možnou stacionarizací a ve třetí kapitole jsme pak navázali výkladem o zdánlivé regresi. Jelikož z tohoto výkladu plyne, že se zdánlivou regresí úzce souvisí pojem kointegrace, je právě tento předmětem další kapitoly, ve které jsme se dostali i k testování přítomnosti jednotkového kořene v generujícím procesu a k testování přítomnosti zdánlivé regrese. V poslední části této práce jsme pak předvedli vybrané metody v praxi.

Největším přínosem této práce je třetí kapitola, která obsahuje nejnovější a významné objevy v oblasti modelování vztahů více časových řad. Je zde podrobně vysvětlen pojem zdánlivé regrese, příčiny jeho vzniku, důsledky v případě neodhalení tohoto problému a metody jeho včasného objevení. V poslední kapitole je většina teoretických tvrzení aplikována na reálné ekonomické ukazatele, díky čemuž je prokázána nutnost nezanedbání problému zdánlivé regrese.

Literatura

- Banerjee, A., Dolado, J., Galbraith, J. W., Hendry, D. F. (1993): *Co-integration, error-correction, and the econometric analysis of non-stationary data*, Oxford University Press, Oxford, 1993.
- Engle, R. F., Granger, C. W. J. (1987): *Co-integration and error-correction: representation, estimation, and testing*, *Econometrica* **55** (1987), 251 - 276.
- Engle, R. F., Granger, C. W. J. (1991): *Long-run economic relationships. Reading in cointegration*, Oxford University Press, Oxford, 1991.
- Granger, C. W. J., Newbold P. (1974): *Spurious regression in econometrics*, *Journal of Econometrics* **2** (1974), 111 - 120.
- Hamilton, J. D. (1994): *Time series analysis*, Princeton University Press, Princeton, 1994.
- Hendry, D. F. (1980): *Econometrics: alchemy or science?*, *Economica* **47** (1980), 387 - 406.
- Phillips, P. C. B. (1986): *Understanding spurious regression in econometrics*, *Journal of Econometrics* **33** (1986), 311 - 340.
- Prášková, Z., Lachout, P. (1998): *Základy náhodných procesů*, Karolinum, Praha, 1998.
- Prášková, Z. (1998): *Základy náhodných procesů. II*, Karolinum, Praha, 2004.
- Yule, G. U. (1926): *Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series*, *Journal of the Royal Statistical Society* **89** (1926), 1-69.