

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

DIPLOMOVÁ PRÁCE

Analýza průběhu nezaměstnanosti v
Olomouckém kraji



Vedoucí diplomové práce:
prof. RNDr. Ing. Lubomír Kubáček, DrSc., Dr. h. c.
Rok odevzdání: 2011

Vypracovala:
Regina Donevska
AME, II. ročník

Prohlášení

Prohlašuji, že jsem vytvořila tuto diplomovou práci samostatně za vedení prof. RNDr. Ing. Lubomíra Kubáčka, DrSc., Dr. h. c. a že jsem v seznamu použité literatury uvedla zdroje použité při zpracování práce.

V Olomouci dne 21. 3. 2011

.....

Poděkování

Děkuji svému vedoucímu diplomové práce prof. RNDr. Ing. Lubomírovi Kubáčkovi, DrSc., Dr. h. c. za vstřícný přístup, za čas strávený konzultacemi, cenné rady, připomínky a celkovou spoluprací, kterou mi poskytoval po celou dobu psaní této práce. Dále bych chtěla poděkovat své rodině a přátelům, kteří mě po celou dobu studia podporovali a také za pomoc Bc. Lence Hudčkové při korektuře českého jazyka.

Obsah

Úvodní slovo.....	4
Použité značení	6
1 DEFINICE A POMOCNÁ TVRZENÍ	8
2 PŘÍSTUPY K ANALÝZE ČASOVÝCH ŘAD A MODEL LINEÁRNÍ REGRESE.....	11
2.1 Základní přístupy k analýze časových řad.....	12
2.2 Obecné lineární regresní modely a metoda nejmenších čtverců	13
2.3 Určení oblastí konfidence a testy hypotéz v lineárních modelech	20
3 POLYNOMICKÝ MODEL TRENDU A ČEBYŠEVOVA METODA	27
4 PERIODOGRAM	32
4.1 Test R. A. Fishera	37
4.2 Metoda skrytých period	39
5 LINEARIZACE NELINEÁRNÍCH REGRESNÍCH MODELŮ.....	41
6 OVĚŘENÍ VHODNOSTI REGRESNÍHO MODELU ZNAMÉNKOVÝM TESTEM.....	46
7 PRAKTICKÝ PŘÍKLAD	49
7.1 Nezaměstnanost a způsoby jejího měření.....	49
7.2 Analýza nezaměstnanosti v Olomouckém kraji	53
Literatura	74
Příloha – tabulky.....	76

Úvodní slovo

Analýza časových řad včetně předpovídání jejího budoucího chování se stává jednou z nejdůležitějších oblastí v rozvoji současné statistiky. Hlavním důvodem rostoucího významu této disciplíny je fakt, že se úspěšně vyrovnává s popisem dynamických systémů, s kterými často přicházíme do styku [3].

Pomocí metod analýzy časových řad lze rozumně popsat většinu skutečných řad z ekonomické praxe. Potom se snažíme ze známých hodnot sledované veličiny z minulosti až do současnosti, které tvoří tzv. časovou řadu, konstruovat odpovídající model, který popisuje, jaký vliv má časový faktor na utváření sledovaných údajů, a dále pak využít tohoto modelu k prognózování budoucího vývoje systému.

Předkládaná diplomová práce se věnuje právě této statistické disciplíně a za pomoci regresní analýzy si klade za cíl vytvořit statistický model pro časovou řadu, která udává míru nezaměstnanosti v Olomouckém kraji od roku 1999 až do roku 2008. Protože budeme modelovat ekonomický jev, o němž víme, že se často vyznačuje cyklickým pohybem s proměnnou amplitudou a proměnnou vzdáleností mezi body zvratu, budeme se snažit kromě trendu kvantifikovat také periodickou složku. Dalším cílem této práce bude určení krátkodobé prognózy vývoje míry nezaměstnanosti a následně zhodnotit kvalitu použitého modelu.

Důvodem, proč jsem si tuto časovou řadu vybrala, je fakt, že nezaměstnanost je jedním ze stále více diskutovaných témat v dnešní době. Vzniká v důsledku nesouladu mezi nabídkou práce lidí a poptávkou po práci firem a zápolí s ní více či méně úspěšně všechny země, neboť je stabilita trhu práce prioritou všech společností.

V České republice je nezaměstnanost poměrně mladým fenoménem a do popředí zájmu vstupuje až po změně politického systému, kdy v centrálně plánovaném hospodářství bylo pracovních míst dostatek pro všechny. Nástup nezaměstnanosti je spojen s přechodem na tržní hospodářství po 17. listopadu 1989.

V rámci kraje již řadu let provádí Český statistický úřad výběrové šetření pracovních sil a tím nám umožňuje sledovat vývoj nezaměstnanosti.

Tato práce na základě stanovených cílů bude rozdělena do dvou částí a to na část teoretickou a praktickou. Teoretická část je tvořena sedmi kapitolami, v nichž bude prezentován použitý matematický aparát pro zpracování časových řad. V první kapitole jsou nadefinovány základní pojmy a tvrzení, které jsou v dalším textu používány. Dále jsou popsány základní přístupy k analýze časových řad a shrnuty obecné lineární regresní modely. Následně se věnujeme polynomickému modelu trendu a podrobněji vysvětlujeme Čebyševovu metodu, která nabízí jednoduchý způsob pro odhad neznámých parametrů. Detailně se zabýváme periodogramem, jakožto dobrým ukazatelem periodicity a Fisherovým testem, na jehož základě jsme rozhodovali o významnosti hodnot periodogramu. Šestá kapitola je věnována nelineárnímu regresnímu modelu a linearizaci pomocí Taylorova rozvoje. Teoretickou část ukončíme znaménkovým testem, pomocí něhož lze ověřit vhodnost navrženého regresního modelu. Na závěr ilustrujeme získané výsledky v praktické části.

Použité značení

\mathbf{R}^n	reálný n - rozměrný euklidovský prostor
\mathbf{I}_n	jednotková matice typu $(n \times n)$
$\mathbf{0}$	nulový vektor nebo nulová matice
$h(\mathbf{A})$	hodnost matice
\mathbf{A}^T	transponovaná matice
\mathbf{A}^{-1}	inverzní matice
$Tr(\mathbf{A})$	stopa matice \mathbf{A}
\mathbf{X}, \mathbf{F}	matice plánu
q_{ij}	prvky matice $(\mathbf{X}^T \mathbf{X})^{-1}$
$\boldsymbol{\beta}$	vektor regresních koeficientů
$\hat{\boldsymbol{\beta}}$	odhad parametru $\boldsymbol{\beta}$
$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	n – rozměrný náhodný vektor má normální rozdělení pravděpodobnosti se střední hodnotou $\boldsymbol{\mu}$ a s kovarianční maticí $\boldsymbol{\Sigma}$
$\mathbf{Y} \sim_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	n – rozměrný náhodný vektor má rozdělení pravděpodobnosti se střední hodnotou $\boldsymbol{\mu}$ a s kovarianční maticí $\boldsymbol{\Sigma}$
α	riziko testu, hladina významnosti
H_0, H_A	nulová hypotéza, alternativní hypotéza
$\varphi_i(x)$	Čebyševův polynom
S_e	reziduální součet čtverců (RSC)
\mathcal{B}^k	třída borelovských množin v prostoru \mathbf{R}^k
$F_{k,n-k}(0; 1 - \alpha)$	$(1 - \alpha)$ - kvantil Fisherova-Snedecorova rozdělení s k a $n - k$ stupněmi volnosti.
$t_{n-k}(1 - \alpha)$	$(1 - \alpha)$ - kvantil Studentovy náhodné veličiny s $n - k$ stupni volnosti
$g_p^F(\alpha)$	kritická hodnota Fisherova testu na hladině významnosti α

$u_{1-\alpha}$	kvantil normovaného normálního rozdělení na hladin významnosti α
$I(\chi)$	periodogram
λ	úhlová frekvence
T_r	délka periody r - té frekvence
$WN(0, \sigma^2)$	bílý šum (white noise) s nulovou střední hodnotou a s rozptylem σ^2
$IID(0, \sigma^2)$	<i>IID</i> (independent identical defined) proces, bílý šum s nezávislými náhodnými veličinami

1 DEFINICE A POMOCNÁ TVRZENÍ

Věta 1.1 Je-li A matice typu $n \times n$ idempotentní s hodnotí $h(A) = r$, je

$$h(I_n - A) = n - r.$$

Důkaz 1.1 Je-li A idempotentní, je $I_n - A$ rovněž idempotentní. Protože víme, že hodnost idempotentní matice se rovná její stopě, je tím dokázáno tvrzení věty.

Věta 1.2 Necht' X je n - rozměrný náhodný vektor s konečnými druhými momenty. Označíme $E(X) = \mu$, $var(X) = V$. Pak pro libovolnou matici A typu $n \times n$ platí

$$E(X^T A X) = Tr(AV) + \mu^T A \mu. \quad (1)$$

Důkaz 1.2 Platí

$$var(X) = E(X - \mu)(X - \mu)^T = E(XX^T) - \mu\mu^T$$

Proto

$$E(X^T A X) = ETr(AX^T X) = Tr(AE[XX^T]) = Tr[A(V + \mu\mu^T)] = Tr(AV) + \mu^T A \mu.$$

Definice 1.1 Řekneme, že náhodný vektor $X_{p \times 1}$ má p - rozměrné normální rozdělení N_p tehdy a jen tehdy, když každá lineární funkce složek X má jednorozměrné normální rozdělení.

Věta 1.3 (o lineární transformaci) Necht' $X \sim N_p(\mu, \Sigma)$ a $d_{q \times 1}, B_{q \times p}$ jsou konstanty. Potom

$$Y = d + BX \sim N_q(d + B\mu, B\Sigma B^T).$$

Důkaz 1.3 Dle Definice 1.1 potřebujeme dokázat, že $\forall c_{q \times 1}: c^T Y \sim N_1$ tj.

$$c^T Y = c^T (d + BX) = c^T d + c^T B X = a + e^T X,$$

kde $a = c^T d$ je konstanta a $e^T = c^T B$ je vektor konstant. Protože je celý výraz ve tvaru lineární kombinace posunutý o konstantu a z předpokladu věty platí, že $c^T Y \sim N_1$.
Dopočítáme

$$E(c^T Y) = E(a + e^T X) = a + e^T \mu = c^T d + c^T B \mu,$$

$$\text{var}(\mathbf{c}^T \mathbf{Y}) = \text{var}(a + \mathbf{e}^T \mathbf{X}) = \mathbf{e}^T \boldsymbol{\Sigma} \mathbf{e} = \mathbf{c}^T \mathbf{B} \boldsymbol{\Sigma} \mathbf{B}^T \mathbf{c}.$$

Potom

$$E(\mathbf{Y}) = \mathbf{d} + \mathbf{B}\boldsymbol{\mu} \text{ a } \text{var}(\mathbf{Y}) = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T.$$

Věta 1.4 (*Pearsonova*) Necht' $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \mathbf{V})$, matice \mathbf{V} je známá pozitivně definitní matice a $\sigma^2 > 0$ neznámé. Potom platí

$$\frac{1}{\sigma^2} (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2_n(0).$$

Důkaz 1.4 Pro pozitivně definitní matice $\mathbf{V} = \mathbf{C}\mathbf{C}^T$ platí, že

$$\mathbf{V}^{-1} = (\mathbf{C}^{-1})^T \mathbf{C}^{-1}.$$

Dále nadefinuujeme náhodný vektor $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ ve tvaru

$$\mathbf{Z} = \frac{1}{\sigma} \mathbf{C}^{-1} (\mathbf{Y} - \boldsymbol{\mu}), \quad (2)$$

Protože jde o lineární transformaci náhodného vektoru $\mathbf{Y} \sim N_n$, víme z věty o lineární transformaci, že platí $\mathbf{Z} \sim N_n$. Dopočítáme

$$E(\mathbf{Z}) = E \left[\frac{1}{\sigma} \mathbf{C}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] = \mathbf{0},$$

$$\begin{aligned} \text{var}(\mathbf{Z}) &= \text{var} \left[\frac{1}{\sigma} \mathbf{C}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] = \frac{1}{\sigma^2} \mathbf{C}^{-1} \text{var}[\mathbf{Y} - \boldsymbol{\mu}] (\mathbf{C}^{-1})^T = \\ &= \frac{1}{\sigma^2} \mathbf{C}^{-1} \mathbf{C} \mathbf{C}^T (\mathbf{C}^{-1})^T = \frac{1}{\sigma^2} \mathbf{I}_n. \end{aligned}$$

Potom pro složky $Z_i, i = 1, \dots, n$ platí, že jsou nezávislé s rozdělením $\frac{1}{\sigma^2} Z_i \sim N(0,1)$.

Jestliže vyjádříme \mathbf{Y} ze vztahu (2), potom dostaneme

$$\frac{1}{\sigma} (\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{C} \mathbf{Z},$$

a tedy

$$\frac{1}{\sigma^2}(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) = \frac{1}{\sigma^2} \mathbf{Z}^T \mathbf{C}^T (\mathbf{C} \mathbf{C}^T)^{-1} \mathbf{C} \mathbf{Z} = \frac{1}{\sigma^2} \mathbf{Z}^T \mathbf{Z} = \frac{1}{\sigma^2} \sum_{i=1}^n \left(\frac{Z_i}{\sigma} \right)^2 \sim \chi_p^2(\delta),$$

$$\delta = \sum_{i=1}^n E \left(\frac{Z_i}{\sigma} \right)^2 = \frac{1}{\sigma^2} E(\mathbf{Z})^T E(\mathbf{Z}) = 0.$$

Definice 1.2 Borelovská množina je prvkem takové třídy množin, která je uzavřená na spočetném sjednocení množin a na operaci komplement, generována otevřenými množinami příslušného prostoru.

2 PŘÍSTUPY K ANALÝZE ČASOVÝCH ŘAD A MODEL LINEÁRNÍ REGRESE

Statistická data popisující společenské a ekonomické jevy v čase zapisujeme pomocí tzv. časových řad. Zápis těchto jevů pomocí časových řad umožňuje provádět nejen kvantitativní analýzu zákonitostí v jejich dosavadním průběhu, ale dává zároveň možnost prognózovat jejich vývoj. (Převzato z [9].)

Pod pojmem časová řada rozumíme posloupnost věcně a prostorově srovnatelných dat, která jsou jednoznačně uspořádána z hlediska času ve směru minulost - přítomnost. Je nutné zdůraznit, že jde o statistickou (stochastickou) časovou řadu, jejíž chování je zatíženo nejistotou, nikoliv o deterministickou časovou řadu, jejíž chování lze jednoznačně popsat nějakým matematickým vzorcem.

Z matematického hlediska můžeme říct, že časová řada je posloupnost náhodných veličin indexovaných časem t . Abychom byli přesní, je třeba odlišovat pojem „náhodný proces“ od pojmů „časová řada“. Teorie náhodných procesů časovou řadu chápe jako konkrétní realizace náhodného procesu, a tedy je analogický i vztah mezi pojmy „pozorovaná hodnota“ a „náhodná veličina“. Pojem náhodného procesu je zobecněním pojmů náhodné veličiny. Zatímco náhodná veličina je reálná funkce jedné proměnné - elementárního jevu, je náhodný proces reálnou funkcí dvou proměnných - elementárního jevu a jedné reálné proměnné, kterou obvykle bývá čas. V tomto aspektu budeme v dalším výkladu pro jednoduchost používat oba pojmy ekvivalentně.

Rozlišujeme dlouhodobé a krátkodobé časové řady. Dlouhodobé časové řady jsou řady s hodnotami sledovanými v ročních či delších obdobích. Hodnoty krátkodobé časové řady jsou sledované v obdobích kratších než je rok, například čtvrtletní, měsíční, týdenní období apod.

2.1 Základní přístupy k analýze časových řad

Volba přístupu k analýze časových řad (výběr metody) závisí na mnoha faktorech, z nichž nedůležitější jsou účel analýzy, kterým je většinou rozpoznání mechanismu generování hodnot časové řady, typ časové řady, protože ne pro každou lze uplatnit libovolnou metodu, zkušenost statistika, jakož i dostupnost výpočetní techniky a statistického softwaru.

Nejčastěji se v analýze časových řad setkáváme s následujícími základními přístupy [4]:

1. Analýza časových řad v časové doméně:
 - a) klasická dekompozice časových řad je založená na regresní analýze,
 - b) neoklasická dekompozice časových řad, tzv. Box - Jenkinsonova metodologie je založená na korelační analýze.
2. Analýza časových řad ve spektrální doméně:
 - je spektrální analýza časových řad založená na Fourierově analýze.

V následujícím výkladu zaměříme svou pozornost především na analýzu v časové doméně a to konkrétně na klasickou dekompozici.

Dekompozicí časové řady rozumíme rozklad časové řady na deterministickou a náhodnou složku. Deterministická složka dále rozkládá časovou řadu na trend, sezónní a cyklickou složku, zatímco náhodná složka je tvořena náhodnými pohyby (fluktuacemi), které modelují drobné a v jednotlivostech nepostižitelné příčiny kolísání časových řad. Trend odráží dlouhodobé změny v chování časových řad. Sezónní složka popisuje periodické změny, které se odehrávají během kalendářního roku a každý rok se pravidelně opakují. Cyklickou složku chápeme jako fluktuace kolem trendu, při nichž se pravidelně střídají fáze růstu s fázemi poklesu, přitom délka cyklu i intenzita jednotlivých fází se mohou v průběhu času měnit. Potom můžeme časovou řadu vnímat jako trend, na který se „nabalují“ periodická a náhodná složka.

Jak už bylo naznačeno, dekompoziční metody modelují pomocí regresních modelů deterministickou složku trendu, do kterého se obvykle zahrnují i cyklické složky s dlouhou periodou. O tom pojednává následující podkapitola.

2.2 Obecné lineární regresní modely a metoda nejmenších čtverců

Regresní analýza je nejpoužívanějším způsobem popisu vývoje časových řad, protože jednak umožňuje vyrovnaní pozorovaných dat časové řady a navíc tím lze prognózovat její další vývoj.

Regresní analýza je statistická metoda pro modelování závislostí jedné nebo několika (nejlépe měřitelných spojitých) vysvětlovaných náhodných veličin Y_1, \dots, Y_n na jedné nebo několika (klasicky nenáhodných v aplikacích skoro libovolných) vysvětlujících veličinách x_1, \dots, x_k . Základní motivace regresní analýzy je snaha a potřeba nepřímo působit na vysvětlované proměnné volbou, ovlivňováním nebo aspoň snadnější odhadem hodnot vysvětlujících proměnných. Hlavním požadavkem je, aby mezi vysvětlovanou proměnnou a vysvětlujícími proměnnými existoval kvantifikovatelný vztah, přesněji aby existovala matematicky popsatelná závislost vysvětlovaných proměnných na vysvětlujících proměnných. Zobrazením této závislosti je regresní model, jehož rozhodující součástí je regresní funkce. Snaha o nalezení vhodného regresního modelu je ústřední náplní vykladu o regresní analýze.

Označily jsme k počet proměnných vstupu, ale ve skutečnosti nelze uvažovat všechny významem skoro zanedbatelné vlivy a nutně se soustředíme jen na ty, které se jeví jako důležité a rozhodující. Neuvažované či nezařazené vlivy tak způsobují, že i ten nejlepší funkční předpis je zatížen chybou. Souhrnný efekt, všech neuvažovaných či svým vlivem drobných činitelů x_{k+1}, x_{k+2}, \dots , jakož i zanedbatelná nedokonalost zvoleného funkčního předpisu, bývá zvykem nazývat chybovou či poruchovou složku ε . Tato nepozorovatelná a tudíž i nezměřitelná složka v uvažované úloze způsobuje, že každé kombinaci hodnot z \mathbf{x} neodpovídá jediná kombinace hodnot z \mathbf{Y} . Rušivá složka ε je označována za náhodnou veličinu způsobující náhodné odchylky modelů od skutečnosti a nikoliv za systematicky zkreslující faktor. Běžně se v této souvislosti říká „jediný pravděpodobnostní předpoklad o náhodné veličině ε umožňuje převést neřešitelný věcný problém měření závislosti \mathbf{Y} na \mathbf{x} na dobře známou statistickou úlohu odhadu neznámých parametrů předpokládaného pravděpodobnostního rozdělení“. (Převzato z [6].) Převedení problému modelování závislostí na úlohu odhadu parametrů známého nebo předpokládaného pravděpodobnostního rozdělení vyžaduje jasnou specifikaci všech vysvětlujících proměnných a obecného tvaru regresní funkce i z tohoto důvodu, že ve

správném regresním modelu jsou důvodem existence rušivé složky ε výhradně nepodstatné (drobné, náhodné) vlivy.

Zvláštní postavení mezi regresními modely mají lineární modely.

V dalším výkladu jsem čerpala z [1], [6], [10].

Mějme náhodný vektor závisle proměnných $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ a matici daných čísel tzv. matici plánu $\mathbf{X} = (x_{ij})$ typu $n \times k$, kde $k < n$. Model ve tvaru

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3)$$

kde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$ je vektor neznámých parametrů a $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ je vektor náhodných veličin splňující podmínky

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n \quad (4)$$

se nazývá model lineární regrese. Přitom σ^2 je rovněž neznámý parametr.

Tento model se snaží vysvětlit vznik většího počtu náhodných veličin Y_1, \dots, Y_n pomocí známých vlivů x_{ij} a pomocí poměrně malého počtu parametrů β_1, \dots, β_k . Této regresi se říká lineární, protože jde o lineární závislost na $\boldsymbol{\beta}$. Přitom je pro regresní modely typické to, že se nezavádějí nadbytečné vysvětlující proměnné x_{ij} a požaduje se, aby sloupce matice \mathbf{X} byly lineárně nezávislé.

Předpoklad, že počet regresních parametrů k není větší než počet pozorování n ve spojení s předpokladem lineárně nezávislých sloupců matice \mathbf{X} (vyjadřujícím neexistenci netriviální nulové lineární kombinace sloupců této matice) znamená, že hodnost $h(\mathbf{X}) = k < n$. Potom čtvercová, symetrická a pozitivně definitní matice $\mathbf{X}^T \mathbf{X}$ je regulární s hodnotou $h(\mathbf{X}^T \mathbf{X}) = k$, ke které existuje inverzní matice $(\mathbf{X}^T \mathbf{X})^{-1}$. Požadavek, aby počet pozorování n (výrazně) převyšoval počet neznámých parametrů k , je běžný pro statistické úlohy a zvláště žádoucí pro rozumné užití regresní analýzy.

Ze vztahů (3) a (4) plyne

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \text{ a } \text{var}(\mathbf{Y}) = \sigma^2 \mathbf{I}_n.$$

Parametry β_1, \dots, β_k se odhadují metodou nejmenších čtverců, která je v regresní analýze nejužívanější, tedy z podmínky nejmenšího reziduálního součtu čtverců, která znamená minimalizovat funkce

$$(Y - X\beta)^T(Y - X\beta).$$

Věta 2.2.1 Odhady metodou nejmenších čtverců jsou daný vztahem

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (5)$$

Důkaz 2.2.1 Pro odhad parametrů β daný vztahem (5) platí

$$X^T(Y - X\hat{\beta}) = X^T Y - X^T X(X^T X)^{-1} X^T Y = 0.$$

Vzhledem k tomu dostáváme

$$\begin{aligned} (Y - X\beta)^T(Y - X\beta) &= (Y - X\hat{\beta} + X\hat{\beta} - X\beta)^T(Y - X\hat{\beta} + X\hat{\beta} - X\beta) \\ &= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (\hat{\beta} - \beta)^T X^T X(\hat{\beta} - \beta) \\ &\quad + (\hat{\beta} - \beta)^T [X^T(Y - X\hat{\beta})] + [(Y - X\hat{\beta})^T X] (\hat{\beta} - \beta) \\ &= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (\hat{\beta} - \beta)^T X^T X(\hat{\beta} - \beta) \geq (Y - X\hat{\beta})^T(Y - X\hat{\beta}), \end{aligned}$$

přičemž rovnosti může být dosaženo tehdy a jen tehdy, je-li $\beta = \hat{\beta}$. To je zřejmé proto, že matice $X^T X$ (která je vždy pozitivně semidefinitní) je regulární, takže musí být pozitivně definitní.

Následující věty uvádí některé základní vlastnosti odhadů $\hat{\beta}$.

Věta 2.2.2 Odhad $\hat{\beta}$ je nestranný tj.

$$E(\hat{\beta}) = \beta,$$

a má varianční matici

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

Důkaz 2.2.2 Podle Věty 2.2.1 dostaneme

$$E(\hat{\beta}) = E[(X^T X)^{-1} X^T Y] = (X^T X)^{-1} X^T E[Y] = (X^T X)^{-1} X^T X\beta = \beta,$$

$$\begin{aligned}\text{var}(\widehat{\boldsymbol{\beta}}) &= \text{var}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{var}[\mathbf{Y}] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\sigma^2 \mathbf{I}_n) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.\end{aligned}$$

Soustavě lineárních rovnic $\mathbf{X}^T \mathbf{X} \widehat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}$, z nichž se vektor $\widehat{\boldsymbol{\beta}}$ většinou prakticky počítá, se říká normální rovnice.

Vektor vyrovnaných hodnot

$$\widehat{\mathbf{Y}} = \mathbf{X} \widehat{\boldsymbol{\beta}} = \mathbf{H} \mathbf{Y},$$

kde projekční matice

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

a vektor reziduí

$$\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{H}) \mathbf{Y}.$$

Vektor $\widehat{\mathbf{Y}} = \mathbf{X} \widehat{\boldsymbol{\beta}}$ může být považován za nejlepší aproximaci vektoru \mathbf{Y} , jaká se dá vytvořit lineární kombinací sloupců matice \mathbf{X} .

Vyrovnáním napozorovaných n bodů se rozumí nahrazení hodnot Y_i vyrovnanými (teoretickými, odhadnutými, vypočítanými) hodnotami \widehat{Y}_i , získanými dosazením odpovídajících kombinací hodnot vysvětlujících proměnných do odhadnuté regresní funkce. Aproximativní charakter každého vyrovnaní vyplývá ze samotné podstaty stochastických vztahů mezi veličinami a je přirozeným důsledkem existence rušivé složky $\boldsymbol{\varepsilon}$ i všech zjednodušení a chyb, kterých jsme se dopustili při konstrukci regresního modelu. Dobré vyrovnaní je takové, při kterém je celková chyba nejmenší. Chyby vyrovnaní jsou běžně vyjádřené rozdílem (rezidui) $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$. Nemá-li regresní model vážnější nedostatky, je možné rezidua $\widehat{\varepsilon}_i$ považovat za odhady hodnot rušivé složky $\boldsymbol{\varepsilon}$. Požadavek nulového součtu reziduí nestačí k nalezení vhodné regresní funkce (viz Hebák str. 32, [6]). Existují různé možnosti, jak odstranit rušící se kladné a záporné chyby, jakož i různá kritéria kvality vyrovnaní. Nejpoužívanějším vyrovnávacím požadavkem i kritériem je co nejmenší reziduální součet čtverců.

Definice 2.2.1 Veličině $S_e = (\mathbf{Y} - \widehat{\mathbf{Y}})^T (\mathbf{Y} - \widehat{\mathbf{Y}})$ se říká reziduální součet čtverců (RSC).

Věta 2.2.3 Pro reziduální součet čtverců platí

$$S_e = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \mathbf{Y}, \quad (6)$$

$$S_e = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y}. \quad (7)$$

Důkaz 2.2.3 Rozepíšeme si rozdíl

$$\mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{Y}. \quad (8)$$

Nyní ukážeme, že $[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]$ je idempotentní.

$$\begin{aligned} & [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T][\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T. \end{aligned}$$

Potom protože je matice $[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]$ idempotentní a symetrická, dostáváme dosazením do (6) ihned vzorec (7). Roznásobením pak z (7) plyne

$$S_e = \mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y}.$$

Věta 2.2.4 Náhodná veličina

$$\hat{\sigma}^2 = \frac{S_e}{n - k}$$

je nestranným odhadem parametru σ^2 .

Důkaz 2.2.4 Vyjdeme ze vzorce (7). Matice $[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]$ je idempotentní. Proto její stopa je rovna její hodnoti. Nejprve dokážeme, že $h(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) = k$.

Platí

$$h(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) = h[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}] = h(\mathbf{I}_k) = k.$$

Podle Věty 1.1 dostáváme

$$h[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] = n - k.$$

Z Věty 1.2 vyplývá

$$E(S_e) = \text{Tr}[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \sigma^2 \mathbf{I}_n + \boldsymbol{\beta}^T \mathbf{X}^T [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{X} \boldsymbol{\beta}$$

$$= \sigma^2 h[I_n - X(X^T X)^{-1} X^T] = \sigma^2(n - k).$$

Nechť $\boldsymbol{\varepsilon}$ má normální rozdělení, pak \mathbf{Y} má rovněž normální rozdělení a vzhledem k předpokladům (3) a (4) platí

$$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad \mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n).$$

Předpoklad normality ponecháme až do konce této kapitoly.

Platí, že při splnění požadavku normality a dalších požadavků, vede metoda nejmenších čtverců k nejlepším odhadům parametrů regresní funkce. Nejsou-li splněny tyto podmínky, je-li kvalita výchozích údajů nízká, neplatí-li předpoklady o pravděpodobnostním chování rušivé složky, jsou-li odchylky od normality poměrně značné nebo vyskytují-li se v souboru vybočující či extrémní pozorování, je ohrožena kvalita odhadů pořízených minimalizací RSC.

Kritérium RSC dává všem pozorováním stejnou váhu. Má-li však náhodná složka $\boldsymbol{\varepsilon}$ rozdělení s těžšími konci, tedy s konci, které mají větší pravděpodobnosti než v případě normálního rozdělení, není dobré dávat všem reziduím stejnou váhu a je lepší dávat větším reziduím menší váhu a tím znevýhodňovat výjimečná (možná z hlediska dané úlohy chybná) pozorování. Samozřejmě záleží na oprávněnosti učiněných předpokladů, ale důležitá je i volba vhodného vyrovnávacího kritéria. [6]

Věta 2.2.5 Jestliže $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, potom platí $\hat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$.

Důkaz 2.2.5 Protože \mathbf{Y} má normální rozdělení a $\hat{\boldsymbol{\beta}}$ je podle (5) získáno z \mathbf{Y} lineární transformací, podle Věty 1.3 má vektor $\hat{\boldsymbol{\beta}}$ rovněž normální rozdělení. Jeho parametry již byly vypočítány ve Větě 2.2.2.

Věta 2.2.6 Jestliže $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ a $h(\mathbf{X}) = k < n$, veličina $\frac{S_e}{\sigma^2}$ má rozdělení χ^2_{n-k} .

Důkaz 2.2.6 Je jasné, že

$$E(\mathbf{Y} - \hat{\mathbf{Y}}) = E(\mathbf{Y}) - E(\hat{\mathbf{Y}}) = \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}.$$

Protože $\text{var}(\mathbf{Y}) = \sigma^2 \mathbf{I}_n$, ze vzorce (8) dostaneme

$$\text{var}(\mathbf{Y} - \hat{\mathbf{Y}}) = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \sigma^2 \mathbf{I}_n [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] = \sigma^2 [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T].$$

Protože vektor $\mathbf{Y} - \hat{\mathbf{Y}}$ má normální rozdělení, podle Věty 1.4 má náhodná veličina

$$(\mathbf{Y} - \hat{\mathbf{Y}})^T \frac{1}{\sigma^2} \mathbf{I}_n (\mathbf{Y} - \hat{\mathbf{Y}}) = \frac{S_e}{\sigma^2}$$

rozdělení χ^2 s počtem stupňů volnosti rovným hodnotě matice $\text{var}(\mathbf{Y} - \hat{\mathbf{Y}})$. V důkazu Věty 2.2.4 bylo vypočteno, že hodnota matice $[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]$ je rovna $n - k$.

Věta 2.2.7 Jestliže je vektor \mathbf{Y} normálně rozdělený, vektor $\hat{\boldsymbol{\beta}}$ a veličina $\hat{\sigma}^2$ jsou nezávislé.

Důkaz 2.2.7 V případě normality, budou náhodný vektor $\hat{\boldsymbol{\beta}}$ a $\hat{\sigma}^2$ nezávislé právě tehdy, jestliže

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\beta}}, \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= \text{cov}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{Y}] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{var}(\mathbf{Y}) [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I}_n [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \\ &= \sigma^2 [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] = \mathbf{0}. \end{aligned}$$

Věta 2.2.8 Označme q^{jj} prvky matice $(\mathbf{X}^T \mathbf{X})^{-1}$. Potom má veličina

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 q^{jj}}}$$

pro libovolné $j = 1, \dots, k$ Studentovo t - rozdělení o $n - k$ stupních volnosti.

Důkaz 2.2.8 Víme z Věty 2.2.5, že $\hat{\boldsymbol{\beta}} \sim N_k[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}]$, potom pro jednotlivé složky $\hat{\beta}_j$ platí

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 q^{jj}),$$

kde q^{jj} je j -tým diagonálním prvkem matice $(\mathbf{X}^T \mathbf{X})^{-1}$. Znormujeme a dostaneme

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 q^{jj}}} \sim N(0, 1),$$

Navíc z Věty 2.2.6. víme, že $\hat{\sigma}^2 \sim \frac{\sigma^2}{n-k} \chi^2_{n-k}$ a že náhodný vektor $\hat{\boldsymbol{\beta}}$ a $\hat{\sigma}^2$ jsou nezávislé.

Využijeme-li definice Studentovy náhodné veličiny, platí

$$\frac{\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 q^{jj}}}}{\sqrt{\frac{\hat{\sigma}^2(n-k)}{\sigma^2}}/(n-k)} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 q^{jj}}} \sim t_{n-k}.$$

Je třeba poznamenat, že velikost diagonálních prvků matice $(\mathbf{X}^T \mathbf{X})^{-1}$ je výhradně dána zvolenými, zjištěnými či napozorovanými hodnotami vysvětlujících proměnných. Na druhé straně reziduální směrodatná odchylka $\sqrt{\hat{\sigma}^2}$ je závislá na typu a kvalitě regresního modelu, na oprávněnosti předpokladů náhodné složky $\boldsymbol{\varepsilon}$ a samozřejmě i na počtu pozorování n a počtu parametrů k .

Věta 2.2.9 Náhodná veličina

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{k \hat{\sigma}^2}$$

má Fisherovo-Snedecerova rozdělení o $(k, n - k)$ stupních volnosti.

Důkaz 2.2.9 Nejprve si uvědomíme, že z Věty 2.2.5 a Pearsonovy věty bude mít kvadratická forma

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

rozdělení $\sigma^2 \chi_k^2$.

Dále využijeme platnosti tvrzení Věty 2.2.6 a Věty 2.2.7, potom na základě definice Fisherovy-Snedecorovy náhodné veličiny platí

$$\frac{\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{k \sigma^2}}{\frac{\hat{\sigma}^2(n-k)}{\sigma^2(n-k)}} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{k \hat{\sigma}^2} \sim F_{k, n-k}.$$

2.3 Určení oblastí konfidence a testy hypotéz v lineárních modelech

V dalším výkladu bude pozornost soustředěna na určení $(1 - \alpha)$ - konfidenční oblasti k - rozměrného regresního parametru $\boldsymbol{\beta}$.

Stále uvažujeme lineární regresní model

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \quad (9)$$

kde hodnost $h(\mathbf{X})$ matice \mathbf{X} typu $n \times k$ je $k < n$. Odhad $\hat{\boldsymbol{\beta}}$ je

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (10)$$

a odhad $\hat{\sigma}^2$ je

$$\frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k}. \quad (11)$$

Nechť dále \mathcal{B}^k je třída borelovských množin v prostoru \mathbf{R}^k . Nejprve uvedeme obecnou definici konfidenční oblasti.

Definice 2.3.1 Nechť k n - rozměrnému náhodnému vektoru \mathbf{Y} je přiřazena třída rozdělení pravděpodobností

$$\mathcal{F} = \{F(\cdot, \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Theta\}.$$

Nechť zobrazení

$$C(\cdot) : \mathbf{R}^n \rightarrow \mathcal{B}^k$$

má následující vlastnost

$$\forall \{\boldsymbol{\beta} \in \Theta\} P_{\boldsymbol{\beta}}\{\boldsymbol{\beta} \in C(\mathbf{Y})\} = 1 - \alpha.$$

Potom toto zobrazení, tj. náhodná množina $C(\mathbf{Y})$, se nazývá $(1 - \alpha)$ - konfidenční oblast pro parametr $\boldsymbol{\beta}$. Číslo $(1 - \alpha)$, kde α je dostatečně malé číslo z intervalu $(0,1)$, se nazývá konfidenční hladina nebo úroveň konfidence.

V praxi konfidenční oblastí nazýváme realizaci $C(\mathbf{y})$ náhodné množiny $C(\mathbf{Y})$ a má-li určitý geometrický tvar, například úsečky anebo elipsoidu, pak se pro její označení používá termín konfidenční interval, konfidenční elipsoid apod.

Nyní tedy hledáme oblast $C(\mathbf{Y})$ s náhodnými hranicemi v \mathbf{R}^k , která pokrývá neznámý vektor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$ s danou spolehlivostí $1 - \alpha$.

Věta 2.3.1 Necht' v modelu (9) má náhodný vektor \mathbf{Y} normální rozdělení pravděpodobnosti. Potom v případě, že parametr σ^2 odhadujeme statistikou (11),

$(1 - \alpha)$ - konfidenční elipsoid má tvar

$$C_{1-\alpha}(\boldsymbol{\beta}) = \left\{ \boldsymbol{\beta} \mid \boldsymbol{\beta} \in \mathbf{R}^k: (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq k \hat{\sigma}^2 F_{k, n-k}(0; 1 - \alpha) \right\},$$

kde $F_{k, n-k}(0; 1 - \alpha)$ je $(1 - \alpha)$ - kvantil Fisherova-Snedecorova rozdělení s k a $n - k$ stupni volnosti.

Důkaz 2.3.1 Rozdělení pravděpodobnosti pro Fisherovo- Snedecorovo rozdělení jsme si odvodili ve Větě 2.2.9. Pak

$$\forall \{\boldsymbol{\beta} \in \Theta\} P_{\boldsymbol{\beta}} \left\{ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq k \hat{\sigma}^2 F_{k, n-k}(0; 1 - \alpha) \right\} = 1 - \alpha,$$

a pro k - rozměrný elipsoid dostáváme

$$C_{1-\alpha}(\boldsymbol{\beta}) = \left\{ \boldsymbol{\beta} \mid \boldsymbol{\beta} \in \mathbf{R}^k: (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq k \hat{\sigma}^2 F_{k, n-k}(0; 1 - \alpha) \right\}.$$

V předchozím odstavci jsme získali $(1 - \alpha)$ - konfidenční elipsoid pro parametr $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$, nyní se budeme zajímat o konfidenční elipsoidy pro jednotlivé parametry. Výklad začneme případem, kdy se zajímáme o jediný neznámý parametr a protože uvažujeme parametry v \mathbf{R}^1 (parametrický prostor je přímka, polopřímka nebo úsečka), jde o určení $(1 - \alpha)$ - konfidenčního intervalu pro $\beta_j, j = 1, \dots, k$.

Omezíme se na oboustranné intervaly spolehlivosti (T_d, T_h) , kde počítáme hodnoty dvou statistik – dolní meze t_d a horní meze t_h určených tak, aby interval mezi nimi s vysokou pravděpodobností pokrýval neznámý parametr, tedy

$$P_{\beta_j} \{T_d \leq \beta_j \leq T_h\} = 1 - \alpha.$$

Při konstrukci konfidenčního intervalu pro $\beta_j, j = 1, \dots, k$ využijeme Větu 2.2.8.

Potom platí

$$P_{\beta_j} \left\{ -t_{n-k} \left(1 - \frac{\alpha}{2} \right) \leq \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 q^{jj}}} \leq t_{n-k} \left(1 - \frac{\alpha}{2} \right) \right\} = 1 - \alpha,$$

kde $t_{n-k}\left(1 - \frac{\alpha}{2}\right)$ je $\left(1 - \frac{\alpha}{2}\right)$ - kvantil Studentovy náhodné veličiny s $n - k$ stupni volnosti, přičemž hledaný konfidenční interval bude mít po úpravě tvar

$$\left[\hat{\beta}_j - \sqrt{\hat{\sigma}^2 q^{jj}} t_{n-p}\left(1 - \frac{\alpha}{2}\right), \hat{\beta}_j + \sqrt{\hat{\sigma}^2 q^{jj}} t_{n-p}\left(1 - \frac{\alpha}{2}\right)\right]. \quad (12)$$

Je vhodné ještě tento výklad doplnit o informace týkající se délky konfidenčních intervalů, která závisí na variabilitě sledované náhodné veličiny v populaci, na rozsahu výběru a na zvolené spolehlivosti odhadu. Přesnost odhadů a jejich funkcí v lineárním regresním modelu je, kromě rozsahu výběru a spolehlivosti odhadu, závislá na velikosti neznámého rozptylu σ^2 rušivé složky ε , na variabilitě vysvětlujících proměnných, ale v neposlední řadě i na intenzitě vzájemné lineární závislosti vysvětlujících proměnných. Při použití metody nejmenších čtverců k odhadu regresních parametrů vysoká vzájemná lineární závislost vysvětlujících proměnných nepříznivě ovlivňuje přesnost odhadu, stejně jako nesplněný předpoklad o konstantnosti rozptylů či nulových kovariancích, ale i přehlédnutí či nevyužití nevýběrových informací, které mohou mít zcela různou podobu a mohou být odlišné.

Zvýšení rozsahu náhodného výběru má při použití konzistentních odhadů za následek zvýšení přesnosti odhadů a tedy pravděpodobností menších výběrových chyb. Zvláště pro malé rozsahy výběru (ale nejen pro ně) jsou intervaly spolehlivosti velmi důležitou informací o kvalitě použitého bodového odhadu, neboť výrazně rozšiřují představu získanou výpočtem (spíše odhadem) směrodatné chyby odhadu. Je vidět, že zvětšení $\hat{\sigma}^2$ nebo q^{jj} má za následek prodloužení intervalů spolehlivosti a tím snížení přesností odhadu j -tého parametru.

Při vyjádření výsledků zkonstruovaných konfidenčních oblastí se projevuje sklon k usuzování, který je charakteristický pro testování hypotéz. Přitom platí skutečnost, že mezi oběma variantami usuzování je úzká souvislost. Lze říci, že jak při určení konfidenční oblasti, tak při testování hypotéz odpovídáme na otázku, které hodnoty parametru β jsou přijatelné s ohledem na získaná pozorování. Podaří-li se nám množinu přijatelných hodnot popsat, pak jsme tím určili konfidenční oblast, zatímco probíráme-li jednotlivé možné hodnoty β a sledujeme, zda jsou v souladu s daty, provádíme testování hypotéz.

V následující části této kapitoly se budeme zabývat testy hypotéz. Jedná se o postup, který umožňuje na základě rozhodovacího pravidla rozhodnout mezi testovanou hypotézou H_0 a alternativní hypotézou H_A , která ji popírá. Rozhodnutí vyplývá z hodnoty testového kritéria, což je statistika zvolená tak, aby zaručovala výhodné vlastnosti testu. Množina přípustných hodnot testového kritéria, tj. výběrový prostor, se rozpadá na dvě části: kritický obor W_α a akceptační obor A_α , přičemž platí, že akceptační obor A_α představuje komplement kritického oboru W_α . Pro kritický obor W_α platí, že když H_0 je pravdivá, potom testové kritérium se realizuje v množině W_α s pravděpodobností nejvýše α , $0 < \alpha < 1$, kterému říkáme hladina významnosti. Číslo α se nazývá též pravděpodobnost chyby prvního druhu anebo rizikem testu, protože s touto pravděpodobností zamítáme nulovou hypotézu v případě, že je pravdivá. Naopak platí, že když se hodnota testového kritéria realizuje v W_α a vyslovíme se ve prospěch H_0 , můžeme se dopustit chyby druhého druhu, platí-li ve skutečnosti H_A . Pravděpodobnost výskytu testového kritéria ve W_α za předpokladu, že platí H_A , se nazývá silou testu, se kterou je kvantitativně vyjádřena schopnost testu rozpoznat platnost alternativní hypotézy. Snahou je provést test tak, aby pravděpodobnost obou chyb byla minimální, přičemž nastává problém vyvážení rizik chyb obou druhů. Z tohoto důvodu se postupuje tak, že se předem volí hladina významnosti, čímž zůstává pod kontrolou chyba prvního druhu.

V následující sekci si uvedeme větu, kterou využijeme v rámci testování lineárních hypotéz v lineárních modelech.

Věta 2.3.2 Necht' $Y \sim N_n(X\beta, \sigma^2 I_n)$, $\beta \in R^k$, $\sigma^2 > 0$, $h(X) = k < n$. Necht' nulová hypotéza o parametru β je

$$H_0: H\beta + h = 0,$$

kde H je matice typu $h \times k$ s hodnotí $h(H) = h < k$ a alternativní hypotéza necht' má tvar

$$H_A: H\beta + h \neq 0.$$

Definujeme-li náhodné veličiny vztahy

$$R_0^2 = \min\{(Y - X\beta)^T(Y - X\beta): \beta \in R^k\},$$

$$R_1^2 = \min\{(Y - X\beta)^T(Y - X\beta): H\beta + h = 0\},$$

potom

1. $R_0^2 \sim \sigma^2 \chi^2_{n-k}(0)$,
2. $R_1^2 - R_0^2 \sim \sigma^2 \chi^2_h(0)$ v případě platnosti nulové hypotézy,
 $R_1^2 - R_0^2 \sim \sigma^2 \chi^2_h(\delta)$ v případě platnosti alternativní hypotézy, kde

$$\delta = (\mathbf{H}\boldsymbol{\beta} + \mathbf{h})^T [\mathbf{H}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{H}^T]^{-1} (\mathbf{H}\boldsymbol{\beta} + \mathbf{h}) / \sigma^2,$$
3. náhodné veličiny R_0^2 a $R_1^2 - R_0^2$ jsou nezávislé.

Důkaz 2.3.2

1. ve Větě 2.2.1 jsme dokázali, že

$$R_0^2 = \min_{\boldsymbol{\beta} \in \mathbb{R}^k} \{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

kde za předpokladu normality $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \sim N_k[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}]$.

Ve Větě 2.2.4 bylo dokázáno, že

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k} \sim \frac{\sigma^2}{n - k} \chi^2_{n-k}(0),$$

odkud je tvrzení 1. zřejmé.

2. a 3. viz [10], str. 88.

Důsledek 2.3.1 Jestliže je σ^2 neznámé a odhadujeme-li ho statistikou $\hat{\sigma}^2$ tvaru (11), potom hypotézu H_0 zamítáme s rizikem α , když

$$\frac{\frac{R_1^2 - R_0^2}{h}}{\frac{R_0^2}{n - k}} = \frac{(\mathbf{H}\hat{\boldsymbol{\beta}} + \mathbf{h})^T [\mathbf{H}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{H}^T]}{h\hat{\sigma}^2} \geq F_{h, n-k}(0; 1 - \alpha).$$

Jestliže testujeme hypotézu

$$H_0: \beta_j = \beta_j^0,$$

kde β_j^0 je hypotetická hodnota, a odhadujeme-li neznámé σ^2 statistikou $\hat{\sigma}^2$ tvaru (11), a použijeme-li Větu 2.2.8, potom hypotézu H_0 zamítáme s rizikem α , když

$$\frac{|\hat{\beta}_j - \beta_j^0|}{\sqrt{\hat{\sigma}^2 q^{jj}}} \geq t_{n-k} \left(1 - \frac{\alpha}{2}\right). \quad (13)$$

Poznámka 2.3.1 Pro $\beta_j^0 = 0$ provádíme tzv. Studentův t-test o významnosti jednotlivých parametrů β_j . Platí-li, že obsahuje-li $(1 - \alpha)100\%$ konfidenční interval pro β_j nulu, nemůžeme při zvoleném riziku α vyloučit, že neexistuje potřeba j -té vysvětlující proměnné v regresním modelu. Pak ovšem nezamítneme na hladině významnosti α testovanou nulovou hypotézu. Neobsahuje-li konfidenční interval nulu, pak nulovou hypotézu zamítneme a tvrdíme, že s pravděpodobností nejméně $1 - \alpha$ se prokázala užitečnost příslušné vysvětlující proměnné pro regresní rovnici.

3 POLYNOMICKÝ MODEL TRENDU A ČEBYŠEVOVA METODA

Při tvorbě této kapitoly jsem vycházela z [10], kde je podrobněji popsána celá problematika Čebyševovy metody.

Uvažujeme lineární regresní model

$$Y = X\beta + \varepsilon, h(X) = h(X^T X) = p + 1 = k, n > p + 1, E(\varepsilon) = \mathbf{0}, \text{var}(\varepsilon) = \sigma^2 I_n.$$

Předpokládejme dále, že trendová funkce $f(t)$ patří do třídy funkcí, které jsou určeny konečným počtem parametrů

$$f(t) = \beta_0 + \beta_1 \varphi_1(t) + \dots + \beta_p \varphi_p(t),$$

kde β_0, \dots, β_p jsou neznámé parametry a $\varphi_0(t), \dots, \varphi_p(t)$ jsou známé funkce, jež neobsahují regresní parametry.

Z velkého okruhu trendových funkcí, které vedou k lineárnímu regresnímu modelu, se zaměříme na polynomický trend tj.

$$f(t) = \beta_0 + \beta_1 t + \dots + \beta_p t^p,$$

kde v případě, že uvažujeme časovou řadu o n pozorováních v časových okamžicích

$t_1 < \dots < t_n$ je matice plánu ve tvaru

$$X = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^p \end{pmatrix}.$$

Víme, že k odhadu regresních parametrů β se použije rovnice (10). Přitom ale v případě polynomického modelu hlavně pro větší stupeň polynomu je vyčíslení matice $(X^T X)^{-1}$ nepohodlné a hlavně numericky málo stabilní. Tím vycházejí často odhady bez fyzikálního smyslu. Regresní křivka sice prochází v těsné blízkosti pozorovaných hodnot, ale buď mezi nimi silně osciluje, nebo je systematicky posunuta. Proto by bylo vhodné pro další analýzu, aby matice $X^T X$ byla diagonální, a tím skalární součiny všech různých dvojic sloupců matice X byly rovny nule. Diagonalita matice $X^T X$ vyžaduje, aby

$$\sum_{i=1}^n \varphi_j(t_i) \varphi_k(t_i) = \begin{cases} 0 & \text{pro } k \neq j \\ \sum_{i=1}^n \varphi_j^2(t_i) & \text{pro } k = j \end{cases} \quad (14)$$

Z rovnice (14) plyne, že diagonalita matice $\mathbf{X}^T \mathbf{X}$ při polynomicke regresi dosáhneme použitím tzv. Čebyševovy metody, což je případ konstrukcí ortogonálních funkcí. Užívání ortogonálních funkcí patří k nejčastějším postupům vedoucím ke značnému zjednodušení statistické analýzy.

Definice 3.1 Čebyševovými polynomy nazveme takovou n -tici polynomů

$$\varphi_i(\cdot), \quad i = 0, 1, \dots, p.$$

Pro danou n - tici bodů x_1, \dots, x_n , pro které platí

$$\sum_{s=1}^n \varphi_i(x_s) \varphi_j(x_s) = 0$$

pro $i \neq j$, $i, j = 0, 1, \dots, p$.

Následující věta ukazuje postup, jak lze Čebyševovy polynomy $\varphi_0(\cdot), \varphi_1(\cdot), \dots$ na R^1 určit.

Věta 3.1 Necht' $\varphi_0(x) = 1$, $x \in R^1$. Čebyševův polynom s -tého stupně je dán vztahem

$$\varphi_s(x) = x^s + a_{s,s-1} \varphi_{s-1}(x) + \dots + a_{s,0} \varphi_0(x),$$

kde

$$\sum_{i=1}^n \varphi_s(x_i) \varphi_k(x_i) = 0, \quad k = 0, 1, \dots, s-1,$$

a pro výpočet koeficientu platí

$$a_{s,k} = -\frac{\sum_{i=1}^n x_i^s \varphi_k(x_i)}{\sum_{i=1}^n \varphi_k^2(x_i)}.$$

Důkaz 3.1 Z předpokladu věty je funkce $\varphi_1(x)$ určena vztahem

$$\varphi_1(x) = x + a_{1,0} \varphi_0(x), \quad (15)$$

kde pro výpočet koeficientu $a_{1,0}$ využijeme podmínku ortogonality, podle které proto, aby funkce byly ortogonální pro dané hodnoty $x_i, i = 1, \dots, n$ musí být jejich skalární součin nulový, tj. $\varphi_1(x)$ splňuje podmínku $\sum_{i=1}^n \varphi_0(x_i) \varphi_1(x_i) = 0$. Z této rovnice se určí

$$a_{1,0} = -\frac{\sum_{i=1}^n x_i \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} = -\frac{1}{n} \sum_{i=1}^n x_i = -\bar{x}. \quad (16)$$

Dosazením vzorce pro $a_{1,0}$ z (16) do pravé strany rovnice (15) dostaneme

$$\varphi_1(x) = x - \frac{\sum_{i=1}^n x_i \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x) = x - \bar{x}.$$

Dále určíme funkci $\varphi_2(x)$ ze vztahu

$$\varphi_1(x) = x^2 + a_{2,1} \varphi_1(x) + a_{2,0} \varphi_0(x),$$

tak, že postupujeme analogicky, tedy pro koeficienty $a_{2,1}$ a $a_{2,0}$ plyne že

$$\begin{aligned} \sum_{i=1}^n \varphi_2(x_i) \varphi_0(x_i) &= 0, & a_{2,0} &= -\frac{\sum_{i=1}^n x_i^2 \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)}, \\ \sum_{i=1}^n \varphi_2(x_i) \varphi_1(x_i) &= 0, & a_{2,1} &= -\frac{\sum_{i=1}^n x_i^2 \varphi_1(x_i)}{\sum_{i=1}^n \varphi_1^2(x_i)}. \end{aligned}$$

Konečně pro $\varphi_2(x)$ dostaneme

$$\varphi_2(x) = x^2 - \frac{\sum_{i=1}^n x_i^2 \varphi_1(x_i)}{\sum_{i=1}^n \varphi_1^2(x_i)} \varphi_1(x) - \frac{\sum_{i=1}^n x_i^2 \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x).$$

Protože

$$\sum_{i=1}^n \varphi_s(x_i) \varphi_k(x_i) = \sum_{i=1}^n x_i^s \varphi_k(x_i) + a_{s,k} \sum_{i=1}^n \varphi_k^2(x_i),$$

je zřejmé

$$a_{s,k} = -\frac{\sum_{i=1}^n x_i^s \varphi_k(x_i)}{\sum_{i=1}^n \varphi_k^2(x_i)}.$$

Lineární regresní model lze nyní vyjádřit ve tvaru

$$\mathbf{Y} = \begin{pmatrix} \varphi_0(x_1), & \dots, & \varphi_{k-1}(x_1) \\ \varphi_0(x_2), & \dots, & \varphi_{k-1}(x_2) \\ \vdots & \ddots & \vdots \\ \varphi_0(x_n), & \dots, & \varphi_{k-1}(x_n) \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{k-1} \end{pmatrix} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \text{var}(\mathbf{Y}) = \sigma^2 \mathbf{I}.$$

Zřejmě

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} \sum_{i=1}^n \varphi_0^2(x_i), & 0, & \dots, & 0 \\ 0, & \sum_{i=1}^n \varphi_1^2(x_i), & \dots, & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots, & \sum_{i=1}^n \varphi_{k-1}^2(x_i) \end{pmatrix},$$

protože jsou $\varphi_j(x)$ ortogonální, lze odhady regresních parametrů γ_j získat přímo dosazením do vztahu

$$\hat{\gamma}_j = \frac{\sum_{i=1}^n y_i \varphi_{j-1}(x_i)}{\sum_{i=1}^n \varphi_{j-1}^2(x_i)}, \quad j = 1, \dots, k, \quad (17)$$

a pro odhad rozptylu platí

$$\text{var}(\hat{\gamma}_j) = \frac{\hat{\sigma}^2}{\sum_{i=1}^n \varphi_j^2(x_i)}. \quad (18)$$

Využitím tzv. Čebyševovy metody lze jednoduše nalézt odhady parametrů lineárních modelů, tedy znalost ortogonálních funkcí vede vyhodnocení dat lineární regrese na pouhé dosazování do výše uvedených vztahů.

Kromě neznámých parametrů $\gamma_j, j = 1, \dots, k$, zbývá určit optimální stupeň regresního polynomu. V tomto smyslu je třeba podotknout další výhodu ortogonálních funkcí, a to že při vynechání některých funkcí se hodnota γ_j u zbývajících funkcí $\varphi_j(x)$ nezmění.

Pro odhad stupně polynomu lze využít dvě intuitivní metody [4]. Metoda „od nejnižšího stupně k nejvyššímu“ spočívá v tom, že testujeme významnost regresních

parametrů na základě Věty 2.2.8 a Důsledku 2.3.1 tak, že začneme s polynomem nultého stupně a postupně stupeň zvyšujeme, je-li regresní parametr významný. Další metoda je „od maximálního stupně dolů“, kde postupujeme stejně jako u první metody tak, že začneme od maximálního stupně a postupně snižujeme stupeň polynomu, je-li regresní parametr nevýznamný.

4 PERIODOGRAM

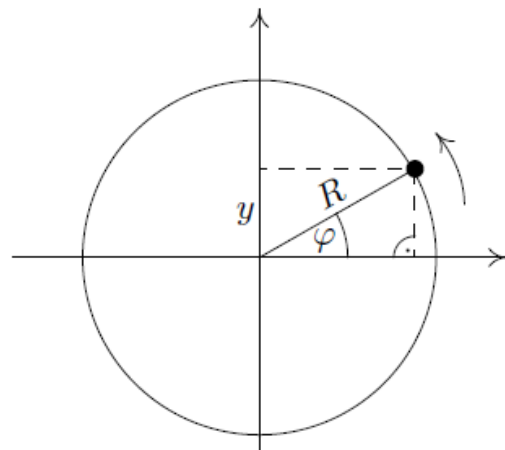
Cílem analýzy časové řady Míra nezaměstnanosti v Olomouckém kraji je rozpoznat její cyklické chování a následně předpovědět její budoucí vývoj. Proto budeme na zkoumanou časovou řadu pohlížet jako na směs periodických složek o různých frekvencích. Tento přístup nese označení spektrální analýza časových řad nebo také analýza časových řad ve spektrální doméně.

Nejprve si uvedeme základní pojmy a vztahy, které budeme potřebovat při zpracování periodických jevů.

Základní veličina, která charakterizuje periodické děje, je perioda T . Je to čas, po kterém se periodicky pohyb opakuje, tj. udává trvání jednoho kompletního cyklu. Počet opakování za jednotku času je frekvence $f = 1/T$, její jednotka je $\text{Hz} = \text{s}^{-1}$. Upozorníme ale, že v souvislosti s analýzou časových řad se budeme zabývat tzv. úhlovou frekvencí λ , která se udává v radiánech za uvažovanou časovou jednotku, což je časový interval mezi dvěma sousedními pozorováními řady.

Pro jasnější představu pojmu úhlová frekvence si pomůžeme s tzv. rovnoměrným pohybem po kružnici (viz Obrázek 1). Velikost úhlu φ , který nazýváme okamžitou fází, se s pohybem kuličky po kružnici zvětšuje. Okamžitou fázi vyjádříme pomocí úhlové rychlosti λ , která udává úhel, jenž kulička uběhne, dělený příslušným časem. Je tedy zřejmé, že při rovnoměrném pohybu po kružnici je λ konstantní, tj. že za každou časovou jednotku uběhne stejně velký oblouk.

Platí, že uběhnutý úhel se rovná úhlové rychlosti násobené časem



Obrázek 1

$$\varphi = \lambda \cdot t . \quad (19)$$

Víme, že jedné otočce odpovídá plný úhel 2π a že pro její uskutečnění je potřebný čas rovný jedné periodě. Použijeme vztah (19) a můžeme zapsat, že $2\pi = \lambda \cdot T$, tedy že

$$\lambda = \frac{2\pi}{T}. \quad (20)$$

Využijeme-li vztah mezi periodou a frekvencí, tj. $f = \frac{1}{T}$, pak můžeme (20) přepsat ve tvaru $\lambda = 2\pi f$. Platí, že frekvence f a úhlová frekvence λ jsou svázány přímo úměrou - když roste jedna, roste i druhá a zároveň platí, že perioda T je s f a λ nepřímo úměrná.

Potom pro libovolnou periodickou funkci času frekvence 2π radiánů za časovou jednotku znamená, že za uvažovanou časovou jednotku se uskuteční právě jeden z opakujících se cyklů dané funkce, zatímco při frekvenci π radiánů za časovou jednotku se cyklus uskuteční až během dvou časových jednotek. Je zřejmé, že čím je frekvence funkce větší, tím častěji se v jejím průběhu střídají jednotlivé cykly.

Uvažujeme nyní takové časové řady, které můžeme rozložit na součet harmonických frekvencí, jejichž délky period lze vyjádřit jako podíl $T_r = \frac{N}{r}$, kde N je počet pozorování a $0 < r \leq N$. Označíme dále

$$f_r = \frac{1}{T_r} = \frac{r}{N} \quad r - \text{tá frekvence,}$$

$$\lambda_r = 2\pi f_r = 2\pi \frac{r}{N} \quad r - \text{tá úhlová frekvence (tzv. Fourierovy frekvence).}$$

Maximální délka periody, kterou jsme schopni určit, je rovna počtu pozorování, tj. $T_{max} = N$, tedy $r = 1$, a minimální frekvence má velikost $\lambda_{min} = \frac{2\pi}{N}$. Nyquistova frekvence je největší frekvence, která poukazuje na charakter periodického chování dané funkce. Tato frekvence má hodnotu π radiánů za časovou jednotku a k ní odpovídající délka periody je $T_{min} = 2$, což je nejkratší zjištěitelná perioda. Je zřejmé, že když bude délka časové řady N malá ve srovnání s délkou skutečné periody $T_r = \frac{2\pi}{\lambda_r}$, bude se tato perioda jevit spíše jako trend a na druhou stranu pak příliš krátké periody nelze správně rozeznat. Z předchozího plyne, že

$$r = \begin{cases} 1, 2, \dots, \frac{N}{2} & \text{pro } N \text{ sudé} \\ 1, 2, \dots, \frac{N-1}{2} & \text{pro } N \text{ liché} \end{cases}$$

Pro nalezení významných periodických složek provedeme analýzu pomocí periodogramu. Periodogram byl jedním z nejstarších statistických nástrojů pro zkoumání periodicity v časových řadách. Zavedl ho Arthur Schuster (1897), který ukázal, že periodogram může poskytnout informaci o periodických komponentech v časové řadě. Je založen na vyjádření původních hodnot časové řady ve formě goniometrických funkcí při zahrnutí interference vlnění. Funguje na principu, že když údaje časové řady obsahují periodický člen s jistou periodou, pak v tomto bodě dosahuje jednak hlavního extrému a jednak dalších dodatečných vrcholů, které jsou odezvami extrému hlavního. Protože periodogram udává nadbytečné množství subperiod, jež často neexistují, a je potřeba otestovat, zda sledovaná časová řada významnou periodu obsahuje. K tomuto testování použijeme Fisherův test.

V následujícím výkladu budeme věnovat pozornost významu periodogramu pouze v diskrétním případě, budou odvozeny a vysvětleny veškeré pojmy a vztahy s ním související.

Definice 4.1 Necht' Y_1, \dots, Y_N je konečná posloupnost náhodných veličin. Funkce $I(\lambda)$ definovaná vztahem

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^N Y_t e^{-it\lambda} \right|^2, \lambda \in \langle -\pi, \pi \rangle \quad (21)$$

se nazývá periodogram posloupnosti Y_1, \dots, Y_N .

Lemma 4.1 Položíme

$$A(\lambda) = \sqrt{\frac{2}{N}} \sum_{t=1}^N Y_t \cos t\lambda, B(\lambda) = \sqrt{\frac{2}{N}} \sum_{t=1}^N Y_t \sin t\lambda.$$

Pak platí

$$I(\lambda) = \frac{1}{4\pi} [A^2(\lambda) + B^2(\lambda)]. \quad (22)$$

Důkaz 4.1 Přepíšeme si funkci $I(\lambda)$ ze vztahu (21) s využitím Eulerova vzorce pro formální zápis tzv. komplexní exponenciální funkci $e^{i\lambda}$ definovanou jako

$$e^{i\lambda} = \cos \lambda + i \sin \lambda.$$

Potom dostaneme

$$\begin{aligned} I(\lambda) &= \frac{1}{2\pi N} \left| \sum_{t=1}^N Y_t e^{-it\lambda} \right|^2 = \frac{1}{2\pi N} \left| \sum_{t=1}^N Y_t \cos t\lambda - i \sum_{t=1}^N Y_t \sin t\lambda \right|^2 = \\ &= \frac{1}{2\pi N} \left[\left(\sum_{t=1}^N Y_t \cos t\lambda \right)^2 + \left(\sum_{t=1}^N Y_t \sin t\lambda \right)^2 \right] = \frac{1}{4\pi} [A^2(\lambda) + B^2(\lambda)]. \end{aligned}$$

Toto Lemma použijeme k numerickému výpočtu hodnot periodogramu.

Dále si ukážeme, že periodogram je dobrým ukazatelem periodicity. Předpokládejme, že náhodná posloupnost $\{Y_t\}$, $t \in \mathbf{Z}$ vyhovuje modelu

$$Y_t = \sum_{r=1}^p a_r e^{it\lambda_r} + X_t, t \in \mathbf{Z} \quad (23)$$

kde $\lambda_r \in \langle -\pi, \pi \rangle$, $\lambda_i \neq \lambda_j$, $\forall i \neq j$, $i, j, r = 1, \dots, p$, $a_r \neq 0$, $\forall r = 1, \dots, p$ a $X_t \sim IID(0, \sigma^2)$. Jedná se o situaci, kdy periodická složka $\sum_{r=1}^p a_r e^{it\lambda_r}$ je překrytá posloupností náhodných veličin $\{Y_t\}$. Když do definice periodogramu (21) dosadíme tento předpoklad, pak je

$$\begin{aligned} \frac{1}{2\pi N} \left| \sum_{t=1}^N Y_t e^{-it\lambda} \right|^2 &= \frac{1}{2\pi N} \left| \sum_{t=1}^N \left(\sum_{r=1}^p a_r e^{it\lambda_r} + X_t \right) e^{-it\lambda} \right|^2 = \\ &= \frac{1}{2\pi} \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N \sum_{r=1}^p a_r e^{it(\lambda_r - \lambda)} + \frac{1}{\sqrt{N}} \sum_{t=1}^N X_t e^{-it\lambda} \right|^2. \end{aligned}$$

Pro přehlednost označíme členy tohoto vztahu jako

$$I_N^{(1)}(\lambda) = \frac{1}{\sqrt{N}} \sum_{t=1}^N \sum_{r=1}^p a_r e^{it(\lambda_r - \lambda)} \quad \text{a} \quad I_N^{(2)}(\lambda) = \frac{1}{\sqrt{N}} \sum_{t=1}^N X_t e^{-it\lambda}.$$

Jestliže je $\lambda \neq \lambda_r$, $\forall r = 1, \dots, p$, pak

$$I_N^{(1)}(\lambda) = \frac{1}{\sqrt{N}} \sum_{t=1}^N \sum_{r=1}^p a_r e^{it(\lambda_r - \lambda)} = \frac{1}{\sqrt{N}} \sum_{r=1}^p a_r \sum_{t=1}^N e^{it(\lambda_r - \lambda)}.$$

Vypočítáme součet geometrické řady tj.

$$\sum_{t=1}^N e^{it(\lambda_r - \lambda)} = e^{i(\lambda_r - \lambda)} \frac{e^{iN(\lambda_r - \lambda)} - 1}{e^{i(\lambda_r - \lambda)} - 1}$$

a dostaneme, že

$$I_N^{(1)}(\lambda) = \frac{1}{\sqrt{N}} \sum_{r=1}^p a_r e^{i(\lambda_r - \lambda)} \frac{e^{iN(\lambda_r - \lambda)} - 1}{e^{i(\lambda_r - \lambda)} - 1}.$$

Je přitom zřejmé, že platí

$$\lim_{N \rightarrow \infty} I_N^{(1)}(\lambda) = 0.$$

Nyní vezmeme případ když $\lambda = \lambda_r$, $r = 1, \dots, p$, potom

$$\begin{aligned} I_N^{(1)}(\lambda) &= \frac{1}{\sqrt{N}} N a_k + \frac{1}{\sqrt{N}} \sum_{\substack{r=1 \\ r \neq k}}^p a_r e^{i(\lambda_r - \lambda)} \frac{e^{iN(\lambda_r - \lambda)} - 1}{e^{i(\lambda_r - \lambda)} - 1} = \\ &= \sqrt{N} a_k + \frac{1}{\sqrt{N}} \sum_{\substack{r=1 \\ r \neq k}}^p a_r e^{i(\lambda_r - \lambda)} \frac{e^{iN(\lambda_r - \lambda)} - 1}{e^{i(\lambda_r - \lambda)} - 1}, \end{aligned}$$

a pro $N \rightarrow \infty$ vzrůstá jeho absolutní hodnota nade všechny meze.

Druhý člen $I_N^{(2)}(\lambda)$ je náhodná veličina s nulovou střední hodnotou a s rozptylem

$$\begin{aligned} \text{var} I_N^{(2)}(\lambda) &= E \left| I_N^{(2)}(\lambda) \right|^2 = E \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N X_t e^{-it\lambda} \right|^2 = \frac{1}{N} E \left(\sum_{t=1}^N X_t e^{-it\lambda} \right) \left(\sum_{s=1}^N X_s e^{-is\lambda} \right) \\ &= \frac{1}{N} \sum_{t=1}^N \sum_{s=1}^N e^{-i(t-s)\lambda} E(X_t X_s) = \frac{1}{N} \sum_{t=1}^N E X_t^2 = \frac{1}{N} N \sigma^2 = \sigma^2. \end{aligned}$$

Odtud vidíme, že rozptyl na N nezávisí a pokud bude náhodná posloupnost splňovat model (23), pak pro velká N bude periodogram mít v bodech $\lambda_1, \dots, \lambda_p$ výrazně velké hodnoty, jinde pak jeho hodnoty budou relativně malé a budou kolísat kolem $\sigma^2/2\pi$. V průběhu

periodogramu budou významná lokální maxima identifikovat periodickou strukturu uvažovaného modelu, ale jak už jsme zmínili, je potřeba použít vhodný statistický test, abychom rozhodli, které hodnoty periodogramu lze považovat za významně velké ve srovnání s hodnotami ostatními.

4.1 Test R. A. Fishera

Test, který byl navržen v roce 1929 R. A. Fisherem, se úspěšně vyrovnal se zmíněným problémem. Tento test nyní stručně popíšeme.

Uvažujeme posloupnost náhodných veličin Y_1, \dots, Y_N . Budeme testovat hypotézu

$$H_0: Y_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2) \quad (24)$$

tj. že tato posloupnost neobsahuje žádnou periodickou složku a je přímo rovna bílému šumu s rozdělením $N(0, \sigma^2)$ proti alternativní hypotéze typu (23).

Testová statistika ve Fisherově testu je založena na hodnotách periodogramu $I(\lambda_r)$ vypočítaných ve Fourierových frekvencích $\lambda_r = \frac{2\pi r}{N}$. Za předpokladu platnosti nulové hypotézy by žádná z hodnot periodogramu neměla být významně větší než zbývající hodnoty.

Postup tohoto testu je následující [13]:

1. Srovnáme hodnoty periodogramu sestupně podle velikostí tj.

$$I_{(1)} \geq I_{(2)} \geq \dots \geq I_{(p)},$$

2. Sestavíme tzv. Fisherovou testovou statistiku

$$F_{(1)} = \frac{I_{(1)}}{\sum_{k=1}^p I_{(k)}},$$

kteřá nabývá hodnoty mezi nulou a jedničkou. Jsou-li všechny veličiny $I_{(1)}, I_{(2)}, \dots, I_{(p)}$ téměř stejné, pak bude hodnota Fisherovy statistiky blízká číslu $\frac{1}{p}$. Pro $I_{(1)}$ s mnohem větší hodnotou než nabývají ostatní veličiny $I_{(2)}, \dots, I_{(p)}$, bude mít testová statistika $F_{(1)}$ hodnotu blízkou číslu jedna. Z toho plyne, že

kritický obor naší hypotézy bude tvořen velikými hodnotami. Nulovou hypotézu zamítáme (tzn., že perioda je významná), jestliže

$$F_{(1)} > g_p^F(\alpha),$$

kde $g_p^F(\alpha)$ je kritická hodnota Fisherova testu na hladině významnosti α . Znamená to, že když Fisherův test prokáže významnou periodickou složku určité frekvence λ_{r_0} (pro tuto frekvence je $I(\lambda_{r_0}) = \max_{r=1,\dots,p} I(\lambda_r)$), pak je třeba testovat významnost další velké hodnoty periodogramu, a to tak, že se hodnota $I(\lambda_{r_0})$ vynechá, a se zbylými hodnotami pracujeme analogicky jako předtím tj. jdeme na krok 3.

3. Pokud v 2. zamítneme H_0 , test pokračuje dalším největším vrcholem a počítáme

$$F_{(2)} = \frac{I_{(2)}}{\sum_{k=2}^p I_{(k)}}.$$

Je-li $F_{(2)} > g_{p-1}^F(\alpha)$ (volíme $p-1$, protože se nám zmenšil počet period), opět jsme narazili na významnou periodu. A tak dále pokračujeme výše uvedeným postupem, dokud hypotézu nelze zamítnout.

R. A. Fisher odvodil distribuční funkce statistiky F za předpokladu, že uvažujeme gaussovský bílý šum, tj. za platnosti H_0 , která je ve tvaru

$$1 - F_{F|H_0}(x) = P(F > x) = p(1-x)^{p-1} - \binom{p}{2}(1-2x)^{m-1} + \dots, \quad (25)$$

kde se sčítá dokud $(1-kx) > 0$ pro $0 < x < 1$, což lze zapsat takto

$$P(F > x) = \sum_{r=1,2,\dots} \binom{r}{k} [\max(0, 1-kx)]^{m-1}.$$

Alternativně, v mnoha případech se dává přednost vypočtení hodnoty $P(F > x)$ před volbou kritické hodnoty Fisherova testu. Platí, že pro $p \leq 50$ dostaneme poměrně přesnou aproximaci kritické hodnoty pouze prvním členem na pravé straně ve vztahu (25), tj.

$$P(F > x) \cong p(1-x)^{p-1}. \quad (26)$$

Shoda kritické hodnoty Fisherova testu pro $\alpha = 0,05$ a její aproximace pomocí (26) je uvedena v Tabulce 1, kterou jsem převzala z [2], str.82, z níž je patrna kvalita této aproximace.

počet period r	5%-ní kritická hodnota	aproximace kritické hodnoty
5	0,68377	0,68377
10	0,44495	0,44495
15	0,33462	0,33463
20	0,27040	0,27046
25	0,22805	0,22813
30	0,19784	0,19794
35	0,17513	0,17525
40	0,15738	0,15752
45	0,14310	0,14324
50	0,13135	0,13149

Tabulka 1

4.2 Metoda skrytých period

V předchozí sekci jsme si ukázali způsob, jak lze stanovit významné frekvence pomocí periodogramu a Fisherova testu. Potom za předpokladu (23) platí

$$E(Y_t) = \sum_{r=1}^p (\alpha_r \cos t\lambda_r' + \beta_r \sin t\lambda_r'),$$

kde koeficienty α_r a β_r se odhadují metodou nejmenších čtverců obvyklým způsobem a $\lambda_r', r = 1, \dots, p$ jsou již známé frekvence.

Proto pak minimalizace výrazu

$$\sum_{t=1}^N \left[Y_t - \sum_{r=1}^p (\alpha_r \cos t\lambda_r' + \beta_r \sin t\lambda_r') \right]^2$$

vzhledem k α_r a β_r vede k soustavě $2p$ lineárních rovnic

$$\sum_{r=1}^p \alpha_r \sum_{t=1}^N \cos t\lambda_j' \cos t\lambda_r' + \sum_{r=1}^p \beta_r \sum_{t=1}^N \cos t\lambda_j' \sin t\lambda_r' = \sum_{t=1}^N Y_t \cos t\lambda_j',$$

$$\sum_{r=1}^p \alpha_r \sum_{t=1}^N \sin t\lambda_j' \cos t\lambda_r' + \sum_{r=1}^p \beta_r \sum_{t=1}^N \sin t\lambda_j' \sin t\lambda_r' = \sum_{t=1}^N Y_t \sin t\lambda_j',$$

kde $j = 1, \dots, p$. Hledané odhady jsou ve tvaru

$$\alpha_r = \frac{2}{N} \sum_{t=1}^N Y_t \cos t\lambda_r', \quad \beta_r = \frac{2}{N} \sum_{t=1}^N Y_t \sin t\lambda_r', \quad r = 1, \dots, p. \quad (27)$$

Poznámka 4.2.1 Jestliže neznáme frekvence $\lambda_1, \dots, \lambda_p$, potom jde o nelineární regresní model, kterému se budeme věnovat v následující kapitole.

5 LINEARIZACE NELINEÁRNÍCH REGRESNÍCH MODELŮ

V této kapitole se budeme zabývat modelem popisující závislosti mezi vysvětlovanou proměnnou a vysvětlujícími proměnnými funkcí nelineární vzhledem k neznámým regresním parametrům. Protože je regresní funkce nelineární, můžeme ji zlinearizovat a následně použít teorii lineárních regresních modelů, která je ve statistické literatuře mnohem propracovanější než teorie nelineárních modelů. Hlavním důvodem však je to, že lineární zjednodušení úlohy je často nejen rozumným, ale i zcela dostatečným řešením řady nelineárních úloh. Konkrétně si zde uvedeme linearizace nelineárního statistického modelu pomocí Taylorova rozvoje. Tato metoda za předpokladu apriorní znalosti přibližné hodnoty β_0 odhadovaného parametru β lokálně nahrazuje nelineární regresní funkci funkcí lineární pomocí rozvoje modelu do Taylorovy řady kolem přibližného bodu β_0 se zanedbáním derivace druhého a vyšších řádů.

Obecně budeme uvažovat regresní model, který vykazuje nelineární vazbu střední hodnoty náhodného vektoru Y na cílový parametr β , tzn.

$$E(Y) = f(\beta),$$

kde $f(\cdot): \mathbf{R}^k \rightarrow \mathbf{R}^n$ je nějaká funkce známého analytického tvaru. Navíc, budeme předpokládat spojitost jejích parciálních derivací libovolného řádu. Pro zjednodušení dalších postupů budou na chybový vektor $\varepsilon = Y - f(\beta)$ kladeny předpoklady klasického lineárního modelu, tedy $E(\varepsilon) = \mathbf{0}$, $\text{var}(\varepsilon) = \sigma^2 \mathbf{I}_n$ a složky vektoru ε jsou nezávislé normálně rozdělené náhodné veličiny.

Regresní model považujeme za nelineární, jestliže je aspoň pro jeden parametr $\beta_j, j = 1, \dots, k$ parciální derivace g_j jeho funkcí tj.

$$g_j = \frac{\partial f(x, \beta)}{\partial \beta_j} \neq \text{konst.}$$

Jestliže jsou všechny parciální derivace funkcí parametrů, hovoří se o tzv. neseparabilním modelu, v opačném případě o modelu separabilním, ve kterém závislost na některých parametrech je lineární a na jiných nelineární. Další členění nelineárních

regresních modelů představuje tzv. vnitřně lineární modely, které jsou sice nelineární, ale lze je konkrétní transformací (tzv. reparametrizací) převést na lineární regresní model. Příkladem je model, který je nelineární v parametrech, ale modelovým průběhem je přímka.

Reparametrizaci lze chápat jako změnu parametrického popisu regresní funkce, která je vyjádřena prostřednictvím různých množin parametrů, v nichž staré a nové parametry jsou spjaty jednoznačnými funkčními vztahy, které neobsahují ani vysvětlující proměnné ani náhodné chyby. Důsledkem toho i dělení na modely separabilní a neseperabilní není jednoznačné, protože z neseperabilních modelů lze často reparametrizací získat separabilní modely a opačně.

Pro popis stupně nelinearity regresní funkce se používají míry křivosti. Pro každý model je možné kvantifikovat parametrickou křivost, která je závislá na parametrech a lze ji ovlivnit vhodnou reparametrizací a tzv. vnitřní křivostí, jež je vlastností modelu (více viz [10], str. 226). Potom v případě slabě nelineárních modelů (s malou mírou křivosti) lze při linearizaci Taylorovým rozvojem zanedbat členy druhého a vyššího řádu, přičemž tento postup bude dávat uspokojivé výsledky a je přitom možné očekávat dobré numerické i statistické vlastnosti podobné modelům lineárním.

Definice 5.1 Model ve tvaru $Y \sim_n [f(\beta), \Sigma]$, kde $\beta \in \Theta \subset \mathbb{R}^k$ je vektor neznámých parametrů, $\Theta \subset \mathbb{R}^k$ je parametrický prostor, \mathbb{R}^k je k - dimenzionální lineární vektorový prostor, $f(\cdot): \Theta \rightarrow \mathbb{R}^n$ je známá funkce, která má spojité druhé derivace a Σ je známá pozitivně definitní matice, nazveme nelineární regresní model.

Předpokládáme-li, že známe hodnotu β_0 ležící „dostatečně blízko“ ke skutečné hodnotě β^* , pak můžeme funkci f v nelineárním regresním modelu rozvinout do Taylorovy řady v bodě β_0

$$f(\beta) = f(\beta_0) + F(\beta_0)\delta\beta + \dots,$$

$$\text{kde } F(\beta_0) = \left. \frac{\partial f(\beta)}{\partial \beta^T} \right|_{\beta=\beta_0} \text{ a } \delta\beta = \beta - \beta_0.$$

Zanedbáním členů druhého a vyšších řádů dostaneme místo nelineárního regresního modelu linearizovaný model

$$Y \sim_n \left[f(\beta_0) + \frac{\partial f(\beta)}{\partial \beta^T} \Big|_{\beta=\beta_0} \delta\beta, \Sigma \right], \quad (28)$$

kde $\Sigma = \text{var}(Y)$ je známá pozitivně definitní (p. d.) matice.

V dalším výkladu budeme používat zkrácené označení pro

$$f_0 = f(\beta_0),$$

$$F = F(\beta_0) = \frac{\partial f(\beta)}{\partial \beta^T} \Big|_{\beta=\beta_0}.$$

Obečně, při odhadu neznámých parametrů v nelineárním regresním modelu se omezíme na metodu nejmenších čtverců (MNČ).

Definice 5.2 Odhad neznámého vektoru β , který je řešením $\hat{\beta}(Y)$ rovnice

$$\frac{\partial}{\partial \beta} [Y - f(\beta)]^T \Sigma^{-1} [Y - f(\beta)] = 0, \quad (29)$$

se nazývá nelineární MNČ odhad.

Je důležité podotknout, že rovnice (29) mají i pro jednoduché regresní funkce relativně složitý tvar, takže při řešení normálních rovnic je třeba použít nějakou vhodnou numerickou metodu [4] a odhady parametrů postupně hledat pomocí iteračního postupu. Navíc řešení rovnice (29) nemusí být jediné, jak je tomu v případě lineárního regresního modelu. Dále nulovost parciálních derivací ještě nezaručuje, že v daném bodě existuje minimum, protože případně nalezený extrém funkce nemusí být globální minimum, ale může to dokonce být i lokální minimum nebo sedlový bod. Právě těmto problémům se vyhneme při znalosti přibližné hodnoty β_0 „s dostatečně vysokou přesností“ neznámého parametru β a dále linearizací modelu ve tvaru (28).

Nyní soustředíme naši pozornost na statistickou analýzu modelu (28).

Definice 5.3 Necht' je dán model tvaru $Y - f_0 \sim_n [F\delta\beta, \Sigma]$, kde $\delta\beta$ reprezentuje k - rozměrný vektorový parametr, Σ je známá pozitivně definitní matice (p.d.) a $h(F) = k < n$. Potom tento model nazveme regulární lineární regresní model.

V dalším se budeme věnovat regulárním lineárním regresním modelům.

Věta 5.1 Nejlepší lineární nestranný odhad vektorového parametru $\delta\beta$ (tzv. BLUE - best linear unbiased estimator) je

$$\widehat{\delta\beta} = (F^T \Sigma^{-1} F)^{-1} F^T \Sigma^{-1} (Y - f_0).$$

Důkaz 5.1 Pro přehlednost budeme používat zkrácené označení pro $C = F^T \Sigma^{-1} F$ a $y = Y - f_0$ tj. $y \sim_n [F\delta\beta, \Sigma]$ a tvrzení $\widehat{\delta\beta} = C^{-1} F^T \Sigma^{-1} y$ odvodíme pomocí metody nejmenších čtverců.

Metodou nejmenších čtverců minimalizujeme funkce

$$\phi(\delta\beta) = (y - F\delta\beta)^T \Sigma^{-1} (y - F\delta\beta) = y^T \Sigma^{-1} y - 2\delta\beta^T F^T \Sigma^{-1} y + \delta\beta^T C \delta\beta.$$

Je-li bod minima vnitřním bodem parametrického prostoru Θ , ze kterého neznámé parametry vybíráme, je možné minimum funkce ϕ nalézt jako bod, ve kterém jsou parciální derivace podle parametrů nulové tj.

$$\frac{\partial \phi(\delta\beta)}{\partial \delta\beta^T} = 2C\delta\beta - 2F^T \Sigma^{-1} y = 0.$$

Derivováním vztahu $\phi(\delta\beta)$ podle parametrů $\delta\beta$ a jeho anulováním získáme soustavu normálních rovnic pro parametr $\delta\beta$, která má tvar

$$2C\delta\beta - 2F^T \Sigma^{-1} y = 0,$$

upravíme na

$$C\delta\beta = F^T \Sigma^{-1} y$$

a obdržíme odhad parametru $\widehat{\delta\beta}$, tj.

$$\widehat{\delta\beta} = C^{-1} F^T \Sigma^{-1} y.$$

Dále aby byl odhad nestranný, musí platit, že $E(\widehat{\delta\beta}) = \delta\beta$ a tedy

$$\begin{aligned} E(\widehat{\delta\beta}) &= E[(F^T \Sigma^{-1} F)^{-1} F^T \Sigma^{-1} (Y - f_0)] = (F^T \Sigma^{-1} F)^{-1} F^T \Sigma^{-1} E(Y - f_0) = \\ &= (F^T \Sigma^{-1} F)^{-1} F^T \Sigma^{-1} F \delta\beta = \delta\beta. \end{aligned}$$

Ukážeme nyní, že se jedná o nejlepší odhad parametru $\delta\beta$. Označme $\widetilde{\delta\beta} = Ty + t$ je jiný nestranný odhad parametru $\delta\beta$. Protože $\widetilde{\delta\beta}$ je nestranný, platí že

$$E(\widetilde{\delta\beta}) = E(Ty + t) = TF\delta\beta + t \Leftrightarrow TF = I \text{ a } t = 0.$$

Dále vyjádříme

$$\begin{aligned} \text{var}(\widetilde{\delta\beta} - \widehat{\delta\beta}) &= \text{var}(Ty - C^{-1}F^T\Sigma^{-1}y) = \text{var}[(T - C^{-1}F^T\Sigma^{-1})y] \\ &= (T - C^{-1}F^T\Sigma^{-1})\text{var}(y)(T - C^{-1}F^T\Sigma^{-1})^T \\ &= T\Sigma T^T - C^{-1}F^T T^T - TFC^{-1} + C^{-1} = T\Sigma T^T - C^{-1} \geq 0, \end{aligned}$$

což je pozitivně semidefinitní pro každý jiný nestranný lineární odhad $\widetilde{\delta\beta}$.

Jestliže ještě jednou zderivujeme vztah $\phi(\delta\beta)$ podle parametrů $\delta\beta$ tj.

$$\frac{\partial(2C\delta\beta - 2F^T\Sigma^{-1}y)}{\partial\delta\beta} = 2C = 2F^T\Sigma^{-1}F,$$

potom na základě uvedených předpokladů je matice $F^T\Sigma^{-1}F$ pozitivně definitní, a to je podmínkou pro to, aby funkce $\phi(\delta\beta)$ v bodě, který je řešením anulované první derivace, nabývala minima.

Věta 5. 2 Odhad parametru $\widehat{\delta\beta}$ má kovarianční matici tvaru

$$\text{var}(\widehat{\delta\beta}) = (F^T\Sigma^{-1}F)^{-1}.$$

Důkaz 5.2

$$\begin{aligned} \text{var}(\widehat{\delta\beta}) &= \text{var}[C^{-1}F^T\Sigma^{-1}y] = C^{-1}F^T\Sigma^{-1}\text{var}(y)\Sigma^{-1}FC^{-1} = \\ &= C^{-1}F^T\Sigma^{-1}\Sigma\Sigma^{-1}FC^{-1} = C^{-1}. \end{aligned}$$

6 OVĚŘENÍ VHODNOSTI REGRESNÍHO MODELU ZNAMÉNKOVÝM TESTEM

Po provedení dekompozice časové řady je nutné ověřit, zda jsme skutečně všechny systematické (nenáhodné) složky eliminovali, tedy že zbytek po eliminaci jsou pouze náhodné fluktuace ve tvaru bílého šumu, tj.

$$\varepsilon_t \sim WN(0, \sigma^2),$$

kde $E(\varepsilon_t) = 0$, $var(\varepsilon_t) = \sigma^2$, $cov(\varepsilon_t, \varepsilon_s) = 0$ pro $s \neq t$.

V případě, že regresní model není navržen správně, pak se v modelu projeví nenáhodnost reziduí, tímto lze říci, že jejich jakákoliv systematická (nenáhodná) indikuje nějaký nedostatek odhadnutého regresního modelu. Tuto nenáhodnost lze testovat tzv. znaménkovým testem [4], který vychází přímo z reziduí $\hat{\varepsilon}_t$. V tomto testu jako nulovou hypotézu testujeme

H_0 : časová řada je realizací vzájemně nezávislých, stejně rozdělených náhodných veličin

tj. že rezidua tvoří bílý šum s nulovou střední hodnotou a nebo bílý šum kolísající kolem nenulové úrovně. Pokud při testování vyjde, že zamítneme nulovou hypotézu, pak je v reziduích trend a model je pro data nesprávně navržen. Dále si ukážeme odvození testovací statistiky znaménkového testu.

Nejprve položíme

$$Z_t = \begin{cases} 0 & \text{pro } \hat{\varepsilon}_{t+1} > \hat{\varepsilon}_t \\ 1 & \text{pro } \hat{\varepsilon}_{t+1} < \hat{\varepsilon}_t \end{cases} \Rightarrow Z_t \sim \text{Alt}\left(\frac{1}{2}\right) \Rightarrow E(Z_t) = \frac{1}{2}, \quad var(Z_t) = \left(1 - \frac{1}{2}\right)\frac{1}{2} = \frac{1}{4}.$$

Definujme náhodnou veličinu S , která udává počet kladných prvních diferencí posloupností $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$, předpisem

$$S = \sum_{t=2}^n Z_t. \quad (30)$$

Věta 6.1 Za platnosti H_0 má statistika S tyto vlastnosti:

$$1. E(S) = \frac{n-1}{2},$$

$$2. \text{var}(S) = \frac{n+1}{12},$$

$$3. U = \frac{S - E(S)}{\sqrt{\text{var}(S)}} \xrightarrow{A} N(0,1).$$

Důkaz 6.1 Pomocí vztahu (30) vyjádříme

$$S = \sum_{t=2}^n Z_t \Rightarrow E(S) = E\left(\sum_{t=2}^n Z_t\right) = \sum_{t=2}^n \left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0\right) = \frac{n-1}{2}$$

$$\text{var}(S) = \sum_{t=2}^n \text{var}(Z_t) + 2 \sum_{t=3}^n \text{cov}(Z_t, Z_{t-1}).$$

Platí, že dvě sousední hodnoty Z_t jsou korelované, a proto počítejme

$$\text{cov}(Z_{t-1}, Z_t) = E(Z_{t-1}Z_t) - E(Z_{t-1})E(Z_t).$$

Nejdříve ale proto, abychom uvedený vztah mohli spočítat, musíme si odvodit sdruženou pravděpodobnostní funkci $p(z_{t-1}, z_t)$ pro výpočet

$$\begin{aligned} E(Z_{t-1}Z_t) &= \sum_{(z_{t-1}, z_t) \in \{0,1\}^2} z_{t-1}z_t P(Z_{t-1} = z_{t-1}, Z_t = z_t) \\ &= \sum_{(z_{t-1}, z_t) \in \{0,1\}^2} z_{t-1}z_t p(z_{t-1}, z_t). \end{aligned}$$

Vyjdeme z toho, že pouze kovariance $\text{cov}(Z_{t-1}, Z_t)$ spočítané pro $n-2$ trojic $(\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_t, \hat{\varepsilon}_{t+1})$ jsou nenulové. Označíme si případ $\hat{\varepsilon}_{t-1} < \hat{\varepsilon}_t < \hat{\varepsilon}_{t+1}$ trojici $(1, 2, 3)$, pokud platí $\hat{\varepsilon}_{t-1} < \hat{\varepsilon}_{t+1} < \hat{\varepsilon}_t$, píšeme $(1, 3, 2)$. Všechny možné uspořádání, kterých je $3! = 6$ jsou uvedené v následující tabulce převzaté z [4], str. 103.

$\hat{\varepsilon}_{t-1}$	$\hat{\varepsilon}_t$	$\hat{\varepsilon}_{t+1}$	Z_{t-1}	Z_t
1	2	3	0	0
1	3	2	0	1
2	1	3	1	0
3	1	2	1	0
2	3	1	0	1
3	2	1	1	1

Tabulka 2

Pravděpodobnost každé z těchto možností je za předpokladu platnosti nulové hypotézy stejná, tj. platí $P(Z_{t-1} = z_{t-1}, Z_t = z_t) = \frac{1}{6}$. Potom dostaneme

$$E(Z_{t-1}Z_t) = 1 \cdot 1 \cdot \frac{1}{6} = \frac{1}{6},$$

$$\text{cov}(Z_{t-1}, Z_t) = \frac{1}{6} - \frac{1}{2} \cdot \frac{1}{2} = -\frac{1}{12},$$

$$\text{var}(S) = \frac{n-1}{4} - 2 \frac{n-2}{12} = \frac{n+1}{12}.$$

Odvození asymptotického rozdělení je uvedeno například viz [15].

Hypotézu H_0 zamítneme, když $|U| \geq u_{1-\frac{\alpha}{2}}$, tj. k zamítnutí H_0 dojde, je-li počet kladných prvních diferencí příliš velký nebo malý. Hladina tohoto testu se s rostoucím n blíží číslu α a v praxi se tohoto postupu používá pro $n \geq 20$.

Znaménkový test řadíme mezi neparametrické testy, protože k jeho odvození nebylo nutné předpokládat přesný typ rozdělení pro daný výběr.

7 PRAKTICKÝ PŘÍKLAD

V této části práce budeme aplikovat teoretické poznatky z předchozích kapitol na časovou řadu, která popisuje míru nezaměstnanosti v Olomouckém kraji v letech 1999 až 2008. Data jsem čerpala z internetových stránek Českého statistického úřadu [17].

Než si ukážeme vlastní tvorbu modelu, uvedeme si základní poznatky týkající se nezaměstnanosti a způsobu jejího měření.

7.1 Nezaměstnanost a způsoby jejího měření

Práce je každodenní součástí našeho života. Ta zajišťuje kulturní úroveň společnosti a pro každého člena společnosti biologické i společenské potřeby. Proto se trh práce stává stále více diskutovaným tématem a prioritou řady společností je jeho stabilita a rozvoj.

Zaměstnání bývá označováno jako druh práce, který je hlavním zdrojem hmotného zabezpečení. Ztráta zaměstnání může mít, pokud se jedná o dlouhodobý stav, závažné důsledky, přičemž nezřídka kdy se člověk ztrátou zaměstnání dostává na samý pokraj společnosti (jedním z následků může být například i trestná činnost).

Na trhu práce se navzájem vyhledávají zaměstnanci a zaměstnavatelé, kteří „se k sobě hodí“. Ale než se najdou, zůstávají někteří lidé určitou dobu bez práce a některá pracovní místa zůstávají určitou dobu neobsazena. Tento nesoulad na pracovních trzích se projevuje jako nezaměstnanost.

„Definice nezaměstnanosti je založena na tom, že osoba schopná práce je z možnosti pracovat v placeném zaměstnání vyřazena, ale i na tom, že se se svým vyřazením nespokojuje a hledá nové placené zaměstnání, někdy jen na částečný úvazek.“

„Ne každý, kdo nepracuje, je nezaměstnaný.“ Souvisí to s odlišením práce, především práce konané pro vlastní potěšení (či v rámci domácnosti) od práce (zaměstnání) na smluvním základě zahrnujícím materiální odměnu za její výkon.

Nezaměstnanost tedy neznamená nemít práci, ale nemít placené zaměstnání, a tedy nemít příjem ze zaměstnání. Proto nezaměstnaným nazveme jen toho, kdo nemá práci a nějakou hledá, přičemž je hledání zaměstnání nezbytným znakem nezaměstnaného [18].

Podle Mezinárodní organizace práce (ILO-International Labour Organization) jsou za nezaměstnané považovány ty osoby, které v určitém období neměly žádné zaměstnání, neodpracovaly ani jednu hodinu za mzdu nebo odměnu a aktivně hledaly práci, do které by byly schopny nastoupit nejpozději do dvou týdnů [19].

Nezaměstnanost má mnoho podob i příčin a bojují s ní více či méně úspěšně všechny země. Vyjadřuje se ukazatelem míry nezaměstnanosti. Údaje o nezaměstnanosti neboli přesněji o míře nezaměstnanosti sleduje v ČR, jak Ministerstvo práce a sociálních věcí, které vychází z informací od úřadů práce, tak také Český statistický úřad.

Míra nezaměstnanosti je ukazatel, který v procentech vyjadřuje kolik nezaměstnaných připadá na 100 pracovně schopných lidí (pracovní síle). Podle toho, jak se definují nezaměstnaní a pracovní síla, se potom v statistické praxi setkáváme s více typy míry nezaměstnanosti. V rámci ekonomické teorie a statistické praxe se setkáváme s třemi základními míry nezaměstnanosti [23]:

- Míra registrované nezaměstnanosti,
- Obecná míra nezaměstnanosti a
- Přirozená míra nezaměstnanosti.

Míra registrované nezaměstnanosti (u_r) zachycuje procentuální podíl dosažitelných uchazečů o zaměstnání evidovaných na úřadech práce (U_d) na celkovém počtu dosažitelných ekonomicky aktivních obyvatel (EO_d), tj.

$$u_r = \frac{U_d}{EO_d} 100.$$

Za dosažitelné uchazeče o zaměstnání evidované na úřadech práce jsou považováni registrovaní nezaměstnaní, kterým žádná objektivní překážka nebrání v nástupu do nového zaměstnání. Jedná se tedy o všechny osoby, které aktivně hledají práci, přičemž se v daném časovém okamžiku neúčastní rekvalifikačních kurzů, nevykonávají krátkodobé

zaměstnání, nejsou v pracovní neschopnosti, popř. ve výkonu trestu či nepobírají peněžitou podporu v mateřství.

Ekonomicky aktivní obyvatelstvo (EAO) je tvořeno jednotlivci, kteří dosáhli minimálně patnácti let věku a současně splňují výše uvedená kritéria, na jejichž základě mohou být zařazeni mezi zaměstnané či nezaměstnané. Na druhou stranu, pak mezi ekonomicky neaktivní obyvatelstvo jsou zařazeni ti jednotlivci, kteří ve sledovaném období nebyli zaměstnáni a současně nesplňovali kritéria nezaměstnaných, a osoby mladší patnácti let. Její součástí jsou například jednotlivci navštěvující předškolní zařízení a vzdělávací instituce, osoby na rodičovské dovolené, dlouhodobě nemocní, invalidní a starobní důchodci a lidé, kteří nejsou připraveni nastoupit do nového zaměstnání nejpozději do dvou týdnů.

V České republice je takto definovaná míra nezaměstnanosti zveřejňována pravidelně jednou za měsíc Ministerstvem práce a sociálních věcí (MPSV), a to na základě údajů vycházejících ze statistik vedených jednotlivými úřady práce. V této souvislosti je však zapotřebí poznamenat, že výše uvedenou metodiku začalo MPSV ČR využívat teprve od července roku 2004 a před tímto datem byla hodnota míry registrované nezaměstnanosti stanovována jako procentuální podíl všech uchazečů o zaměstnání evidovaných na úřadech práce k celkovému počtu ekonomicky aktivních osob.

Obecná míra nezaměstnanosti (u_o) vyjadřuje procentuální podíl nezaměstnaných zjištěných na základě Výběrového šetření pracovních sil ($U_{VŠPS}$) na celkovém počtu ekonomicky aktivních obyvatel stanovených na základě stejného šetření ($EAO_{VŠPS}$), tj.

$$u_o = \frac{U_{VŠPS}}{EAO_{VŠPS}} 100.$$

Při výpočtu obecné míry nezaměstnanosti se již nepracuje s údaji z úřadů práce, ale z Výběrového šetření pracovních sil (VŠPS), jež každé čtvrtletí provádí u náhodně vybraných respondentů Český statistický úřad (ČSÚ) a který výsledky tohoto šetření také zveřejňuje. Předmětem šetření jsou všechny osoby obvykle bydlící v soukromých domácnostech na území České republiky. V současné době je tento soubor tvořen téměř 25 tisíci bytovými jednotkami, v nichž bylo šetřeno přes 58 tisíc respondentů, přičemž více než 50 tisíc je představováno dospělou populací. Potom tento rozsah souboru umožňuje

získat spolehlivé odhady charakteristik trhu práce. Využití dat získaných z VŠPS vede k tomu, že při výpočtu této míry nezaměstnanosti jsou zahrnuti jednak uchazeči evidovaní na úřadech práce, ale také i ti uchazeči, kteří sice aktivně hledají práci, ale nejsou součástí této evidence. Protože rozsah i metodika VŠPS odpovídá jak definicím Mezinárodní organizace práce, tak doporučením Eurostatu, je obecná míra nezaměstnanosti ideálním ukazatelem pro komparaci vývoje na jednotlivých národních trzích práce.

Přirozená míra nezaměstnanosti je míra, při které je trh práce ve stavu dlouhodobé rovnováhy, tedy stav, kdy všichni jednotlivci, kteří chtějí při dané mzdové sazbě pracovat, jsou zaměstnáni. Tento stav je v literatuře jinými slovy označován jako plná zaměstnanost (více například viz Holman, str. 159, [8]).

Přestože míra nezaměstnanosti představuje obecně uznávané kritérium pro hodnocení vývoje na trhu práce, existují mezi ekonomy a analytiky takoví, kteří považují tento způsob měření nezaměstnanosti za problematický. Za hlavní nedostatky této metody se považuje to, že jsou do skupiny nezaměstnaných zahrnuti i takoví jedinci, kteří mají nereálné mzdové požadavky, a proto nejsou schopni nalézt zaměstnání, a na druhé straně nejsou zahrnuty ty osoby, které z důvodu dlouhodobé nezaměstnanosti opustily oficiální trh práce, čímž se staly součástí ekonomicky neaktivního obyvatelstva, anebo ty osoby, které nedobrovolně pracují na částečný pracovní úvazek. Zatímco první skupina obyvatel skutečnou míru nezaměstnanosti nadhodnocuje, druhé dvě skupiny tuto míru naopak významně podhodnocují.

V České republice se nezaměstnanost stala závažným problémem ve druhé polovině devadesátých let. Konkrétně je to rok 1989, který představuje významný mezník pro českou společnost, když po sametové revoluci nastal přechod z centrálně řízeného hospodářství na hospodářství tržní. S touto transformací je spojen i logický nástup nezaměstnanosti. (Čerpáno z [8] a [23].)

Nyní konkrétně zaměříme naši pozornost na situaci na trhu práce v Olomouckém kraji, který se řadí mezi kraje s nadprůměrnou mírou nezaměstnanosti.

7.2 Analýza nezaměstnanosti v Olomouckém kraji

Olomoucký kraj se rozkládá ve střední části Moravy a zasahuje i do její severní části. Z hlediska územně-správního tvoří spolu se Zlínským krajem oblast Střední Moravy. Člení se na pět okresů (Jeseník, Olomouc, Prostějov, Přerov a Šumperk). Od 1. 1. 2005 došlo k územnímu rozšíření Olomouckého kraje o tři obce z kraje Moravskoslezského. Na území Olomouckého kraje bylo stanoveno 13 správních obvodů obcí s rozšířenou působností a 20 správních obvodů obcí s pověřeným obecním úřadem. Olomoucký kraj má na severu 104 km dlouhou mezistátní hranici s Polskem, na východě sousedí s Moravskoslezským krajem, na jihu se Zlínským a Jihomoravským krajem a na západě s krajem Pardubickým. [21]



Charakteristiky pracovního trhu v Olomouckém kraji

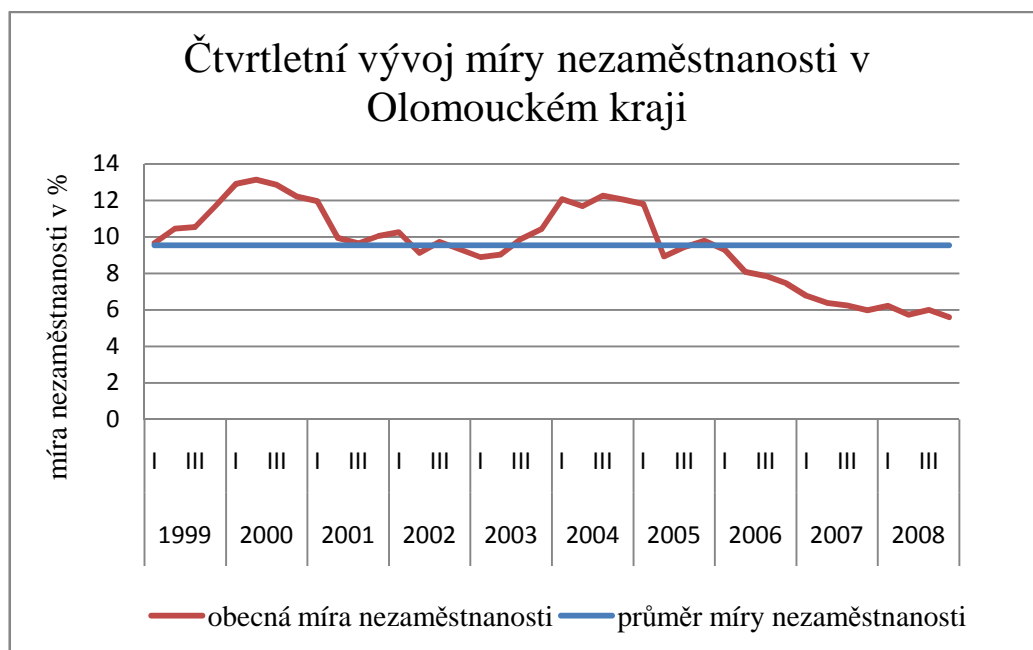
Geografické rozdíly mezi centrální a severní částí Olomouckého kraje se promítají do ekonomiky, infrastruktury a zaměstnanosti. Hanácké okresy (Olomouc, Prostějov, Přerov) jsou více stabilní a diverzifikované. Pracovní trh v horských okresech (Jeseník, Šumperk) je silněji ovlivněn sezónností a horší dopravní obslužností. V centrech kraje jsou dobré podmínky pro rozvoj služeb a v celém regionu pro rozvoj cestovního ruchu. Z ekonomických odvětví je nejsilněji zastoupeno strojírenství a elektrotechnika. K nejvýznamnějším zaměstnavatelům patří UNEX Uničov, Mora Moravia, Moravské železářny, Pars nova, MIELE technika, Pramet tools, AŽD Praha, PSP Engineering, Hanácké železářny a pérovny, EPCOS, Honeywell Aerospace Olomouc, Lindab-Astrom, SSI Schäfer, MUBEA-HZP, Koyo Bearings ČR, SIGMA PUMPY HRANICE apod. K silně zastoupeným odvětvím dále patří elektrotechnický průmysl, jehož vliv posílil v

posledních letech díky příchodu několika zahraničních investorů. Mezi nejvýznamnější patří Siemens Elektromotory a HELLA AUTOTECHNIK NOVA v Mohelnici, M. L. S. Holice v Olomouci a Invensys Appliance Controls ve Šternberku. V Přerově patří mezi nejvýznamnější zaměstnavatele Meopta – optika.

Potravinářství s textilním a oděvnickým průmyslem patří mezi odvětví s dlouholetou tradicí, ale snižujícím se významem. V potravinářství zůstávají největšími zaměstnavateli Nestlé Česko, OLMA Olomouc a pekárny Penam. Textilní průmysl je na ústupu. Poměrně vysoký podíl zaměstnanců pracuje ve stavebnictví a zemědělství, konkrétně ve stavebnictví patří mezi významné zaměstnavatele Skanska DS a GEMO Olomouc. Poptávka zaměstnavatelů po pracovní síle je nevyrovnaná. Zejména lokality na hranicích okresů a okrajové oblasti nabízejí minimum pracovních příležitostí a ve většině případů málo atraktivních. Okres s nejvyšší nezaměstnaností je Jeseník a naopak s nejnižší nezaměstnaností v rámci Olomouckého kraje se potýkají okresy Prostějov a Olomouc [22].

Vývoj obecné míry nezaměstnanosti v Olomouckém kraji

V této kapitole se zabýváme vývojem nezaměstnanosti v Olomouckém kraji, a to v období od roku 1999 do roku 2008. Jako ukazatel vývoje jsme vybrali obecnou míru nezaměstnanosti, o níž bylo pojednáno v kapitole 7.1. Detailnější pohled na vývoj nezaměstnanosti je uveden v Grafu č. 1, kde je zaznamenáno celkem 40 údajů (v procentech) ke konci každého čtvrtletí.



Graf 1

V tomto grafu si můžeme povšimnout, že nárůst nezaměstnanosti nastal během roku 1999 a trval přes celou první polovinu roku 2000, kdy byla dosažena maximální hodnota míry nezaměstnanosti a jejíž hodnota činila přibližně 13,4 %. V dalším období, od druhé poloviny roku 2001 až do konce roku 2003, lze zaznamenat mírné kolísání kolem průměrné hodnoty, která pro tyto data činí kolem 9,212 %. V následujícím období opět nezaměstnanost prudce roste a přes celý rok 2004 se pohybuje kolem 12 %. Od roku 2004 se Česká republika stala členem Evropské unie a díky lepšímu mezinárodnímu porovnávání byla zavedena nová metodika spočívající v zavedení pojmu dosažitelný neumístěný uchazeč o zaměstnání, díky které se dosahuje nižší nezaměstnanosti. Teprve od konce roku 2005 až do konce sledovaného období, což je rok 2008, se nezaměstnanost neustále snižovala.

Nyní budeme hledat model, který dostatečně přesně popíše předloženou časovou řadu, která obsahuje čtvrtletní údaje o vývoji míry nezaměstnanosti v Olomouckém kraji od 1999 do 2008, vyjádřené v procentech. Pro větší přehlednost jsem získaná data z internetových stránek Českého statistického úřadu [17] uspořádala do následující tabulky (Tabulka č. 3).

Rok	Čtvrtletí	Míra nezaměstnanosti (v %)
1999	<i>I</i>	9,670466
	<i>II</i>	10,45748
	<i>III</i>	10,54540
	<i>IV</i>	11,69587
2000	<i>I</i>	12,91948
	<i>II</i>	13,14165
	<i>III</i>	12,86142
	<i>IV</i>	12,20358
2001	<i>I</i>	11,96827
	<i>II</i>	9,948531
	<i>III</i>	9,655959
	<i>IV</i>	10,04837
2002	<i>I</i>	10,26155
	<i>II</i>	9,121792
	<i>III</i>	9,742171
	<i>IV</i>	9,323546
2003	<i>I</i>	8,897474
	<i>II</i>	9,033261
	<i>III</i>	9,896945
	<i>IV</i>	10,44183
2004	<i>I</i>	12,07772
	<i>II</i>	11,69293
	<i>III</i>	12,26320
	<i>IV</i>	12,06066
2005	<i>I</i>	11,80257
	<i>II</i>	8,935512
	<i>III</i>	9,446708
	<i>IV</i>	9,795011
2006	<i>I</i>	9,289022
	<i>II</i>	8,089053
	<i>III</i>	7,863689
	<i>IV</i>	7,462359
2007	<i>I</i>	6,774868
	<i>II</i>	6,392044
	<i>III</i>	6,245071
	<i>IV</i>	5,982594
2008	<i>I</i>	6,235479
	<i>II</i>	5,730874
	<i>III</i>	5,994629
	<i>IV</i>	5,598327

Tabulka 3

Postup

Časová řada Míra nezaměstnanosti v Olomouckém kraji, jakožto časová řada popisující ekonomický jev lze rozložit na několik složek. Dekompoziční rozklad provedeme, abychom pronikli hlouběji do podstaty historického průběhu řady, navíc nám umožní provést extrapolaci, a tím zjistit budoucí vývoj jednotlivých složek anebo konstruovat předpověď v celé časové řadě tak, že ji složíme z předpovědí jednotlivých složek, které se obvykle sestojí poměrně jednoduše a přesně. Dále víme, že si časovou řadu můžeme představit jako trend, na kterou jsou „nabaleny“ ostatní složky.

Při tomto přístupu budeme nejprve předpokládat, že analyzovaná časová řada je ve tvaru

$$Y_t = T_t + \varepsilon_t, \quad t = 1, \dots, n. \quad (31)$$

Trendovou funkci $T_t = T(t)$ popíšeme polynomičtým modelem trendu a díky předpokladu (31) můžeme použít pro odhad parametru trendových křivek aparát lineární regresní analýzy.

V rámci teoretické části jsme se zabývali tzv. Čebyševovou metodou, kterou nyní použijeme pro konstrukci polynomičtého modelu trendu.

Polynomičtý model trendu a Čebyševova metoda

Využijeme Větu 3.1, na základě které určíme Čebyševovy polynomy. Podle této věty víme, že pro $x = 1, \dots, n$ je

$$\varphi_0(x) = 1,$$

$$\varphi_1(x) = x - \frac{\sum_{i=1}^n x_i \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x) = x - \bar{x},$$

$$\varphi_2(x) = x^2 - \frac{\sum_{i=1}^n x_i^2 \varphi_1(x_i)}{\sum_{i=1}^n \varphi_1^2(x_i)} \varphi_1(x) - \frac{\sum_{i=1}^n x_i^2 \varphi_0(x_i)}{\sum_{i=1}^n \varphi_0^2(x_i)} \varphi_0(x),$$

...

Protože pracujeme s ortogonálním systémem funkcí je matice normální soustavy diagonální a jak již bylo ukázáno, v modelu

$$Y = X\beta + \varepsilon$$

se výpočet regresních parametrů zjednoduší

$$\hat{\beta}_j = \frac{\sum_{i=1}^n Y_i \varphi_{j-1}(x_i)}{\sum_{i=1}^n \varphi_{j-1}^2(x_i)}, \quad \forall j = 1, \dots, k.$$

Konkrétně pro naše data byly výpočty provedeny v programu MS EXCEL a výsledky jsou uvedené viz příloha, Tabulka č. 10 a Tabulka č. 11.

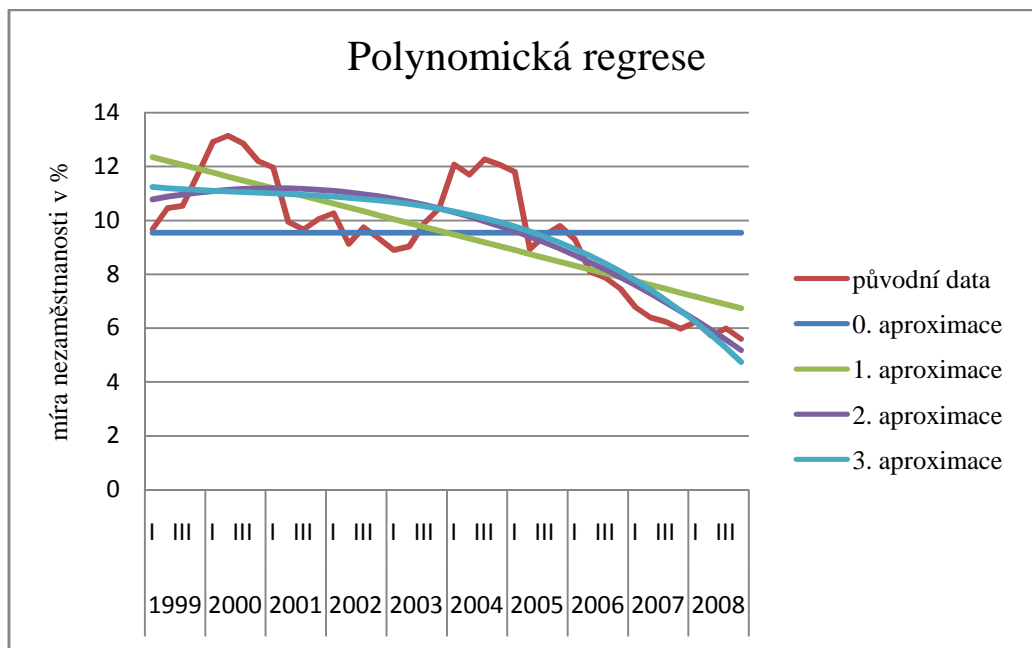
Protože bylo třeba určit optimální stupeň polynomu, porovnávali jsme jednak RSC pro jednotlivé aproximace a také jsme použili výše popsanou metodu „od nejnižšího stupně k nejvyššímu“. Tím náš výpočet polynomů skončil u polynomu 3. stupně.

Potom odhady regresních parametrů a odhady jejich rozptylů jsme obdrželi dosazením do vzorce (10) a (11) na str. 21. Výsledky spolu s t-testem významnosti a RSC pro každý model zvlášť jsou uvedeny v Tabulce č. 4.

	0. aproximace	1. aproximace	2. aproximace	3. aproximace
<i>k</i>	1	2	3	4
<i>n – k</i>	39	38	37	36
$\hat{\beta}_j$	9,539184	-0,14373	-0,00629	-0,00016
<i>RSČ</i>	192,0664	81,95304	59,51154	<u>57,96706</u>
$\hat{\sigma}^2$	4,924779	2,156659	1,60842	1,610196
$t_{n-p, 1-\frac{\alpha}{2}}$	2,0227	2,0244	2,0262	2,0281
t	27,1861	7,1454	3,7353	0,9793
ROZHODNUTÍ	VÝZNAMNÝ	VÝZNAMNÝ	VÝZNAMNÝ	<u>NEVÝZNAMNÝ</u>

Tabulka 4

V této Tabulce č.4 je vidět, že sice reziduální součet čtverců u 3. aproximace je nejmenší, ale protože nám t-test prokázal nevýznamnost parametru kubického trendu, rozhodli jsme se, že pro optimální popis trendovou funkcí zvolíme funkci kvadratickou.



Graf 2

Analýza periodické složky

V dalším textu se budeme věnovat analýze periodické složky. Z tohoto důvodu potřebujeme naši časovou řadu očistit od trendu, a tedy budeme analyzovat hodnoty reziduí vypočítané z kvadratického trendu (viz Graf 3). Nyní předpokládáme, že

$$\hat{\varepsilon}_t = Y_t - \hat{T}_t = P_t + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2).$$



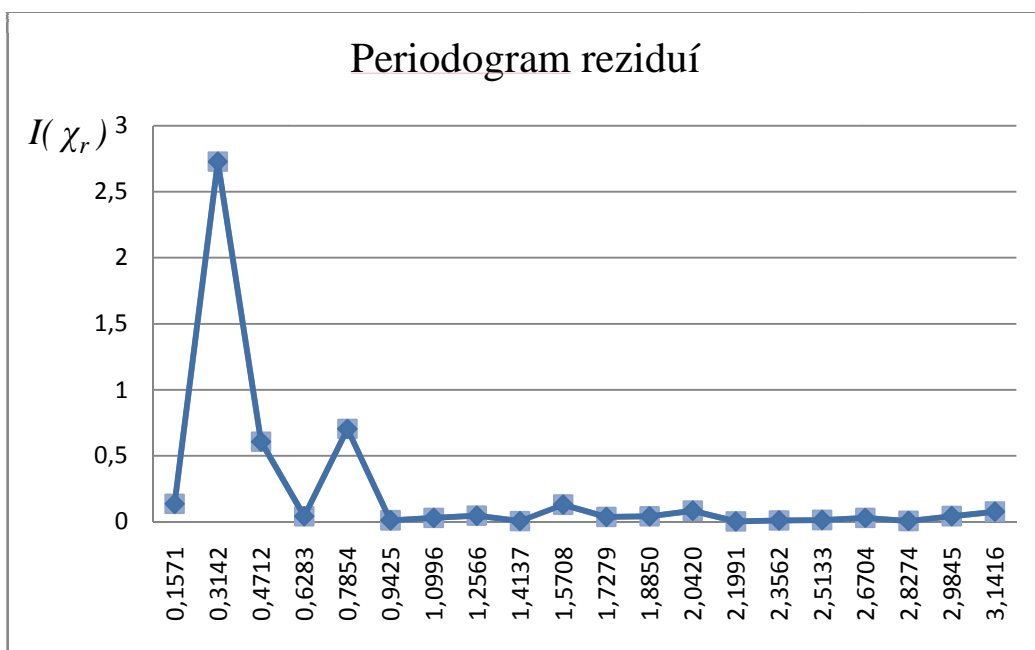
Graf 3

Pro nalezení významných periodických složek provedeme analýzu periodogramu, jemuž byla věnována kapitola 4. Výpočet hodnot periodogramu jsme provedli na základě Lemma 4.1.

Protože $N = 40$, je $r = 1, \dots, 20$. Na základě výše uvedené teorie jsou získané výsledky shrnuty v Tabulce č. 5.

r	λ	$A(\chi_r)$	$A^2(\chi_r)$	$B(\chi_r)$	$B^2(\chi_r)$	$I(\chi_r)$
1	0,15707963	0,029316	0,000859	-1,30312	1,698115	0,1352
2	0,31415927	0,523668	0,274228	5,830999	34,00055	2,7275
3	0,4712389	-1,68249	2,830776	-2,18885	4,791084	0,606528
4	0,62831853	-0,67604	0,457028	-0,23759	0,056451	0,040861
5	0,78539816	0,712751	0,508014	-2,88492	8,322766	0,702731
6	0,9424778	0,366486	0,134312	0,022648	0,000513	0,010729
7	1,09955743	0,604423	0,365327	0,063448	0,004026	0,029392
8	1,25663706	0,189484	0,035904	-0,7317	0,535384	0,045462
9	1,41371669	-0,25553	0,065298	-0,03225	0,00104	0,005279
10	1,57079633	1,133571	1,284983	0,588722	0,346594	0,129837
11	1,72787596	-0,38918	0,151459	-0,55755	0,310857	0,03679
12	1,88495559	-0,65286	0,426223	-0,29155	0,085	0,040682
13	2,04203522	0,997672	0,99535	-0,232	0,053824	0,083491
14	2,19911486	-0,03237	0,001048	-0,18798	0,035336	0,002895
15	2,35619449	0,153452	0,023548	-0,30972	0,095925	0,009507
16	2,51327412	0,299059	0,089436	-0,27439	0,075288	0,013108
17	2,67035376	0,047236	0,002231	-0,60386	0,364652	0,029196
18	2,82743339	0,255385	0,065221	0,107445	0,011545	0,006109
19	2,98451302	0,725031	0,52567	-0,07479	0,005594	0,042277
20	3,14159265	-0,97988	0,960164	-2,6E-15	6,82E-30	0,076407

Tabulka 5



Graf 4

Dále abychom rozhodli, které hodnoty periodogramu lze považovat za významně velké ve srovnání s ostatními hodnotami, provedeme Fisherův test. Při provádění Fisherova testu jsme postupovali dle kroků uvedených na str. 37 a dosažené výsledky jsou prezentovány v Tabulce č. 6.

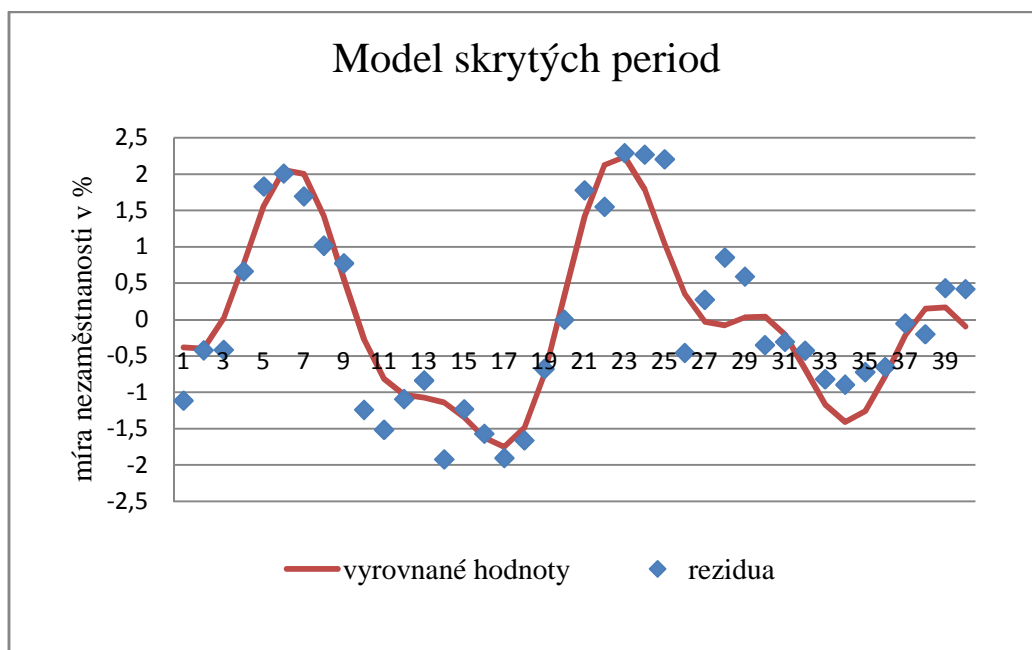
FISHERŮV TEST				
$I_{(j)}$	$F_{(j)}$	$g_j^F(0,01)$	ROZHODNUTÍ	T_r
2,727499	0,571326	0,330284	VÝZNAMNÁ	20
0,702731	0,343385	0,345638	VÝZNAMNÁ	8
0,606529	0,45137	0,360992	VÝZNAMNÁ	13,33333
0,1352	0,183392	0,376347	NEVÝZNAMNÁ	

Tabulka 6

Na základě tohoto testu jsme vyvodili závěr, že v modelu existují tři významné periody, a to perioda 20., 8. a 13. čtvrtletí. Proto jsme tyto tři významné periody zahrnuli do MSP tvaru

$$\hat{\varepsilon}_t = \beta_1 \cos\left(\frac{2\pi t}{T_1}\right) + \beta_2 \sin\left(\frac{2\pi t}{T_1}\right) + \beta_3 \cos\left(\frac{2\pi t}{T_2}\right) + \beta_4 \sin\left(\frac{2\pi t}{T_2}\right) + \beta_5 \cos\left(\frac{2\pi t}{T_3}\right) + \beta_6 \sin\left(\frac{2\pi t}{T_3}\right) + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2), \quad t = 1, \dots, 40.$$

Odhady parametrů dostaneme využitím vztahu (10), přičemž výpočty provedené v MS Excelu jsem opět shrnula do Tabulky č. 12, která je uvedena v příloze. Nyní se můžeme podívat na Graf 5, který nám udává vyrovnané hodnoty.



Graf 5

Výsledný model

Výsledný model, který kromě trendu obsahuje i periodickou složku, dostaneme přičtením kvadratického trendu k modelu skrytých period, tj.

$$Y_t = T_t + P_t + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2).$$

Získáme pak lineární model ve tvaru

$$Y_t = \beta_0 \varphi_0(t) + \beta_1 \varphi_1(t) + \beta_2 \varphi_2(t) + \beta_3 \cos\left(\frac{2\pi t}{T_1}\right) + \beta_4 \sin\left(\frac{2\pi t}{T_1}\right) + \beta_5 \cos\left(\frac{2\pi t}{T_2}\right) + \beta_6 \sin\left(\frac{2\pi t}{T_2}\right) + \beta_7 \cos\left(\frac{2\pi t}{T_3}\right) + \beta_8 \sin\left(\frac{2\pi t}{T_3}\right) + \varepsilon_t, \quad (31)$$

$$\varepsilon_t \sim WN(0, \sigma^2), \quad t = 1, \dots, 40,$$

nebo ekvivalentně maticově

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim WN(\mathbf{0}, \sigma^2 I_n) \quad (32)$$

kde X je matice plánu ve tvaru

$$X = \begin{pmatrix} \varphi_0(1) & \varphi_1(1) & \varphi_2(1) & \cos\left(\frac{2\pi}{T_1}\right) & \sin\left(\frac{2\pi}{T_1}\right) & \cos\left(\frac{2\pi}{T_2}\right) & \sin\left(\frac{2\pi}{T_2}\right) & \cos\left(\frac{2\pi}{T_3}\right) & \sin\left(\frac{2\pi}{T_3}\right) \\ \varphi_0(2) & \varphi_1(2) & \varphi_2(2) & \cos\left(\frac{4\pi}{T_1}\right) & \sin\left(\frac{4\pi}{T_1}\right) & \cos\left(\frac{4\pi}{T_2}\right) & \sin\left(\frac{4\pi}{T_2}\right) & \cos\left(\frac{4\pi}{T_3}\right) & \sin\left(\frac{4\pi}{T_3}\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_0(40) & \varphi_1(40) & \varphi_2(40) & \cos\left(\frac{80\pi}{T_1}\right) & \sin\left(\frac{80\pi}{T_1}\right) & \cos\left(\frac{80\pi}{T_2}\right) & \sin\left(\frac{80\pi}{T_2}\right) & \cos\left(\frac{80\pi}{T_3}\right) & \sin\left(\frac{80\pi}{T_3}\right) \end{pmatrix}$$

Na základě teoretických poznatků v kapitole o lineárních regresních modelech jsme vypočetli MNČ odhady pro regresní parametry dle Věty 2.2.1, dále jsme určili hodnotu RSČ dle Definice 2.2.1 a dopočítali jsme odhad rozptylu dle Věty 2.2.4 tj.

$$S_e = (Y - \hat{Y})^T (Y - \hat{Y}) = 6,53598$$

$$\hat{\sigma}^2 = \frac{S_e}{n - k} = \frac{6,53598}{40 - 9} = 0,21084$$

Sestrojili jsme také 95 % -ní konfidenční intervaly dle vzorce (12) a navíc jsme dopočítali hodnotu testového kritéria Studentovým t-testem o významnosti jednotlivých parametrů β_j podle vzorce (13). Pro zvolené $\alpha = 0,05$ je kvantil Studentova rozdělení $t_{31; 0,975} = 2,0395$, jež jsme vypočítali v MS EXCEL pomocí funkce TINV. Výsledky pro naši časovou řadu jsou uvedeny níže v Tabulce č. 7.

j	$\hat{\beta}_j$	95 % -ní konfidenční intervaly		$ t_j $
		t_d	t_h	
0	9,5392	9,3911	9,6873	131,393
1	-0,1201	-0,1349	-0,1053	16,5624
2	-0,0064	-0,0077	-0,0051	10,1711
3	0,0991	-0,1171	0,3152	0,9349
4	1,4539	1,2245	1,6833	12,9246
5	0,1366	-0,0735	0,3466	1,3260
6	-0,5877	-0,8002	-0,3753	5,6420
7	-0,3974	-0,6085	-0,1863	3,8392
8	-0,3905	-0,6088	-0,1721	3,6471

Tabulka 7

Dále jsme provedli F-test dle Důsledku 2.3.1, pro nějž hodnota testového kritéria vyšla

$$F = 1,2779 < 2,199$$

Protože je hodnota statistiky F menší než hodnota kvantilu Fisherova-Snedecerova rozdělení o (9, 31) stupních volnosti na hladině $\alpha = 0,05$, který jsme v MS Excelu dopočítali pomocí funkce FINV, můžeme konstatovat, že navržený model je vhodný.

Na základě Tabulky č. 7 je vidět, že parametry β_3 a β_5 se významně neliší od nuly, což plyne z Poznámky 2.3.1 i ze struktury jejich konfidenčních intervalů, které obsahují nulu. Protože F-test vychází významný, dospěli jsme takto k závěru, že model můžeme považovat za vhodný, ale protože t-testy ukazují nevýznamnost některých regresních parametrů, je třeba z modelu vypustit 3. a 5. vysvětlující proměnnou, neboť pro ně jsou parametry nevýznamně odlišné od nuly.

Znaménkový test

Nyní potřebujeme ověřit, zda jsme skutečně všechny systematické (nenáhodné) složky eliminovali. Testujeme

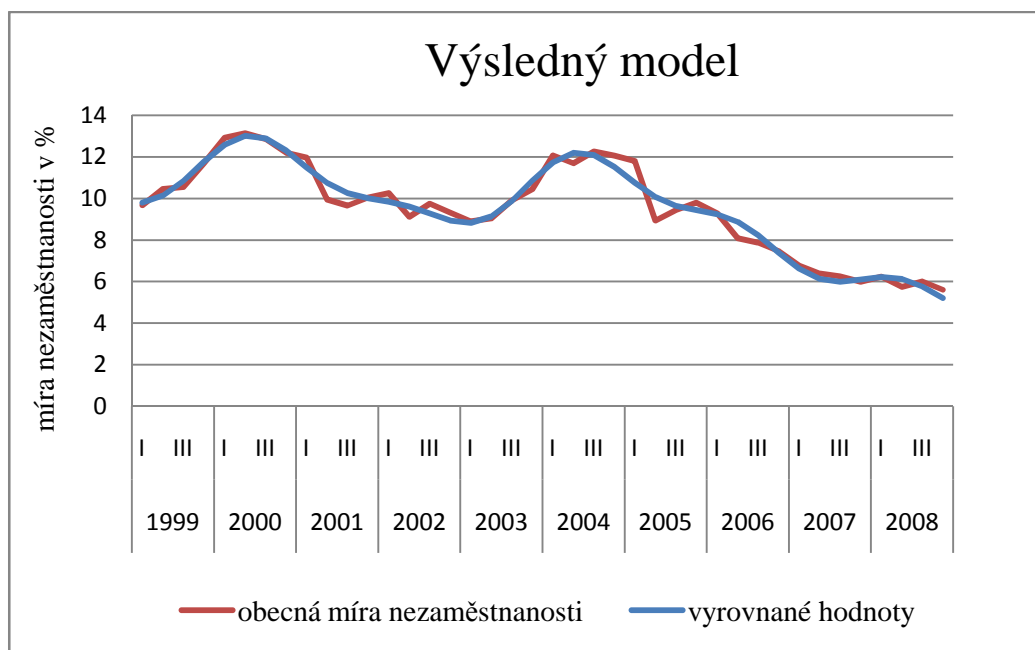
H_0 : časová řada je realizací vzájemně nezávislých, stejně rozdělených náhodných veličin

tj. že rezidua tvoří bílý šum s nulovou střední hodnotou a nebo bílý šum kolísající kolem nenulové úrovně. Vypočítáme

$$U = \frac{S - E(S)}{\sqrt{\text{var}(S)}} = \frac{23 - \frac{40 - 1}{2}}{\sqrt{\frac{40 + 1}{12}}} = 1,8935 < 1,96,$$

kde 1,96 je hodnota kvantilu normovaného normálního rozdělení na hladině $\alpha = 0,05$. Závěr je, že nulovou hypotézu nezamítneme a že model (32) je navržen správně.

Výsledný model bez nevýznamných parametrů je znázorněn v Grafu 5.



Graf 5

Aniž bychom nějak komentovali tento výsledný model (je z grafu patrna velká shoda modelu s daty), vrhneme se k jeho zpřesnění. Způsob, který nám to umožní, byl popsán v kapitole 6. Proto v dalším budeme uvažovat, že výsledky periodogramu a Fisherova testu jsou pouze přibližné hodnoty délky period, nikoliv skutečné (známé). Na základě této úvahy a Poznámky 4.2.1 je model (31) nelineární vzhledem k regresním parametrům. Linearizaci provedeme pomocí rozvoje do Taylorovy řady, kdy zanedbáme druhé a vyšší derivace.

Linearizace

Protože zpřesnění se týká periodické složky řady a konkrétně modelu skrytých period vytvořeném na základě výsledků periodogramu a Fisherova testu, budeme tímto vlastně opět analyzovat model bez trendu a tedy hodnoty reziduí vypočítané z kvadratického trendu tj.

$$\begin{aligned}\hat{\varepsilon}_t = Y_t - \hat{T}_t = P_t + \varepsilon_t = & \beta_1 \cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right) + \beta_2 \sin\left(\frac{2\pi t}{\beta_7^{(0)}}\right) + \beta_3 \cos\left(\frac{2\pi t}{\beta_8^{(0)}}\right) \\ & + \beta_4 \sin\left(\frac{2\pi t}{\beta_8^{(0)}}\right) + \beta_5 \cos\left(\frac{2\pi t}{\beta_9^{(0)}}\right) + \beta_6 \sin\left(\frac{2\pi t}{\beta_9^{(0)}}\right) + \varepsilon_t\end{aligned}\quad (33)$$

Nyní si uvedeme postup, jakým jsme uvedenou problematiku řešili.

Pro linearizaci ovšem potřebujeme znát přibližné hodnoty parametru $\boldsymbol{\beta}$. Prakticky ji můžeme zjistit následujícím způsobem. Odhady neznámých parametrů v MSP (33) položíme rovny přibližným hodnotám tj. $\boldsymbol{\beta}^0 = \hat{\boldsymbol{\beta}}$. Jinými slovy, již získané výsledky z periodogramu a Fisherova testu jsme položili za body $\beta_7^{(0)} = T_1$, $\beta_8^{(0)} = T_2$ a $\beta_9^{(0)} = T_3$, které leží „dostatečně“ blízko (víme, že periodogram je dobrým ukazatelem periodicity v časových řadách, viz kapitola 4) hodnoty skutečné délky period, jež označíme β_7^* , β_8^* a β_9^* .

Linearizovaný model na základě teoretických poznatků na str. 43 této práce je ve tvaru

$$Y - \mathbf{f}_0 \sim_{40} [\mathbf{F} \boldsymbol{\delta\beta}, \sigma^2 \mathbf{I}_{40}],$$

kde $\boldsymbol{\delta\beta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$,

$$\boldsymbol{\beta}^0 = [0,1171; 1,3039; 0,1594; -0,6451; -0,3762; -0,4894; 20; 8; 13,3333]^T.$$

Matice v linearizovaném modelu jsou následující

$$\begin{aligned}\{\mathbf{f}_0\}_t = & \left\{ \beta_1^{(0)} \cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right) + \beta_2^{(0)} \sin\left(\frac{2\pi t}{\beta_7^{(0)}}\right) + \beta_3^{(0)} \cos\left(\frac{2\pi t}{\beta_8^{(0)}}\right) + \beta_4^{(0)} \sin\left(\frac{2\pi t}{\beta_8^{(0)}}\right) \right. \\ & \left. + \beta_5^{(0)} \cos\left(\frac{2\pi t}{\beta_9^{(0)}}\right) + \beta_6^{(0)} \sin\left(\frac{2\pi t}{\beta_9^{(0)}}\right) \right\}, \\ \{\mathbf{F}\}_t = & \left\{ \cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right), \sin\left(\frac{2\pi t}{\beta_7^{(0)}}\right), \dots, \beta_5^{(0)} \frac{2\pi t}{(\beta_9^{(0)})^2} \sin\left(\frac{2\pi t}{\beta_9^{(0)}}\right) - \beta_6^{(0)} \frac{2\pi t}{(\beta_9^{(0)})^2} \cos\left(\frac{2\pi t}{\beta_9^{(0)}}\right) \right\},\end{aligned}$$

kde $t = 1, \dots, 40$. Výsledky jsou uvedeny v příloze, Tabulka č. 15 a Tabulka č. 16.

Odhady parametrů $\delta\beta_7$, $\delta\beta_8$ a $\delta\beta_9$ jsme následně otestovali, zda-li se významně liší od nuly. Potom ty, které se lišit budou, přičteme k přibližné hodnotě délky period a následně je zahrneme do výsledného modelu. Výsledky tohoto postupu jsme shrnuli do Tabulky č. 8.

j	$\widehat{\delta\beta_j}$	$ t_j $	ROZHODNUTÍ
7	2,034536	1,822499094	NEVÝZNAMNÝ
8	0,227781	1,393314812	NEVÝZNAMNÝ
9	1,903476	1,985074376	VÝZNAMNÝ

Tabulka 8

Při rozhodování o významnosti parametrů jsme hodnotu testového kritéria porovnali s kvantilem Studentova rozdělení $t_{31; 0,975} = 2,0395$, jež jsme vypočítali v MS EXCEL pomocí funkce TINV. Výsledkem je, že všechny parametry jsou nevýznamně odlišné od nuly. Tento fakt by ovšem znamenal, že na takto zvolené hladině významnosti $\alpha = 0,05$ můžeme parametry vyloučit a pracovat s původním modelem. Protože ale naším cílem je zjistit, jaký dopad by měla provedená linearizace na náš výsledný model, a přesvědčit se o účinnosti této metody, rozhodli jsme se do modelu tyto výsledky zahrnout.

Nyní na základě výše uvedených výsledků zahrneme periody

$$T_1 + \widehat{\delta\beta_7} \doteq 20 + 2,0345 = 22,0345,$$

$$T_2 + \widehat{\delta\beta_8} \doteq 8 + 0,2278 = 8,2278,$$

$$T_3 + \widehat{\delta\beta_9} \doteq 13,3333 + 1,9035 = 15,2368$$

do modelu skrytých period. Přitom je matice plánu ve tvaru

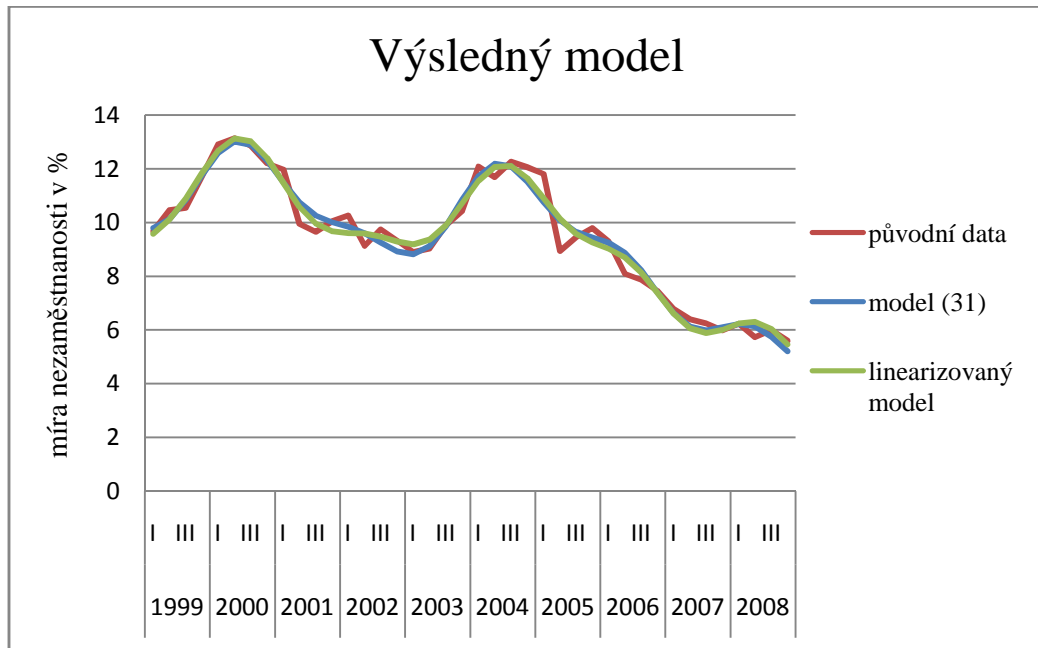
$$\mathbf{X} = \begin{pmatrix} \varphi_0(1), & \dots, & \cos\left(\frac{2\pi}{T_3 + \widehat{\delta\beta_3}}\right) & \sin\left(\frac{2\pi}{T_3 + \widehat{\delta\beta_3}}\right) \\ \varphi_0(2), & \dots, & \cos\left(\frac{4\pi}{T_3 + \widehat{\delta\beta_3}}\right) & \sin\left(\frac{4\pi}{T_3 + \widehat{\delta\beta_3}}\right) \\ \vdots & \ddots & \vdots & \vdots \\ \varphi_0(40), & \dots, & \cos\left(\frac{80\pi}{T_3 + \widehat{\delta\beta_3}}\right) & \sin\left(\frac{80\pi}{T_3 + \widehat{\delta\beta_3}}\right) \end{pmatrix}.$$

Dále jsme postupovali stejně jako u modelu (31) a dosažené výsledky jsou následující

$$S_e = (Y - \hat{Y})^T (Y - \hat{Y}) = 6,2806,$$

$$\hat{\sigma}^2 = \frac{S_e}{n - k} = \frac{7,7988}{40 - 7} = 0,2026.$$

Porovnáme-li tyto hodnoty s výsledky z modelu bez linearizace, pak můžeme tvrdit, že v našem případě linearizace splnila cíl, tedy model se nám skutečně zpřesnil. Následující Graf 6 porovnává tyto dva postupy.



Graf 6

Předpovědi

Uvažujeme lineární regresní model (32). Dále budeme předpokládat normalitu, tj.

$$Y \sim N_n(X\beta, \sigma^2 I_n).$$

Označíme budoucí pozorování

$$Y_0 = x_0^T \beta + \varepsilon.$$

Predikujeme Y_0 pomocí $\hat{\beta}$ (BLUE), tedy

$$\hat{Y}_0 = x_0^T \hat{\beta}.$$

Zajímá nás rozdělení $Y_0 - \hat{Y}_0$. Platí, že Y_0 a \hat{Y}_0 jsou nezávislé, neboť odhad \hat{Y}_0 je dosažen na základě n nezávislých pozorování a Y_0 je budoucí pozorování na něm také nezávislé. Dále platí, že

$$E(Y_0 - \hat{Y}_0) = \mathbf{x}_0^T \boldsymbol{\beta} - E(\mathbf{x}_0^T \hat{\boldsymbol{\beta}}) = \mathbf{x}_0^T \boldsymbol{\beta} - \mathbf{x}_0^T \boldsymbol{\beta} = 0,$$

$$\begin{aligned} \text{var}(Y_0 - \hat{Y}_0) &= \text{var}(Y_0) + \text{var}(\hat{Y}_0) = \sigma^2 + \text{var}(\mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}) \\ &= \sigma^2 + \sigma^2 \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0 = \sigma^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0), \end{aligned}$$

tj. $Y_0 - \hat{Y}_0 \sim N[0, \sigma^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)]$.

Potřebujeme také ukázat, že $\hat{\sigma}^2$ a $Y_0 - \hat{Y}_0$ jsou nezávislé. Protože z Věty 2.2.7 víme že $\hat{\sigma}^2$ a $\hat{\boldsymbol{\beta}}$ jsou nezávislé, a dále víme, že je $\hat{\sigma}^2$ funkce $\hat{\boldsymbol{\varepsilon}}$, stačí dopočítat

$$\text{cov}(Y_0, \hat{\boldsymbol{\varepsilon}}) = \text{cov}(Y_0, \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \text{cov}(Y_0, [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{Y}) = 0.$$

Potom využitím Věty 2.2.6 a po znormování má

$$\frac{Y_0 - \hat{Y}_0}{\sqrt{\sigma^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)}} \sim N(0, 1),$$

dostáváme

$$\frac{Y_0 - \hat{Y}_0}{\sqrt{\hat{\sigma}^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)}} \sim t_{n-k}.$$

Konečně můžeme určit tvar predikčního intervalu, který je dán vztahem

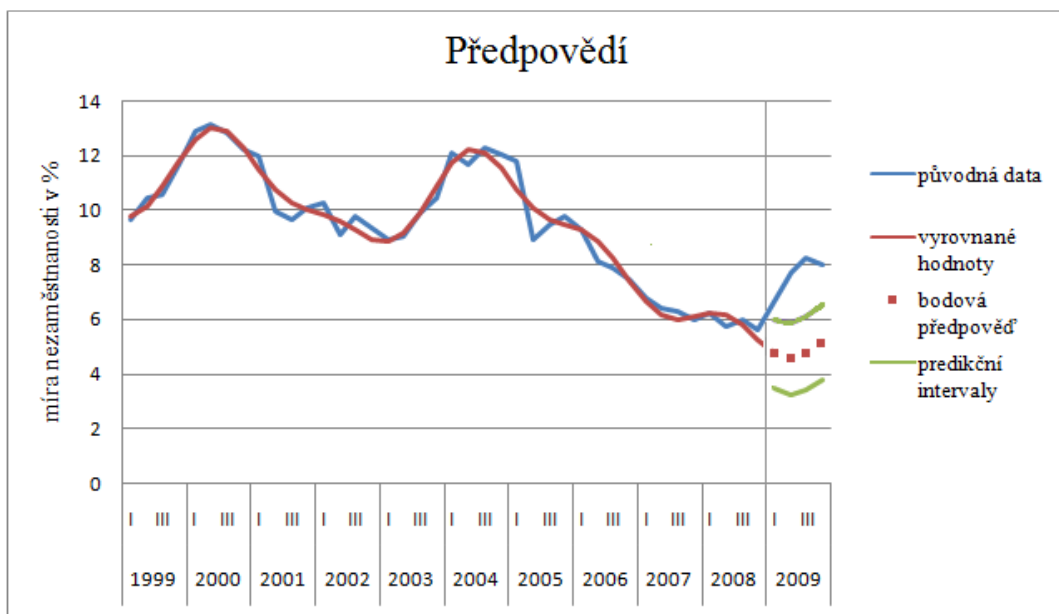
$$\begin{aligned} P_{1-\alpha}(Y_0) &= \left[\hat{Y}_0 - \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)} t_{n-k} \left(1 - \frac{\alpha}{2} \right), \right. \\ &\quad \left. \hat{Y}_0 + \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)} t_{n-k} \left(1 - \frac{\alpha}{2} \right) \right]. \end{aligned}$$

Pro naši časovou řadu jsme sestrojili předpověď na následující rok 2009. Výpočty byly provedené v MS EXCEL a výsledky jsou uvedené v následující Tabulce č.9.

Rok 2009 čtvrtletí	Bodová předpověď	Predikční intervaly		Skutečná hodnota
		dolní mez	horní mez	
I.	4,7336	3,4773	5,9899	6,638001
II.	4,558431	3,2349	5,882	7,705761
III.	4,750798	3,3876	6,114	8,207043
IV.	5,156219	3,7639	6,5485	7,991476

Tabulka 9

Je důležité poznamenat, že tyto předpovědi míry nezaměstnanosti v Olomouckém kraji platí při zachování stávajících podmínek, které se vyskytly během sledovaného období od roku 1998 do roku 2008. Dále je nutné upozornit, že na vývoj nezaměstnanosti působí mnoho vlivů. Například, konkrétně v našem případě můžeme za faktor znehodnocující kvalitu předpovědí považovat celosvětovou krizi, která v ČR vypukla koncem roku 2008 a která ovšem do modelu není zahrnuta. Je samozřejmé, že tyto předpovědi nejsou schopné takovéto změny postřehnout, ale i tak předpovědi orientačně dobře zachycují budoucí průběh řady, což je znázorněno na následujícím Grafu 7.



Graf 7

Závěr

Cílem této práce bylo analyzovat časový průběh nezaměstnanosti v Olomouckém kraji od roku 1998 do roku 2008. Stěžejní část této práce byla právě tvorba statistického modelu, který by měl zkoumáným datům co nejlépe vyhovovat. Součástí tvorby tohoto modelu byla identifikace periodické složky, kterou jsme do výsledného modelu zahrnuli. Konečně, poté co jsme model vytvořili, bylo dalším cílem stanovit prognózu na další období, porovnat model se skutečností a zhodnotit ho.

Teoretická část této práce za výše vytyčených cílů měla za úkol představit teorii potřebnou ke konstrukci odpovídajícího modelu. Za tímto účelem jsme věnovali pozornost klasické dekompozici časových řad a regresní analýze, protože tak lze snadněji proniknout do historického vývoje řady. Protože dalším cílem bylo identifikovat periodicitu, popsali jsme podrobněji i metody, které jsme v rámci zpracování naší časové řady následně použili. Dále jsme se pokoušeli náš model zdokonalit, a proto jsme se zabývali linearizací nelineárního statistického modelu pomocí Taylorova rozvoje. Teoretickou část jsme uzavřeli znaménkovým testem, který nám umožnil ověřit, zda jsme skutečně všechny systematické (nenáhodné) složky eliminovali.

Jak si čtenář mohl pošimnout, je tato práce založena spíše matematicky než ekonomicky. Ačkoliv hlavní pozornost této práce je věnována právě použitému matematickému aparátu, nechybí zde teoretické poznatky o nezaměstnanosti a charakteristiky pracovního trhu v Olomouckém kraji, který představuje naši hlavní motivaci. Nezaměstnaností jsme se zabývali jen stručně, protože hlavním cílem bylo popsat její základní rysy a způsoby jejího měření. Čtenáři jsme seznámili se situací na trhu práce v Olomouckém kraji, protože představuje předmět zkoumání této práce.

Zdrojová data pro tuto analýzu představují obecnou míru nezaměstnanosti v Olomouckém kraji publikovaná na internetových stránkách Českého statistického úřadu. Při práci s daty jsem používala programy MS EXCEL a MATLAB.

V rámci praktické části byl podrobněji uveden postup tvorby hledaného modelu. Nejprve jsme rozdělili naši časovou řadu na složku trendovou a periodickou, a to tak, že jsme data proložili polynomickým trendem, a na základě porovnávání RSC a t-testem

významnosti regresních parametrů jsme určili optimální stupeň polynomu. Výsledkem tohoto postupu pro naše data byl kvadratický trend. Potom jsme od původních dat trend odečetli, abychom mohli analyzovat periodickou složku. Model periodické složky jsme získali využitím periodogramu, Fisherova testu a modelu skrytých period. Periodickou složku jsme následně k trendové funkci přičetli, a tak jsme získali výsledný model. Poté co se prokázalo, že je model vhodný, jsme se rozhodli pokusit se o jeho zpřesnění. Postupovali jsme tak, že jsme výsledky Fisherova testu určili za přibližné hodnoty délky period, a po této úvaze už byl analyzovaný model nelineární. Proto jsme model linearizovali výše uvedeným postupem. Po vyrovnaní hodnot řady jsme se zajímali, zda-li se nám skutečně podařilo model zlepšit, a proto jsme porovnali RSČ s modelem, v němž jsme výsledky Fisherova testu přímo zahrnuli. Závěrem tohoto postupu je, že pro naši časovou řadu očekávané zlepšení skutečně nastalo. Protože ale nám t - test prokázal, že všechny parametry jsou nevýznamně odlišné od nuly, v dalším jsme parametry vyloučili a pracovali jsme s původním modelem, který věrně popisuje zkoumaná data, a tím jsme si postačili s periodogramem.

Dále jsme vytvořili předpovědi na další rok, v našem případě je to rok 2009, a následně jsme sestrojili predikční intervaly. Ve srovnání s již známými údaji jsme zjistili, že nelze takto složitou veličinu předpovídat pouze za nápomocí použitých matematických nástrojů, ale je ovlivněna mnoha faktory. Například pro předvídání budoucnosti trhu práce se české ekonomické autority řídí dle předpovědí určených Českou národní bankou, která ovšem všechny rozhodující faktory zohledňuje a nevychází pouze z matematických modelů (jak to je v mém případě), proto samozřejmě i úspěšněji předvídá budoucí vývoj nejen nezaměstnanosti, ale i dalších ekonomických ukazatelů.

Věřím, že tato práce přispěla k vytvoření komplexního pohledu na situaci na trhu práce v Olomouckém kraji. Celá tato práce se může stát podkladovým materiálem pro jiné statistické analýzy, které se týkají ekonomických charakteristik. Práce především prezentuje univerzální postup pro zpracování ekonomických jevů.

Literatura

- [1] Anděl, J.: Matematická statistika, Praha, SNTL/ALFA, 1978
- [2] Anděl, J.: Statistická analýza časových řad, 1. vydání Praha, SNTL, 1976
- [3] Cípra, T.: Analýza časových řad s aplikacemi v ekonomii, SNTL, Praha, 1976
- [4] Forbelská, M.: Stochastické modelování jednorozměrných časových řad, Brno 2009
- [5] Fišerová, E.: Pravděpodobnost a statistika 3 (texty k přednáškám), 2010,
- [6] Hebák, P., Hustopecký, J., Malá, I.: Vícerozměrné statistické metody (2), Informatorium, Praha, 2005,
- [7] Hindls, R., Seger, J.: Statistické metody v tržním hospodářství, Victoria Publishing, Praha, 1995,
- [8] Holman, R.: Makroekonomie středně pokročilý kurz, 1. vydání Praha, C. H. Beck, Praha, 2004,
- [9] Kropač, J.: Statistika B: jednorozměrné a dvourozměrné datové soubory, regresní analýza, časové řady, VUT, Brno, 2007,
- [10] Kubáček, L., Kubáčková, L.: Statistika a metrologie, Vydavatelství UP Olomouc, 2000
- [11] Kunderová, P.: Základy pravděpodobnosti a matematické statistiky, Vydavatelství UP Olomouc, 2004,
- [12] Meloun, M., Militký, J.: Statistické zpracování experimentálních dat v chemometrii, biometrii, ekonometrii a v dalších oborech přírodních, technických a společenských věd, Praha, ARS MAGNA, 1998,
- [13] Müller, I.: Časové řady (texty k přednáškám), 2010,
- [14] Rao, R. C.: Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha, 1978,
- [15] Stuart, A.: The power of two Difference – Sign Tests, Journal of the American Statistical Association. Vol. 47, Issue 259, 1952, 416 – 424.

Internetové zdroje

[16] Bloomfield, P.: Fourier Analysis of Time Series: An Introduction [online], dostupné z http://www.google.com/books?hl=cs&lr=&id=zQsupRg5rrAC&oi=fnd&pg=PA1&dq=Peter+Bloomfield&ots=SvuhfmwyVl&sig=VqajWoKN5eZ8jy_wxso2stjqj4Q#v=onepage&q&f=false [citováno 15. 3. 2011]

[17] Český statistický úřad - nezaměstnanost dle oblastí a krajů, <http://www.czso.cz/csu/2010edicniplan.nsf/p/3103-10> [citováno 15. 3. 2011]

[18] Beneš, R.: Vývoj nezaměstnanosti [online], dostupné z <http://www.e-polis.cz/nezarazene-clanky/28-vyvoj-nezamestnanosti.html> [citováno 15. 3. 2011]

[19] Finance.cz: Nezaměstnanost v ČR klesá [online], dostupné z <http://www.finance.cz/zpravy/finance/106299-nezamestnanost-v-cr-klesa/> [citováno 15. 3. 2011]

[20] Horová, Z.: Demonstrace základních vlastností šíření vln na datech umělých družic [online], diplomová práce, Praha, 2007 dostupné z http://fyzweb.cz/materialy/hvizdy/hvizdy_cele.pdf [citováno 15. 3. 2011]

[21] BussinessInfo.cz: Charakteristika Olomouckého kraje [online], dostupné z <http://www.businessinfo.cz/cz/clanek/olomoucky-kraj/charakteristika-olomouckeho-kraje/1000930/41072/> [citováno 15. 3. 2011]

[22] EUROPA-EURES: Informace o pracovním trhu – Olomoucký kraj [online], dostupné z <http://ec.europa.eu/eures/main.jsp?catId=2759&acro=lmi&lang=cs&countryId=CZ®ionId=CZ0&nuts2Code=CZ07&nuts3Code=CZ071> [citováno 15. 3. 2011]

[23] Trh práce a zaměstnanost [online], dostupné z http://knihy.cpress.cz/DataFiles/Book/00002405/Download/ke0325_ukazkova%20kap.pdf [citováno 15. 3. 2011]

PŘÍLOHA – TABULKY

Tabulka 10: Čebyševovy polynomy

$\varphi_0(x)$	$\varphi_1(x)$	$\varphi_2(x)$	$\varphi_3(x)$
1	-19,5	247	-2741,7
1	-18,5	209	-1898,1
1	-17,5	173	-1165,5
1	-16,5	139	-537,9
1	-15,5	107	-9,3
1	-14,5	77	426,3
1	-13,5	49	774,9
1	-12,5	23	1042,5
1	-11,5	-1	1235,1
1	-10,5	-23	1358,7
1	-9,5	-43	1419,3
1	-8,5	-61	1422,9
1	-7,5	-77	1375,5
1	-6,5	-91	1283,1
1	-5,5	-103	1151,7
1	-4,5	-113	987,3
1	-3,5	-121	795,9
1	-2,5	-127	583,5
1	-1,5	-131	356,1
1	-0,5	-133	119,7
1	0,5	-133	-119,7
1	1,5	-131	-356,1
1	2,5	-127	-583,5
1	3,5	-121	-795,9
1	4,5	-113	-987,3
1	5,5	-103	-1151,7
1	6,5	-91	-1283,1
1	7,5	-77	-1375,5
1	8,5	-61	-1422,9
1	9,5	-43	-1419,3
1	10,5	-23	-1358,7
1	11,5	-1	-1235,1
1	12,5	23	-1042,5
1	13,5	49	-774,9
1	14,5	77	-426,3
1	15,5	107	9,3
1	16,5	139	537,9
1	17,5	173	1165,5
1	18,5	209	1898,1
1	19,5	247	2741,7

Tabulka 11: Polynomická regrese

t	Y_t	$\hat{Y}_t^{(0)}$	$\hat{\varepsilon}_t^{(0)}$	$\hat{Y}_t^{(1)}$	$\hat{\varepsilon}_t^{(1)}$	$\hat{Y}_t^{(2)}$	$\hat{\varepsilon}_t^{(2)}$	$\hat{Y}_t^{(3)}$	$\hat{\varepsilon}_t^{(3)}$
1	9,6705	9,5392	0,1313	12,3420	-2,6715	10,7882	-1,1177	11,2356	-1,5651
2	10,4575	9,5392	0,9183	12,1982	-1,7408	10,8835	-0,4260	11,1932	-0,7358
3	10,5454	9,5392	1,0062	12,0545	-1,5091	10,9662	-0,4208	11,1564	-0,6110
4	11,6959	9,5392	2,1567	11,9108	-0,2149	11,0364	0,6595	11,1242	0,5717
5	12,9195	9,5392	3,3803	11,7670	1,1524	11,0940	1,8255	11,0955	1,8240
6	13,1417	9,5392	3,6025	11,6233	1,5183	11,1389	2,0027	11,0694	2,0723
7	12,8614	9,5392	3,3222	11,4796	1,3818	11,1713	1,6901	11,0449	1,8165
8	12,2036	9,5392	2,6644	11,3358	0,8677	11,1912	1,0124	11,0210	1,1825
9	11,9683	9,5392	2,4291	11,1921	0,7762	11,1984	0,7699	10,9969	0,9714
10	9,9485	9,5392	0,4093	11,0484	-1,0998	11,1931	-1,2445	10,9714	-1,0228
11	9,6560	9,5392	0,1168	10,9046	-1,2487	11,1751	-1,5192	10,9435	-1,2876
12	10,0484	9,5392	0,5092	10,7609	-0,7125	11,1446	-1,0963	10,9125	-0,8641
13	10,2616	9,5392	0,7224	10,6172	-0,3556	11,1016	-0,8400	10,8771	-0,6156
14	9,1218	9,5392	-0,4174	10,4734	-1,3517	11,0459	-1,9241	10,8365	-1,7147
15	9,7422	9,5392	0,2030	10,3297	-0,5875	10,9776	-1,2355	10,7897	-1,0475
16	9,3235	9,5392	-0,2156	10,1860	-0,8624	10,8968	-1,5733	10,7357	-1,4122
17	8,8975	9,5392	-0,6417	10,0422	-1,1448	10,8034	-1,9059	10,6735	-1,7761
18	9,0333	9,5392	-0,5059	9,8985	-0,8653	10,6974	-1,6642	10,6022	-1,5689
19	9,8969	9,5392	0,3578	9,7548	0,1422	10,5789	-0,6819	10,5207	-0,6238
20	10,4418	9,5392	0,9026	9,6111	0,8308	10,4477	-0,0059	10,4282	0,0137
21	12,0777	9,5392	2,5385	9,4673	2,6104	10,3040	1,7738	10,3235	1,7542
22	11,6929	9,5392	2,1537	9,3236	2,3693	10,1477	1,5453	10,2058	1,4872
23	12,2632	9,5392	2,7240	9,1799	3,0834	9,9788	2,2844	10,0740	2,1892
24	12,0607	9,5392	2,5215	9,0361	3,0245	9,7973	2,2634	9,9272	2,1335
25	11,8026	9,5392	2,2634	8,8924	2,9102	9,6032	2,1993	9,7643	2,0382
26	8,9355	9,5392	-0,6037	8,7487	0,1869	9,3966	-0,4611	9,5845	-0,6490
27	9,4467	9,5392	-0,0925	8,6049	0,8418	9,1774	0,2693	9,3867	0,0600
28	9,7950	9,5392	0,2558	8,4612	1,3338	8,9456	0,8494	9,1700	0,6250
29	9,2890	9,5392	-0,2502	8,3175	0,9716	8,7012	0,5878	8,9334	0,3557
30	8,0891	9,5392	-1,4501	8,1737	-0,0847	8,4442	-0,3552	8,6758	-0,5868
31	7,8637	9,5392	-1,6755	8,0300	-0,1663	8,1747	-0,3110	8,3964	-0,5327
32	7,4624	9,5392	-2,0768	7,8863	-0,4239	7,8925	-0,4302	8,0941	-0,6317
33	6,7749	9,5392	-2,7643	7,7425	-0,9677	7,5978	-0,8230	7,7680	-0,9931
34	6,3920	9,5392	-3,1471	7,5988	-1,2067	7,2906	-0,8985	7,4170	-1,0250
35	6,2451	9,5392	-3,2941	7,4551	-1,2100	6,9707	-0,7256	7,0402	-0,7952
36	5,9826	9,5392	-3,5566	7,3113	-1,3287	6,6382	-0,6556	6,6367	-0,6541
37	6,2355	9,5392	-3,3037	7,1676	-0,9321	6,2932	-0,0577	6,2054	0,0301
38	5,7309	9,5392	-3,8083	7,0239	-1,2930	5,9356	-0,2047	5,7454	-0,0145
39	5,9946	9,5392	-3,5446	6,8801	-0,8855	5,5654	0,4292	5,2557	0,7390
40	5,5983	9,5392	-3,9409	6,7364	-1,1381	5,1826	0,4157	4,7352	0,8631

$\hat{Y}_t^{(i)}$ vyrovnané hodnoty pro i -té aproximace, $i = 0, 1, 2, 3$,

$\hat{\varepsilon}_t^{(i)}$ odhad rezidua pro i -té aproximace.

Tabulka 12: Model skrytých period

t	$\hat{\varepsilon}_t^{(2)} \cos\left(\frac{2\pi t}{T_1}\right)$	$\hat{\varepsilon}_t^{(2)} \sin\left(\frac{2\pi t}{T_1}\right)$	$\hat{\varepsilon}_t^{(2)} \cos\left(\frac{2\pi t}{T_2}\right)$	$\hat{\varepsilon}_t^{(2)} \sin\left(\frac{2\pi t}{T_2}\right)$	$\hat{\varepsilon}_t^{(2)} \cos\left(\frac{2\pi t}{T_3}\right)$	$\hat{\varepsilon}_t^{(2)} \sin\left(\frac{2\pi t}{T_3}\right)$
1	-1,0630	-0,3454	-0,7904	-0,7904	-0,9959	-0,5074
2	-0,3447	-0,2504	0,0000	-0,4260	-0,2504	-0,3447
3	-0,2474	-0,3405	0,2976	-0,2976	-0,0658	-0,4157
4	0,2038	0,6272	-0,6595	0,0000	-0,2038	0,6272
5	0,0000	1,8255	-1,2908	-1,2908	-1,2908	1,2908
6	-0,6189	1,9047	0,0000	-2,0027	-1,9047	0,6189
7	-0,9934	1,3673	1,1951	-1,1951	-1,6693	-0,2644
8	-0,8191	0,5951	1,0124	0,0000	-0,8191	-0,5951
9	-0,7322	0,2379	0,5444	0,5444	-0,3495	-0,6860
10	1,2445	0,0000	0,0000	-1,2445	0,0000	1,2445
11	1,4448	0,4695	1,0742	-1,0742	-0,6897	1,3536
12	0,8869	0,6444	1,0963	0,0000	-0,8869	0,6444
13	0,4937	0,6796	0,5940	0,5940	-0,8297	0,1314
14	0,5946	1,8299	0,0000	1,9241	-1,8299	-0,5946
15	0,0000	1,2355	-0,8736	0,8736	-0,8736	-0,8736
16	-0,4862	1,4963	-1,5733	0,0000	-0,4862	-1,4963
17	-1,1203	1,5419	-1,3477	-1,3477	0,2982	-1,8825
18	-1,3463	0,9782	0,0000	-1,6642	0,9782	-1,3463
19	-0,6485	0,2107	0,4822	-0,4822	0,6076	-0,3096
20	-0,0059	0,0000	0,0059	0,0000	0,0059	0,0000
21	1,6869	0,5481	-1,2542	-1,2542	-1,5804	-0,8053
22	1,2502	0,9083	0,0000	-1,5453	-0,9083	-1,2502
23	1,3428	1,8482	1,6153	-1,6153	-0,3574	-2,2563
24	0,6994	2,1526	2,2634	0,0000	0,6994	-2,1526
25	0,0000	2,1993	1,5552	1,5552	1,5552	-1,5552
26	0,1425	-0,4385	0,0000	-0,4611	-0,4385	0,1425
27	-0,1583	0,2179	-0,1905	0,1905	0,2660	0,0421
28	-0,6872	0,4993	-0,8494	0,0000	0,6872	0,4993
29	-0,5591	0,1817	-0,4157	-0,4157	0,2669	0,5238
30	0,3552	0,0000	0,0000	0,3552	0,0000	-0,3552
31	0,2958	0,0961	-0,2199	0,2199	0,1412	-0,2771
32	0,3480	0,2529	-0,4302	0,0000	0,3480	-0,2529
33	0,4837	0,6658	-0,5819	-0,5819	0,8128	-0,1287
34	0,2777	0,8545	0,0000	-0,8985	0,8545	0,2777
35	0,0000	0,7256	0,5131	-0,5131	0,5131	0,5131
36	-0,2026	0,6235	0,6556	0,0000	0,2026	0,6235
37	-0,0339	0,0467	0,0408	0,0408	-0,0090	0,0570
38	-0,1656	0,1203	0,0000	0,2047	-0,1203	0,1656
39	0,4082	-0,1326	0,3035	-0,3035	0,3825	-0,1949
40	0,4157	0,0000	0,4157	0,0000	0,4157	0,0000
Σ	2,3419	26,0770	3,1875	-12,9018	-7,5243	-9,7889
j	1	2	3	4	5	6
$\hat{\beta}_j$	0,1171	1,3039	0,1594	-0,6451	-0,3762	-0,4894

Tabulka 13: Výsledný model

t	\hat{Y}_t	$\hat{\varepsilon}_t$
1	9,79515	-0,12469
2	10,13447	0,32300
3	10,84133	-0,29593
4	11,76124	-0,06537
5	12,58712	0,33236
6	13,01330	0,12835
7	12,89091	-0,02949
8	12,29820	-0,09463
9	11,48887	0,47939
10	10,75097	-0,80244
11	10,25937	-0,60341
12	10,00581	0,04256
13	9,84310	0,41845
14	9,61148	-0,48968
15	9,26676	0,47541
16	8,92950	0,39405
17	8,82244	0,07503
18	9,13162	-0,09836
19	9,87380	0,02315
20	10,85198	-0,41015
21	11,73067	0,34705
22	12,19326	-0,50033
23	12,09529	0,16792
24	11,52825	0,53241
25	10,75883	1,04374
26	10,07873	-1,14322
27	9,65078	-0,20407
28	9,43715	0,35786
29	9,24711	0,04191
30	8,87196	-0,78291
31	8,22479	-0,36110
32	7,40180	0,06056
33	6,62955	0,14532
34	6,13076	0,26128
35	5,99005	0,25503
36	6,10079	-0,11819
37	6,22630	0,00918
38	6,14014	-0,40926
39	5,76255	0,23208
40	5,21119	0,38713

Tabulka 14: Znaménkový test

t	$\hat{\varepsilon}_t$	$\hat{\varepsilon}_{t+1} - \hat{\varepsilon}_t$	Z_t
1	-0,12469		
2	0,32300	0,4477	1
3	-0,29593	-0,6189	0
4	-0,06537	0,2306	1
5	0,33236	0,3977	1
6	0,12835	-0,2040	0
7	-0,02949	-0,1578	0
8	-0,09463	-0,0651	0
9	0,47939	0,5740	1
10	-0,80244	-1,2818	0
11	-0,60341	0,1990	1
12	0,04256	0,6460	1
13	0,41845	0,3759	1
14	-0,48968	-0,9081	0
15	0,47541	0,9651	1
16	0,39405	-0,0814	0
17	0,07503	-0,3190	0
18	-0,09836	-0,1734	0
19	0,02315	0,1215	1
20	-0,41015	-0,4333	1
21	0,34705	0,7572	1
22	-0,50033	-0,8474	0
23	0,16792	0,6682	1
24	0,53241	0,3645	1
25	1,04374	0,5113	1
26	-1,14322	-2,1870	0
27	-0,20407	0,9391	1
28	0,35786	0,5619	1
29	0,04191	-0,3160	0
30	-0,78291	-0,8248	0
31	-0,36110	0,4218	1
32	0,06056	0,4217	1
33	0,14532	0,0848	1
34	0,26128	0,1160	1
35	0,25503	-0,0063	0
36	-0,11819	-0,3732	0
37	0,00918	0,1274	1
38	-0,40926	-0,4184	0
39	0,23208	0,6413	1
40	0,38713	0,1551	1
S			23

Tabulka 15: Matice v linearizovaném modelu

t	$\{f_0\}_t$
1	-0,386587
2	-0,401073
3	0,012553
4	0,767614
5	1,567238
6	2,055496
7	2,003003
8	1,423081
9	0,554993
10	-0,272741
11	-0,817820
12	-1,037171
13	-1,075233
14	-1,140182
15	-1,347123
16	-1,626220
17	-1,754024
18	-1,491573
19	-0,747380
20	0,333936
21	1,415140
22	2,123307
23	2,234777
24	1,784827
25	1,040464
26	0,352207
27	-0,030982
28	-0,079777
29	0,028102
30	0,038549
31	-0,210733
32	-0,685063
33	-1,172096
34	-1,412259
35	-1,260579
36	-0,781482
37	-0,217997
38	0,148269
39	0,164285
40	-0,099745

Tabulka 16: Matice v linearizovaném modelu – pokračování

t	$\cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$	$\cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$	$\cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$	$\cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$	$\cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$	$\cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$	$f_{.7}^*)$	$f_{.8}^{**})$	$f_{.9}^{***})$
1	0,9511	0,3090	0,7071	0,7071	0,8910	0,4540	-0,0189	0,0558	0,0094
2	0,8090	0,5878	0,0000	1,0000	0,5878	0,8090	-0,0310	0,0313	-0,0012
3	0,5878	0,8090	-0,7071	0,7071	0,1564	0,9877	-0,0317	-0,1012	-0,0313
4	0,3090	0,9511	-1,0000	0,0000	-0,3090	0,9511	-0,0183	-0,2533	-0,0720
5	0,0000	1,0000	-0,7071	-0,7071	-0,7071	0,7071	0,0092	-0,2792	-0,1082
6	-0,3090	0,9511	0,0000	-1,0000	-0,9511	0,3090	0,0485	-0,0939	-0,1234
7	-0,5878	0,8090	0,7071	-0,7071	-0,9877	-0,1564	0,0947	0,2360	-0,1050
8	-0,8090	0,5878	1,0000	0,0000	-0,8090	-0,5878	0,1412	0,5067	-0,0494
9	-0,9511	0,3090	0,7071	0,7071	-0,4540	-0,8910	0,1804	0,5026	0,0359
10	-1,0000	0,0000	0,0000	1,0000	0,0000	-1,0000	0,2048	0,1565	0,1330
11	-0,9511	-0,3090	-0,7071	0,7071	0,4540	-0,8910	0,2080	-0,3709	0,2167
12	-0,8090	-0,5878	-1,0000	0,0000	0,8090	-0,5878	0,1859	-0,7600	0,2617
13	-0,5878	-0,8090	-0,7071	-0,7071	0,9877	-0,1564	0,1372	-0,7260	0,2491
14	-0,3090	-0,9511	0,0000	-1,0000	0,9511	0,3090	0,0641	-0,2191	0,1728
15	0,0000	-1,0000	0,7071	-0,7071	0,7071	0,7071	-0,0276	0,5058	0,0424
16	0,3090	-0,9511	1,0000	0,0000	0,3090	0,9511	-0,1293	1,0133	-0,1168
17	0,5878	-0,8090	0,7071	0,7071	-0,1564	0,9877	-0,2300	0,9494	-0,2692
18	0,8090	-0,5878	0,0000	1,0000	-0,5878	0,8090	-0,3177	0,2817	-0,3766
19	0,9511	-0,3090	-0,7071	0,7071	-0,8910	0,4540	-0,3809	-0,6406	-0,4075
20	1,0000	0,0000	-1,0000	0,0000	-1,0000	0,0000	-0,4096	-1,2667	-0,3459
21	0,9511	0,3090	-0,7071	-0,7071	-0,8910	-0,4540	-0,3971	-1,1728	-0,1969
22	0,8090	0,5878	0,0000	-1,0000	-0,5878	-0,8090	-0,3408	-0,3443	0,0130
23	0,5878	0,8090	0,7071	-0,7071	-0,1564	-0,9877	-0,2427	0,7755	0,2398
24	0,3090	0,9511	1,0000	0,0000	0,3090	-0,9511	-0,1099	1,5200	0,4318
25	0,0000	1,0000	0,7071	0,7071	0,7071	-0,7071	0,0460	1,3962	0,5408
26	-0,3090	0,9511	0,0000	1,0000	0,9511	-0,3090	0,2100	0,4069	0,5345
27	-0,5878	0,8090	-0,7071	0,7071	0,9877	0,1564	0,3652	-0,9104	0,4051
28	-0,8090	0,5878	-1,0000	0,0000	0,8090	0,5878	0,4942	-1,7733	0,1730
29	-0,9511	0,3090	-0,7071	-0,7071	0,4540	0,8910	0,5814	-1,6196	-0,1158
30	-1,0000	0,0000	0,0000	-1,0000	0,0000	1,0000	0,6144	-0,4695	-0,3989
31	-0,9511	-0,3090	0,7071	-0,7071	-0,4540	0,8910	0,5862	1,0452	-0,6107
32	-0,8090	-0,5878	1,0000	0,0000	-0,8090	0,5878	0,4956	2,0266	-0,6979
33	-0,5878	-0,8090	0,7071	0,7071	-0,9877	0,1564	0,3482	1,8430	-0,6324
34	-0,3090	-0,9511	0,0000	1,0000	-0,9511	-0,3090	0,1557	0,5321	-0,4196
35	0,0000	-1,0000	-0,7071	0,7071	-0,7071	-0,7071	-0,0644	-1,1801	-0,0990
36	0,3090	-0,9511	-1,0000	0,0000	-0,3090	-0,9511	-0,2908	-2,2800	0,2628
37	0,5878	-0,8090	-0,7071	-0,7071	0,1564	-0,9877	-0,5005	-2,0664	0,5860
38	0,8090	-0,5878	0,0000	-1,0000	0,5878	-0,8090	-0,6707	-0,5947	0,7951
39	0,9511	-0,3090	0,7071	-0,7071	0,8910	-0,4540	-0,7819	1,3150	0,8365
40	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	-0,8193	2,5333	0,6919

$$*) f_{.7} = \beta_1^{(0)} \frac{2\pi t}{(\beta_7^{(0)})^2} \sin\left(\frac{2\pi t}{\beta_7^{(0)}}\right) - \beta_2^{(0)} \frac{2\pi t}{(\beta_7^{(0)})^2} \cos\left(\frac{2\pi t}{\beta_7^{(0)}}\right)$$

$$**) f_{.8} = \beta_3^{(0)} \frac{2\pi t}{(\beta_8^{(0)})^2} \sin\left(\frac{2\pi t}{\beta_8^{(0)}}\right) - \beta_4^{(0)} \frac{2\pi t}{(\beta_8^{(0)})^2} \cos\left(\frac{2\pi t}{\beta_8^{(0)}}\right)$$

$$***) f_{.9} = \beta_5^{(0)} \frac{2\pi t}{(\beta_9^{(0)})^2} \sin\left(\frac{2\pi t}{\beta_9^{(0)}}\right) - \beta_6^{(0)} \frac{2\pi t}{(\beta_9^{(0)})^2} \cos\left(\frac{2\pi t}{\beta_9^{(0)}}\right)$$

Tabulka 17: Výsledný model po linearizaci

t	\hat{Y}_t	$\hat{\varepsilon}_t$
1	9,68510	-0,01464
2	10,14413	0,31335
3	10,92983	-0,38443
4	11,87416	-0,17829
5	12,68185	0,23763
6	13,06883	0,07282
7	12,90018	-0,03876
8	12,25463	-0,05105
9	11,38189	0,58638
10	10,57714	-0,62861
11	10,04092	-0,38496
12	9,79944	0,24893
13	9,72508	0,53647
14	9,64096	-0,51917
15	9,44668	0,29549
16	9,19135	0,13220
17	9,05100	-0,15353
18	9,22217	-0,18891
19	9,79204	0,10491
20	10,66027	-0,21844
21	11,56031	0,51741
22	12,17307	-0,48014
23	12,27474	-0,01154
24	11,84047	0,22019
25	11,04879	0,75378
26	10,18511	-1,24960
27	9,49668	-0,04997
28	9,07697	0,71804
29	8,83990	0,44912
30	8,59400	-0,50495
31	8,17137	-0,30768
32	7,53672	-0,07437
33	6,81415	-0,03928
34	6,21660	0,17544
35	5,91850	0,32657
36	5,94477	0,03782
37	6,14119	0,09429
38	6,24571	-0,51484
39	6,02415	-0,02952
40	5,39649	0,20184