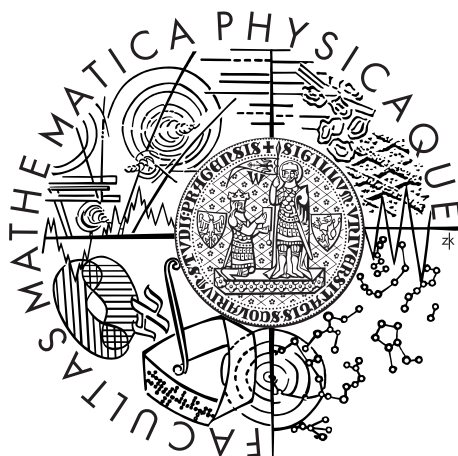


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Bc. František Kalibán

Nelinearita v modelech časových řad

Katedra pravděpodobnosti a matematické statistiky (32-KPMS)

Vedoucí diplomové práce: prof. RNDr. Jiří Anděl, DrSc.

Studijní program: Matematika

Studijní obor: Ekonometrie

Praha 2013

Rád bych poděkoval prof. RNDr. Jiřímu Andělovi, DrSc. za odborné vedení práce,
cenné připomínky a rady.

Prohlašuji, že jsem tuto diplomovou práci vypracoval 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 1.12.2013

Podpis autora

Název práce: Nelinearita v modelech časových řad

Autor: Bc. František Kalibán

Katedra: Katedra pravděpodobnosti a matematické statistiky (32-KPMS)

Vedoucí diplomové práce: prof. RNDr. Jiří Anděl, DrSc.

E-mail vedoucího: andel@karlin.mff.cuni.cz

Abstrakt: Práce pojednává o vlastnosti linearity v modelech časových řad, možnostech jejího definování a testování. Jsou zde představeny především testy v časové doméně založené na různých statistických teoriích, jako jsou regrese, neuronové sítě nebo náhodná pole. Důraz je kladen na jejich implementaci v softwarovém prostředí R. U testů, které jsou implementovány ve více verzích, jsou srovnány jejich výhody a nevýhody. Dále se práce věnuje definici aditivity v nelineárních časových řadách a možnostem jejího testování. Následuje simulační studie testů linearity i aditivity. Za účelem těchto srovnání byly některé testy naprogramovány ve statistickém software R. V závěru práce jsou testy aplikovány na reálná data.

Klíčová slova: nelineární časové řady, testy nelinearity, testy aditivity

Title: Nonlinearity in time series models

Author: Bc. František Kalibán

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Jiří Anděl, DrSc.

Supervisor's e-mail address: andel@karlin.mff.cuni.cz

Abstract: The thesis concentrates on property of linearity in time series models, its definitions and possibilities of testing. Presented tests focus mainly on the time domain; these are based on various statistical methods such as regression, neural networks and random fields. Their implementation in R software is described. Advantages and disadvantages for tests, which are implemented in more than one package, are discussed. Second topic of the thesis is additivity in nonlinear models. The definition is introduced as well as tests developed for testing its presence. Several test (both linearity and additivity) have been implemented in R for purposes of simulations. The last chapter deals with application of tests to real data.

Keywords: Nonlinear Time Series, Nonlinearity Tests, Nonadditivity Tests

Obsah

Úvod	1
1 Základní pojmy	2
2 Definice linearity v modelech časových řad	6
3 Testování linearity v časových řadách	8
3.1 Keenanův test	8
3.1.1 Keenan.test	11
3.2 Tsayův test	12
3.3 Tsayův test na současnou nelinearitu	14
3.4 RESET test	15
3.5 Testy založené na neuronových sítích	16
3.5.1 Whiteův test	18
3.5.2 Teräsvirtaův test	19
3.6 Skórové testy	20
3.7 Testy založené na náhodných polích	21
3.7.1 Hamiltonův test	23
3.7.2 Rozšíření Hamiltonova testu	25
3.8 Testy bez specifické alternativy	25
3.8.1 BDS test	25
3.8.2 Korelační test	26
3.9 Další testy linearity	27
3.9.1 Testy ve spektrální doméně	27
3.9.2 Testy na konkrétní nelineární modely	27
3.10 Vyjádření modelu Taylorovým rozvojem	28
4 Aditivita	29
4.1 Definice aditivity	29
4.2 Testování aditivity	29
4.2.1 Test podmíněných středních hodnot	30
4.2.2 Test založený na Lagrangeově multiplikátoru	32
4.2.3 Permutační test	33
5 Simulační studie	35
5.1 Testy nelinearity	35

5.1.1	Taylorův rozvoj spojitých funkcí	35
5.1.2	Aplikace RESET testu na určení typu nelinearity	38
5.1.3	Mocninný nelineární model	40
5.2	Testy aditivity	43
6	Užití testů na reálných datech	48
	Závěr	51
	Dodatky	52
A	Pomocná tvrzení	52
B	Oprava Keenanova testu z TSA	53
C	Implementace testů v R	55
C.1	Hamiltonův test	55
C.2	Test aditivity podmíněných středních hodnot	56
C.3	Permutační test aditivity	57
D	Obsah příloženého CD	58

Úvod

S rozvojem teorie časových řad se v 90. letech minulého století objevila myšlenka popisovat zákonitosti chování některých náhodných procesů nelineárně. Struktura závislosti různých přírodních, ekonomických, popř. finančních jevů na vnějších faktorech je totiž většinou mnohem komplikovanější, např. vývoj počasí nemusí systematicky záviset jen na stavu počasí v předchozích dnech (ať už mluvíme o teplotách, oblačnosti nebo jiných ukazatelích), ale například i na určité specifické kombinaci těchto stavů, což vnáší do procesu charakter nelinearity.

V souvislosti s tím se objevily otázky, jak tuto skutečnost testovat, a v případě odůvodněného podezření na nelineární charakter řady upozornit na nutnost popisu řady nelineárními modely.

V kapitole 1 jsou popsány základní definice a pojmy nutné pro porozumění obsahu této práce. Kapitola 2 popisuje možné definice linearity v modelech časových řad, kapitola 3 představuje testy linearity a hovoří o jejich implementaci v prostředí R. Důraz je kladen na testy v časové doméně, testy ve spektrální doméně jsou uvedeny výčtem. V kapitole 3.10 je popsána možnost, jak by se dal blíže určit tvar nelinearity v případě, že je linearita zamítnuta. Je zde nastíněna myšlenka, která je v simulační práci ověřena.

Kapitola 4 definuje aditivitu v nelineárních modelech a popisuje možnosti jejího testování. Shrnuje několik testů a pojednává o jejich implementaci.

Kapitola 5 srovnává jak testy linearity, tak testy aditivity, simulačně na některých konkrétních modelech. Závěrečná kapitola pak ukazuje použití testů linearity i aditivity na reálných datech. V dodatcích práce nalezneme především implementační skripty pro prostředí R.

1. Základní pojmy

Nejprve připomeneme několik základních definic. S drobnými úpravami pro přizpůsobení značení jsou převzaté ze skript Prášková (2007).

Definice 1.1 (náhodný proces). Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní prostor a nechť $T \subset \mathbb{R}$. Rodina reálných náhodných veličin $\{Y_t, t \in T\}$ definovaných na (Ω, \mathcal{A}, P) se nazývá *náhodný proces*.

Definice 1.2 (striktně stacionární proces). Řekneme, že náhodný proces $\{Y_t, t \in T\}$ je *striktně stacionární*, jestliže pro libovolné $n \in \mathbb{N}$, pro libovolná reálná y_1, \dots, y_n a pro libovolná t_1, \dots, t_n a h taková, že $t_k \in T, t_k + h \in T, 1 \leq k \leq n$, platí

$$F_{t_1, \dots, t_n}(y_1, \dots, y_n) = F_{t_1+h, \dots, t_n+h}(y_1, \dots, y_n).$$

Pokud je náhodný proces $\{Y_t, t \in T\}$ striktně stacionární, pak všechny náhodné veličiny Y_t mají stejné rozdělení a základní charakteristiky, jako je střední hodnota, rozptyl nebo kovariance. Opačná implikace neplatí, vede ale k následující definici:

Definice 1.3 (slabá stacionarita). Má-li náhodný proces $\{Y_t, t \in T\}$ s konečnými druhými momenty konstantní střední hodnotu $\mu_t = \mu$ pro všechna $t \in T$ a je-li jeho autokovarianční funkce $R(s, t)$ funkcí pouze $s - t$, pak o takovém procesu říkáme, že je *slabě stacionární*. Pokud v následujícím textu nebude uvedeno jinak, pak pojmem *stacionarita* myslíme právě slabou stacionaritu.

Definice 1.4 (časová řada). Pokud platí $T \subset \mathbb{Z}$, tedy náhodný proces má *diskrétní čas*, pak jej označujeme také jako *náhodnou posloupnost*, případně jako *časovou řadu*.

V literatuře je často pojem náhodný proces a časová řada zaměňován. V této práci se budeme zabývat pouze procesy s diskrétním časem, pojem proces má tedy v kontextu této práce stejný význam jako časová řada.

Definice 1.5 (bílý šum). Posloupnost náhodných veličin $\{e_t, t \in \mathbb{Z}\}$ s vlastnostmi

- $E(e_t) = 0$,
- $E(e_t^2) = \sigma^2$,
- $E(e_t e_s) = 0$, pro $t \neq s$,

kde $\sigma^2 \in \mathbb{R}$, nazýváme *bílý šum*. Pokud jsou náhodné veličiny navíc nezávislé a stejně rozdělené, pak jde o *striktní bílý šum*.

Definice 1.6 (obecný lineární proces). Nechť posloupnost $\{e_t, t \in \mathbb{Z}\}$ je bílý šum a $\{c_j, j \in \mathbb{N}_0\}$ je posloupnost obecně komplexních konstant taková, že $\sum_{j=-\infty}^{\infty} |c_j| < \infty$. Náhodná posloupnost $\{Y_t, t \in \mathbb{Z}\}$ definovaná jako

$$Y_t = \text{l.i.m.}_{n \rightarrow \infty} \sum_{j=-n}^n c_j e_{t-j}$$

se nazývá *obecný lineární proces*.

Poznámka. Symbol l.i.m. značí limitu podle kvadratického středu, tedy Y je limitou posloupnosti Y_n , pokud je její limitou v prostoru L_2 , tedy

$$\|Y_n - Y\|^2 = \mathbb{E}|Y_n - Y|^2 \rightarrow 0 \text{ pro } n \rightarrow \infty.$$

Definice 1.7 (kauzální lineární proces). Nechť $\{e_t, t \in \mathbb{Z}\}$ je bílý šum a $\{c_j, j \in \mathbb{N}_0\}$ je posloupnost komplexních konstant nyní taková, že platí $\sum_{j=0}^{\infty} |c_j| < \infty$. Pak náhodné posloupnosti $\{Y_t, t \in \mathbb{Z}\}$ definované jako

$$Y_t = \text{l.i.m.}_{n \rightarrow \infty} \sum_{j=0}^n c_j e_{t-j}$$

říkáme *kauzální lineární proces*.

Pro kauzální náhodný proces je podstatné, že hodnoty posloupnosti Y_t závisí v čase t pouze na „minulosti“ procesu e_s , tedy pouze na hodnotách e_s takových, že $s \leq t$.

Definice 1.8 (AR proces). Náhodná posloupnost $\{Y_t, t \in \mathbb{Z}\}$ se nazývá *autoregresní posloupnost řádu n* (značíme $\text{AR}(n)$), pokud lze vyjádřit ve tvaru

$$Y_t = \sum_{j=1}^n a_j Y_{t-j} + e_t, \quad t \in \mathbb{Z}, \quad (1.1)$$

kde $a_1, \dots, a_n \in \mathbb{R}$, $a_n \neq 0$ a $\{e_t, t \in \mathbb{Z}\}$ je bílý šum.

Definice 1.9 (MA proces). Náhodná posloupnost $\{Y_t, t \in \mathbb{Z}\}$ definovaná předpisem

$$Y_t = \sum_{j=0}^n b_j e_{t-j}, \quad t \in \mathbb{Z},$$

kde $b_0, \dots, b_n \in \mathbb{R}$, $b_0 \neq 0$, $b_n \neq 0$ a $\{e_t, t \in \mathbb{Z}\}$ je bílý šum, se nazývá *posloupnost klouzavých součtů řádu n* a značíme ji $\text{MA}(n)$.

Kauzální lineární proces se někdy značí jako $MA(\infty)$, neboť jej lze vyjádřit ve tvaru nekonečné řady klouzavých součtů.

Poznámka. Název posloupnost klouzavých součtů může být matoucí při studiu problematiky v anglicky psané literatuře, neboť tam je tato posloupnost značena jako *posloupnost klouzavých průměrů* (angl. moving average).

Za určitých podmínek je AR proces kauzálním lineárním procesem a můžeme jej tedy vyjádřit ve tvaru $MA(\infty)$. Podmínky uvádí následující věta.

Věta 1.1. Nechť $\{Y_t, t \in \mathbb{Z}\}$ je autoregresní posloupnost n -tého řádu definovaná vztahem (1.1). Jestliže polynom $a(z) = 1 + a_1z + \dots + a_nz^n$ má všechny kořeny vně jednotkového kruhu, potom $\{Y_t, t \in \mathbb{Z}\}$ je kauzální lineární proces, tj. platí

$$Y_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad t \in \mathbb{Z},$$

kde koeficienty c_j jsou určeny vztahem

$$c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{1}{a(z)}, \quad |z| \leq 1.$$

Důkaz. Viz Prášková (2007). □

Definice 1.10 (ARMA proces). Nechť je náhodná posloupnost $\{Y_t, t \in \mathbb{Z}\}$ definovaná předpisem

$$Y_t = \sum_{i=1}^m a_i Y_{t-i} + \sum_{j=0}^n b_j e_{t-j}, \quad t \in \mathbb{Z},$$

kde $a_i \in \mathbb{R}$, $i = 1, \dots, m$, $b_j \in \mathbb{R}$, $j = 1, \dots, n$, $a_m \neq 0$, $b_n \neq 0$ a $\{e_t, t \in \mathbb{Z}\}$ je bílý šum. Pak říkáme, že se tato posloupnost řídí modelem $ARMA(m, n)$.

Za určitých podmínek lze ARMA proces přepsat do tvaru kauzálního lineárního procesu. Označme $a(z) = 1 + a_1z + a_2z^2 + \dots + a_mz^m$ a $b(z) = 1 + b_1z + b_2z^2 + \dots + b_nz^n$. Užitím operátoru zpětného¹ posunu B můžeme ARMA proces vyjádřit ve tvaru

$$a(B)Y_t = b(B)e_t. \tag{1.2}$$

Podmínky, za kterých lze ARMA proces převést do tvaru lineárního procesu shrnuje následující věta.

¹Někdy též nazývaný jako operátor časového posunu

Věta 1.2. Nechť $\{Y_t, t \in \mathbb{Z}\}$ je náhodná posloupnost ARMA(m, n) daná vztahem (1.2). Nechť polynomy $a(z)$ a $b(z)$ nemají společné kořeny a nechť polynom $a(z)$ má všechny kořeny vně jednotkového kruhu. Potom Y_t lze psát ve tvaru

$$Y_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad t \in \mathbb{Z},$$

kde koeficienty c_j jsou určeny vztahem

$$c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)}, \quad |z| \leq 1.$$

Důkaz. Viz Prášková (2007). □

Poznámka. Náhodné procesy dle výše uvedených definic mají nulovou střední hodnotu. Definice je možné zobecnit přičtením $\mu \in \mathbb{R}$.

Definice 1.11. Řekneme, že náhodná posloupnost $\{Y_t, t \in \mathbb{Z}\}$ se střední hodnotou μ je ergodická, neboli splňuje zákon velkých čísel v $L_2(\Omega, \mathcal{A}, P)$, jestliže pro $n \rightarrow \infty$ platí

$$\frac{1}{n} \sum_{t=1}^n Y_t \xrightarrow{P} \mu,$$

neboli je splněn slabý zákon velkých čísel pro stacionární posloupnost.

2. Definice linearity v modelech časových řad

Definice 2.1 (lineární časová řada). *Lineární časová řada* je taková časová řada, kterou lze převést na kauzální lineární proces (viz 1.7).

Uvažujme stacionární časové řady, které lze zapsat ve tvaru

$$\begin{aligned}
 Y_t = \mu + \sum_{k=0}^{\infty} a_k e_{t-k} + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{jk} e_{t-j} e_{t-k} \\
 + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{ijk} e_{t-i} e_{t-j} e_{t-k} + \dots
 \end{aligned}
 \tag{2.1}$$

Takový zápis se nazývá Volterrův rozvoj („Volterra expansion“) a zavádí jej např. Keenan (1985). Vyjádření zahrnuje modely lineárních stacionárních řad jako jsou řady klouzavých součtů, ale i AR a ARMA procesy (za podmínek dle vět 1.1 a 1.2). Mimo to ale vyhovuje i některým nelineárním časovým řadám. O lineární řady se zřejmě jedná v případě, že všechny koeficienty řádu 2 a vyššího, tj. a_{jk}, a_{ijk}, \dots , jsou nulové (pak se jedná o *kauzální lineární proces*).

V článku Priestley (1980) je uvedena ještě jedna verze Volterrova rozvoje. Priestley ji označuje jako „duální“ Volterrův rozvoj. Pokud je daná časová řada invertibilní, pak ji lze zapsat ve tvaru

$$e_t = \mu' + \sum_{k=0}^{\infty} b_k Y_{t-k} + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} b_{jk} Y_{t-j} Y_{t-k} + \dots
 \tag{2.2}$$

Tento zápis je pro studium linearity v modelech časových řad vhodnější, neboť ve většině testů se pracuje s fitováním AR procesů do časové řady. Řada je potom lineární, pokud jsou všechny členy druhého a vyšších řádů (b_{jk}, \dots) nulové.

Alternativně se dá linearita časové řady definovat jako tzv. *podmíněná linearita ve střední hodnotě*. Uvažujme pro každý časový okamžik t vektor proměnných \mathbf{X}_t . Pak je daná časová řada lineární, pokud pro nějaké $\boldsymbol{\theta}^*$ platí

$$P[E(Y_t | \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\theta}^*] = 1.
 \tag{2.3}$$

Vektor \mathbf{X}_t může být tvořen exogenními proměnnými nebo může jít o zpožděné hodnoty časové řady Y_t . Tato definice vyhovuje např. i modelům ARCH, či GARCH, které převést do tvaru lineárního procesu nelze, přesto však za určitých

podmínek mohou splňovat vztah (2.3). Tuto definici uvádí například Lee a kol. (1993).

Další možností, jak můžeme chápat linearitu v časové řadě, je linearita nejlepší předpovědi o jeden krok. Tedy proces je lineární, pokud mezi všemi jednokrokovými předpověďmi má nejmenší střední čtvercovou chybu taková predikce, která je lineární kombinací zpožděných hodnot nebo exogenních proměnných. Tuto definici můžeme nalézt např. v knize Cryer a Chan (2008).

V této práci budeme linearitu chápat ve smyslu definice 2.1.

3. Testování linearity v časových řadách

V polovině 80. let minulého století (Keenan, 1985) se ukazuje, že žádný z dosavadních testů linearity není dostatečně vhodný pro praktické použití. Keenan tedy prezentuje test, založený na Tukeyově testu aditivity (Tukey, 1949), a hypotézu linearity testuje proti alternativě přítomnosti členu druhého řádu (aspoň jedno nenulové b_{jk} ve vyjádření (2.2)).

3.1 Keenanův test

Předpokládejme, že $\{Y_t\}$ je stacionární časová řada. Na základě veličin Y_1, \dots, Y_n zkonstruujeme test nulové hypotézy linearity proti alternativě nelinearity ve třech krocích:

1. Regresí Y_t na $\{1, Y_{t-1}, \dots, Y_{t-M}\}$, kde M je předem zvolený řád autoregresního procesu, získáme rezidua $\{\hat{e}_t\}$ a fitované hodnoty $\{\hat{Y}_t\}$. Z důvodu zahrnutí zpožděných hodnot regresorů do M -tého řádu provádíme regresi pro $t = M + 1, \dots, n$, kde n je délka časové řady.
2. Regresí čtverců fitovaných hodnot řady z předchozího kroku, tj. $\{\hat{Y}_t^2\}$ na $\{1, Y_{t-1}, \dots, Y_{t-M}\}$ pro $t = M + 1, \dots, n$, určíme rezidua $\{\hat{\xi}_t\}$.
3. Konečně regresí $\hat{e} = (\hat{e}_{M+1}, \dots, \hat{e}_n)$ na $\hat{\xi} = (\hat{\xi}_{M+1}, \dots, \hat{\xi}_n)$, tedy $\hat{e}_t = \eta_0 \hat{\xi}_t + e_t$, dostaneme odhad regresního koeficientu $\hat{\eta}_0$. Povšimněme si, že v tomto kroku jde již o regresi bez interceptu. Z odhadu regresního koeficientu postupně napočítáme

$$\hat{\eta} = \hat{\eta}_0 \left(\sum_{t=M+1}^n \hat{\xi}_t^2 \right)^{\frac{1}{2}}$$

a testovou statistiku

$$\hat{F}_{1, n-2M-2} = \frac{\hat{\eta}^2 (n - 2M - 2)}{\left(\sum_{t=M+1}^n \hat{e}_t^2 \right) - \hat{\eta}^2}. \quad (3.1)$$

Ta se dle věty uvedené ve zmíněném článku za platnosti nulové hypotézy řídí asymptoticky F rozdělením s 1 a $n - 2M - 2$ stupni volnosti.

Zde není na první pohled zcela zřejmý záměr, proč provádět regresi fitovaných hodnot \hat{Y}_t na získaných reziduích. Proč závislost kvadrátů vyrovnaných hodnot

na vypočtených reziduích z prvního kroku znamená přítomnost nelinearity? Dobré vysvětlení tohoto záměru poskytuje článek Tsay (1986), který uvádí rozšířenou verzi testu. Veličina \hat{Y}_t^2 vznikne totiž váženým součtem kvadratických členů z (2.1), podrobněji je tento vztah popsán v kapitole 3.2. Nejprve ale dokážeme platnost konvergence testové statistiky Keenanova testu k F rozdělení za platnosti nulové hypotézy.

Důkaz. Vztah (3.1) můžeme napsat ve tvaru

$$\hat{F}_{1,n-2M-2} = \frac{\hat{\eta}^2}{\frac{RSS - \hat{\eta}^2}{n-2M-2}},$$

kde $RSS = \sum_{t=M+1}^n \hat{e}_t^2$ reziduální součet čtverců s $n - 2M - 1$ stupni volnosti¹. Platí tedy $\frac{RSS}{\sigma_e^2} \sim \chi_{n-2M-1}^2$, kde σ_e^2 je teoretický rozptyl náhodné složky regresního modelu zkonstruovaného v 1. kroku testu.

Ukážeme, že $\hat{\eta} \sim N(0, \sigma_e^2)$, pak $\frac{\hat{\eta}^2}{\sigma_e^2} \sim \chi_1^2$ a $\frac{RSS - \hat{\eta}^2}{\sigma_e^2} \sim \chi_{n-2M-2}^2$.

Označme $\mathbf{W}_t = (1, Y_{t-1}, \dots, Y_{t-M})'$ a $\boldsymbol{\theta} = (a_0, a_1, a_2, \dots, a_M)'$. Pak regresní model z prvního kroku testu můžeme zapsat jako

$$\begin{bmatrix} Y_{M+1} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \mathbf{W}'_{M+1} \\ \vdots \\ \mathbf{W}'_n \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} e_{M+1} \\ \vdots \\ e_n \end{bmatrix}. \quad (3.2)$$

Definujeme matici \mathbf{A}_n a vektor \mathbf{V}_n jako

$$\mathbf{A}_n = (n - M)^{-1} \sum_{t=M+1}^n \mathbf{W}_t \mathbf{W}'_t, \quad \mathbf{V}_n = (n - M)^{-1} \sum_{t=M+1}^n \mathbf{W}_t Y_t.$$

Pak OLS odhad modelu (3.2) můžeme podle věty 5.1 z Anděl (2007) vyjádřit jako

$$\hat{\boldsymbol{\theta}} = \mathbf{A}_n^{-1} \mathbf{V}_n.$$

Pro $t = M + 1, \dots, n$ zřejmě platí

$$\hat{Y}_t = \mathbf{W}'_t \hat{\boldsymbol{\theta}} \quad (3.3)$$

a

$$\hat{e}_t = Y_t - \hat{Y}_t = \mathbf{W}'_t (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + e_t. \quad (3.4)$$

Odhad reziduí z druhého kroku testu můžeme vyjádřit jako

$$\hat{\xi}_t = \hat{Y}_t^2 - \mathbf{W}'_t \mathbf{A}_n^{-1} \mathbf{Q}_n, \quad (3.5)$$

¹Z teorie lineární regrese, např. v Anděl (2007), věta 5.4.; zde máme model s $n - M$ pozorováními a $M + 1$ parametry.

kde

$$\begin{aligned}\mathbf{Q}_n &= (n - M)^{-1} \sum_{t=M+1}^n \mathbf{W}_t \hat{Y}_t^2 \\ &= (n - M)^{-1} \sum_{t=M+1}^n \mathbf{W}_t \hat{\boldsymbol{\theta}}' \mathbf{W}_t \mathbf{W}_t' \hat{\boldsymbol{\theta}}.\end{aligned}\tag{3.6}$$

Dle Fuller (1996) (věta 8.2.1) platí

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = o_p(1)\tag{3.7}$$

a

$$\mathbf{A}_n^{-1} \mathbf{V}_n - \mathbf{A}^{-1} \mathbf{V} = o_p(1),$$

kde \mathbf{A} je matice, jejíž prvky jsou limitami v pravděpodobnosti prvků matice \mathbf{A}_n . Vektor \mathbf{V} se skládá z prvků, které jsou limitou v pravděpodobnosti složek vektoru \mathbf{V}_n . Zbývá najít limitu vektoru \mathbf{Q}_n .

Definujme \mathbf{Q} jako sloupcový vektor, jehož první složkou je

$$a_0^2 + 2a_0\mu \sum_{j=1}^M a_j + \sum_{i=1}^M \sum_{j=1}^M a_i a_j \text{cov}(Y_i, Y_j)$$

a jeho k -tá složka, pro $k \in (2, \dots, M + 1)$, je

$$a_0^2 + 2a_0\mu \sum_{j=1}^M a_j \text{cov}(Y_j, Y_k) + \sum_{i=1}^M \sum_{j=1}^M a_i a_j \mathbb{E}[(Y_i - \mu)(Y_j - \mu)(Y_k - \mu)],$$

kde $\mu = \mathbb{E}Y_t$. Pak z (3.7) ale plyne

$$\mathbf{Q}_n - \mathbf{Q} = o_p(1).$$

Dále platí

$$\begin{aligned}\hat{\eta}_0 &= \frac{\sum_{t=M+1}^n \hat{e}_t \hat{\xi}_t}{\sum_{t=M+1}^n \hat{\xi}_t^2}, \\ \hat{\eta} &= \frac{n^{-1/2} \sum_{t=M+1}^n \hat{e}_t \hat{\xi}_t}{n^{-1/2} \sqrt{\sum_{t=M+1}^n \hat{\xi}_t^2}}.\end{aligned}\tag{3.8}$$

Dosazením (3.3), (3.4) a (3.5) do čitatele (3.8) dostaneme, že se čítec rovná

$$n^{-1/2} \sum_{t=M+1}^n e_t (\boldsymbol{\theta}' \mathbf{W}_t \mathbf{W}_t' \boldsymbol{\theta} - \mathbf{W}_t' \mathbf{A}^{-1} \mathbf{Q}) + o_p(1) = n^{-1/2} \sum_{t=M+1}^n \mathbf{Z}_t.\tag{3.9}$$

Posloupnost \mathbf{Z}_t tvoří stacionární ergodický proces martingalových diferencí s konstantním rozptylem $\sigma_e^2 \delta$, kde $\delta = \mathbb{E}[(\hat{\boldsymbol{\theta}}' \mathbf{W}_t \mathbf{W}_t' \hat{\boldsymbol{\theta}} - \mathbf{W}_t' \mathbf{A}^{-1} \mathbf{Q})^2]$. Z centrální limitní věty pro martingalové diference potom plyne, že čítec (3.8) konverguje

k $N(0, \sigma_e^2 \delta)$. Druhá mocnina jmenovatele (3.8) je po dosazení (3.3) a (3.5) rovna

$$\begin{aligned} n^{-1} \sum_{t=M+1}^n \hat{\xi}_t^2 &= n^{-1} \sum_{t=M+1}^n (\hat{\boldsymbol{\theta}}' \mathbf{W}_t \mathbf{W}_t' \hat{\boldsymbol{\theta}} - \mathbf{W}_t' \mathbf{A}^{-1} \mathbf{Q})^2 + o_p(1) \\ &= \delta + o_p(1). \end{aligned}$$

Odtud $\hat{\eta} \sim N(0, \sigma^2)$. □

Důkaz je s drobnými úpravami ve značení převzat z Tong (2003) a Keenan (1985).

Poznámka. Keenan v článku výslednou testovou statistiku neporovnává se zmíněným $F_{1, n-2M-2}$ rozdělením, které pro $n \rightarrow \infty$ konverguje k χ_1^2 , ale právě s χ_1^2 rozdělením. Tato konvergence je poměrně pomalá, a tak je dobré především pro krátké časové řady používat skutečně rozdělení $F_{1, n-2M-2}$. Důkaz konvergence je popsán v dodatku A1. Vzhledem k výkonu dnešních počítačů a přítomnosti F rozdělení v každém běžně užívaném statistickém softwaru není k používání χ^2 rozdělení důvod, a tak je například v testu `Keenan.test`, který je součástí balíku TSA pro software R, již naprogramováno použití F rozdělení.

3.1.1 Keenan.test

Funkce `Keenan.test`, která je implementována v balíku TSA (verze 1.01) obsahuje chybu, regrese v prvním a druhém kroku Keenanova testu je totiž implementována bez interceptu. To způsobuje špatný výpočet testové statistiky a zamítání nulové hypotézy linearity pro řady, jejichž střední hodnota není rovna nule. Čím je střední hodnota procesu od nuly vzdálenější, tím je hodnota testové statistiky větší a χ^2 test hypotézu zamítá. Dle simulací ale závislost není monotónní.

Uvažujme lineární časovou řadu generovanou autoregresním modelem

$$(X_t - \mu) = 0.5 * (X_{t-1} - \mu) + e_t, \quad e_t \sim N(0, 1).$$

Řada o délce 1000 pozorování byla vygenerována v R tisíckrát a vždy na ni byla aplikována původní (z balíčku TSA) a opravená verze Keenanova testu. V tabulce 3.1 nalezneme podíl případů, kdy test zamítl nulovou hypotézu na 5% hladině (jelikož řada lineární je, měl by podíl dosahovat přibližně 5 %).

Dle tabulky již pro malé odchýlení střední hodnoty od nuly test poměrně často hypotézu zamítá (náhodná složka použitá pro generování procesu má rozptyl roven 1). Pro hodnoty větší než 2 již původní test vždy nulovou hypotézu zamítl, a tak v tabulce nejsou uvedeny.

μ \ Test	Původní	Opravený
0	0.047	0.045
0.25	0.555	0.062
0.5	0.911	0.046
0.75	0.793	0.052
1	0.263	0.041
1.25	0.328	0.041
1.5	0.950	0.051
1.75	1.000	0.043
2	1.000	0.039

Tabulka 3.1: Četnost zamítnutí lineárního modelu na 5% hladině pro původní a opravenou verzi testu

V simulační části této práce je použit opravený test `Keenan.test.corrected`, jehož výsledky jsou v tabulce 3.1 ve sloupci „Opravený“. Balíček `TSA`, z něhož původní verze testu pochází, byl vytvořen jako doplněk ke knize Cryer a Chan (2008). V ní jsou naprogramované testy popsány a využívány. Autor byl na chybu upozorněn s kladnou odezvou, je tedy pravděpodobné, že v následující verzi bude test opraven i v repozitářích softwaru R. Samotný zdrojový kód původní i opravené funkce je uveden v dodatku B.

3.2 Tsayův test

V článku Tsay (1986) z roku 1986 je představena rozšířená verze Keenanova testu. Myšlenka je založena na rozdělení závisle proměnné z druhého kroku Keenanova testu na sčítance a testování jejich závislosti na reziduích z prvního kroku separátně. Platí totiž:

$$\hat{Y}_t = \mathbf{W}_t' \boldsymbol{\theta}, \quad (3.10)$$

kde $\mathbf{W}_t = (1, Y_{t-1}, \dots, Y_{t-M})'$ a $\boldsymbol{\theta} = (a_0, a_1, a_2, \dots, a_M)'$ je vektor konstant, které získáme odhadem metodou nejmenších čtverců v AR modelu. Potom

$$\begin{aligned} \hat{Y}_t^2 &= (\mathbf{W}_t' \boldsymbol{\theta})^2 = (a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_M Y_{t-M})^2 \\ &= a_0^2 + \sum_{i=1}^M a_0 a_i Y_{t-i} + \sum_{i,j=1}^M a_i a_j Y_{t-i} Y_{t-j}. \end{aligned} \quad (3.11)$$

Je zřejmé, že v (3.11) první sčítanec a první sumu tvoří právě člen \hat{Y}_t násobený konstantou a_0 , který tvoří lineární část časové řady. Veškerá případná nelinearita

v řadě se bude vykazovat korelaci členů poslední sumy s rezidui, která vzniknou „extrakcí“ AR procesu. Můžeme tedy zkoumat závislost jednotlivých členů této sumy na reziduích. Opět testujeme hypotézu H_0 : řada je lineární proti alternativě H_1 : řada obsahuje alespoň jeden kvadratický nelineární člen. Test se tedy provede následovně:

1. Regresí Y_t na $\{1, Y_{t-1}, \dots, Y_{t-M}\}$, kde M je předem zvolený řád autoregresního procesu, získáme rezidua $\{\hat{e}_t\}$. Z důvodu zahrnutí zpožděných hodnot regresorů do M -tého řádu provádíme regresi pro $t = M + 1, \dots, n$, kde n je délka časové řady.
2. Sestavíme vektor závisle proměnné \mathbf{Z}_t jako $\mathbf{Z}_t' = \text{vech}(\mathbf{U}_t' \mathbf{U}_t)$, kde $\mathbf{U}_t = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-M})$ pro $t = M + 1, \dots, n$. Symbol vech značí tzv. „poloviční vektorizaci“ matice, jde o poskládání prvků, které leží na nebo pod diagonálou do jednoho vektoru. Vektor \mathbf{Z}_t má tedy rozměr $m = \frac{1}{2}M(M+1)$. Nyní odhadneme matici parametrů \mathbf{H} mnohorozměrného regresního modelu

$$\mathbf{Z}_t = \mathbf{W}_t \mathbf{H} + \mathbf{X}_t.$$

3. Konečně odhadneme regresní model

$$\hat{e}_t = \hat{\mathbf{X}}_t' \boldsymbol{\beta} + e_t \quad (3.12)$$

a sestavíme testovou statistiku

$$\hat{F} = \frac{(\sum_{M+1}^n \hat{\mathbf{X}}_t' \hat{e}_t)(\sum_{M+1}^n \hat{\mathbf{X}}_t' \hat{\mathbf{X}}_t)^{-1}(\sum_{M+1}^n \hat{\mathbf{X}}_t' \hat{e}_t)/m}{\sum_{M+1}^n \hat{e}_t^2 / (n - M - m - 1)}, \quad (3.13)$$

kde \hat{e}_t jsou rezidua z OLS odhadu modelu (3.12).

Ta se v případě platnosti hypotézy H_0 řídí asymptoticky F rozdělením o $\frac{1}{2}M(M+1)$ a $n - \frac{1}{2}M(M+3) - 1$ stupních volnosti.

Tsayův test je dle knihy Tong (2003) silnější a vyžaduje více pozorování než Keenanův, který má pouze 1 stupeň volnosti při libovolné délce M testovaného AR procesu. Nabízí ale navíc kromě identifikace přítomnosti nelinearity i bližší určení, jaký kvadratický člen tuto nelinearitu způsobuje. Jednotlivé složky odhadnutého vektoru $\hat{\boldsymbol{\beta}}$ můžeme t -testem porovnávat s nulou a pozorovat signifikanci jednotlivých složek.

Věta 3.1 (Tsayův test). Nechť Y_t je stacionární autoregresní posloupnost řádu M splňující model

$$(Y_t - \mu) = \sum_{i=1}^M a_i (Y_{t-i} - \mu) + e_t,$$

kde e_t je bílý šum s konečným čtvrtým kumulantem². Pak výraz (3.13) asymptoticky (vzhledem k n) konverguje k rozdělení $F_{\frac{1}{2}M(M+1), n-\frac{1}{2}M(M+3)-1}$.

Důkaz. Při zachování značení z (3.10) konvergují OLS odhady $\hat{\boldsymbol{\theta}}$ skoro jistě k $\boldsymbol{\theta} = (a_0, a_1, \dots, a_M)'$, kde $a_0 = (1 - a_1 - \dots - a_M)\mu$. Podobně jako v (3.1) ukážeme, že náhodná veličina

$$n^{-\frac{1}{2}} \sum_{t=M+1}^n \mathbf{X}'_t e_t,$$

kde \mathbf{X}_t je definováno v druhém kroku testu, konverguje v distribuci k mnohorozměrné náhodné veličině. Je tomu tak proto, že $\mathbf{X}'_t e_t$ tvoří stacionární ergodickou posloupnost martingalových diferencí, neboť \mathbf{X}_t závisí pouze na zpožděných hodnotách $Y_{t-j}, j > 0$, tedy \mathbf{X}_t je nezávislé s e_t .

Pak podle standardních postupů analýzy rozptylu konverguje (3.13) k F rozdělení s $\frac{1}{2}M(M+1)$ a $n - \frac{1}{2}M(M+3) - 1$ stupni volnosti, protože $\frac{1}{2}M(M+1)$ odpovídá počtu omezení v nelineárním modelu, $n - \frac{1}{2}M(M+3) - 1$ počtu pozorování zmenšených o počet parametrů a matice $(\sum_{t=M+1}^n \hat{\mathbf{X}}'_t \hat{\mathbf{X}}_t)$ je limitou kovarianční matice $n^{-1} \sum_{t=M+1}^n \mathbf{X}'_t \mathbf{X}_t$. \square

Test je naprogramován v R v balíku TSA pod jménem `Tsay.test`, případně v knihovně `nlts` jako `lin.test`. Pro praktické použití je zřejmě lepší `Tsay.test`, neboť je implementován obecněji a není omezen maximálním řádem zahrnutých zpoždění. Proti tomu `lin.test` dovoluje řád nejvýše 5. Další výhodou implementace `Tsay.test` je, že v případě nespecifikovaného řádu testu jej sám doplní dle optimální volby standardní funkce `ar`.

3.3 Tsayův test na současnou nelinearitu

Jak vyplynulo z konstrukce Tsayova testu, můžeme jej použít k identifikaci nelinearity způsobené kvadratickými členy zpožděných hodnot časové řady. Co když je ale řada například tvaru: $Y_t = e_t e_{t-1} + e_t$? Pak test nedokáže dobře zachytit přítomnou nelinearitu, neboť testuje pouze kvadratické členy se zpožděnými hodnotami, tedy Y_{t-i} , kde $i \geq 1$. Pro odhalení nelinearity tohoto typu se tedy nabízí zahrnout do testu členy tvaru $Y_t Y_{t-i}$, kde $i \geq 1$.

Toho docílíme tak, že ve druhém kroku testu zkonstruujeme proměnnou $\hat{\mathbf{A}}_t = (Y_{t-1}\hat{e}_t, \dots, Y_{t-M}\hat{e}_t)$ a testovou statistiku definujeme jako

$$\hat{C} = \frac{(\sum_{t=M+1}^n \hat{\mathbf{R}}_t)(\sum_{t=M+1}^n \hat{\mathbf{R}}'_t \hat{\mathbf{R}}_t)^{-1}(\sum_{t=M+1}^n \hat{\mathbf{R}}'_t)/M}{(\sum_{t=M+1}^n \hat{e}_t)/(n-M-1)}$$

²Kumulant $\kappa_4 = \mu_4 - 3\mu_2^2$, kde μ_k je centrální moment, tedy $\mu_k = E(e_t - Ee_t)^k$.

kde $\hat{\mathbf{R}}_t = \hat{\mathbf{A}}_t \hat{e}_t - \mathbf{U}_t \sigma_e^2$ a $\mathbf{U}_t = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-M})$. Pak bude mít za platnosti hypotézy linearity asymptoticky F rozdělení s M a $n - M - 1$ stupni volnosti.

Důkaz. Podobně jako v předchozím testu plyne konvergence k F rozdělení z omezení v nelineárním modelu (a standardní analýzy rozptylu). Je tedy třeba ukázat asymptotickou normalitu statistiky

$$n^{-\frac{1}{2}} \sum_{t=M+1}^n \mathbf{R}'_t = n^{-\frac{1}{2}} \sum_{t=M+1}^n \{Y_{t-1}(e_t^2 - \sigma_e^2), \dots, Y_{t-M}(e_t^2 - \sigma_e^2)\},$$

kde \mathbf{R}_t je teoretická veličina odpovídající odhadu $\hat{\mathbf{R}}_t$.

Opět předpokládáme platnost autoregresního procesu jako v předchozím důkazu. Nechť $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, \dots, Y_{t-M}\}$ je σ -algebra nesoucí informaci o procesu do času $t-1$. Je zřejmé, že $\mathbf{E}(\mathbf{R}'_t | \mathcal{F}_{t-1}) = 0$. Každá složka vektoru \mathbf{R}_t tvoří stacionární ergodickou posloupnost martingalových diferencí s konstantním rozptylem $(\kappa - 1)\sigma_e^4\{\text{var}(Y_t) + \mu^2\}$, kde κ je čtvrtý kumulant e_t . Konvergence je tedy zajištěna martingalovou centrální limitní větou.

Jelikož proces $\{Y_{t-j}e_t^2$ pro $j = 1, \dots, M\}$ je stacionární a ergodický, pak $n^{-1} \sum_{t=M+1}^n \hat{\mathbf{R}}'_t \hat{\mathbf{R}}_t$ konverguje v pravděpodobnosti k varianční matici vektoru \mathbf{R}_t . □

3.4 RESET test

Název tohoto testu vznikl jako zkratka slov „Regression specification error test“, někdy je dle autora (Ramsey, 1968) též nazýván Ramseyův test. Tento test je v současnosti používán v ekonometrických softwarech (Eviews).

Test je založen opět na testování lineárního podmodelu pomocí analýzy rozptylu, tentokrát v modelu

$$Y_t = \mathbf{X}'_t \boldsymbol{\theta} + a_2 f_t^2 + \dots, a_k f_t^k + v_t \text{ pro nějaké } k \in \mathbb{N}, \quad (3.14)$$

kde $f_t = \mathbf{X}'_t \boldsymbol{\theta}$, v_t je náhodná složka a $\mathbf{X}_t = (Y_{t-1}, \dots, Y_{t-M})$ vektor zpožděných hodnot časové řady Y_t . Obecněji může jít o libovolný regresní model pro vysvětlování dané časové řady, odtud název testu. Model je tedy lineární, pokud $a_2 = \dots = a_k$. Test tedy provádíme pomocí testové statistiky

$$RESET = \frac{\hat{e}' \hat{e} - \hat{v}' \hat{v}}{\frac{\hat{v}' \hat{v}}{n-k}},$$

kde \hat{v} je vektor reziduí odhadnutých v modelu (3.14) a \hat{e} je vektor reziduí odhadnutých v lineárním podmodelu.

Představená statistika se dle standardních postupů analýzy rozptylu řídí za předpokladu normality rozdělením $F_{n-1, n-k}$ (Zvára, 2008, věta 3.1)³.

Mezi jednotlivými regresory v modelu (3.14) je ale zpravidla poměrně vysoká korelace. Proto se k odstranění multikolinearity užívá hlavních komponent vektoru (f_t^2, \dots, f_t^k) , případně pouze několika jejich prvních složek (výpočet hlavních komponent je popsán např. v Jolliffe (2002), kapitola 1).

RESET test je v R naprogramován v balíčku `lmtest` pod jménem `resettest` nebo totožně pod zkráceným jménem `reset`. Tento test je ale naprogramován obecně pro regresní modely a vyžaduje tak vstup ve tvaru vzorce regresního modelu. V souvislosti s testováním linearity v časových řadách jej popsali Lee a kol. (1993). Pro nulovou hypotézu autoregresního modelu řádu 1 tedy tento test provedeme jako:

```
reset(Y[-1] ~ Y[-n]),
```

kde Y je požadovaná časová řada. Tato verze testu přímo umožňuje pomocí parametru `power` zvolit vektor mocnin pro alternativní hypotézu (vektor $(1, \dots, k)$, obecně nemusí být souvislý) a pomocí parametru `type` zda má volit mocniny fitované řady, mocniny regresorů nebo hlavní komponenty spočítané z regresorů. Možnost užití hlavních komponent fitovaných hodnot ale bohužel chybí.

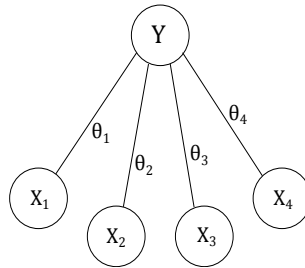
3.5 Testy založené na neuronových sítích

Další sada testů vychází s teorie neuronových sítí. Pravděpodobně první z nich byl popsán v článku Lee a kol. (1993), někdy je nazývaný jako Whiteův test. Ve stejném roce byl ale publikován článek Teräsvirta a kol. (1993), kde je blíže popsána síla Whiteova testu proti konkrétním alternativám a navíc je zde představena také jeho upravená verze (známá pod jménem Teräsvirta test). Před uvedením samotných testů stručně připomeňme princip neuronových sítí.

Neuronová síť je systém, který na základě vstupů (angl. inputs) vyhodnotí dle předem dané struktury sítě vstupní signály a určí výstup (angl. output). Jednoduchým příkladem může být model klasické lineární regrese s odhadnutými regresními koeficienty $\theta = \{\theta_1, \dots, \theta_n\}$. Za vstupy považujeme hodnoty regresorů, zkonstruováním jejich lineární kombinace (skalární součin s θ) získáme odhad střední hodnoty regresandu. Tato primitivní síť je znázorněna na obr. 3.1, odpo-

³Analýza rozptylu předpokládá jednotlivá pozorování nezávislá. Zde máme pozorování generovaná časovou řadou, a tak nelze nezávislost zajistit. Tento problém se v analýze časových řad řeší testováním autokorelace reziduí a závěry vycházející z teorie regrese se považují za platné.

vídačící model je tedy $EY = \mathbf{X}'\boldsymbol{\theta}$ se čtyřmi regresory X_1, \dots, X_4 .

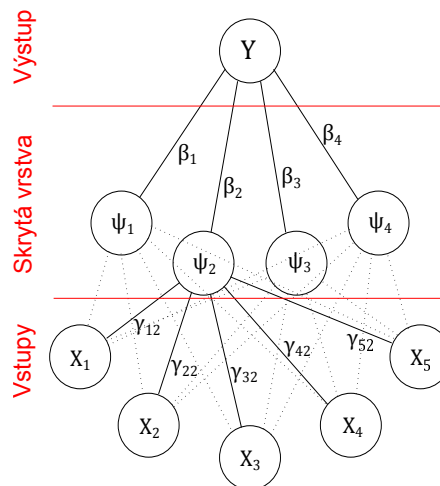


Obrázek 3.1: Jednoduchý regresní model

Neuronové sítě myšlenku zpracování vstupů zobecňují přidáním tzv. skrytých vrstev (angl. hidden layers), které signál ze vstupu agregují na několika meziúrovních a výsledkem je pak výstup získaný lineární kombinací poslední úrovně. Pro účely testování linearity si vystačíme s neuronovou sítí s jednou skrytou vrstvou. Vstupy X_i předávají signál do skryté vrstvy skrze konstanty γ_{ij} , j -tý prvek ve skryté vrstvě se vyjádří jako $\mathbf{X}'\boldsymbol{\gamma}_j$. Hodnoty skryté vrstvy jsou zpravidla transformovány tzv. aktivační funkcí ψ a společně pak tvoří výstup Y , tedy

$$Y = \sum_j \beta_j \psi(\mathbf{X}'\boldsymbol{\gamma}_j),$$

kde $\psi(\lambda)$ je zmíněná aktivační funkce. Názorně můžeme popsanou síť vidět na obrázku 3.2, kde jsme označili $\psi_j = \psi(\mathbf{X}'\boldsymbol{\gamma}_j)$ a $\boldsymbol{\gamma}_j = (\gamma_{1j}, \dots, \gamma_{5j})$ (uvažujeme pět vstupů).



Obrázek 3.2: Neuronová síť s jednou skrytou vrstvou

Popsaný model je zřejmě přeparametrizován, na základě vstupů \mathbf{X}_i tak není možné určit jednoznačné odhady koeficientů β_j i γ_{ij} . Samotný odhad probíhá tak, že se koeficienty γ_{ij} určí předem, aktivační funkce ψ je též dána předem a na základě dat se již odhadují pouze koeficienty β_j .

3.5.1 Whiteův test

White navrhl v roce 1993 test, který uvažuje právě popsanou neuronovou síť s jednou skrytou vrstvou. Za funkci ψ doporučuje volit logistickou funkci, tedy funkci $\psi(\lambda) = \frac{1}{1+e^{-\lambda}}$, která dle autora dokáže případnou nelinearitu dobře zachytit.

Koeficienty γ_{ij} doporučuje volit náhodně, konkrétně z rovnoměrného rozdělení na intervalu $[0, 2]$, přičemž regresory X_i doporučuje škálovat na interval $[0, 1]$.

Vycházíme z toho, že pokud je řada lineární, pak by ve výrazu

$$o = \mathbf{X}_t \boldsymbol{\theta} + \sum_{j=1}^q \beta_j \psi(\mathbf{X}' \boldsymbol{\gamma}_j)$$

měly být koeficienty β_j rovny nule. Samotný test pro testování linearity časové řady probíhá následovně:

1. Odhadneme autoregresní model $Y_t = \theta_0 + \sum_{k=1}^M \theta_k Y_{t-k}$ pro $t = 1, \dots, n - M$. Konstanta M je předem zvolený řád autoregresního modelu (můžeme jej určit např. dle Akaikeho informačního kritéria) a spočítáme vektor reziduí $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_{n-M})$.
2. Vygenerujeme náhodně koeficienty γ_{ij} pro $i = 0, \dots, M$, a $j = 1, \dots, q$, kde q je předem určený počet prvků ve skryté vrstvě neuronové sítě.
3. Pro $t = M + 1, \dots, n$ určíme pomocný vektor $\boldsymbol{\psi}_t$, tedy hodnoty ψ_{t1} až ψ_{tq} , kde $\psi_{tj} = \psi(\tilde{\mathbf{Y}}_t' \boldsymbol{\gamma}_j)$, $\tilde{\mathbf{Y}}_t = (1, Y_{t-1}, \dots, Y_{t-M})$ a $\boldsymbol{\gamma}_j = (\gamma_{1j}, \dots, \gamma_{Mj})$.
4. Odhadneme regresní model $\hat{u}_t = \theta_0 + \sum_{k=1}^M \theta_k Y_{t-k} + \boldsymbol{\psi}_t' \boldsymbol{\beta}$. Sestavíme testovou statistiku

$$F = \frac{(SSR_0 - SSR)/q}{SSR/(n - M - 1 - q)},$$

kde SSR_0 je reziduální součet čtverců modelu z prvního kroku a SSR je reziduální součet čtverců modelu z posledního kroku. Ta se za předpokladu nulové hypotézy linearity řídí rozdělením $F_{q, n-M-1-q}$.

Prakticky jde o test lineárního podmodelu v nelineárním modelu. Ten lze provést i metodou maximální věrohodnosti, kdy testová statistika

$$C = n * \log \left(\frac{SSR_0}{SSR} \right)$$

má za platnosti nulové hypotézy rozdělení $\chi^2(q)$.

V článku Lee a kol. (1993) se doporučuje volit $q = 10$. Vzhledem k velké korelaci proměnných ψ_1, \dots, ψ_q navrhuje počítat hlavní komponenty a pro samotný regresní model použít první dvě.

V softwarovém prostředí R je naprogramován pod názvem `white.test` v balíku `tseries` nebo také v balíku `fNonlinear` pod jménem `wnnTest`. Obě verze jsou téměř totožné, test z balíku `tseries` nabízí možnost volby mezi Waldovým testem a testem poměrem věrohodnosti ⁴ (parametr `type`, který se nastaví buď na `F`, nebo `Chi`), test v balíku `fNonlinear` vypisuje vždy obě verze.

Podstatný rozdíl v testech je ale v tom, jak přistupují ke škálování vstupů do neuronové sítě. White doporučuje škálovat proměnné na interval $[0, 1]$, toho se ani jeden z testů nedrží. Test z balíku `fNonlinear` neškáluje data vůbec, test z balíku `tseries` umožňuje volbu zapnutí nebo vypnutí škálování, přičemž implicitně škáluje dle defaultního nastavení funkce `scale`, tedy tak, aby měla data nulový výběrový průměr a jednotkový výběrový rozptyl. V nápovědě k testům se ale bohužel autoři o odlišném naprogramování, než uvádí článek Lee a kol. (1993), nezmiňují.

Pro časové řady užívané jako modelové v mnoha člancích nemá tento rozdíl přílišný vliv (jelikož řady jsou zpravidla centrovány a rozptyl není příliš vzdálen od jedné). Pro praktické použití může ale vést k odlišným závěrům. Navíc pokud data neškálujeme, nemá u obou testů parametr `range`, který určuje délku intervalu pro generování konstant γ_{ij} , smysl, jelikož součin $\mathbf{X}'\boldsymbol{\gamma}_j$ bude záviset na měřítku regresoru \mathbf{X} .

3.5.2 Teräsvirtaův test

V článku Teräsvirta a kol. (1993) se ukazuje, že síla Whiteova testu není příliš velká z důvodu náhodné volby parametrů ve skryté vrstvě. Logistickou funkci $\psi(\lambda)$, kterou volil White, studuje Teräsvirta blíže rozvojem do Taylorova polynomu. Ukazuje, že pro vhodné zachycení nelinearity stačí několik prvních členů Taylorova rozvoje této funkce. Konkrétně pro sestavení nelineárního modelu užívá jako regresory kvadratické členy řady, tedy $Y_{t-i}Y_{t-j}$, pro $i, j = 1, \dots, M$, kubické členy $Y_{t-i}Y_{t-j}Y_{t-k}$ pro $i, j, k = 1, \dots, M$ a obdobně členy čtvrtého řádu. Jako maximální zpoždění M volí řád dva. Jednotlivým regresorům, které mají zachytit případnou nelinearitu řady, říká *pomocné regresory* (angl. auxiliary regressors).

V článku je uvedeno několik modelů, v nichž se užívá různých kombinací těchto regresorů, následně se pak testuje lineární podmodel. Mimo jiné se zaměřuje také na zahrnutí interceptu uvnitř logistické funkce. Ten White automaticky zahrnul vždy, Teräsvirta ale ukazuje, že na některé nelineární modely je vhodnější volit argument logistické funkce bez interceptu. Výhodou přístupu zahrnování

⁴Popis testu poměrem věrohodnosti lze najít např. v knize Anděl (2007), str. 177, test označený jako LR (likelihood ratio).

pomocných regresorů namísto celé nelineární funkce je, že dokáží simulovat logistickou funkci jak bez interceptu, tak s ním, pouze záleží na tom, jaké pomocné regresory zvolíme.

Konkrétní modely, které Teräsvirta představuje, shrnuje tabulka 3.2. Model V2 odpovídá Tsayově testu, jelikož zahrnuje všechny kvadratické členy. Model *RES-M* je modifikovaný *RESET* test. Od standardního *RESET* testu z kapitoly 3.4 se liší tím, že jako pomocné regresory vkládá mocniny zpožděných hodnot zvláště, ne pouze jejich lineární kombinaci určenou koeficienty modelu AR(2).

Test	pomocné regresory									
	Y_{t-1}^2	$Y_{t-1}Y_{t-2}$	Y_{t-2}^2	Y_{t-1}^3	$Y_{t-1}^2Y_{t-2}$	$Y_{t-1}Y_{t-2}^2$	Y_{t-2}^3	Y_{t-1}^4	$Y_{t-1}^3Y_{t-2}$	Y_{t-2}^4
LSTAR4	×	×		×	×			×	×	
LSTAR2	×	×								
V2	×	×	×							
ESTAR				×	×					
V23	×	×	×	×	×	×	×			
V3				×	×	×	×			
RES-M	×		×	×			×	×		×

Tabulka 3.2: Tabulka testů představených v Teräsvirta a kol. (1993)

Teräsvirtaův test je v prostředí R naprogramován opět v balících `tseries` a `fNonlinear` pod jménem `terasvirta.test`, respektive `tnnTest`. V obou případech se jedná o test proti alternativě nelinearity určené nelineárním modelem V23. Tato informace ale bohužel není uvedena v nápovědě prostředí R ani u jednoho z testů, a tak může být pro uživatele matoucí, jaký test z článku Teräsvirta a kol. (1993) se vlastně používá.

3.6 Skórové testy

Zřejmě první testy založené na Fisherově informační matici normálního rozdělení byly popsány v článku White (1996), ve zkrácené verzi jsou uvedeny v článku Lee a kol. (1993):

Pro normální lineární model platí

$$Y_t = \mathbf{X}_t' \boldsymbol{\theta} + e_t, \text{ kde } e_t \sim N(0, \sigma^2), \quad (3.15)$$

hustota e_t lze tedy vyjádřit jako

$$f_t(\mathbf{x}_t, \boldsymbol{\theta}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_t - \mathbf{X}_t' \boldsymbol{\theta})^2}{2\sigma^2}\right)$$

a logaritmická věrohodnost je tedy

$$\log f_t(\mathbf{x}_t, \boldsymbol{\theta}, \sigma) = -\frac{1}{2} \log(2\pi) - \log \sigma - \frac{(Y_t - \mathbf{X}_t' \boldsymbol{\theta})^2}{2\sigma^2}.$$

Označením $u_t = (Y_t - \mathbf{X}_t' \boldsymbol{\theta})/\sigma$ dostaneme podmíněnou skórovou funkci

$$\mathbf{s}_t(\mathbf{X}_t, \boldsymbol{\theta}, \sigma) = \nabla \log f_t(\mathbf{x}_t, \boldsymbol{\theta}, \sigma) = \sigma^{-1}(u_t, u_t \mathbf{X}_t', u_t^2 - 1),$$

kde ∇ je gradient vzhledem k vektoru $(\boldsymbol{\theta}, \sigma)$. Dále označíme $\mathbf{s}_t^* = \mathbf{s}_t(\mathbf{X}_t, \boldsymbol{\theta}^*, \sigma^*)$, kde $\boldsymbol{\theta}^*$ a σ^* značí skutečné hodnoty parametrů $\boldsymbol{\theta}$ a σ . Pak za předpokladu linearit (správné specifikace (3.15)) bude platit $E(\mathbf{s}_t^*) = 0$ a $E(\mathbf{s}_t^* \mathbf{s}_{t-\tau}^{*\prime}) = 0$ pro $\tau = 1, 2, \dots, t$.

Na základě těchto vztahů lze odvodit několik testů, kde figuruje Fisherova informační matice. Označme ještě $\mathbf{m}_t = \mathbf{S} \text{vec}(\mathbf{s}_t \mathbf{s}_{t-1}')$, kde \mathbf{S} je deterministická selekční matice zaměřená na nějaké konkrétní odchylení od lineární specifikace modelu (vybírání některých konkrétních složek $\text{vec}(\mathbf{s}_t \mathbf{s}_{t-1}')$). Symbol vec značí vektorizaci, tedy rozložení matice do sloupcového vektoru, po sloupcích. Nakonec ještě označme $\hat{\boldsymbol{\theta}}$ a $\hat{\sigma}$ odhady modelu metodou maximalizace kvazivěrohodnosti (QMLE) a jim odpovídající $\hat{\mathbf{s}}_t$ a $\hat{\mathbf{m}}_t$. Testová statistika je pak ve tvaru

$$White_1 = n \hat{\boldsymbol{\mu}}_n' \hat{\mathbf{J}}_n^{-1} \hat{\boldsymbol{\mu}}_n,$$

kde

$$\hat{\mathbf{J}}_n = n^{-1} \sum \hat{\mathbf{m}}_t \hat{\mathbf{m}}_t' - \left(n^{-1} \sum \hat{\mathbf{m}}_t \hat{\mathbf{s}}_t' \right) \left[n^{-1} \sum \hat{\mathbf{s}}_t \hat{\mathbf{s}}_t' \right]^{-1} \left(n^{-1} \sum \hat{\mathbf{s}}_t \hat{\mathbf{m}}_t' \right)$$

a $\hat{\boldsymbol{\mu}} = n^{-1} \sum_{t=1}^n \hat{\mathbf{m}}_t$. Dalším testem je

$$White_2 = nR^2,$$

kde R^2 je koeficient determinace z regrese konstantní $\mathbf{1}$ na vysvětlujících proměnných $\hat{\mathbf{m}}_t$ a $\hat{\mathbf{s}}_t$. Obě uvedené testové statistiky se za platnosti nulové hypotézy řídí asymptoticky rozdělením $\chi^2(q)$, kde q je rozměr vektoru \mathbf{m}_t .

Konkrétní volba selekční matice \mathbf{S} , která byla v simulační studii použita, ale článek Lee a kol. (1993) nespecifikuje. Skórové testy se používají i v dalších modifikacích, např. v testech využívajících teorii náhodných polí.

3.7 Testy založené na náhodných polích

V článku Hamilton (2001) je uveden test nelinearity založený na teorii náhodných polí (dále v této práci označen jako Hamiltonův test). O dva roky později byl tento test rozšířen (Dahl a González-Rivera, 2003).

Před samotným představením testu uvedme nejprve podstatu gaussovského náhodného pole. Náhodné pole je funkce $m(\omega, \mathbf{x}) : \Omega \times A \rightarrow \mathbb{R}$ pro každé $\mathbf{x} \in A$, kde $A \in \mathbb{R}^k$, někdy ji značíme pouze jako $m(\mathbf{x})$. Pokud je $m(\mathbf{x})$ systém náhodných veličin s konečněrozměrným normálním rozdělením, pak říkáme, že jde o gaussovské náhodné pole. To je jednoznačně určeno střední hodnotou a kovarianční funkcí, tedy funkcemi $\mu(\mathbf{x}) = \mathbf{E}[m(\mathbf{x})]$ a $\mathbf{C}(\mathbf{x}, \mathbf{z}) = \mathbf{E}[(m(\mathbf{x}) - \mu(\mathbf{x}))(m(\mathbf{z}) - \mu(\mathbf{z}))]$, pro $\mathbf{x}, \mathbf{z} \in A$. Náhodné pole je homogenní (nebo také stacionární), pokud $\mu(\mathbf{x}) = \mu$ a $\mathbf{C}(\mathbf{x}, \mathbf{z})$ je funkcí pouze rozdílu $\mathbf{x} - \mathbf{z}$. Pak také funkce \mathbf{C} můžeme značit jako funkci pouze jedné proměnné, a sice $\mathbf{C}(\mathbf{x} - \mathbf{z})$. Pokud je kovariance navíc závislá pouze na vzdálenosti vektorů \mathbf{x} a \mathbf{z} , tedy $\mathbf{C}(\mathbf{x}, \mathbf{z}) = \mathbf{C}(d(\mathbf{x}, \mathbf{z}))$, kde $d(\cdot, \cdot)$ je skalární funkce, pak říkáme, že náhodné pole je *izotropické*.

Kovarianční funkce \mathbf{C} náhodného pole je podstatou testování linearity. Test předpokládá, že v případě nelinearity spolu budou blízká pozorování *interferovat* (blízká ve smyslu vzdálenosti v náhodném poli, v kontextu autoregresních modelů tedy hodnoty časové řady v takových časových okamžicích, kdy zpožděné hodnoty nabývají blízkých hodnot v euklidovské metrice). Čím blíže sobě prvky budou, tím bude jejich korelace vyšší. Odhad kovarianční matice \mathbf{C} závisí na dimenzi k prostoru náhodného pole, dále ji tedy budeme značit jako \mathbf{C}_k .

Testy jsou konstruovány jako skórové testy (viz Anděl (2007), str. 177) nulovosti parametrů λ , resp. \mathbf{g} v modelu

$$Y_t = \mathbf{X}_t' \boldsymbol{\beta} + \lambda m(\mathbf{g} \odot \mathbf{X}_t) + e_t, \quad (3.16)$$

kde e_t je normální bílý šum s rozptylem σ^2 , $Y_t \in \mathbb{R}$ a $\mathbf{X}_t \in \mathbb{R}^k$, obě posloupnosti jsou stacionární a ergodické. Symbol \odot značí násobení vektorů po složkách a $m(\mathbf{z})$ je gaussovské homogenní náhodné pole s nulovou střední hodnotou. Vektor \mathbf{X}_t zastupuje buď exogenní proces nebo může jít o zpožděné hodnoty procesu Y_t . Pro test linearity budeme obdobně jako v předchozích testech za \mathbf{X}_t uvažovat zpožděné hodnoty Y_t a testovat tak linearitu ve smyslu AR procesu. Lineární část procesu je tedy zastoupena ve výrazu $\mathbf{X}_t' \boldsymbol{\beta}$ a případná nelinearita je zachycena ve výrazu $\lambda m(\mathbf{g} \odot \mathbf{X}_t)$.

Aby byl model identifikovaný, je nutné fixovat buď parametr λ , nebo parametr \mathbf{g} . Nulová hypotéza linearity časové řady pak platí, pokud druhý z parametrů je roven nule.

3.7.1 Hamiltonův test

Hamilton navrhl test tak, že vektor \mathbf{g} zafixujeme a nulovou hypotézu linearity pak formulujeme jako $H_0 : \lambda^2 = 0$. Spočítáme tedy skórový vektor pro testované parametry a odpovídající Fisherovu informační matici za platnosti nulové hypotézy H_0 . Označíme $\mathbf{Y} = (Y_1, \dots, Y_n)$ a \mathbf{C}_k kovarianční matici odpovídající náhodnému poli $m(\mathbf{z})$. Předpokládáme-li platnost (3.16), pak $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \lambda^2\mathbf{C}_k + \sigma^2\mathbf{I}_n)$ a logaritmická věrohodnostní funkce má tvar

$$l(\lambda^2, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\lambda^2\mathbf{C}_k + \sigma^2\mathbf{I}_n| - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\lambda^2\mathbf{C}_k + \sigma^2\mathbf{I}_n)^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Její derivováním a položením $\lambda^2 = 0$ dostaneme pro parametr λ^2 skóre

$$s(\lambda^2)|_{\lambda^2=0} = \frac{\partial l(\boldsymbol{\beta}, \lambda^2, \sigma^2)}{\partial \lambda^2} = -\frac{1}{2\sigma^4} [\sigma^2 \text{tr}(\mathbf{C}_k) - \mathbf{e}'\mathbf{C}_k\mathbf{e}].$$

Pro parametrický vektor $\boldsymbol{\zeta} = (\boldsymbol{\beta}, \lambda^2, \sigma^2)$ za platnosti nulové hypotézy $\boldsymbol{\zeta}_0 = (0, \boldsymbol{\beta}_0, \sigma_0^2)$ dostaneme Fisherovu informační matici

$$\mathbf{J}(\boldsymbol{\zeta}) = -\mathbf{E} \left(\frac{\partial^2 l(\boldsymbol{\beta}, \lambda^2, \sigma^2)}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}'} \Big|_{\boldsymbol{\zeta}=\boldsymbol{\zeta}_0} \right) = \begin{bmatrix} \frac{\text{tr}(\mathbf{C}_k^2)}{2\sigma^4} & \frac{\text{tr}(\mathbf{C}_k)}{2\sigma^4} & \mathbf{0}' \\ \frac{\text{tr}(\mathbf{C}_k)}{2\sigma^4} & \frac{n}{2\sigma^4} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \end{bmatrix}.$$

Prvek v levém horním rohu matice $\mathbf{J}(\boldsymbol{\zeta})^{-1}$ je roven

$$\frac{(2\sigma^4)^{-1} * n}{(2\sigma^4)^{-2} [n * \text{tr}(\mathbf{C}_k^2) - \text{tr}(\mathbf{C}_k)^2]} = \frac{2\sigma^4}{\text{tr}(\mathbf{C}_k^2) - n^{-1} \text{tr}(\mathbf{C}_k)^2}.$$

Testová statistika skórového testu je tedy

$$Hamilton_n = \frac{[\hat{\mathbf{e}}\mathbf{C}_k\hat{\mathbf{e}} - \hat{\sigma}_n^2 \text{tr}(\mathbf{C}_k)]^2}{2\hat{\sigma}_n^4 [\text{tr}(\mathbf{C}_k^2) - n^{-1} \text{tr}(\mathbf{C}_k)^2]},$$

kde $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ a $\hat{\sigma}_n^2 = n^{-1}\hat{\mathbf{e}}'\hat{\mathbf{e}}$.

Tato statistika se řídí za platnosti H_0 asymptoticky rozdělením $\chi^2(1)$, neboť ve skórovém testu máme právě jedno omezení (na λ^2).

Pro konstrukci testu zbývá uvést, jak spočítat odhad kovarianční matice \mathbf{C}_k . Hamilton navrhuje v náhodném poli fixovat vektor \mathbf{g} jako

$$g_i = \frac{2}{\sqrt{k}s_i^2},$$

kde s_i je směrodatná odchylka i -té složky \mathbf{X}_t a k je počet složek (bez interceptu). Tato volba vychází z teoretických vlastností lognormálního rozdělení (Hamilton,

2001, str. 553). Dále využijeme vlastností gaussovského izotropního náhodného pole. V něm totiž platí, že pokud je vzdálenost prvků $\mathbf{x} \in A$ a $\mathbf{z} \in A$ větší než 1, pak jsou nekorelované, pokud je rovna 0, pak je jejich korelace rovna 1 a mezi hodnotami 0 a 1 je kovariance rovna objemu průniku k -rozměrných jednotkových koulí opsaných prvky \mathbf{x} a \mathbf{z} . Analytický vzorec ale závisí na dimenzi k . Lze jej napočítat rekurentně, pro $k = 1, \dots, 5$ je roven:

- $H_1(h) = 1 - h$,
- $H_2(h) = 1 - (2/\pi)[h(1 - h^2)^{1/2} + \sin^{-1}(h)]$,
- $H_3(h) = 1 - (3h/2) + (h^3/2)$,
- $H_4(h) = 1 - (2/\pi)[(2/3)h(1 - h^2)^{3/2} + h(1 - h^2)^{1/2} + \sin^{-1}(h)]$,
- $H_5(h) = 1 - (3h/2)h + (h^3/2) - (3h/8)(1 - h^2)^2$,

kde h je polovina vzdálenosti prvků \mathbf{x} a \mathbf{z} . Jinými slovy, z uvedených vztahů dostaneme číslo

$$H_k(h) = \text{cov}(m(\mathbf{x}), m(\mathbf{z})) / [(\mathbf{x} - \mathbf{z})'(\mathbf{x} - \mathbf{z})]^{1/2} = 2h).$$

Samotný prvek na pozici (t, s) v rozptylové matici \mathbf{C}_k pak určíme jako

$$\mathbf{C}_k^{(t,s)} = H_k \left(\frac{1}{2} [g_1^2(x_{1t} - x_{1s})^2 + g_2^2(x_{2t} - x_{2s})^2 + \dots + g_k^2(x_{kt} - x_{ks})^2]^{1/2} \right),$$

kde x_{ij} je i -tá složka vektoru regresorů v čase j , tedy vektoru \mathbf{X}_j .

V závěru představení testu uvádí Hamilton ještě penalizaci pro krátké časové řady, zde ji uvedeme bez odvození. Výsledkem je statistika

$$\nu^2 = \frac{[\hat{\mathbf{e}}' \mathbf{H} \hat{\mathbf{e}} - \tilde{\sigma}^2 \text{tr}(\mathbf{M} \mathbf{H} \mathbf{M})]^2}{\tilde{\sigma}^4 (2 \text{tr}\{[\mathbf{M} \mathbf{H} \mathbf{M} - (n - k - 1)^{-1} \mathbf{M} \text{tr}(\mathbf{M} \mathbf{H} \mathbf{M})]^2\})},$$

kde \mathbf{M}_n je projekční matice do prostoru reziduí, tedy $\mathbf{M}_n = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Tato statistika se za platnosti nulové hypotézy řídí také asymptoticky rozdělením $\chi^2(1)$.

Test byl pro účely simulační části naprogramován v prostředí R ve verzi s úpravou pro krátké časové řady.

3.7.2 Rozšíření Hamiltonova testu

Testy založené na teorii náhodných polí byly dále rozvedeny v článku Dahl a González-Rivera (2003). Vycházejí taktéž z modelu (3.16) a jako nulovou hypotézu linearity formulují jednak $H_0 : \lambda^2 = 0$, ale i $H'_0 : \mathbf{g} = \mathbf{0}$. Jednotlivé verze testů jsou pak označeny jako λ -testy, resp. \mathbf{g} -testy. Náhodné pole je zobecněno a pro odhad kovarianční matice jsou použita i neizotropická náhodná pole. Nově popsané testy jsou simulačně srovnány s Hamiltonovým a Tsayovým testem. Ukazuje se, že při aplikaci na většinu jednoduchých nelineárních modelů mají větší sílu nové testy založené na náhodných polích. Pro složitější nelineární modely je ale v některých případech Tsayův test silnější. V prostředí R tyto testy naprogramovány nejsou.

3.8 Testy bez specifické alternativy

Doposud uvedené testy byly založeny na podobné myšlence, a sice na testování lineárního podmodelu v nelineárním modelu. Všechny tedy předpokládaly konkrétní tvar nelinearity v alternativní hypotéze, kterou byly testy schopné detekovat. Následující testy nepředpokládají specifický nelineární model, vycházejí z faktů, které pro lineární modely platí, a ty testují.

3.8.1 BDS test

Tento test není čistě testem nelinearity, přesto se ale v některých člancích (Lee a kol., 1993) nebo knihách (Zivot a Wang, 2007, str. 654) pro testování nelinearity modelu doporučuje. Jedná se o neparametrický test, jehož nulová hypotéza je, že časová řada X_t je nezávislá a stejně rozdělená. Pro „ověření linearity“ test aplikujeme na časovou řadu reziduí, které získáme extrakcí autoregresního modelu z dané časové řady. Jejich nezávislost a stejné rozdělení jsou postačující podmínky linearity procesu (v takovém případě AR model dobře popisuje danou časovou řadu). Název dostal podle iniciálů autorů, kteří test publikovali (Broock a kol., 1996). Zkráceně je popsán například v Lee a kol. (1993).

Pro časovou řadu X_t definujeme

$$C_m(\epsilon) = n^{-2} * [\text{počet dvojic } i \text{ a } j, \text{ pro které platí}$$

$$|X_i - X_j| < \epsilon, |X_{i+1} - X_{j+1}| < \epsilon, \dots, |X_{i+m-1} - X_{j+m-1}| < \epsilon],$$

tedy X_i, \dots, X_{i+m-1} a X_j, \dots, X_{j+m-1} jsou dva úseky časové řady délky m takové, že hodnoty řady se v úsecích neliší o více než ϵ . Testová statistika se sestaví

jako $BDS = n^{1/2}[C_m(\epsilon) - C_1(\epsilon)^m]$. Za předpokladu nulové hypotézy, že je časová řada nezávislá a stejně rozdělená, se statistika BDS řídí asymptoticky normálním rozdělením s nulovou střední hodnotou a známým, komplikovaným rozptylem, a to pro každou kombinaci m a ϵ . Ve statistických programech pak test probíhá vypočtením několika BDS statistik pro kombinace hodnot m a ϵ a výpisem p -hodnot pro každou z kombinací.

Tento test je v R naprogramován v balíku `tseries` pod jménem `bds.test` a také v balíku `fNonlinear` pod jménem `bdsTest`. Oba testy dávají stejné výsledky, test z balíku `tseries` vypisuje přehlednější tabulku.

3.8.2 Korelační test

Dle autorů bývá tento test někdy označován jako McLeod-Li test (McLeod a Li, 1983). Vychází z předpokladu, že pokud je náhodný proces $\{Y_t\}$ stacionární a gaussovský (jednotlivé veličiny Y_t se řídí normálním rozdělením), pak platí

$$\rho_\tau(Y_t^2) = \{\rho_\tau(Y_t)\}^2 \text{ pro všechna } \tau,$$

kde

$$\rho_\tau(X_t) = \text{corr}(X_{t+\tau}, X_t).$$

Autoři navrhli portmonteau test založený na tomto faktu, tedy nechť $\{\hat{\epsilon}_1, \dots, \hat{\epsilon}_N\}$ jsou rezidua časové řady po fitování ARMA modelu. Nechť r_k značí výběrovou autokorelaci čtveřců reziduí, tedy

$$r_k = \sum_{j=1}^{n-k} \frac{(\hat{\epsilon}_j^2 - \hat{\sigma}^2)(\hat{\epsilon}_{j+k}^2 - \hat{\sigma}^2)}{\sum_{j=1}^n (\hat{\epsilon}_j^2 - \hat{\sigma}^2)^2},$$

kde

$$\hat{\sigma}^2 = \sum_{j=1}^n \hat{\epsilon}_j^2 / n.$$

Analogicky k Box-Ljungově portmonteau testu navrhli McLeod a Li testovou statistiku

$$Q = n(n+2) \sum_{k=1}^m r_k^2 / (n-k),$$

kde m je pevná předem zvolená konstanta (parametr testu). Testová statistika se za předpokladu nezávislých reziduí řídí asymptoticky rozdělením χ_m^2 .

Test detekuje nejen případnou nelinearitu, ale i případnou špatnou specifikaci ARMA, proto nemůžeme zamítnutí nulové hypotézy ihned považovat za nelinearitu.

Tento test je naprogramován v R pod jménem `McLeod.Li.test` v knihovně `TSA`. Jako vstupní parametr m (pokud není předem zadáno uživatelem) volí $10 \cdot \log_{10} n$, kde n je délka časové řady.

3.9 Další testy linearity

3.9.1 Testy ve spektrální doméně

Mimo testů v časové doméně, které shrnuje předchozí kapitola, byly popsány i testy ve spektrální doméně. Zřejmě nejstarší byl prezentován v článku Rao a Gabr (1980) a v literatuře je často nazýván jako *bispektrální test*. Tong (2003) jej hodnotí jako slabý a překonaný Keenanovým a Tsayovým testem.

Bispektrální test je založen na vztazích mezi spektrální a bispektrální hustotou lineárního procesu. Spektrální funkce náhodného procesu Y_t je definována jako

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s)e^{-is\omega}, \quad |\omega| \leq \pi,$$

kde $R(s)$ je autokovarianční funkce procesu Y_t , bispektrální funkce pak jako

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} c(t_1, t_2)e^{-it_1\omega_1}e^{-it_2\omega_2}, \quad |\omega_1| \leq \pi, |\omega_2| \leq \pi,$$

kde $c(t_1, t_2) = \mathbf{E}[(Y_t - \mu)(Y_{t+t_1} - \mu)(Y_{t+t_2} - \mu)]$.

Autoři v článku ukazují, že pokud je proces Y_t lineární, pak pro uvedené hustoty platí

$$\frac{|f(\omega_i\omega_j)|^2}{f(\omega_i)f(\omega_j)f(\omega_i + \omega_j)} = \frac{\mu_3^2}{2\pi\sigma_e^2},$$

pro všechna i, j . Pokud je navíc proces gaussovský, pak je výraz roven nule, neboť šikmost μ_3 je rovna nule.

Tato vlastnost je dále rozvedena v článku Berg a kol. (2010), kde se autoři zaměřují jednak na možné odhady spektrální a bispektrální hustoty, ale také na aplikaci metody bootstrap na testování.

3.9.2 Testy na konkrétní nelineární modely

Závěrem zbývá dodat, že vzhledem k širokému spektru nelineárních modelů časových řad byly představeny i testy zaměřující se na konkrétní nelineární model. Pro model prahové autoregrese (TAR), který je v literatuře často diskutován, bylo představeno několik testů. Popsali je např. Tong, Chan, Gospodinov a další.

Dále byly popsány testy pro model STAR, SETAR a jiné. Vzhledem k rozsahu práce se těmito testům nebudeme blíže věnovat.

3.10 Vyjádření modelu Taylorovým rozvojem

Jak bylo řečeno v kapitole 3.5.2, na alternativní hypotézu nelineárního modelu je možné nahlížet lokálně, rozvojem do Taylorova polynomu. Pokud je tedy daná nelineární časová řada vygenerována modelem tvořeným diferencovatelnými funkcemi, lze takovou řadu lokálně vyjádřit ve tvaru (2.2). Blíže teorii popisuje článek Teräsvirta a kol. (1993), my se zaměříme na možné aplikace této myšlenky.

Uvažujme časovou řadu tvaru

$$Y_t = b * f(Y_{t-1}) + e_t, \quad (3.17)$$

kde e_t je bílý šum, b je koeficient menší než jedna a $f(x)$ je omezená funkce nekonečně diferencovatelná na nějakém okolí nuly. Pak můžeme využít aproximace takové časové řady Taylorovým polynomem. Dle tvaru rozvoje je možné určit, které pomocné regresory budou v modelu popsáném Teräsvirtaou významné při testování konkrétních alternativ nelineárního modelu.

Pokud tedy za funkci $f(x)$ budeme volit takové funkce, jejichž Taylorův rozvoj sestává pouze z lichých mocnin (např. $\sin(x)$), pak by neměly být takové časové řady snadno detekovatelné Keenanovým nebo Tsayovým testem, neboť tyto testy jsou založené na pomocných regresorech ve druhé mocnině. Naopak tyto testy by měly být schopné detekovat nelinearitu v modelu, kde za $f(x)$ budeme volit funkci $\cos(x)$. Tyto vlastnosti budou ověřeny v simulační části.

Když se zamyslíme dále nad možností využití informace, který pomocný regresor nabývá jaké významnosti, můžeme na základě znalosti Taylorova rozvoje diferencovatelných funkcí odvodit, z jakého modelu byla původní řada vygenerována. V simulační části této práce se pokusíme na základě dané časové řady ve tvaru (3.17) určovat, jaká vstupní funkce $f(x)$ figuruje v modelu, který tuto řadu vygeneroval.

4. Aditivita

V předchozích kapitolách jsme se zabývali otázkou jak testovat linearitu v modelech časových řad. Co když ale testy linearitu zamítnou? Pak musíme hledat nelineární modely, které řadu dobře popisují. Pokud chceme odhadovat tvar modelu z dat generovaných tímto modelem, pak často potřebujeme mnoho pozorování, vzhledem k množství možných variant nelineárních modelů.

Celá situace se ale poměrně zjednoduší, pokud je daný nelineární model aditivní. Upřesněme nejprve, co pojmem aditivita myslíme.

4.1 Definice aditivity

Rozšířme definici autoregresního procesu 1.8 pro nelineární modely.

Definice 4.1 (Nelineární AR proces). Náhodnou posloupnost $\{Y_t, t \in \mathbb{Z}\}$ nazýváme *nelineární autoregresní posloupnost řádu n* , pokud ji lze vyjádřit ve tvaru

$$Y_t = f(Y_{t-1}, \dots, Y_{t-n}) + e_t, \quad t \in \mathbb{Z}, \quad (4.1)$$

kde $\{e_t, t \in \mathbb{Z}\}$ je bílý šum a $f(\cdot)$ je n -rozměrná nelineární reálná funkce.

Tento model budeme dále značit NAR z anglického *nonlinear AR*.

Definice 4.2 (Aditivní nelineární AR proces). Náhodnou posloupnost $\{Y_t, t \in \mathbb{Z}\}$ nazýváme *aditivní nelineární autoregresní posloupnost řádu n* , pokud lze vyjádřit ve tvaru

$$Y_t = f_1(Y_{t-1}) + f_2(Y_{t-2}) + \dots + f_n(Y_{t-n}) + e_t, \quad t \in \mathbb{Z}, \quad (4.2)$$

kde $\{e_t, t \in \mathbb{Z}\}$ je bílý šum a $f_1(\cdot), \dots, f_n(\cdot)$ jsou měřitelné reálné funkce.

V dalším textu budeme tento model značit NAAR z anglického *nonlinear additive AR*.

Poznámka. Modely typu NAAR jsou zřejmě podmnožinou modelů NAR.

Pokud splňuje nelineární autoregresní proces předpoklad aditivity, pak je snazší interpretace odhadnutého modelu a navíc nepotřebujeme takové množství pozorování pro odhad modelu.

4.2 Testování aditivity

Chen a kol. (1995) uvádí tři testy, které v modelech typu NAR testují nulovou hypotézu H_0 : Model je typu NAAR. Konkrétně jde o test podmíněných středních hodnot, test Lagrangeovým multiplikátorem a permutační test.

4.2.1 Test podmíněných středních hodnot

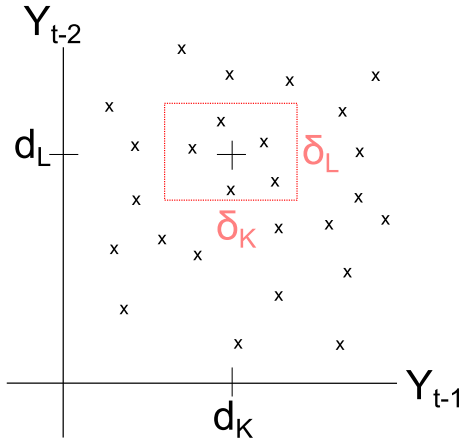
Test vychází z předpokladu blízkých hodnot procesu Y_t za podmínky podobných zpožděných hodnot procesu. Test pro přehlednost uvedeme pro autoregresní řád 2, pro vyšší řády je test podobný. Předpokládejme tedy stacionární ergodický autoregresní proces Y_t , v němž budeme testovat nulovou hypotézu

$$H_0 : Y_t = f_1(Y_{t-1}) + f_2(Y_{t-2}) + e_t.$$

Stavový prostor procesu rozdělíme posloupností bodů (d_1, \dots, d_m) . Na každém průsečíku mřížky, která rozděluje prostor generovaný veličinami Y_{t-1} a Y_{t-2} , odhadneme podmíněnou střední hodnotu Y_t . V bodě (d_i, d_j) je tedy

$$\hat{f}(d_i, d_j) = \frac{1}{n_{ij}} \sum_{t \in I_{ij}} Y_t,$$

kde $I_{ij} = \{t : d_i - \delta_i/2 < Y_{t-1} < d_i + \delta_i/2 \text{ \& } d_j - \delta_j/2 < Y_{t-2} < d_j + \delta_j/2\}$, n_{ij} je počet pozorování v I_{ij} a δ_i, δ_j jsou velikosti obdélníkového okénka okolo bodu (d_i, d_j) . Okénko je obecně obdélníkové, aby bylo možné docílit podobného počtu pozorování v jednotlivých okénkách. Odhad střední hodnoty Y_t za podmínky,



Obrázek 4.1: Okénko okolo bodu (d_K, d_L)

že bod Y_{t-1} leží v okolí d_K a Y_{t-2} leží v okolí bodu d_L , dostaneme jako průměr z pozorování Y_t ležících uvnitř červeného obdélníku na obrázku 4.1. Dá se ukázat, že pro $n \rightarrow \infty$ a $\delta_i, \delta_j \rightarrow 0$ platí

$$(n\delta_i\delta_j)^{1/2}[\hat{f}(d_i, d_j) - f(d_i, d_j)] \sim N(0, \sigma^2/p(d_i, d_j)), \quad (4.3)$$

kde σ^2 je rozptyl bílého šumu e_t a $p(d_i, d_j)$ je sdružená hustota náhodného vektoru (Y_{t-1}, Y_{t-2}) v bodě (d_i, d_j) . Navíc $\hat{f}(d_i, d_j)$ jsou vzájemně nezávislé, viz

Cryer a Chan (2008). Optimální velikost okénka δ_i by měla být řádově velikosti $n^{-1/4}$. Sdruženou hustotu $p(d_i, d_j)$ můžeme konzistentně odhadnout jako $\hat{p}(d_i, d_j) = n_{ij}/(n\delta_i, \delta_j)$.

Odtud a z (4.3) plyne asymptotická platnost

$$n^{1/2}[\hat{f}(d_i, d_j) - f(d_i, d_j)] \sim N(0, \sigma^2).$$

Na základě těchto teoretických poznatků zkonstruujeme test použitím teorie nevyváženého dvojného třídění. Neparametrické metody odhadů často trpí problémem vychýlení kvůli použití okrajových hodnot (viz např. Levine a Li (2012)). Proto extrémní hodnoty uvažované časové řady nebudeme používat a pro konstrukci testu použijeme pouze část intervalu $[Y_{min}, Y_{max}]$. Pro tento účel budeme volit konstantu δ , která bude určovat, jak velkou část intervalu použijeme. Test tedy zkonstruujeme takto:

1. Zvolíme konstantu $0 < \delta < 1$ a interval zkrácený o okrajové hodnoty, jehož délka je $\delta(Y_{max} - Y_{min})$, rozdělíme do m ekvivalentních částí. Určíme tedy body a_0, \dots, a_m jako

$$a_i = Y_{min} + (1 - \delta) \frac{(Y_{max} - Y_{min})}{2} + i\delta \frac{(Y_{max} - Y_{min})}{m}.$$

2. Pozorování Y_t , pro $t = 3, \dots, n$, rozdělíme do buněk $B_{(j,k)}$ sítě určené dělicími body a_i . Tedy $Y_t \in B_{(j,k)} \Leftrightarrow Y_{t-1} \in (a_{j-1}, a_j] \& Y_{t-2} \in (a_{k-1}, a_k]$. Pokud je Y_{t-1} nebo Y_{t-2} menší než a_0 nebo větší než a_m , pak takové pozorování vyřadíme. Pozorování Y_t v buňce $B_{(j,k)}$ označíme X_{jkl} , kde l značí index pozorování (v jedné buňce může být více pozorování).

3. Model dvojného třídění má tvar

$$X_{jkl} = \mu + \alpha_j + \beta_k + \gamma_{jk} + e_{jkl}$$

a za platnosti nulové hypotézy aditivity zřejmě platí $\gamma_{jk} = 0$. Test tedy provedeme jako test nulovosti parametru γ_{jk} . V R lze test snadno provést procedurou **anova**.

Vzhledem k nevyváženosti vzorku dvojného třídění může docházet k problému s nízkým počtem pozorování v některých buňkách $B_{(j,k)}$ a pro $n \rightarrow \infty$ může být konvergence počtu pozorování v jednotlivých buňkách $n_{jk} \rightarrow \infty$ velice pomalá. K tomu dochází poměrně často, neboť pozorování generované jedním modelem jsou prostorově korelované (angl. spatially correlated). Pokud má celý řádek nebo sloupec 0 pozorování, doporučují jej autoři testu vyřadit. Tento problém se zřejmě

zvětšuje s rostoucím m , volba malého m ale naopak zřejmě způsobuje problém s „lokálností“ odhadu střední hodnoty. Volba optimálního m je tak poměrně důležitá a silně ovlivňuje výsledek testu.

Levine a Li (2012) řeší problém s malým počtem pozorování imputací celotabulkových průměrů a variancí do buněk, kde je méně než 2 % pozorování. Dále článek Levine a Li (2012) test rozšiřuje pro modely s nekonstantním rozptylem, tedy pro modely tvaru

$$Y_t = f(Y_{t-1}, \dots, Y_{t-n}) + v(Y_{t-1}, \dots, Y_{t-n})^{1/2} e_t, \quad t \in \mathbb{Z},$$

kde $v(\cdot)$ je obecná spojitá rozptylová funkce, v nichž testuje podmodel

$$Y_t = f_1(Y_{t-1}) + f_2(Y_{t-2}) + \dots + f_n(Y_{t-n}) + v(Y_{t-1}, \dots, Y_{t-n})^{1/2} e_t, \quad t \in \mathbb{Z}.$$

Pro účely simulací byl test ve verzi vhodné pro modely s konstantním rozptylem implementován v R, více v sekci 5.2.

4.2.2 Test založený na Lagrangeově multiplikátoru

Tento test aditivity spočívá v odhadu aditivního nelineárního modelu a následném testu dobré shody, dále jej v textu budeme značit jako LM test. Vyjdeme z duálního Volterrova rozvoje nelineární časové řady

$$\begin{aligned} Y_t &= f(Y_{t-1}, \dots, Y_{t-p}) + e_t \\ &= \mu + \sum_{u=1}^p \phi_u Y_{t-u} + \sum_{u \leq v}^p \phi_{uv} Y_{t-u} Y_{t-v} + \sum_{u \leq v \leq w}^p \phi_{uvw} Y_{t-u} Y_{t-v} Y_{t-w} + \dots + e_t \\ &= \mu + \sum_{u=1}^p f_u(Y_{t-u}) + \sum_{u < v}^p \phi_{uv} Y_{t-u} Y_{t-v} + \sum_{u \leq v \leq w}^p \phi_{uvw} Y_{t-u} Y_{t-v} Y_{t-w} + \dots + e_t, \end{aligned}$$

kde

$$f_u(Y_{t-u}) = \phi_u Y_{t-u} + \phi_{uv} Y_{t-u}^2 + \phi_{uvw} Y_{t-u}^3 + \dots$$

je soubor všech mocnin členu Y_{t-u} . Pokud je model aditivní, pak všechny koeficienty vyšších řádů $\psi_{uu}, \psi_{uuu}, \dots$ budou rovny nule. Na základě tohoto pozorování zkonstruujeme test v následujících krocích:

1. Do časové řady fitujeme aditivní model tvaru (4.2). K odhadu takového modelu použijeme algoritmus ACE (Breiman a Friedman, 1985), zkratka je odvozena z anglického „alternating conditional expectations“. Algoritmus je založen na hledání transformace regresandu či regresorů k zajištění optimálního fitování a v R je implementován v balíčku `acepack`. Jako parametr omezíme transformaci závisle proměnné na lineární. Takto získané

odhady funkcí $f_1(\cdot), \dots, f_p(\cdot)$ označíme $\hat{f}_1(\cdot), \dots, \hat{f}_p(\cdot)$ a spočteme rezidua $\hat{e}_t = Y_t - \sum_{k=1}^p \hat{f}_k(Y_{t-k})$.

2. Vyhodnotíme sadu regresí součinných výrazů druhého řádu $Y_{t-j_1} \cdot Y_{t-j_2}$, kde $1 \leq j_1 < j_2 \leq p$ na zpožděných hodnotách Y_{t-1}, \dots, Y_{t-p} užitím ACE algoritmu opět s restrikcí na lineární transformaci závisle proměnné. Totéž provedeme pro součinné výrazy třetího řádu $Y_{t-k_1} \cdot Y_{t-k_2} \cdot Y_{t-k_3}$, kde $j \leq k_1 \leq k_2 \leq k_3 \leq p$, vyjma kombinace $k_1 = k_2 = k_3$. Takto dostaneme $\frac{p(p-1)}{2}$ reziduálních řad z regresí druhého řádu a $\frac{p(p+1)(p+2)}{6} - p$ reziduálních řad z regresí třetího řádu, celkem tedy $K = \frac{p(p-1)(p+7)}{6}$ takových řad. Ty označíme e_t^1, \dots, e_t^K .
3. Nakonec provedeme lineární regresi reziduí \hat{e}_t z prvního kroku na reziduích e_t^1, \dots, e_t^K . Výsledkem je testová statistika nR^2 , kde n je délka časové řady a R^2 koeficient determinace poslední regrese. Tato statistika se za platnosti nulové hypotézy aditivity řídí asymptoticky rozdělením $\chi_{p(p-1)(p+7)/6}^2$, důkaz viz Chen a kol. (1995).

Pro zesílení testu by bylo možné přidat regrese součinných výrazů vyšších řádů v druhém kroku testu. Po této úpravě by test ale vyžadoval více pozorování pro dosažení asymptotických vlastností. V R je test implementován pod jménem `add.test` v balíčku `nlt`s dle autorova popisu, tedy ve verzi se součinnými členy druhého a třetího řádu.

4.2.3 Permutační test

Poslední zde uvedený test vychází opět z autoregresního Volterrova rozvoje. Označíme

$$f_{uv}(Y_{t-u} Y_{t-v}) = \phi_{uv} Y_{t-u} Y_{t-v} + \phi_{uuv} (Y_{t-u} Y_{t-v})^2 + \phi_{uuuv} (Y_{t-u} Y_{t-v})^3 + \dots,$$

rozvoj přepíšeme jako

$$Y_t = \mu + \sum_{u=1}^p f_u(Y_{t-u}) + \sum_{u<v}^p f_{uv}(Y_{t-u} Y_{t-v}) + \sum_{u \leq v \leq w}^p f_{uvw}(Y_{t-u} Y_{t-v} Y_{t-w}) + \dots + e_t,$$

a nulovou hypotézu

$$H_0 : Y_t = \mu + \sum_{u=1}^p f_u(Y_{t-u}) + e_t$$

budeme testovat proti alternativě

$$H_1 : Y_t = \mu + \sum_{u=1}^p f_u(Y_{t-u}) + \sum_{u<v}^p f_{uv}(Y_{t-u} Y_{t-v}) + e_t.$$

Test opět založíme na fitování aditivního modelu algoritmem ACE, nyní nebudeme aplikovat F test přímo, ale získaná rezidua budeme permutovat. Konstrukce testu bude následující:

1. Jako u předchozího testu fitujeme aditivní model algoritmem ACE a získáme rezidua \hat{e}_t .
2. Provedeme regresi reziduí z předchozího kroku na součinných výrazech druhého řádu $Y_{t-j_1} \cdot Y_{t-j_2}$, pro $1 \leq j_1 < j_2 \leq p$ užitím ACE algoritmu a získáme reziduální součet čtverců RSS_0 .
3. Rezidua \hat{e}_t náhodně permutujeme, dostaneme řadu reziduí \hat{e}_t^P a provedeme tutéž regresi na výrazech druhého řádu $Y_{t-j_1} \cdot Y_{t-j_2}$ užitím ACE algoritmu. Opět spočítáme reziduální součet čtverců, který označíme jako RSS_i a celý proces opakujeme N -krát.
4. Výslednou p -hodnotu určíme jako

$$\frac{\sum_{i=1}^N \mathbb{I}[RSS_i < RSS_0]}{N},$$

kde \mathbb{I} značí indikátorovou funkci.

Test je založen na faktu, že za platnosti nulové hypotézy jsou rezidua \hat{e}_t asymptoticky nezávislá a stejně rozdělená, každá jejich permutace by se tedy měla řídit stejným rozdělením. Reziduální součet čtverců by měl mít tedy také stejné rozdělení pro všechny permutace. Pokud je ale model neaditivní, pak permutování reziduí vede ke ztrátě informace o $f_{uv}(Y_{t-u} Y_{t-v})$ obsažené v řadě reziduí a v důsledku ke zvýšení RSS .

Test byl pro účely této práce implementován v **R**, více v sekci 5.2.

5. Simulační studie

V článcích se často opakují simulace testů nelinearity na modelech TAR, EXPAR, LSTAR a dalších. My se v této práci zaměříme na nelineární modely, které nejsou v literatuře hojně testovány, nicméně testování jejich nelinearity si pozornost zaslouží.

5.1 Testy nelinearity

5.1.1 Taylorův rozvoj spojitých funkcí

Jak bylo řečeno v kapitole 3.10, v modelu (3.17) s použitím funkce $\sin(x)$ by neměla být detekce nelinearity použitím Tsayova či Keenanova testu úspěšná.

Pro ověření této domněnky byly provedeny simulace pro všechny kombinace $b \in (0.1, 0.25, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95)$ a $\sigma^2 \in (0.1, 1, 10, 100)$, kde σ^2 je rozptyl bílého šumu v daném modelu. Pro každý takový model byla vygenerována řada délky 1000 a provedeny oba testy v R, tedy `Keenan.test.corrected` a `Tsay.test`. Oba testy měly nastavenou automatickou volbu řádu fitovaného autoregresního modelu, tedy tak, jak ji volí funkce `ar`. Simulace měla 1000 opakování.

Výsledné p -hodnoty každého testu byly porovnány s hodnotou 5 % a stanoveny četnosti zamítnutí testu. Pro Keenanův test výsledky shrnuje tabulka 5.1. Ve všech případech je četnost zamítnutí blízká 5 %, test tedy opravdu nelinearitu nezachytí. Výjimkou je pouze kombinace parametrů $\sigma^2 = 1$ a b blízkí se jedné.

$b \backslash \sigma^2$	0.1	1	10	100
0.1	0.057	0.047	0.046	0.049
0.25	0.056	0.053	0.045	0.044
0.5	0.045	0.075	0.060	0.049
0.6	0.032	0.091	0.037	0.053
0.7	0.044	0.100	0.062	0.047
0.8	0.027	0.113	0.035	0.050
0.9	0.049	0.148	0.059	0.049
0.95	0.049	0.159	0.038	0.052

Tabulka 5.1: Četnost zamítnutí nulové hypotézy Keenanovým testem pro model (3.17) s funkcí $\sin(x)$

Pro Tsayův test jsou relativní četnosti zamítnutí zobrazeny v tabulce 5.2. V případě nejmenších hodnot rozptylu ($\sigma^2 = 0.1$) test zamítá přibližně v 5 % případů, nelinearita tedy opět není zachycena. Pro $\sigma^2 = 1$ a rostoucí parameter b ale již v některých případech k detekci nelinearity dojde, podobně jako u Keenanova testu. U Tsayova testu můžeme vidět četnosti zamítnutí vyšší než 5 % i pro $\sigma^2 > 1$.

$b \backslash \sigma^2$	0.1	1	10	100
0.1	0.058	0.099	0.154	0.137
0.25	0.060	0.054	0.131	0.147
0.5	0.043	0.069	0.143	0.160
0.6	0.037	0.084	0.133	0.168
0.7	0.048	0.088	0.153	0.153
0.8	0.039	0.110	0.164	0.143
0.9	0.059	0.132	0.152	0.141
0.95	0.051	0.148	0.137	0.149

Tabulka 5.2: Četnost zamítnutí nulové hypotézy Tsayovým testem pro model (3.17) s funkcí $\sin(x)$

Stejný postup aplikovaný na tentýž model, avšak s funkcí $f(x) = \cos(x)$ dává výsledky popsané tabulkami 5.3 a 5.4. Pro velký rozptyl, tedy hodnoty $\sigma^2 = \{10, 100\}$ jsou výsledky podobné jako v případě funkce $\sin(x)$. Důvodem je to, že náhodná složka způsobuje překročení periody kosinu, a tak testy, které vycházejí z lokální aproximace, nefungují. Pro menší hodnoty rozptylu ale již můžeme vidět, že oba testy mají poměrně velkou sílu, jak se předpokládalo v kapitole 3.10. Pro hodnoty $b \geq 0.6$ je v obou případech síla větší než 90 %. Pro malé hodnoty parametru b se ale ukazuje Tsayův test obecně silnější, jak nasvědčovala teoretická konstrukce testu.

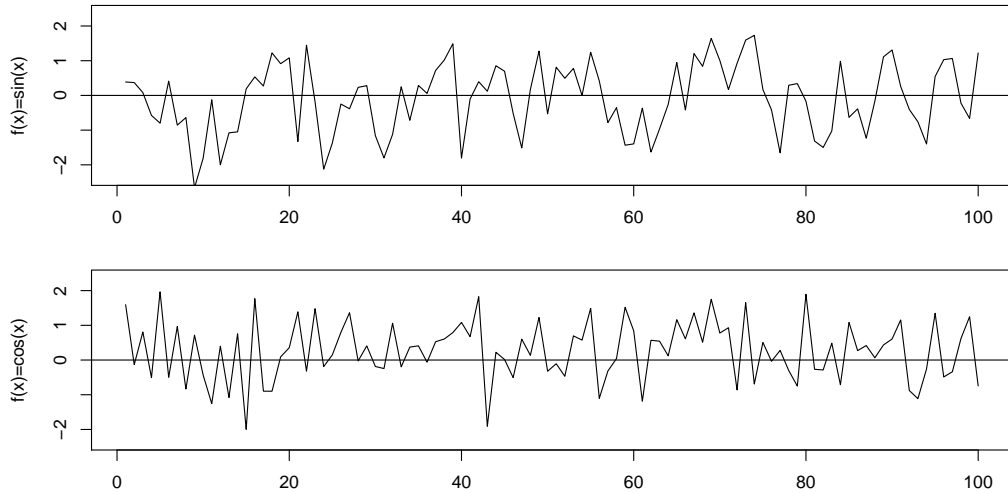
$b \backslash \sigma^2$	0.1	1	10	100
0.1	0.064	0.080	0.047	0.055
0.25	0.243	0.223	0.044	0.039
0.5	0.850	0.829	0.048	0.046
0.6	0.954	0.959	0.046	0.049
0.7	0.991	0.991	0.057	0.042
0.8	0.999	0.999	0.040	0.058
0.9	1.000	1.000	0.051	0.053
0.95	1.000	1.000	0.057	0.059

Tabulka 5.3: Četnost zamítnutí nulové hypotézy Keenanovým testem pro model (3.17) s funkcí $\cos(x)$

$b \backslash \sigma^2$	0.1	1	10	100
0.1	0.174	0.188	0.150	0.154
0.25	0.300	0.411	0.152	0.151
0.5	0.823	0.918	0.146	0.155
0.6	0.904	0.983	0.146	0.161
0.7	0.968	0.999	0.145	0.141
0.8	0.985	1.000	0.137	0.146
0.9	0.989	1.000	0.153	0.164
0.95	0.998	1.000	0.135	0.154

Tabulka 5.4: Četnost zamítnutí nulové hypotézy Tsayovým testem pro model (3.17) s funkcí $\cos(x)$

Pro úplnost ještě doplňme grafickou podobu právě testovaných časových řad. Na obrázku 5.1 můžeme vidět část obou centrovaných řad, konkrétně posledních 100 hodnot z realizace délky 1000.



Obrázek 5.1: Nelineární modely s parametry $b = 0.6$ a $\sigma^2 = 1$

5.1.2 Aplikace RESET testu na určení typu nelinearity

V této sekci rozvineme myšlenku, zda lze pomocí významnosti jednotlivých podpůrných regresorů rozlišovat, která funkce $f(x)$ v modelu (3.17) vygenerovala danou časovou řadu. V předchozí části bylo ukázáno, že Keenanův a Tsayův test dokáží detekovat nelinearitu v modelech, kde Taylorův rozvoj funkce $f(x)$ obsahuje druhou mocninu. Pokusme se naopak aplikovat na modelovanou časovou řadu testy s různými řádů mocnin v podpůrných regresorech a dle p -hodnot usuzovat, které členy rozvoje jsou významné. Odtud pak bude možné určit, o jakou funkci $f(x)$ při generování šlo.

Uvažujme model (3.17) s parametrem $b = 0.8$ a jednotkovým rozptylem v náhodné složce. Jako funkce $f(x)$ volme postupně $\sin(x)$, $\cos(x)$, $\exp(-x)$ a $\text{logist}(x) = \frac{1}{1+e^{-x}}$. Taylorův rozvoj poslední uvedené je popsán v Teräsvirta a kol. (1993). Nyní na každý z modelů aplikujme RESET test a jako parametr testu udávající mocninu v alternativním modelu volme hodnoty od 2 do 10. Prakticky půjde tedy o test lineárního regresního podmodelu

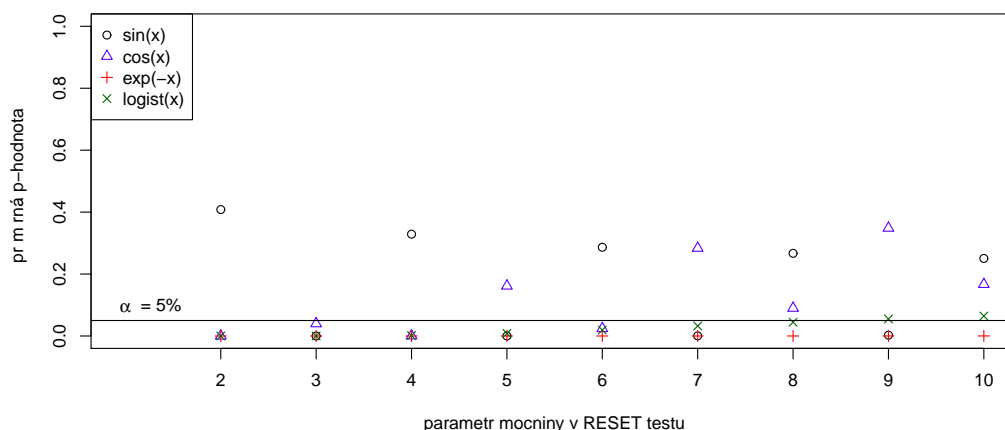
$$Y_t = \alpha + \beta_1 * Y_{t-1}$$

v nelineárním modelu

$$Y_t = \alpha + \beta_1 * Y_{t-1} + \beta_2 Y_{t-1}^i,$$

kde i je zmíněný parametr testu.

Pro každou ze čtyř výše uvedených funkcí byla v R vygenerována časová řada délky 1000 a simulace byla opakována tisíckrát. Průměrná p -hodnota RESET testu pro jednotlivé mocniny je vykreslena v obrázku 5.2.



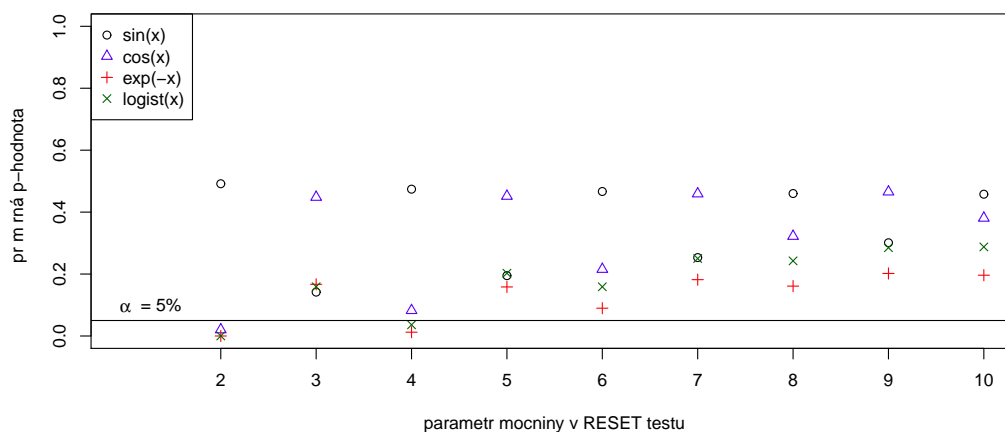
Obrázek 5.2: RESET test pro $b = 0.8$ a $\sigma^2 = 1$

Dle grafu se potvrdila vlastnost funkce $\sin(x)$, jejíž rozvoj obsahuje pouze liché mocniny. Zde RESET test zamítá výrazně silněji v případě parametrů lichých mocnin. Funkce $\cos(x)$ naopak výrazněji zamítá test pro sudé mocniny, nicméně poměrně významně (na 5% hladině) zamítá i test s lichými mocninami nižších řádů, např. 3. Důvodem je, že Taylorův rozvoj funkce $\cos(x)$ obsahuje pouze sudé mocniny, a tak lineární model obsahující mocniny Y_t v prvním řádu nepopisuje časovou řadu Y_t příliš dobře a přidání regresoru s mocninou 3 přinese relativně „hodně informace“. Naproti tomu funkce $\sin(x)$ je již poměrně dobře popsána lineárním autoregresním modelem řádu 1, a tak přidání regresoru ve 2. mocnině nepřináší novou informaci.

Pokud zvolíme $b = 0.25$, je stále efekt vlastností rozvoje do Taylorova polynomu silný a lze dle p -hodnot rozeznat, zda časovou řadu generoval model daný funkcí $\sin(x)$ nebo $\cos(x)$, i když test již hypotézu linearity na 5% hladině nezamítá (viz obrázek 5.3). Zde ale můžeme pozorovat podobný efekt střídajících se p -hodnot i u modelu generovaného funkcí $\exp(-x)$, a tak by bylo možné jej zaměnit s funkcí $\cos(x)$.

Pokud se pokusíme graficky rozlišit funkce $\exp(-x)$ a $\logist(x)$, zjistíme, že ač jsou p -hodnoty RESET testu u exponenciální funkce zpravidla menší, nelze stabilně rozlišit, o kterou z daných funkcí šlo. Taylorův rozvoj funkce $\exp(-x)$ je sice zamítán častěji, nicméně pokud neznáme parametr b , nemůžeme ze síly zamítnutí usuzovat, jak vypadaly koeficienty funkce generující původní model.

Závěrem dodejme, že ač použití RESET testu s jednotlivými mocninami může pomoci utvoření představy o funkci generující model, tak je to pouze ve smyslu přítomnosti či nepřítomnosti konkrétních mocnin v Taylorově rozvoji funkce,



Obrázek 5.3: RESET test pro $b = 0.25$ a $\sigma^2 = 1$

nikoliv o určení přesné specifikace konkrétního modelu.

5.1.3 Mocninný nelineární model

Nyní budeme studovat autoregresní nelineární model prvního řádu, tedy

$$Y_t = b * (Y_{t-1})^{\frac{1}{k}} + e_t,$$

kde b je předem určená konstanta a e_t je náhodná složka. Aby byla taková časová řada stacionární (slabě, v souladu s definicí 1.3), musí platit $b < 1$. Pokusíme se ověřit, které z testů představených v kapitole 3 jsou schopny tuto nelinearitu zachytit. Aby dávala časová řada smysl a odmocnina byla vždy definována i pro sudá k , použijeme bílý šum s exponenciálním rozdělením a parametrem exponenciálního rozdělení λ (tedy se střední hodnotou $Ee_t = 1/\lambda$). Pro některé testy tak dojde k porušení předpokladu normality, přesto se může ukázat, že na síle příliš neztrátí.

b	0.25	0.5	0.75	0.9	0.95
k	2	3	4	5	
λ	1	3	5	10	

Tabulka 5.5: Hodnoty parametrů použité v simulacích

Testy byly opět aplikovány na řady délky 1000 s počtem tisíce opakování. Jako parametry modelu byly voleny všechny kombinace z tabulky 5.5, použité testy jsou uvedeny v tabulce 5.6. Poslední sloupec udává parametry testu, které byly volány v R.

Test	Název v R	Popsán v kapitole
Keenanův	<code>Keenan.test.corrected()</code>	3.1
Tsayův	<code>Tsay.test()</code>	3.2
Whiteův	<code>white.test()</code>	3.5.1
Terrasvirtaův	<code>terasvirta.test()</code>	3.5.2
Hamiltonův	<code>Hamilton.test()</code>	3.7.1
RESET	<code>reset()</code>	3.4
Test	Dostupný v balíčku	Dodatečné parametry
Keenanův	TSA, opraven	<code>order=ar(Y)\$order</code>
Tsayův	TSA	<code>order=ar(Y)\$order</code>
Whiteův	<code>tseries</code>	<code>lag=ar(Y)\$order</code>
Terrasvirtaův	<code>tseries</code>	<code>lag=ar(Y)\$order</code>
Hamiltonův	vlastní implementace	<code>order=ar(Y)\$order</code>
RESET	<code>lmtest</code>	Je třeba definovat <code>reset(Y[-1]~Y[-n])</code>

Tabulka 5.6: Testy užívané v simulacích

k	Test	Keenan	Tsay	Teräsvirta	White	Hamilton	Reset
	2		0.08	0.09	0.07	0.07	0.06
3		0.06	0.06	0.09	0.03	0.04	0.03
4		0.07	0.05	0.09	0.03	0.04	0.03
5		0.04	0.03	0.06	0.03	0.05	0.04

Tabulka 5.7: Výsledky testů pro $b = 0.25$ a $\lambda = 1$

Jak lze předpokládat, síla všech testů roste s parametrem b a λ . Pro malé hodnoty obou parametrů ani jeden z testů nelinearity nedetekuje, naopak pro velké hodnoty jej detekují prakticky všechny. Pro hodnoty $b = 0.25$ a $\lambda = 1$ jsou četnosti zamítnutí nulové hypotézy na hladině 5 % v závislosti na mocnině k uvedeny v tabulce 5.7. Pro parametry $b = 0.95$ a $\lambda = 10$ jsou výsledky těchto testů v tabulce 5.8. Obě řady po ustálení (vynechání několika prvních hodnot) vykresluje obrázek 5.4.

Zřejmě nejslabší je Hamiltonův test, který ani s největšími uvažovanými parametry nedosahuje ve většině případů síly 50 %. Tento test je totiž silně závislý na předpokladu normality. RESET test se zdá být nejsilnější (což je zřejmě způsobeno tím, že alternativu nelinearity testuje ve formě mocnin, tedy ideálně vzhledem

$k \backslash$ Test	Keenan	Tsay	Teräsvirta	White	Hamilton	Reset
2	0.86	0.82	0.82	0.78	0.40	0.97
3	0.93	0.87	0.89	0.81	0.45	0.99
4	0.84	0.85	0.86	0.81	0.59	1.00
5	0.81	0.81	0.83	0.78	0.49	1.00

Tabulka 5.8: Výsledky testů pro $b = 0.95$ a $\lambda = 10$

k tvaru našeho modelu). Ostatní testy jsou, co se týče síly, velice podobné. Závislost na mocnině k není z výsledných četností zřejmá. To se ovšem změní, když budeme volit hodnoty parametru b uvnitř sledovaného intervalu. Například již pro $b = 0.5$ je schopnost detekovat nelinearitu lepší pro modely s malou hodnotou odmocniny k , viz tabulka 5.9. Zde se zdá být již závislost zřejmá, i když jsou všechny testy poměrně slabé.

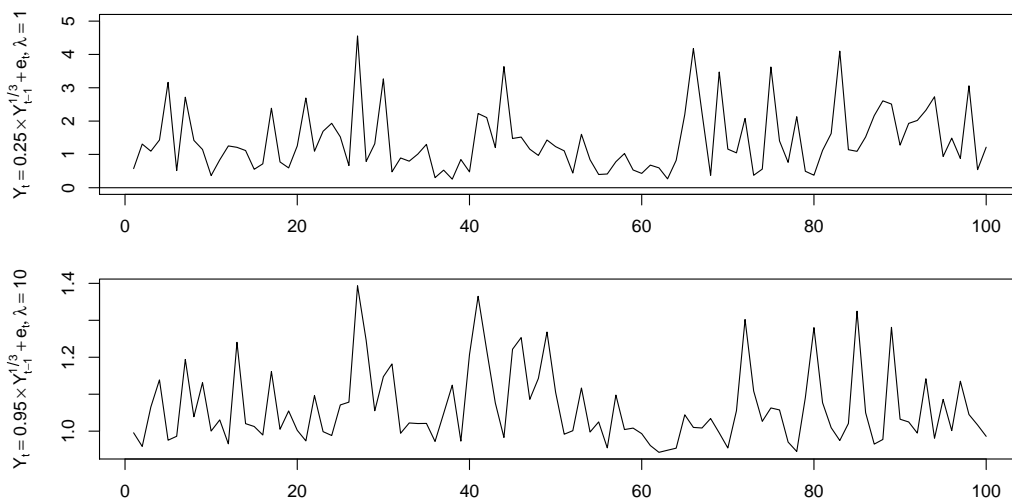
$k \backslash$ Test	Keenan	Tsay	Teräsvirta	White	Hamilton	Reset
2	0.19	0.17	0.19	0.11	0.13	0.16
3	0.10	0.11	0.13	0.06	0.06	0.11
4	0.08	0.10	0.09	0.07	0.05	0.08
5	0.08	0.07	0.09	0.05	0.07	0.04

Tabulka 5.9: Výsledky testů pro $b = 0.5$ a $\lambda = 1$

Zajímavé srovnání ale nastává např. pro hodnoty parametrů $b = 0.9$ a $\lambda = 5$. Tyto jsou uvedeny v tabulce 5.10. Zde se síla testů výrazně liší. Rozdíl mezi testy založenými na Tukeyově testu, tedy Keenanovým a Tsayovým, není příliš velký. Naproti tomu testy neuronových sítí se liší významně, Teräsvirtaův test dosahuje téměř dvojnásobné síly.

$k \backslash$ Test	Keenan	Tsay	Teräsvirta	White	Hamilton	Reset
2	0.49	0.49	0.57	0.37	0.08	0.68
3	0.56	0.53	0.61	0.34	0.04	0.81
4	0.53	0.56	0.64	0.45	0.05	0.83
5	0.47	0.47	0.62	0.47	0.06	0.85

Tabulka 5.10: Výsledky testů pro $b = 0.5$ a $\lambda = 1$



Obrázek 5.4: Průběh řad generovaných mocninným modelem

5.2 Testy aditivity

Z testů aditivity popsanych v této práci je dostupný v R pouze LM test, ostatní byly naprogramovány jako doplněk k této práci.

Pro test podmíněných středních hodnot se ukazuje jako poměrně důležitá volba parametrů δ a m . Vzhledem ke korelovanosti pozorování v prostoru totiž dojde v případě velkých m k řídkosti některých buněk. Pro ilustraci volme neaditivní model tvaru

$$Y_t = Y_{t-1} + 0.5 \cdot \sin(Y_{t-2}) + e_t, e_t \sim N(0, 1)$$

o délce 1000 pozorování a testujeme aditivní závislost na dvou prvních zpožděních. Pro volbu $m = 15$ dostaneme počty pozorování v buňkách $B_{(i,j)}$ tak, jak je ukazuje tabulka 5.11. Je zřejmé, že v takto nevyvážené tabulce nelze efektivně provést test dvojného třídění. Pokud naopak zvolíme např. $m = 3$, pak dostaneme situaci dle tabulky 5.12.

V tomto případě naopak ztrácíme informaci o lokálních vlastnostech funkce a např. středová hodnota tabulky sestává již z více než poloviny všech pozorování. Schopnost rozlišit neaditivitu tím klesá, je tedy třeba volit kompromis. Pro hledání správné volby budeme simulačně testovat aditivní a neaditivní modely pro různé volby parametrů δ a m . Zároveň tyto modely otestujeme zbývajícími dvěma testy.

Testované modely popisuje tabulka 5.13. První tři modely jsou aditivní, další tři neaditivní. Postupně, první model je lineární AR(2), druhý je jednoduchý NAAR(2) model a třetí je taktéž NAAR(2) model, ovšem se složitější nelineární strukturou. Model 4 je již neaditivní, jde o analogii třetího modelu, ovšem se zaměněnými členy prvního a druhého zpoždění. Pátý model je jednoduchý ne-

$i \backslash j$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	2	0	1	0	0	1	0	0	0	0	0	0	1	0
2	0	0	0	1	1	0	1	0	0	2	1	0	1	0	0
3	0	0	1	3	5	7	2	0	0	1	2	1	1	0	1
4	1	1	1	5	2	9	1	5	5	5	2	3	3	2	0
5	0	0	2	3	9	8	12	10	14	12	7	6	0	2	0
6	1	0	1	6	9	16	17	21	15	13	12	5	1	2	1
7	1	0	1	3	8	14	25	37	31	17	3	7	4	1	0
8	0	1	1	2	15	25	28	35	22	19	14	5	2	0	0
9	0	0	1	2	14	17	28	31	36	14	8	1	1	0	0
10	0	0	1	5	12	10	21	15	20	7	6	1	1	0	0
11	1	2	2	4	3	6	12	12	8	5	8	3	0	0	0
12	1	1	6	4	2	6	3	3	2	3	1	3	1	1	0
13	0	0	4	4	0	2	0	1	0	1	0	1	0	1	1
14	0	0	2	2	4	1	1	0	0	0	1	1	0	0	0
15	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0

Tabulka 5.11: Test podmíněných středních hodnot pro $m = 15$

$i \backslash j$	1	2	3
1	38	95	33
2	84	534	75
3	44	66	23

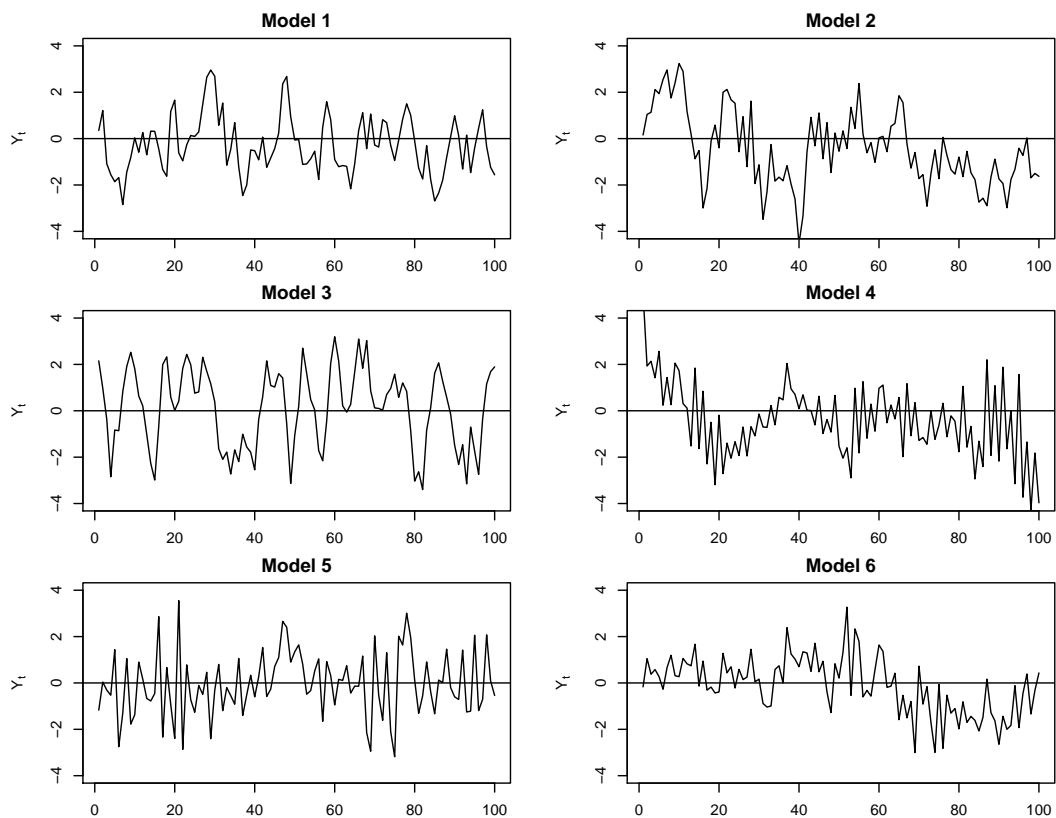
Tabulka 5.12: Test podmíněných středních hodnot pro $m = 3$

lineární model, jehož stacionarita je zajištěna použitím funkce sinus. Podmínky stacionarity a ergodicity pro modely tohoto typu jsou popsány například v Chen a Tsay (1993). Poslední model je podobný, obsahuje navíc „rušivý“ člen. Ve všech modelech byla náhodná složka e_t generována z rozdělení $N(0, 1)$. Části těchto časových řad o délce 100 pozorování po vynechání několika prvních pozorování, jsou znázorněny na obrázku 5.5.

Začněme s výsledky testu podmíněných středních hodnot. Pro modely 1–3 jsou četnosti zamítnutí na 5% hladině uvedeny v tabulce 5.14. Pro $m = 3$ a velká δ je chyba prvního druhu výrazně vyšší než 5 %. Zde zřejmě dochází k problému s hodnotami na okraji pozorovaného okénka. Užitím hodnot ze středu pozorovaného okénka se problém zmenšuje (např. volbou $\delta = 0.7$, kdy odstraníme 15 %

Označení	Tvar modelu
Model 1	$Y_t = 0.8 * Y_{t-1} - 0.3 * Y_{t-2} + e_t$
Model 2	$Y_t = 0.5 * Y_{t-1} + \sin(Y_{t-2}) + e_t$
Model 3	$Y_t = 2 * \exp(-0.1 * Y_{t-1}^2) * Y_{t-1} - \exp(-0.1 * Y_{t-2}^2) * Y_{t-2} + e_t$
Model 4	$Y_t = 2 * \exp(-0.1 * Y_{t-2}^2) * Y_{t-1} - \exp(-0.1 * Y_{t-1}^2) * Y_{t-2} + e_t$
Model 5	$Y_t = Y_{t-1} * \sin(Y_{t-2}) + e_t$
Model 6	$Y_t = 0.5 * Y_{t-1} + 0.5 * \cos(Y_{t-1}) * Y_{t-2} + e_t$

Tabulka 5.13: Modely pro ověření testů aditivity



Obrázek 5.5: Modely pro ověření testů aditivity

délky intervalu na každé straně). Pro složitější charakter nelinearity ale testy pro malé m stále trpí nepřiměřenou chybou prvního druhu. Dle simulovaných hodnot bychom tak doporučili m alespoň 5 pro řady délky 1000.

		m							
		δ	3	5	7	9	11	13	15
Model 1	0.7	0.038	0.024	0.030	0.033	0.040	0.032	0.031	
	0.8	0.114	0.013	0.020	0.039	0.032	0.028	0.034	
	0.9	0.256	0.023	0.019	0.023	0.030	0.036	0.028	
	0.95	0.355	0.018	0.016	0.018	0.035	0.030	0.028	
Model 2	0.7	0.168	0.027	0.035	0.025	0.034	0.031	0.032	
	0.8	0.422	0.033	0.026	0.031	0.042	0.025	0.025	
	0.9	0.747	0.037	0.018	0.029	0.026	0.036	0.038	
	0.95	0.871	0.048	0.019	0.016	0.029	0.032	0.028	
Model 3	0.7	0.842	0.080	0.048	0.023	0.036	0.027	0.030	
	0.8	0.980	0.123	0.042	0.039	0.038	0.028	0.033	
	0.9	1.000	0.188	0.057	0.039	0.024	0.028	0.033	
	0.95	1.000	0.238	0.063	0.035	0.024	0.032	0.029	

Tabulka 5.14: Četnosti zamítnutí aditivity modelů 1, 2 a 3 testem podmíněných středních hodnot na 5% hladině

Tytěž simulace pro neaditivní modely uvádí tabulka 5.15. V závislosti na m zde můžeme sledovat stejný efekt, s rostoucím m klesá četnost zamítnutí aditivity. Pro model 6 ale pozorujeme četnost zamítnutí okolo 5 % již pro $m = 5$. Takový model tedy pro žádnou volbu parametrů nelze tímto testem spolehlivě detekovat tak, abychom zachovali chybu prvního druhu 5 % pro aditivní modely.

Cílem je najít takové hodnoty parametrů m a δ , kdy četnost zamítnutí neaditivních modelů bude co možná nejvyšší. Pokud tedy vezmeme v potaz výsledky simulací pro model 4 a 5 a zohledníme předchozí závěry, dospějeme zřejmě k $m = 5$. Parametr δ bychom volili nejspíše 0.7 nebo 0.8, vzhledem k výsledkům modelu 3.

Pro zbývající dva testy byl použit parametr `order=2`. LM test automatickou volbu nenabízí a v permutačním jsme ji také neimplementovali. Proti testům nelinearity totiž není zřejmé, jaký řád by se měl použít. Automatická volba fitováním AR modelu nemusí fungovat správně, jelikož řady lineární nejsou. Volba řádu dva tak testu pomáhá na síle (neboť všechny modely jsou autoregresní druhého řádu), ale taková volba byla použita u všech testů, a tak jsou výsledky porovnatelné. Test podmíněných středních hodnot byl také fixován na 2 zpoždění, pro zpoždě-

		m	3	5	7	9	11	13	15
		δ							
Model 4	0.7	0.965	0.863	0.469	0.221	0.116	0.068	0.055	
	0.8	0.911	0.932	0.667	0.311	0.158	0.092	0.082	
	0.9	0.780	0.955	0.773	0.439	0.212	0.134	0.090	
	0.95	0.717	0.967	0.817	0.508	0.255	0.143	0.110	
Model 5	0.7	0.988	0.744	0.331	0.142	0.074	0.054	0.056	
	0.8	0.994	0.858	0.450	0.206	0.109	0.065	0.060	
	0.9	0.986	0.919	0.591	0.302	0.158	0.096	0.071	
	0.95	0.981	0.933	0.667	0.367	0.173	0.109	0.073	
Model 6	0.7	0.184	0.052	0.029	0.035	0.033	0.032	0.045	
	0.8	0.320	0.045	0.037	0.027	0.050	0.033	0.028	
	0.9	0.447	0.040	0.031	0.018	0.032	0.036	0.030	
	0.95	0.498	0.042	0.035	0.025	0.038	0.035	0.030	

Tabulka 5.15: Četnosti zamítnutí aditivity modelů 4, 5 a 6 testem podmíněných středních hodnot na 5% hladině

ní většího řádu by musel být zobecněn na trojné třídění (čtverné třídění, ...). Četnosti zamítnutí jsou uvedeny v tabulce 5.16.

Test	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
LM	0.031	0.204	0.268	1.000	0.995	0.999
Permutační	0.015	0.013	0.016	0.326	1.000	0.106

Tabulka 5.16: Četnosti zamítnutí aditivity LM testem a permutačním testem na 5% hladině

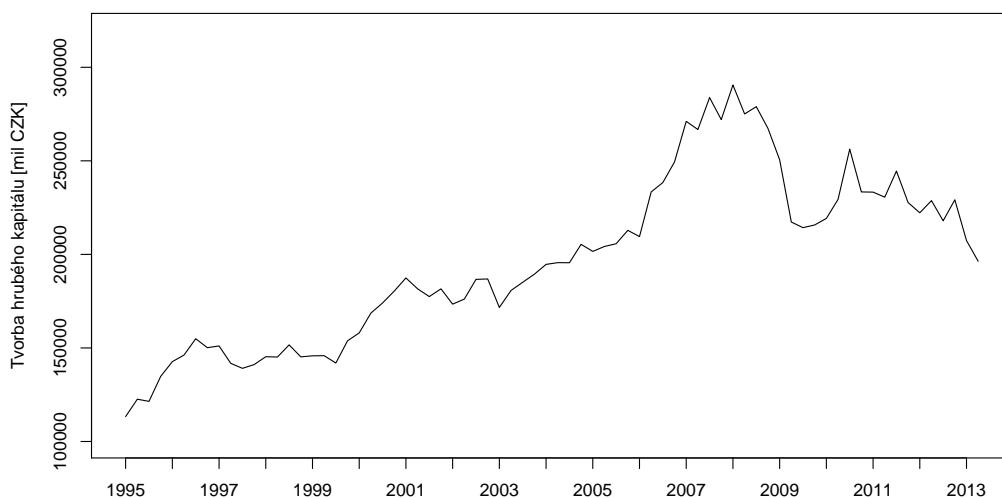
Z výsledků vidíme, že jako nejsilnější se ukazuje LM test. Hypotézu aditivity pro všechny neaditivní modely jasně zamítá, ovšem pro některé aditivní modely zamítá aditivitu také. Permutační test byl implementován s počtem 100 permutací, viz C.3. Pro aditivní modely neodpovídá chyba prvního druhu hladině testu 5 %. Neaditivní modely zamítá s různou silou, záleží na konkrétním tvaru nelinearity v modelu.

Pro praktické použití bychom tedy doporučili pro testování aditivity LM test, protože je mezi popsány testy nejsilnější. Pokud ale aditivitu zamítne, je vhodné ještě tuto skutečnost ověřit dodatečnou aplikací ostatních testů, které tuto domněnku mohou potvrdit.

6. Užití testů na reálných datech

Pro názornou ukázkou toho, že je vhodné linearitu v modelech časových řad testovat, uvedeme na závěr příklad využívající reálná data. Pro studium byla zvolena data vývoje makroekonomických ukazatelů dle Českého statistického úřadu (http://www.czso.cz/csu/redakce.nsf/i/hdp_cr). Tento zdroj poskytuje čtvrtletní časové řady, pro většinu ukazatelů jsou zde data od roku 1995 do současnosti, tedy přibližně 70 pozorování (u jednotlivých řad je délka různá). Analyzováno bylo několik desítek těchto řad, většina z nich má lineární charakter. Pro názornost zde uvedeme příklad řady, u níž zanedbání její nelinearity vede k nevhodnému modelování.

Analyzovat budeme řadu ze souboru¹ shrnujícího výdaje na hrubý domácí produkt v České republice. Ukazatel HDP je tvořen výdaji na konečnou spotřebu, tvorbou hrubého kapitálu a hodnotou zahraničního obchodu (export–import). My se zaměříme na tvorbu hrubého kapitálu. Tuto časovou řadu máme k dispozici v celém časovém okénku od roku 1995 do poloviny roku 2013, celkem tedy 74 pozorování. Graf řady ukazuje obrázek 6.1.

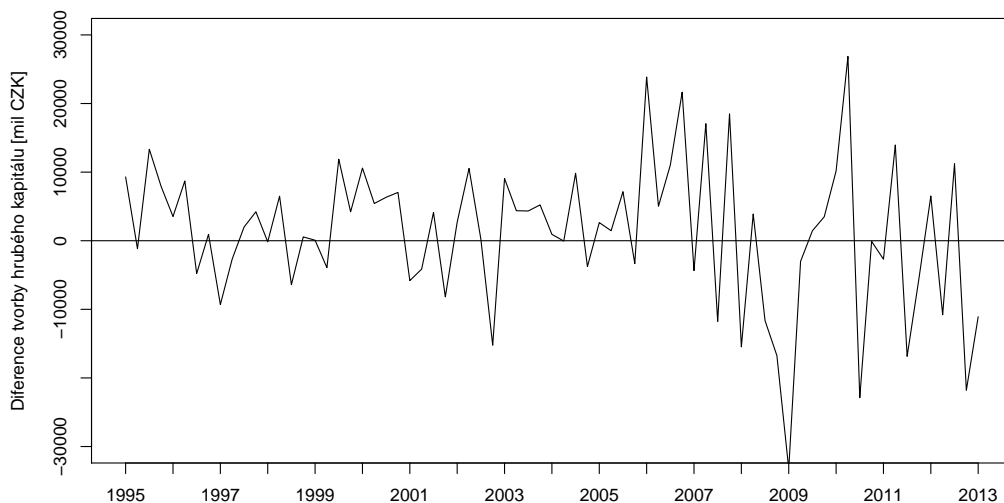


Obrázek 6.1: Čtvrtletní tvorba hrubého kapitálu v letech 1995–2013

Užitím standardní Box-Jenkinsovy metodologie (viz Cipra (2008) str. 339) se pokusíme tuto řadu popsat lineárním modelem. Jelikož řada není stacionární, přejdeme k prvním diferencím a budeme tedy hledat model typu $ARIMA(p,1,q)$. Diferencováním dostaneme řadu na obrázku 6.2.

Tato řada se již zdá býti stacionární, potvrzují to i testy stacionarity. Konkrétně byly pro ověření použity testy KPSS a ADF (Cipra, 2008, str. 355–356).

¹K dispozici na [http://www.czso.cz/csu/csu.nsf/i/tab_vs/\\$File/tab_vs_2q13r.xls](http://www.czso.cz/csu/csu.nsf/i/tab_vs/$File/tab_vs_2q13r.xls)



Obrázek 6.2: Diference čtvrtletní tvorby hrubého kapitálu v letech 1995–2013

Výsledná p -hodnota KPSS testu, jehož nulová hypotéza je stacionarita testované řady, dosahuje hodnoty větší než 0.1 (přesné číslo test z balíčku `tseries` neuvádí, neboť tato statistika se neřídí za platnosti nulové hypotézy žádným standardním rozdělením a její hodnoty jsou tabelovány). Naopak ADF test, jehož nulovou hypotézou je jednotkový kořen (a v důsledku nestacionarita časové řady), dosáhl p -hodnoty 0.034; na 5% hladině hypotézu jednotkového kořenu tedy zamítneme.

Model identifikujeme pomocí informačních kritérií. Akaikeho i Schwarzovo (viz Cipra (2008), str. 342) svědčí pro AR(2) model diferencované řady, v ARIMA zápisu tedy ARIMA(2,1,0). Standardní verifikační postupy (popsané v Cipra (2008)) svědčí pro volbu tohoto modelu. Rezidua korelovaná nejsou, Durbin-Watsonova statistika (jejíž výpočet je implementován např. v balíčku `lmtest`) je rovna 1.95. Normalita reziduí dle Jarque-Bera testu (balíček `tseries`) porušena není, neboť p -hodnota je rovna 0.2. Zdá se tedy, že jsme našli lineární model, který tuto časovou řadu dobře popisuje.

Zaměříme se nyní ale na testy linearity. Tabulka 6.1 zobrazuje p -hodnoty jednotlivých testů představených v kapitole 3. Především Tsayův test a testy založené na neuronových sítích poukazují na možné zanedbání nelineární struktury modelu. Ověříme ještě, zda je nelinearita ve smyslu prvních dvou zpoždění analyzované časové řady aditivní. Aplikací testů představených v kapitole 4 dostaneme p -hodnoty uvedené v tabulce 6.2. Všechny testy naznačují, že předpoklad aditivity porušen není.

Na závěr ještě srovnáme, jak velký vliv má na analyzované řadě zanedbání nelinearity. Model ARIMA(2,1,0) dosahoval $R^2 = 7.4\%$. Pokud budeme předpokládat, že model generující řadu je nelineární a aditivní, můžeme pro jeho odhad

Test	Výsledná p -hodnota
Keenanův	0.821
Tsayův	0.006
Whitův	0.012
Terrasvirtův	0.013
Hamiltonův	0.070

Tabulka 6.1: Testy linearity na diferencované řadě tvorby kapitálu

Test	Výsledná p -hodnota
Podmíněných středních hodnot s parametry $m = 5$ a $\delta = 0.8$	0.203
LM	0.319
Permutační	0.750

Tabulka 6.2: Testy aditivity na diferencované řadě tvorby kapitálu

použít ACE algoritmus (Breiman a Friedman, 1985). Užitím nelineárních transformací zpožděných hodnot řady i samotné vysvětlované proměnné dostaneme $R^2 = 27.4\%$, tedy téměř čtyřnásobný nárůst. Na tomto příkladu tedy použití nelineárního modelu přinese významné zlepšení ve smyslu zmenšení reziduálního součtu čtverců.

Závěr

V práci byly představeny definice a testy týkající se linearity a aditivity v modelech časových řad. U jednotlivých testů je uvedeno, zda a v jakých balíčcích jsou implementovány ve statistickém prostředí R. V Keenanově testu byla popsána chyba v implementaci v balíku TSA (více v sekci 3.1.1). U Whitova testu založeného na neuronových sítích je popsáno, jakým způsobem v R škáluje data, neboť implementace tohoto testu v různých balíčcích uvažuje toto škálování různě a ani v jedné z nalezených verzí neodpovídá původnímu návrhu autora testu. Hamiltonův test založený na náhodných polích byl pro účely této práce naprogramován v R, stejně tak jako testy aditivity, které byly v práci popsány.

Simulační část ukazuje, jak který z testů dokáže detekovat nelinearitu v různých nelineárních modelech. Je zde představen pokus o detekci typu nelinearity z předem dané sady možných nelineárních modelů. Ukazuje se, že studium Taylorova rozvoje nelineární funkce generující nelineární časovou řadu umožňuje některé funkce rozlišit, praktická aplikace je ale omezená z důvodu předem neznámé množiny takových funkcí.

Dále byly provedeny simulace testů linearity na mocninném nelineárním modelu a srovnány jejich síly pro různé parametry modelu. Poslední část simulací popisuje testy aditivity, z nichž některé byly v R implementovány, některé doprogramovány.

Závěrečná kapitola se zabývá aplikací testů na reálných datech. Ukázalo se, že zanedbání nelinearity může vést k odhadu modelu s výrazně větším reziduálním součtem čtverců, než v případě použití nelineárního modelu.

Dodatky

A Pomocná tvrzení

Věta A1 (Konvergence F rozdělení k rozdělení χ^2). Nechť náhodná veličina $X \sim F_{m,n}$. Pak pro $n \rightarrow \infty$ konverguje $X \xrightarrow{D} Y/m$, kde $Y \sim \chi_m^2$.

Důkaz. Veličina X lze napsat jako

$$X = \frac{U/m}{V/n},$$

kde $U \sim \chi_m^2$ a $V \sim \chi_n^2$ a navíc U a V jsou nezávislé. Náhodnou veličinu V můžeme zapsat jako

$$V = \sum_{k=1}^n W_k^2,$$

kde $W_k \sim N(0, 1)$ pro všechna k a jsou nezávislé. Platí, že $W_k^2 \sim \chi_1^2$, má tedy střední hodnotu rovnu jedné a rozptyl roven dvěma. Pak dle zákona velkých čísel (Čebyšev) bude pro $n \rightarrow \infty$ výraz

$$\frac{\sum_{k=1}^n W_k^2}{n} = \frac{V}{n}$$

konvergovat v pravděpodobnosti ke střední hodnotě W_k^2 , tedy k jedné.

Z Cramér-Sluckého věty pak ihned dostáváme

$$X = \frac{U/m}{V/n} \xrightarrow{D} Y/m,$$

kde $Y = U$, tedy $Y \sim \chi_m^2$. □

B Oprava Keenanova testu z TSA

V balíčku TSA, verze 1.01, je Keenanův test naprogramován následovně:

```
> Keenan.test
function (x, order, ...) {
  if (missing(order))
    order = ar(x, ...)$order
  m = order
  myfun = function(y, x, m) {
    X = NULL
    for (i in 1:m) X = cbind(X, zlag(x, i))
    X = cbind(y, X)
    X = na.omit(X)
    lm.fit(y = X[, 1], x = X[, -1, drop = FALSE]) }
  x = as.vector(x)
  n = length(x)
  lm1 = myfun(y = x, x = x, m = m)
  fit1 = lm1$fit^2
  res1 = lm1$residuals
  lm2 = myfun(y = c(rep(NA, m), fit1), x = x, m = m)
  res2 = lm2$residuals
  lm3 = lm(res1 ~ res2 - 1)
  test.stat = summary(lm3)$fstatistic[1] *
    (n - 2 * m - 2)/(n - m - 1)
  names(test.stat) = NULL
  return(list(test.stat = test.stat, p.value = pf(test.stat,
    df1 = 1, df2 = n - 2 * m - 2, lower.tail = FALSE),
    order = order)) }
```

Je zde ale chyba, neboť v průběhu testu se fituje AR model bez interceptu. Samotná chyba je uvnitř funkce `myfun`, kde je použit příkaz `lm.fit()`. Tento příkaz je standardně volán funkcí `lm()` pro fitování lineárního modelu. Jeho vstupní parametry jsou odezva a designová matice regresního modelu, kterou ale R použije tak, jak ji uživatel zadá. V našem případě se z tohoto důvodu fituje model bez interceptu, neboť zadaná matice neobsahuje sloupec jedniček.

Chybu lze opravit nahrazením příkazu `lm.fit` příkazem `lm` a zadáním modelu ve tvaru formule, pak se již odhadne model s interceptem. V našem případě tedy správná implementace může vypadat například takto:

```

> Keenan.test.corrected
function (x, order, ...) {
  require(TSA)
  if (missing(order))
    order = ar(x, ...)$order
  m = order
  myfun = function(y, x, m) {
    X = NULL
    for (i in 1:m) X = cbind(X, zlag(x, i))
    X = cbind(y, X)
    X = na.omit(X)
    lm(X[, 1] ~ X[, -1, drop = FALSE]) }
  x = as.vector(x)
  n = length(x)
  lm1 = myfun(y = x, x = x, m = m)
  fit1 = lm1$fit^2
  res1 = lm1$residuals
  lm2 = myfun(y = c(rep(NA, m), fit1), x = x, m = m)
  res2 = lm2$residuals
  lm3 = lm(res1 ~ res2 - 1)
  test.stat = summary(lm3)$fstatistic[1] *
    (n - 2 * m - 2)/(n - m - 1)
  names(test.stat) = NULL
  return(list(test.stat = test.stat, p.value = pf(test.stat,
    df1 = 1, df2 = n - 2 * m - 2, lower.tail = FALSE),
    order = order)) }

```

C Implementace testů v R

Za účelem simulační studie byly některé testy implementovány v R. V této příloze jsou implementační skripty popsány.

C.1 Hamiltonův test

```
> Hamilton.test
function (x, order) {
  require(TSA)
  if (missing(order))
    order <- ar(x)$order
  k <- order
  if (k>5){k<-5} #stop("AR order is equal ",order, ".
Please insert order <=5",sep="")
  if(k==0) {k<-1}
  x <- as.vector(x)
  n <- length(x)-k
  #Odhad lineárního modelu
  X = NULL
  for (i in 1:k) X = cbind(X, zlag(x, i))
  X = cbind(x, X)
  X = na.omit(X)
  lin<-lm(X[, 1] ~ X[, -1])
  #Koeficienty g
  g<-2/sqrt(k*apply(as.matrix(X[,-1]),2,sd)^2)
  gX<-g*X[,-1]
  #Kovarianční matice
  H<-function(h){
    switch(order,
      T1={1-h},
      T2={1-(2/pi)*(h*sqrt(1-h^2)+asin(h))},
      T3={1-(3*h/2)+(h^3/2)},
      T4={1-(2/pi)*((2/3)*h*sqrt(1-h^2)^3
        +h*sqrt(1-h^2)+asin(h))},
      T5={1-(3*h/2)+(h^3/2)-(3*h/8)*(1-h^2)^2},
    )}
  order<-paste("T",k,sep="")
}
```

```

#Vzdálenost pozorování
D<-as.matrix(pmin((dist(gX,diag=TRUE,upper=TRUE)/2),1))
Hmx<-H(D)
#Projekční matice
P<-qr.Q(lin$qr,complete=TRUE)
N<-as.matrix(P[, (dim(qr.Q(lin$qr))[2]+1):dim(P)[1]])
M<-(N%*%t(N)) # (M%*%X[,1]=lin$res)
res<-lin$res
sig<-summary(lin)$sigma
#Testová statistika
MHM<-M%*%Hmx%*%M
test.stat = as.vector((t(res) %*% Hmx %*% res
- sig^2*sum(diag(MHM)))^2 /
(sig^4*(2*tr(((MHM-M*sum(diag(MHM))/(n-k-1))))%*%
((MHM-M*sum(diag(MHM))/(n-k-1)))))) )
names(test.stat) = NULL
return(list(test.stat = test.stat,
p.value = 1-pchisq(test.stat, df= 1,order=k)) )

```

C.2 Test aditivity podmíněných středních hodnot

```

#pomocná funkce pro dělení intervalu
> bin
function(number,splitter){sum(number>splitter)}
> Additivity.cond.test
function (x, delta=0.8, m=5) {
  require(TSA)
  # delta - oříznutí okaje okénka
  # m - dělení okénka
  n<-length(x)
  range<-max(x)-min(x)
  a<-min(x)+(1-delta)*range/2+(0:m)*delta*range/m
  X<-NULL
  k<-2
  for (i in 1:k) X = cbind(X, zlag(x, i))
  X = cbind(x, X)
  X = na.omit(X)
  Xreg<-data.frame(X[,1])

```

```

for (i in 1:k)
  Xreg = cbind(Xreg, factor(sapply(X[,i],bin,splitter=a)))
names(Xreg)<-c("t","ord1","ord2")
Xused<-subset(Xreg,ord1!=0 & ord1!=m+1 & ord2!=0 & ord2!=m+1)
a<-anova(lm(t~ord1*ord2,data=Xused))
return(list(test.stat = a[3,4], p.value = a[3,5],order=k)) }

```

C.3 Permutační test aditivity

```

> Additivity.perm.test
function (x, order,N=100) {
  #N...počet permutací v testu
  require(acepack)
  resid.ace <- function(aceobj)
  {aceobj$ty - apply(aceobj$tx, 1, sum)}
  nx <- length(x)
  tmp.mkx <- matrix(0, (nx - order), (order + 1))
  for (i in 1:order) tmp.mkx[, i] <- x[(order + 1 - i):(nx - i)]
  tmp.mkx[, (order + 1)] <- x[(order + 1):nx]
  tmp.ace <- ace(tmp.mkx[, 1:order],
  tmp.mkx[, (order + 1)], lin = 0)
  resid1 <- resid.ace(tmp.ace)
  K<-order*(order-1)/2
  h<-1
  cross <- matrix(NA, ncol = K, nrow = dim(tmp.mkx)[1])
  for (i in 1:order) for (j in i:order) if (i != j) {
    cross[,h] <- apply(tmp.mkx[, c(i, j)], 1, prod)
    h <- h + 1}
  res<- resid.ace(ace(cross,resid1, lin = 0))
  RSS0<-sum(res^2)
  RSS<-c()
  for (perm in 1:N){
    residperm<-sample(resid1)
    res<- resid.ace(ace(cross,residperm, lin = 0))
    RSS[perm]<-sum(res^2) }
  return(list(p.value = sum(RSS<RSS0)/N,order=order)) }

```

D Obsah příloženého CD

Součástí CD příloženého k práci jsou zdrojové kódy testů, které byly implementovány v R, zdrojové kódy provedených simulací a aplikace na reálných datech. Na CD jsou obsaženy soubory

1_Keenan_test.R opravená verze Keenanova testu prezentovaná v 3.1.1,

2_Hamilton_test.R implementace Hamiltonova testu ze sekce 3.7.1,

3_Testy_aditivity.R implementace testů aditivity z 4.2.1 a 4.2.3,

4_Simulace_Linearita.R zdroj pro simulace v kapitole 5.1,

5_Simulace_Aditivita.R zdroj pro simulace v kapitole 5.2,

6_Realna_data.R užití testů na reálných datech popsané v kapitole 6,

data.csv zdrojová data časové řady tvorby hrubého kapitálu.

Literatura

- ANDĚL, J. (2007). *Základy matematické statistiky*. MATFYZPRESS, Praha. ISBN 80-7378-001-1.
- BERG, A., PAPANODITIS, E. a POLITIS, D. N. (2010). A Bootstrap Test for Time Series Linearity. *Journal of Statistical Planning and Inference*, **140**(12), 3841–3857. ISSN 0378-3758.
- BREIMAN, L. a FRIEDMAN, J. H. (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. *Journal of the American Statistical Association*, **80**(391), 580–598. ISSN 0162-1459.
- BROOCK, W. A., SCHEINKMAN, J. A., DECHERT, W. D. a LEBARON, B. (1996). A Test for Independence Based on the Correlation Dimension. *Econometric Reviews*, **15**(3), 197–235.
- CHEN, R. a TSAY, R. S. (1993). Functional-Coefficient Autoregressive Models. *Journal of the American Statistical Association*, **88**(421), 298–308. ISSN 0162-1459.
- CHEN, R., LIU, J. S. a TSAY, R. S. (1995). Additivity Tests for Nonlinear Autoregression. *Biometrika*, **82**(2), 369–383. ISSN 1464-3510.
- CIPRA, T. (2008). *Finanční ekonometrie*. Ekopress, Praha. ISBN 978-8-086929-43-9.
- CRYER, J. a CHAN, K. (2008). *Time Series Analysis: With Applications in R*. Springer Texts in Statistics. Springer. ISBN 978-0-387-75958-6.
- DAHL, C. M. a GONZÁLEZ-RIVERA, G. (2003). Testing for Neglected Nonlinearity in Regression Models Based on the Theory of Random Fields. *Journal of Econometrics*, **114**(1), 141 – 164. ISSN 0304-4076.
- FULLER, W. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley. ISBN 978-0-471-55239-0.
- HAMILTON, J. D. (2001). A Parametric Approach to Flexible Nonlinear Inference. *Econometrica*, **69**(3), 537–573. ISSN 1468-0262.
- JOLLIFFE, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer, New York. ISBN 978-0-387-95442-2.

- KEENAN, D. M. (1985). A Tukey Nonadditivity-Type Test for Time Series Nonlinearity. *Biometrika*, **72**, 39–44. ISSN 1464-3510.
- LEE, T.-H., WHITE, H. a GRANGER, C. W. (1993). Testing for Neglected Nonlinearity in Time Series Models. A Comparison of Neural Network Methods and Alternative Tests. *Journal of Econometrics*, **56**(3), 269–290. ISSN 0304-4076.
- LEVINE, M. a LI, J. T. (2012). A Simple Additivity Test for Conditionally Heteroscedastic Nonlinear Autoregression. *Computational Statistics & Data Analysis*, **56**(8), 2421 – 2429. ISSN 0167-9473.
- MCLEOD, A. I. a LI, W. K. (1983). Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. *Journal of Time Series Analysis*, **4**(4), 269–273. ISSN 1467-9892.
- PRIESTLEY, M. B. (1980). State-Dependent Models: A General Approach to Non-Linear Time Series Analysis. *Journal of Time Series Analysis*, **1**(1), 47–71. ISSN 1467-9892.
- PRÁŠKOVÁ, Z. (2007). *Základy náhodných procesů II*. Učební texty. Karolinum, Praha. ISBN 978-80-246-0971-3.
- RAMSEY, J. (1968). Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society*, **31**(2), 350–371. ISSN 1369-7412.
- RAO, T. S. a GABR, M. M. (1980). A Test for Linearity of Stationary Time Series. *Journal of Time Series Analysis*, **1**(2), 145–158. ISSN 1467-9892.
- TERÄSVIRTA, T., LIN, C.-F. a GRANGER, C. W. J. (1993). Power of the Neural Network Linearity Test. *Journal of Time Series Analysis*, **14**(2), 209–220. ISSN 1467-9892.
- TONG, H. (2003). *Non-linear Time Series. A Dynamical System Approach*. Oxford University Press, New York. ISBN 0-09-852300-9.
- TSAY, R. S. (1986). Nonlinearity Tests for Time Series. *Biometrika*, **73**(2), 461–466. ISSN 1464-3510.
- TUKEY, J. W. (1949). One Degree of Freedom for Non-Additivity. *Biometrics*, **5**(3), 232–242. ISSN 0006-341X.

- WHITE, H. (1996). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press. ISBN 978-0-521-57446-4.
- ZIVOT, E. a WANG, J. (2007). *Modeling Financial Time Series with S-Plus*. International Federation for Information Processing. Springer, New York. ISBN 978-0-387-32348-0.
- ZVÁRA, K. (2008). *Regrese*. MATFYZPRESS, Praha. ISBN 978-80-7378-041-8.