Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Tereza Zachařová

Modelování nezaměstnanosti pomocí nehomogenního Poissonova procesu

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Soňa Reisnerová

Studijní program: Matematika, obecná matematika

2006

Děkuji vedoucí práce RNDr. Soně Reisnerové za náměty, rady a připomínky, které mi pomohly při tvorbě této práce.

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

V Praze dne 23. 5. 2006

Tereza Zachařová

Tereza Zachařová

Obsah

Název práce: Modelování nezaměstnanosti pomocí nehomogenního Poissonova procesu
Autor: Tereza Zachařová
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: RNDr. Soňa Reisnerová
e-mail vedoucího: reisnero@utia.cas.cz

Abstrakt: V předložené práci se zabýváme modelováním nezaměstnanosti pomocí nehomogenního Poissonova procesu. Předpokládáme, že přírůstky nezaměstnaných jsou nezávislé náhodné veličiny pozorované v čase a že mají Poissonovo rozdělení s parametrem závislým na čase. Zmíníme základy teorie odhadu a budeme se podrobněji zabývat metodou maximální věrohodnosti, vlastnostmi odhadů získaných touto metodou a základními statistickými testy významnosti parametrů. Dále aplikuje navržený model na reálná data týkající se nezaměstnanosti v České republice.

Klíčová slova: nezaměstnanost, nehomogenní Poissonův proces, teorie odhadu, metoda maximální věrohodnosti

Title: Unemployment Modelling with via nonhomogenous Poisson Process
Author: Tereza Zachařová
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Soňa Reisnerová
Supervisor's e-mail address: reisnero@utia.cas.cz

Abstract: In our work we study the unemployment rate via nonhomogenous Poisson Process. We assume that monthly increases of unemployment are independent random variables that follow Poisson distribution with a parameter depending on time. We also mention fundamentals of the Theory of Estimation, we study Maximum Likelihood Method and properties of the respective estimation. We apply the designed model with real data of the unemployment rate in the Czech Republic.

Keywords: Unemployment, nonhomogenous Poisson Process, Theory of Estimation, Maximum Likelihood estimating

Kapitola 1

ÚVOD

Nezaměstnanost je dnes vnímána jako nepříjemný průvodní jev rozvoje společenských systémů založených na tržní ekonomice. Od dob Velké hospodářské krize je jí v ekonomii věnována zvýšená pozornost a stala se jedním z hlavních témat hospodářské politiky a politických bojů.

Během let 1929 až 1933 vzrostla nezaměstnanost v USA z 1,6 na 12,8 milionu lidí, osobní příjmy Američanů klesly na polovinu. O padesát procent se díky nadúrodě snížily také ceny zemědělských výrobků, což znamenalo katastrofu pro farmáře. V troskách se ocitl americký akciový trh, zaniklo devět tisíc bank, zavíraly se továrny, doly, tisícovky podnikatelů a farmářů zbankrotovaly. Dalo se něco tak strašného předvídat, nebo dokonce tomu zabránit? Velké depresi předcházelo ve Spojených státech období velké prosperity. Prudce vzrostla průmyslová výroba a trh byl zaplaven zbožím. Rozvoj rozhlasu, televize a automobilového průmyslu znamenaly převrat v doposud relativně střídmém životním stylu a lidé začali horečně utrácet. Začaly se však také prohlubovat třídní rozdíly, osmdesát procent obyvatelstva nemělo žádné úspory. Poptávka přestávala stačit stále rostoucí nabídce zboží, stále více lidí kupovalo na dluh[1].

Z hospodářské krize vyvedl Spojené státy až nový prezident Franklin Delano Roosevelt, který při své inauguraci vyhlásil Nový úděl (New Deal), tedy jakýsi plán na obnovu hospodářství. S důsledky krize však bojovaly Spojené státy až do druhé světové války.

Nezaměstnanost se stala jedním z hlavních problémů společnosti, a proto se objevuje zájem tento jev více studovat, učit se ho modelovat, předpovídat a v neposlední řadě i řídit, aby se něco takového nemohlo opakovat. V této práci se budeme zabývat modelováním nezaměstnanosti pomocí nehomogenního Poissonova procesu a pokusíme se ji předpovídat. V první kapitole

[1] více viz. materiál k přednášce ZZZ261 Ekonomie II v příloze

se seznámíme podrobněji s pojmem nezaměstnanost, dále následuje kapitola o základech teorie odhadu a metodě maximální věrohodnosti, kterou jsme použili k odhadnutí modelu. V třetí kapitole představíme samotný model a v závěru je uvedeno numerické řešení představeného modelu pro vývoj nezaměstnanosti v České republice v letech 1999 až 2006.

Kapitola 2

NEZAMĚSTNANOST

2.1 Definice nezaměstnanosti

Existuje několik definic nezaměstnanosti, které jsou používány v různých oborech, neboť ne všechny jsou pro určitý obor stejně vhodné. V této práci je pro nás nezaměstnaným ten, kdo je v právním řádu České republiky nazýván uchazečem o zaměstnání. Dle zákona č. 435/2004 Sb., o zaměstnanosti je to taková fyzická osoba, která osobně požádá o zprostředkování vhodného zaměstnání úřad práce, v jehož správním obvodu má bydliště, a která splňuje zákonem stanovené podmínky (např. není osoba samostatně výdělečně činná, student, nesmí být ani soudcem, poslancem, senátorem, členem vlády... zákon vyjmenovává mnoho dalších významných funkcí včetně prezidenta republiky). Tuto definici používá i Ministerstvo práce a sociálních věcí, na jehož internetových stránkách[1] jsou také zveřejněna data, která v této práci budeme používat.

Srozumitelnější definici nezaměstnanosti lze nalézt na webu Českého statistického úřadu [2]. Nezaměstnaní jsou všechny osoby 15-ti leté a starší, které ve sledovaném období souběžně splňovaly následující tři podmínky:

- nebyly zaměstnané,
- aktivně hledaly práci,
- byly připraveny k nástupu do práce, t.j. během referenčního období byly k dispozici okamžitě nebo nejpozději do 14 dnů pro výkon zaměstnání.

Pokud osoby nesplňují alespoň jednu ze tří uvedených podmínek, jsou klasifikovány jako zaměstnané nebo ekonomicky neaktivní. Jedinou výjimkou je

[1] http://portal.mpsv.cz/sz/stat/nz/mes
[2] http://www.czso.cz/csu/redakce.nsf/i/zam_vsps

skupina osob, které práci nehledají, protože ji již našly, ale nástup je stanoven na dobu nejpozději do 14 dnů. Tyto osoby jsou zařazeny rovněž mezi nezaměstnané. Zařazení do této kategorie nesouvisí s kategorií registrovaných uchazečů o zaměstnání na úřadech práce a ani s faktem, zda tyto osoby pobírají či nepobírají příspěvek v nezaměstnanosti či jiné sociální dávky nebo příspěvky.

V ekonomii se za nezaměstnané považují osoby produktivního věku, které splňují dvě podmínky [2, str. 65]:

- nemají placené zaměstnání ani příjem ze sebezaměstnání, jsou dočasně uvolněny z práce a očekávají, že budou znovu zaměstnány;
- aktivně hledají práci a jsou ochotny do práce nastoupit.

Naopak zaměstnaný může být i člověk momentálně nepracující, ale musí mít vazbu na zaměstnání (např. nemoc, dovolená, mateřská dovolená).

Jako poslední zmiňme sociologickou definici. Nezaměstnaný je ten, kdo nemá práci a aktivně ji hledá, nebo hledal, ale neúspěšně a už nemá síly hledat dál. Sociologie se jako jediná zabývá nedostatečným zaměstnáním. Jako nezaměstnané počítá i osoby pracující nedobrovolně na částečný úvazek, které by daly přednost práci na plný úvazek, ale nemohou ji najít.

2.2 Druhy nezaměstnanosti

Problematika nezaměstnanosti patří již dlouhá desetiletí ke sledovaným problémům a její pojetí se vyvíjí. Neoklasická ekonomie dospěla k *teorii dobrovolné nezaměstnanosti* [2, str. 65]. Práce je vnímána jako jedna z komodit, která je nabízena a poptávána na trhu, speciálně mluvíme o trhu práce. Subjekt nabízející práci má možnost volby, zda bude vynakládat práci nebo upřednostní volný čas. Hlavním faktorem, který ovlivňuje jeho rozhodování je reálná mzda. Dobrovolná nezaměstnanost znamená, že je preferován volný čas před konáním práce. Dobrovolně nezaměstnaní mohou mít nabídky pracovních příležitostí, ale buď aktivně hledají jiné, např. lépe placené místo, nebo práci nehledají vůbec. Termín dobrovolná nezaměstnanost je často užíván pro označení stavu, kdy je počet nezaměstnaných nižší nebo roven počtu volných pracovních míst, tedy situace, kdy ten, kdo hledá práci, ji také může najít. Pokud by byl trh práce dokonale konkurenční, pak by vyrovnávání nabídky a poptávky dospělo ke stavu, kdy by jiná než dobrovolná nezaměstnanost neexistovala. To ovšem vylučuje nepružnost mezd, která je způsobena jakousi uměle (vládou) vytvořenou spodní hladinou (minimální mzda). Snižování platů brání různé kolektivní dohody a tlak odborů. Proto je mnoho lidí, kteří jsou neúspěšní při hledání zaměstnání. Od vydání Keynesovy Obecné teorie zaměstnanosti, úroku a peněz (1936) se teorie ekonomie rozšířila o tzv. *nedobrovolnou nezaměstnanost* [2, str. 67].

Nedobrovolnou nezaměstnanost můžeme dále rozdělit na:

- **Frikční** (stojí na pomezí mezi dobrovolnou a nedobrovolnou), která vzniká v důsledku neustálého pohybu lidí mezi místy či pracovními příležitostmi. Vždy existují na trhu práce ti, kteří byli propuštěni v důsledku skutečnosti, že firmy vznikají a zanikají, dochází k technologickým změnám, které mohou vést k likvidaci pracovišť. Do této skupiny patří i ti, kteří dobrovolně opustili pracovní místo a hledají jinou, zpravidla lépe placenou práci. Lidé opouštějí místo i z důvodu stěhování a hledají příležitost v novém bydlišti. Do této skupiny můžeme zařadit i ty, kteří hledají první zaměstnání. Z výše uvedených důvodů však také vznikají nová pracovní místa. Svou roli zde hraje i nedostatečná informovanost osob hledajících práci o nabídce vhodných pracovních příležitostí. Frikční nezaměstnanost není vnímána jako závažný problém, neboť po určité době nezaměstnaní nalézají uplatnění. V případě frikční nezaměstnanosti se předpokládá, že jak profesní orientace, tak regionální rozmístění je na straně poptávky a nabídky v souladu viz.[2, str. 66].

- **Srukturální**, která představuje složitější poruchu. Postihuje některá odvětví výroby a je vyvolána nedostatečnou poptávkou po určité produkci statků. Důsledkem je klesající poptávka po práci v odvětví produkujícím výše uvedené statky a útlum těchto výrob. Útlum jedněch odvětví či výrob je doprovázen růstem výrob v jiných odvětvích či výrobách. Nezaměstnanost, která vzniká v důsledku útlumu některých výrob je však charakteristická tím, že uvolňovaná pracovní síla nalézá na trhu práce možnost uplatnění na pracovních místech vyžadujících jinou kvalifikaci. Přestože v ekonomice může být počet těch, kteří práci hledají, shodný s počtem volných míst, znamená strukturální nezaměstnanost nerovnováhu regionálních trhů práce. Strukturální nezaměstnanost může být podmíněna existencí bariér v pohybu pracovní síly (dopravní omezení, bydlení atp.) a je hlavní faktorem ovlivňujícím regionální rozdíly míry nezaměstnanosti na trhu práce [2, str. 66].

- **Cyklickou**, která souvisí s ekonomickým cyklem. V období hospodářských poklesů narůstá, naopak při růstu výkonu ekonomiky je potlačována. Vznik cyklické nezaměstnanosti souvisí s poklesem celkové poptávky v ekonomice. Další jev, který prohlubuje cyklickou nezaměstnanost je nepružnost mezd směrem dolů. Cyklická nezaměstnanost bývá dlouhodobá.

- **Sezónní**, která vzniká v souvislosti s ročním obdobím a poklesem poptávky po práci ve specifických oborech, kde provádění určitých činností je možné pouze po nějakou část roku (např. v zemědělství, stavebnictví, cestovním ruchu).

2.3 Nezaměstnanost a hospodářská politika

Nezaměstnanost bývá společností vnímána jako závažný sociální i politický problém. Vlivem historického vývoje a hlavně hospodářské krize ve 30. letech 20. století západoevropské země a USA dospěly k názoru, že je nutné ovlivňovat tržní mechanismus a zmírňovat tak sociální tvrdosti provázející vývoj na trhu práce. Názory těchto zemí výrazně ovlivnily myšlenky britského ekonoma Johna Maynarda Keynese, který zastával názor, že nezaměstnanost je důsledkem poklesu kupní síly. Doporučoval proto zásahy státu, který může zvýšit poptávku a vytvořit tak tlak na vznik nových pracovních míst, protože bude potřeba více práce [4, str.326]. Jedním z cílů hospodářské politiky se stala co nejnižší míra nezaměstnanosti. Ideálem je tzv. plná zaměstnanost, neboli takový stav na trhu práce, kdy neexistuje nedobrovolná nezaměstnanost. Plná zaměstnanost by neměla být vykládána jako povinnost pracovat, protože vždy existují lidé, kteří upřednostní jinou činnost (např. ženy v domácnosti). Plná zaměstnanost odpovídá rovnováze na trhu práce a s ní související stabilizací výkonů a omezení cyklických výkyvů ekonomiky.

Až do 70. let se tato politika osvědčovala, neboť byla provázena víceméně trvalým hospodářským růstem. V posledních desítkách let tento přístup selhává a nezaměstnanost narůstá i přes zásahy státu. Plná zaměstnanost se stává nedosažitelným cílem, a proto je v moderní ekonomii preferován termín přirozená míra nezaměstnanosti, což je do jisté míry analogie. Je to taková míra nezaměstanosti, která by vznikla působením tržních sil (s výše zmíněnými omezeními trhu) a kterou nelze dlouhodobě ovlivňovat nástroji fiskální[3] ani monetární[4] politiky. V současné době se nezaměstnanost nejčastěji hodnotí v souvislosti s inflací. Přirozená míra nezaměstanosti pak odpovídá takové úrovni, kdy je míra inflace stabilní, neroste ani neklesá.

[3]Politika vlády; k ovlivnění ekonomického vývoje se využívá státní rozpočet, jak příjmy v podobě daní, tak výdaje.

[4]Politika centrální banky, která ovlivňuje ekonomiku prostřednictvím měny, hlavně změnami úrokových sazeb.

Kapitola 3

METODA MAXIMÁLNÍ VĚROHODNOSTI

3.1 Základy teorie odhadu

Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení, které závisí na neznámém parametru θ. Množinu hodnot, které může parametr θ nabývat značíme Θ a nazýváme ji *parametrický prostor*.

Rozdělení, z něhož náhodný výběr pochází, může být rozdělením jednorozměrné náhodné veličiny nebo rozdělením k-rozměrného náhodného vektoru; $k \in \mathbb{N}$. Obdobně θ je obecně m-rozměrným vektorovým parametrem $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, $m \in \mathbb{N}$.

Úkolem matematické statistiky je na základě vektoru $\boldsymbol{X}$ získat co možná nejlepší odhad parametru $\boldsymbol{\theta}$, o kterém je předem známo pouze to, že patří do nějakého parametrického prostoru $\Theta \subseteq \mathbb{R}^m$.

Můžeme hledat *bodový odhad*, což znamená najít takové měřitelné zobrazení $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$, aby náhodný vektor $\boldsymbol{T} = g(\boldsymbol{X})$ nezávisel na parametru $\boldsymbol{\theta}$ a v nějakém rozumném smyslu co nejlépe aproximoval jeho hodnotu. Hledáme-li *intervalový odhad*, snažíme se najít takový interval, který s předem danou pravděpodobností pokrývá hodnotu parametru $\boldsymbol{\theta}$. Metodou maximální věrohodnosti se určují zejména bodové odhady, a proto se nadále budeme zabývat pouze těmito.

Při řešení úloh bodového odhadu obvykle nejprve zpracováváme vektor $\boldsymbol{X}$ tak, aby se neztratila žádná informace a přitom se snížila dimenze. Nejčastěji volíme nějaké měřitelné funkce $S_1(\boldsymbol{x}), \ldots, S_k(\boldsymbol{x})$ a vytvoříme náhodný vektor $\boldsymbol{S}(\boldsymbol{X}) = \big(S_1(\boldsymbol{X}), \ldots, S_k(\boldsymbol{X})\big)'$. $\boldsymbol{S}(\boldsymbol{X})$ se mnohdy značí zkráceně $\boldsymbol{S}$. Každému takovémuto vektoru $\boldsymbol{S}$ se říká *statistika*. Pokud $k = m$ může již sama statistika $\boldsymbol{S}$ sloužit jako odhad $\boldsymbol{\theta}$, nebo je $\boldsymbol{S}$ pouze jakýsi mezistupeň a hledaný odhad $\boldsymbol{T}$ získáme až jako funkci složek $\boldsymbol{S}$. Zdůrazněme ještě, že odhad $\boldsymbol{T}$ je funkcí náhodných veličin a jako takový je také náhodnou veličinou.

Definice 1. Říkáme, že odhad $\boldsymbol{T}$ parametru $\boldsymbol{\theta}$ je *nestranný*, nebo také *nevychýlený*, jestliže $\mathsf{E}\boldsymbol{T} = \boldsymbol{\theta}$ pro každé $\boldsymbol{\theta} \in \Theta$.

Místo $\mathsf{E}\boldsymbol{T}$ bychom měli psát přesněji $\mathsf{E}_{\boldsymbol{\theta}}\boldsymbol{T}$, aby bylo vyznačeno, že se střední hodnota počítá při dané hodnotě parametru $\boldsymbol{\theta}$. Pokud nehrozí nedorozumění, dává se přednost stručnějšímu zápisu $\mathsf{E}\boldsymbol{T}$.

Zjednodušeně řečeno požadavek nestrannosti znamená, že odhad je zatížen pouze náhodnou nikoli systematickou chybou, a proto je nestrannost jeden z nejčastějších požadavků na kvalitu odhadu. Někdy se ovšem nedá na této podmínce bezpodmínečně trvat. Nestranný odhad nemusí existovat, nebo existuje jiný odhad, který není nestranný a přitom je z určitého hlediska lepší. Může se také stát, že nestranný odhad sice existuje, ale je naprosto nevhodný, nebo natolik patologický, že se vůbec nepoužívá. Příklady najdeme např. v [1, str. 100].

Definice 2. Nechť $\boldsymbol{T}$ je nestranný odhad parametru $\boldsymbol{\theta}$ a nechť pro všechna $\boldsymbol{\theta} \in \Theta$ a pro jakýkoli jiný nestranný odhad $\boldsymbol{T}^*$ platí nerovnost $\operatorname{var} \boldsymbol{T} \leq \operatorname{var} \boldsymbol{T}^*$. Pak $\boldsymbol{T}$ nazýváme *nejlepší nestranný odhad* parametru $\boldsymbol{\theta}$.

Definice 3. Odhad $\boldsymbol{T}$ se nazývá *asymptoticky nestranný* odhad parametru $\boldsymbol{\theta}$, jestliže $\lim\limits_{n \to \infty} \mathsf{E}\boldsymbol{T} = \boldsymbol{\theta}$ pro každé $\boldsymbol{\theta} \in \Theta$.

Definice 4. Nechť $X_1, X_2, \ldots$ je výběr z nějakého rozdělení, které závisí na $\boldsymbol{\theta}$ a nechť pro každé $n \in \mathbb{N}$ je definován odhad $\boldsymbol{T}_n = g_n(X_1, \ldots, X_n)'$. Říkáme, že odhad $\boldsymbol{T}_n$ je *konzistentní*, jestliže pro každé $\varepsilon > 0$ a pro každé $\boldsymbol{\theta} \in \Theta$ platí:

$$\lim_{n \to \infty} P(\|\boldsymbol{\theta} - \boldsymbol{T}_n\| > \varepsilon) = 0.$$

3.2 Regularita a Fisherova informace

Abychom dále mohli ukázat některé vlastnosti maximálně věrohodných odhadů, zabývejme se chvíli následujícími pojmy.

Definice 5. Nechť θ je jednorozměrný parametr a nechť náhodný vektor $\boldsymbol{X} = (X_1 \ldots X_n)'$ má hustotu $f(\boldsymbol{x}, \theta)$ vzhledem k nějaké σ-konečné míře μ. Říkáme, že systém hustot $\{f(\boldsymbol{x}, \theta), \theta \in \Theta\}$ je *regulární*, platí-li následující podmínky:

(a) Množina Θ je neprázdná a otevřená.

(b) Množina $M = \{\boldsymbol{x} : f(\boldsymbol{x}, \theta) > 0\}$ nezávisí na θ.

(c) Pro μ-skoro všechna $\boldsymbol{x} \in M$ existuje konečná parciální derivace $f'(\boldsymbol{x}, \theta) = \frac{\partial f(\boldsymbol{x}, \theta)}{\partial \theta}$.

(d) Pro všechna $\theta \in \Theta$ platí $\int\limits_M f'(\boldsymbol{x}, \theta) \mathrm{d}\mu(\boldsymbol{x}) = 0$.

(e) Integrál $J_n(\theta) = \int\limits_M \left[\frac{f'(\boldsymbol{x}, \theta)}{f(\boldsymbol{x}, \theta)}\right]^2 f(\boldsymbol{x}, \theta) \mathrm{d}\mu(\boldsymbol{x})$ je konečný a kladný.

Podmínka (d) vlastně říká, že je možné zaměnit pořadí derivace a integrálu při derivování výrazu $\int\limits_M f(\boldsymbol{x}, \theta) \mathrm{d}\mu(\boldsymbol{x}) = 1$. Vzorec pro $J_n(\theta)$ je jednou z možných definic *Fiserovy míry informace*. Můžeme ho také chápat jako integrál funkce $(\frac{f'}{f})^2$ podle míry, která má hustotu f vzhledem k μ. Potom je možné $(\frac{f'}{f})^2$ na doplňku M dodefinovat libovolně a integrovat přes celý prostor $\mathbb{R}^n$. Tím se hodnota $J_n(\theta)$ nezmění. Použijeme-li ještě větu o přenosu integrace dostáváme definici ve tvaru:

Definice 6. Nechť θ je jednorozměrný parametr a nechť náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)'$ má hustotu $f(\boldsymbol{x}, \theta)$ vzhledem k nějaké σ-konečné míře μ. Potom definujeme $J_n(\theta) = \mathsf{E}\left[\frac{f'(\boldsymbol{X}, \theta)}{f(\boldsymbol{X}, \theta)}\right]^2$ a nazýváme *Fisherova míra informace* o parametru θ obsažená v náhodném vektoru $\boldsymbol{X}$.

Je dobré si uvědomit, že $(\ln f)' = \frac{f'}{f}$ a tedy $J_n(\theta) = \mathsf{E}\left[\frac{\partial \ln f(\boldsymbol{X}, \theta)}{\partial \theta}\right]^2$. Pro stručnost označíme $J_1(\theta) = J(\theta)$ a dokážeme si nějaké vlastnosti Fisherovy míry informace.

Nejprve uvedeme zobecnění Fisherovy míry informace pro mnohorozměrný parametr $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$.

Definice 7. Nechť náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)'$ má hustotu $f(\boldsymbol{x}, \boldsymbol{\theta})$ vzhledem k nějaké σ-konečné míře μ a nechť platí:

(a) Množina Θ je neprázdná otevřená v $\mathbb{R}^m$.

(b) $\boldsymbol{\theta} \in \Theta$.

(c) Množina $M = \{\boldsymbol{x} : f(\boldsymbol{x}, \boldsymbol{\theta}) > 0\}$ nezávisí na $\boldsymbol{\theta}$.

(d) Pro μ-skoro všechna $\boldsymbol{x} \in M$ a pro všechna $i = 1, \ldots, m$ existuje parciální derivace $f'_i(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_i}$.

(e) Pro každé i a pro všechna $\boldsymbol{\theta} \in \Theta$ platí $\int\limits_M f'_i(\boldsymbol{x}, \boldsymbol{\theta}) \mathrm{d}\mu(\boldsymbol{x}) = 0$.

(f) Pro každou dvojici (i, j) existuje konečný integrál

$$J_{ij}(\boldsymbol{\theta}) = \int\limits_M \frac{f_i'(\boldsymbol{x},\boldsymbol{\theta}) f_j'(\boldsymbol{x},\boldsymbol{\theta})}{f^2(\boldsymbol{x},\boldsymbol{\theta})} f(\boldsymbol{x},\boldsymbol{\theta}) \mathrm{d}\mu(\boldsymbol{x}).$$

(g) Matice $\boldsymbol{J}_n(\boldsymbol{\theta}) = (J_{ij}(\boldsymbol{\theta}))_{i,j=1}^m$ je pozitivně definitní pro každé $\boldsymbol{\theta} \in \Theta$.

Pak je systém hustot $\{f(\boldsymbol{x},\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ *regulární* a $\boldsymbol{J}_n(\boldsymbol{\theta})$ se nazývá *Fisherova informační matice.*

Opět budeme zkráceně značit $\boldsymbol{J}_1(\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{\theta})$.
Pro výpočet Fisherovy infromace můžeme použít následující větu.

Věta 1. *Nechť je systém hustot $\{f(\boldsymbol{x},\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ regulární a nechť platí:*

(a) Pro μ-skoro všechna $\boldsymbol{x} \in M$ a pro $i, j = 1, \ldots, m$ existují derivace $f_{ij}''(\boldsymbol{x},\boldsymbol{\theta}) = \frac{\partial^2 f(\boldsymbol{x},\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}$.

(b) $\int\limits_M f_{ij}''(\boldsymbol{x},\boldsymbol{\theta}) \mathrm{d}\mu(\boldsymbol{x}) = 0$ pro všechna $\boldsymbol{\theta} \in \Theta$ a $i, j = 1, \ldots, m$.

Pak platí

$$J_{ij}(\boldsymbol{\theta}) = -\int\limits_M \frac{\partial^2 \ln f(\boldsymbol{x},\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} f(\boldsymbol{x},\boldsymbol{\theta}) \mathrm{d}\mu(\boldsymbol{x}) \qquad i, j = 1, \ldots, m.$$

Důkaz. Pro skoro všechna $\boldsymbol{x} \in M$ platí

$$\frac{\partial^2 \ln f}{\partial\theta_i \partial\theta_j} = \frac{f_{ij}''}{f} - \frac{f_i' f_j'}{f^2}$$

Odsud dostáváme

$$J_{ij}(\boldsymbol{\theta}) = \int\limits_M \left(\frac{f_i' f_j'}{f^2}\right) f \mathrm{d}\mu = \int\limits_M \left(\frac{f_{ij}''}{f}\right) f \mathrm{d}\mu(\boldsymbol{x}) - \int\limits_M \frac{\partial^2 \ln f}{\partial\theta_i \partial\theta_j} f \mathrm{d}\mu = -\int\limits_M \frac{\partial^2 \ln f}{\partial\theta_i \partial\theta_j} f \mathrm{d}\mu$$

neboť první z integrálů je podle předpokladu roven nule. □

Ve složitějších praktických příkladech pro výpočet Fisherovy informace Věta 1 obvykle nestačí a jako odhad $\boldsymbol{J}(\boldsymbol{\theta})$ se užívá $-\frac{1}{n} \cdot D$, kde $D = \left(\frac{\partial^2 L(\boldsymbol{x},\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right)_{i,j=1}^m$ je matice druhých derivací logaritmické věrohodnostní funkce (viz. dále), do které se navíc dosazuje odhad $\boldsymbol{\theta}$.

Další užitečná věta pro výpočet Fisherovy informace v mnohorozměrném případě je:

Věta 2. *Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s hustotou $f(x, \theta)$ vzhledem k nějaké σ-konečné míře μ. Nechť je systém hustot $\{f(x, \theta), \theta \in \Theta\}$ regulární a má Fisherovu míru informace $J(\theta)$. Pak náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)'$ má hustotu $f_n(x_1, \ldots, x_n, \theta) = f(x_1, \theta) \ldots f(x_n, \theta)$ vzhledem k míře $\mu_n = \mu \times \cdots \times \mu$. Systém hustotot $\{f_n(\boldsymbol{x}, \theta), \theta \in \Theta\}$ je také regulární a pro jeho Fisherovu míru informace $J_n(\theta)$ platí $J_n(\theta) = nJ(\theta)$.*

Důkaz. Viz. [1, str. 115] □

3.3 Metoda maximální věrohodnosti

Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)'$ je náhodný výběr z rozdělení, kterému přísluší hustota $f(\boldsymbol{x}, \boldsymbol{\theta})$, resp. pravděpodobnost $P(\boldsymbol{x}, \boldsymbol{\theta})$, kde $\boldsymbol{\theta} \in \Theta$. Náhodné veličiny $X_1, \ldots, X_n$ jsou tedy nezávislé stejně rozdělené a vektor $\boldsymbol{X}$ má hustotu $f_n(\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})$ resp. pravděpodobnost $P_n(\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} P(x_i, \boldsymbol{\theta})$, které pro jednoduchost souhrnně označíme $p(\boldsymbol{x}, \boldsymbol{\theta})$. Pro pevnou hodnotu $\boldsymbol{x}$ je $p(\boldsymbol{x}, \boldsymbol{\theta})$ funkcí $\boldsymbol{\theta}$ a nazývá se *věrohodnostní funkce*.

Hodnota $\hat{\boldsymbol{\theta}}$ parametru $\boldsymbol{\theta}$, která maximalizuje věrohodnostní funkci $p(\boldsymbol{x}, \boldsymbol{\theta})$ pro daný výběr $\boldsymbol{X} = \boldsymbol{x}$, se nazývá *maximálně věrohodný odhad* parametru $\boldsymbol{\theta}$. Názorně řečeno, je to taková hodnota parametru $\boldsymbol{\theta}$, pro kterou je získaný výsledek nejpravděpodobnější. Protože ln je prostá rostoucí funkce, hodnota $\hat{\boldsymbol{\theta}}$, která maximalizuje $p(\boldsymbol{x}, \boldsymbol{\theta})$ maximalizuje i funkci $\ln p(\boldsymbol{x}, \boldsymbol{\theta})$. Tuto funkci nazýváme *logaritmická věrohodnostní funkce*, označujeme $L(\boldsymbol{x}, \boldsymbol{\theta})$ stručněji $L(\boldsymbol{\theta})$ a s výhodou ji využíváme, neboť zjednodušuje výraz, který se má maximalizovat.

Pro hledání $\hat{\boldsymbol{\theta}}$ můžeme použít klasický postup diferenciálního počtu, který vede k řešení tzv. systému věrohodnostních rovnic $\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_j} = 0$, $j = 1, \ldots, m$. Je-li $\boldsymbol{\theta}$ mnohorozměrný parametr, pak se často stává, že analytické řešení úlohy je velmi náročné nebo ani není možné a musí se použít nějaká numerická metoda. Maximálně věrohodné odhady mají příznivé vlastnosti, a proto se používají velmi často, i když výpočet bývá složitý.

Věta 3. *Nechť je systém hustot $\{f(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ regulární a má Fisherovu matici $\boldsymbol{J}(\boldsymbol{\theta})$ a nechť jsou dále splněny následující předpoklady:*

1. *$\Theta \subseteq \mathbf{R}^m$ je parametrický prostor, který obsahuje takový neprázdný otevřený interval I, že skutečná hodnota parametru $\boldsymbol{\theta}_0$ patří do I.*

2. $\boldsymbol{X} = (X_1, \ldots, X_n)'$, *kde* X_i *jsou nezávislé stejně rozdělené náhodné veličiny s hustotou* $f(x, \boldsymbol{\theta})$ *vzhledem k nějaké* σ*-konečné míře* μ.

3. $M = \{x : f(x, \boldsymbol{\theta}) > 0\}$ *nezávisí na* $\boldsymbol{\theta}$.

4. Pro každé $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ *platí* $f(x, \boldsymbol{\theta}_1) = f(x, \boldsymbol{\theta}_2)$ μ*-skoro všude právě tehdy, když* $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.

5. Derivace $\frac{\partial^3 f(x,\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j \partial\theta_k}$ *existuje pro skoro všechna* x, *pro všechna* $\boldsymbol{\theta} \in I$ *a pro všechna* $i, j, k = 1, \ldots, m$.

6. Pro všechna $\boldsymbol{\theta} \in I$ *platí* $\int\limits_M f''_{ij}(x, \boldsymbol{\theta}) \mathrm{d}\mu(x) = 0, \ i, j = 1, \ldots, m.$

7. Pro všechna $i, j, k = 1, \ldots, m$ *existují funkce* $M_{ijk}(x) \geq 0$ *tak, že* $\mathsf{E}_{\boldsymbol{\theta}_0} M_{ijk}(X) < \infty$ *a* $\left|\frac{\partial^3 \ln f(x,\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j \partial\theta_k}\right| \leq M_{ijk}(x)$ *pro všechna* $\boldsymbol{\theta} \in I$ *a skoro všechna* $x \in M$.

Pak platí následující tvrzení.

(a) Pro každé $\varepsilon > 0$ *existuje takové řešení* $\hat{\boldsymbol{\theta}}_n$ *systému věrohodnostních rovnic, že* $\lim\limits_{n\to\infty} P(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| < \varepsilon) = 1$.

(b) Položme

$$\boldsymbol{U}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial L(\boldsymbol{\theta})}{\partial\theta_1} \\ \vdots \\ \frac{\partial L(\boldsymbol{\theta})}{\partial\theta_m} \end{pmatrix},$$

pak platí $\frac{1}{\sqrt{n}} \boldsymbol{U}(\boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathsf{d}} \mathsf{N}[\mathbf{0}, \boldsymbol{J}(\boldsymbol{\theta}_0)]$.

(c) Existuje-li pro každé dostatečně velké n *a pro každé* $\boldsymbol{X}$ *takový kořen* $\hat{\boldsymbol{\theta}}_n$ *systému věrohodnostních rovnic, že* $\hat{\boldsymbol{\theta}}_n$ *je konzistentním odhadem parametru* $\boldsymbol{\theta}_0$, *pak* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathsf{d}} \mathsf{N}(\mathbf{0}, [\boldsymbol{J}(\boldsymbol{\theta}_0)]^{-1})$.

Důkaz. Návod lze nalézt např. v [1, str. 159], podrobně je proveden např. v [5] v kapitole 6.4. $\square$

V praxi se $\boldsymbol{J}(\boldsymbol{\theta}_0)$ již zmiňovaným odhadem $-\frac{1}{n} \cdot D$, kde navíc dosazujeme odhad $\hat{\boldsymbol{\theta}}$.

3.4 Statistické testy

V praxi nás vždy zajímá, jestli zvolený model není zbytečně složitý, neboli jestli všechny parametry vychází statisticky významné, tj. testujeme hypotézu $H_0 : \boldsymbol{\theta} = \mathbf{0}$ proti alternativě $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$.

3.4.1 Waldův test

Za předpokladů uvedených ve Větě 3 a za předpokladu spojitosti matice $\boldsymbol{J}(\boldsymbol{\theta})$ v bodě $\boldsymbol{\theta}_0$ lze dokázat, že náhodná veličina

$$W(\boldsymbol{\theta}_0) = n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'\boldsymbol{J}(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

má asymptoticky χ^2 rozdělení o m stupních volnosti, kde m je rozměr parametru $\boldsymbol{\theta}$. Tvrzení zůstává v platnosti, i když se místo matice $\boldsymbol{J}(\boldsymbol{\theta}_0)$ užije nějaký její konzistentní odhad.

Tato veličina se používá jako testová statistika ve Waldově testu. Nulovou hypotézu zamítáme na hladině α, když $W(\boldsymbol{\theta}_0) \geq \chi^2_m(\alpha)$.

3.4.2 Test poměrem věrohodnosti

Test poměrem věrohodnosti, někdy též zvaný Wilksův test je založen na náhodné veličině

$$LR(\boldsymbol{\theta}_0) = 2[L(\hat{\boldsymbol{\theta}}_n) - L(\boldsymbol{\theta}_0)],$$

o které lze za předpokladů uvedených ve Větě 3 dokázat, že má asymptoticky χ^2 rozdělení o m stupních volnosti, kde m je rozměr parametru $\boldsymbol{\theta}$. Nulovou hypotézu zamítáme na hladině α, když $LR(\boldsymbol{\theta}_0) \geq \chi^2_m(\alpha)$. Lze ukázat, že test poměrem věrohodnosti je asymptoticky ekvivalentní Waldově testu, přesto se vlastnosti obou testů mohou na konečném souboru mírně lišit.

Kapitola 4

POISSONŮV PROCES

4.1 Základní vlastnosti Poissonova procesu

Definice 8. Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor a nechť $T \subseteq \mathbb{R}$. Rodina reálných náhodných veličin $\{X_t, t \in T\}$ definovaných na $(\Omega, \mathcal{A}, P)$ se nazývá *náhodný proces*.

V případě, že $T = \mathbb{Z}$ nebo $T = \mathbb{N}_0$ mluvíme o procesu s diskrétním časem, kterým se budeme zabývat.

Definice 9. Nechť $\{X_t, t \geq 0\}$ je proces celočíselných náhodných veličin, kde X_t značí počet událostí, které nastanou v časovém intervalu $(0, t]$. Předpokládejme, že pozorované události jsou téhož typu a nastávají náhodně v čase. Dále předpokládejme, že počty událostí v disjuktních časových intervalech jsou nezávislé a závisí jen na délce příslušného intervalu. Nechť pravděpodobnost, že událost nastane v intervalu $(t, t+h]$, je pro všechna t a nějaké $\lambda > 0$ rovna $\lambda h + o(h)$ a pravděpodobnost, že v $(t, t+h]$ nastane více než jedna událost, nechť je $o(h)$. Nechť navíc $X_0 = 0$, potom $\{X_t, t \geq 0\}$ nazýváme *Poissonův proces s intenzitou* λ.

Z předpokladů učiněných v definici Poissonova procesu mimo jiné plyne, že pro libovolné časové okamžiky $t_1 < t_2 < t_3 < t_4$ jsou počty událostí $X_{t_2} - X_{t_1}$ a $X_{t_4} - X_{t_3}$ v intervalech $(t_1, t_2]$ a $(t_3, t_4]$ nezávislé náhodné veličiny. Říkáme, že proces X_t je *proces s nezávislými přírůstky*. Odtud je možné odvodit různé vlastnosti viz. [6]. Zmiňme jedinou pro nás důležitou vlastnost, a to že přírůstky $X_{s+t} - X_s$ mají Poissonovo rozdělení s parametrem λt pro všechna $s, t > 0$.

V aplikované statistice se běžně setkáváme s řadou dat, která představuje pozorování výskytu nějaké události v časovém intervalu. Jako typické příklady můžeme uvést sledování výskytu nějakých onemocnění, nebo počty dopravních nehod. Takováto data jsou nutně přirozená čísla včetně nuly a bývá vhodné o nich předpokládat, že se řídí Poissonovým procesem. Podle povahy dat můžeme předpokládat i jiná rozdělení, další často používaná rozdělení jsou binomické a negativně binomické, více viz. [3, str. 339].

4.2 Specifikace modelu

V této práci sledujeme počty nově hlášených nezaměstnaných na úřadech práce, nazývejme je pro jednoduchost přírůstky nezaměstnaných. Předpokládáme, že tyto přírůstky jsou nezávislé a mají Poissonovo rozdělení závislé na čase. Data jsou získána s měsíční periodicitou tj. v diskrétním čase, a proto pro modelování přírůstků nezaměstnanosti používáme nehomogenní Poissonův proces s diskrétním časem a intenzitou λ, přesněji $\lambda_{t,s}$, kde t vyjadřuje závislost na čase (v letech) a s sezónnost (měsíce).

Parametr λ předpokládáme ve tvaru

$$\lambda_{t,s} = de^{\boldsymbol{\theta} \boldsymbol{M}'_{(12t-1+s)}} \qquad t = 1, \dots, n \quad s = 1, \dots, 12, \tag{4.1}$$

kde $\boldsymbol{M}'_{(12t-1+s)}$ značí $(12t-1+s)$-tý řádek matice $\boldsymbol{M}$ braný jako sloupcový vektor, d je nějaká známá konstanta (v praktické aplikaci, které je věnována pátá kapitola je d rovno průměrnému přírůstku nezaměstnaných za sledované období), $\boldsymbol{\theta} = (a, b, \boldsymbol{c})$ je neznámý parametr o třinácti složkách, kde a je absolutní člen, b parametr vyjadřující závislost na letech a $\boldsymbol{c} = c_1, \dots, c_{12}$ jsou paramery příslušející sezónám. Odhadovat budeme pouze $c_1, \dots, c_{11}$, dvanáctý parametr položíme roven nule a k příslušnému měsíci ostatní vztáhneme. $\boldsymbol{M}$ je matice tvaru:

$$\boldsymbol{M} = \begin{pmatrix} 1 & 1 & \\ \vdots & \vdots & \boldsymbol{A} \\ 1 & 1 & \\ 1 & 2 & \\ \vdots & \vdots & \boldsymbol{A} \\ 1 & 2 & \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & n & \\ \vdots & \vdots & \boldsymbol{A} \\ 1 & n & \end{pmatrix},$$

kde $\boldsymbol{A}$ je matice indikující sezóny a je tvaru

$$\boldsymbol{A}_{12\times 11} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Z důvodu uvedeného výše má sloupcovou hodnost pouze 11. Rozepsáno po složkách pro dané t a s dostáváme:

$$\lambda_{t,s} = de^{a+bt+c_s} \qquad t = 1, \dots, n \quad s = 1, \dots, 12. \tag{4.2}$$

Odhad parametru $\boldsymbol{\theta}$ získáme metodou maximální věrohodnosti.

4.3 Odvození věrohodnosti

Jak bylo uvedeno výše, mají jednotlivé přírůstky $X_{t,s}$ Poissonovo rozdělení s parametrem $\lambda_{t,s}$ a jejich distribuční funkce je tedy

$$F_{t,s}(k_{t,s}) = P(X_{t,s} = k_{t,s}) = \frac{e^{-\lambda_{t,s}}\lambda_{t,s}^{k_{t,s}}}{k_{t,s}!} \qquad t = 1, \dots, n \quad s = 1, \dots, 12.$$

O pozorováních předpokládáme, že jsou vzájemně nezávislá vzhledem k různým časům i sezónám, a proto pro sdruženou distribuční funkci platí (při použití zkráceného značení $\lambda_{t,s} = \lambda$)

$$G(\boldsymbol{k}) = \prod_{t=1}^{n}\prod_{s=1}^{12} F_{t,s}(k_{t,s}) = \prod_{t=1}^{n}\prod_{s=1}^{12} \frac{e^{-\lambda}\lambda^{k_{t,s}}}{k_{t,s}!} = p(t, s, \boldsymbol{\theta}),$$

zlogaritmováním dostáváme věrohodnostní funkci ve tvaru:

$$L(\boldsymbol{\theta}) = \ln p(t, s, \boldsymbol{\theta}) = \sum_{t=1}^{n}\sum_{s=1}^{12} \ln \frac{e^{-\lambda}\lambda^{k_{t,s}}}{k_{t,s}!} = \sum_{t=1}^{n}\sum_{s=1}^{12} \frac{e^{-de^{a+bt+c_s}} \cdot (de^{a+bt+c_s})^{k_{t,s}}}{k_{t,s}!} =$$

$$= \sum_{t=1}^{n}\sum_{s=1}^{12} (-de^{a+bt+c_s} + \ln(de^{a+bt+c_s})^{k_{t,s}} - \ln k_{t,s}!) =$$

$$= \sum_{t=1}^{n}\sum_{s=1}^{12} (-de^{a+bt+c_s} + k_{t,s}(a + bt + c_s)) + \sum_{t=1}^{n}\sum_{s=1}^{12} \ln \frac{d^{k_{t,s}}}{k_{t,s}!}.$$

Druhá ze sum nezávisí na $\boldsymbol{\theta}$, a proto se nadále budeme zabývat úlohou:

$$\max_{a,b,\boldsymbol{c}}\Big\{\sum_{t=1}^{n}\sum_{s=1}^{12}(-de^{a+bt+c_s}+k_{t,s}(a+bt+c_s))\Big\}. \tag{4.3}$$

Což lze v řeči výpočetního prostředí softwaru MatLab přepsat maticově takto:

$$\max_{\boldsymbol{\theta}}\Big\{-d\cdot\sum\exp\{\boldsymbol{\theta}\boldsymbol{M}'\}+\boldsymbol{\theta}\boldsymbol{M}'\boldsymbol{k}\Big\}, \tag{4.4}$$

kde $\exp\{\boldsymbol{x}\}$ (pro $\boldsymbol{x} = x_1,\dots,x_n$) chápeme jako vektor $(e^{x_1},\dots,e^{x_n})$ a $\sum\boldsymbol{x} = \sum_{i=1}^{n} x_i$. Data $k_{t,s}$ jsme si pro maticový zápis seřadili do vektoru v pořadí, v jakém byla naměřena. Dále byla úloha řešena numericky v MatLabu (verze 6.5), neboť přímé řešení nelze explicitně vyjádřit.

Kapitola 5

NUMERICKÉ VÝSLEDKY

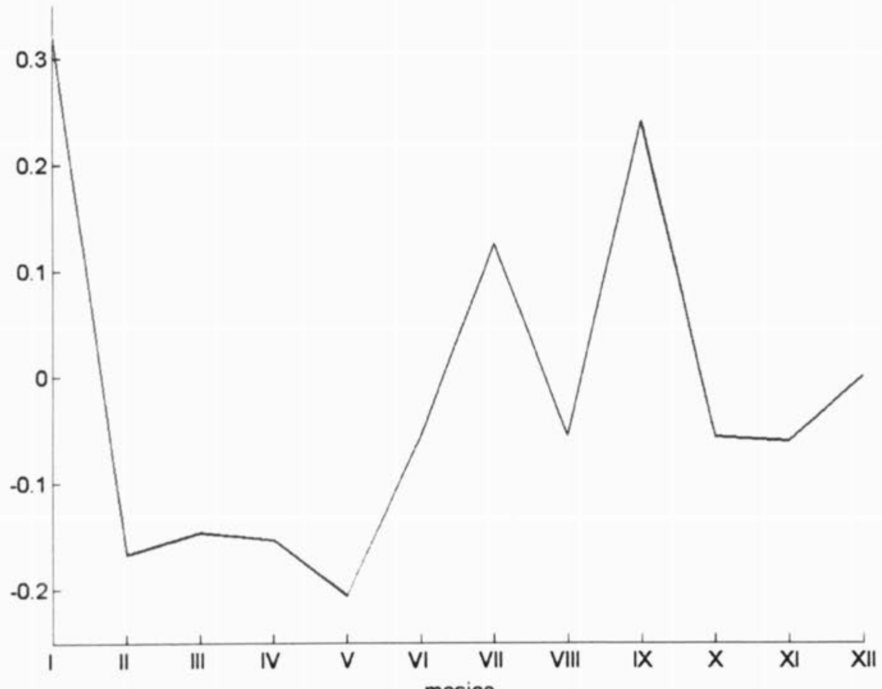
Popsaný model jsme aplikovali na data týkající se vývoje nezaměstnanosti v celé České republice za období leden 1999 až březen 2006, ale k odhadnutí modelu jsme použili pouze data do konce roku 2005, ostatní jsme si nechali pro porovnání s předpovědí získanou na základě modelu. Dostáváme model, který se řídí Poissonovým procesem s intenzitou (4.1) pro $n = 7$. Odhad parametru $\boldsymbol{\theta}$ jsme získali metodou maximální věrohodnosti z (4.4), k nalezení maxima věrohodnostní funkce byla použita funkce fminunc programu MatLab. Fminunc je standardní funkce Matlabu, která hledá globální minimum funkce více proměnných bez jakýchkoli omezení. Tuto funkci jsme mohli použít, neboť platí $\max\{f\} = -\min\{-f\}$. Funkce fminunc má několik volitelných parmetrů a díky nim si můžeme nechat vypsat takové informace jako, která numerická metoda byla použita, zda bylo minima dosaženo nebo k němu mezivýsledky pouze konvergují a nebo také pro nás užitečný gradient a Hessián funkce v bodě minima.

Všechny složky parametru $\boldsymbol{\theta}$ vychází statisticky významné, Waldova statistika je $1,38 \cdot 10^5$, odpovídající hodnota χ^2 statistiky o 13-ti stupních volnosti je 22,36, takže zamítáme hypotézu o nulovosti parametru $\boldsymbol{\theta}$. Test poměrem věrohodnosti dává stejný výsledek. V tabulce 5.1 je uveden 95%-ní interval spolehlivosti pro parametr $\boldsymbol{\theta}$.

Dle očekávání je situace nejhorší v lednu, červenci a září. Potvrzuje se, že podniky nejčastěji propouštějí na konci nebo v polovině roku a také, že nezaměstnanost výrazně ovlivňují absolventi škol. Ve všech ostatních měsících je situace lepší než v prosinci, který byl zvolen jako vztažný měsíc viz. obrázek 5.1.

$\boldsymbol{\theta}$	odhad	interval	spolehlivosti
a	$0,0507$	$0,0465$	$0,0549$
b	$-0,0115$	$-0,0120$	$-0,0111$
c_1	$0,3216$	$0,3171$	$0,3261$
c_2	$-0,1670$	$-0,1720$	$-0,1621$
c_3	$-0,1468$	$-0,1517$	$-0,1419$
c_4	$-0,1535$	$-0,1583$	$-0,1486$
c_5	$-0,2063$	$-0,2112$	$-0,2015$
c_6	$-0,0565$	$-0,0613$	$-0,0518$
c_7	$0,1249$	$0,1203$	$0,1295$
c_8	$-0,0558$	$-0,0604$	$-0,0512$
c_9	$0,2405$	$0,2359$	$-0,2451$
c_{10}	$-0,0567$	$-0,0618$	$-0,0517$
c_{11}	$-0,0609$	$-0,0656$	$-0,0562$

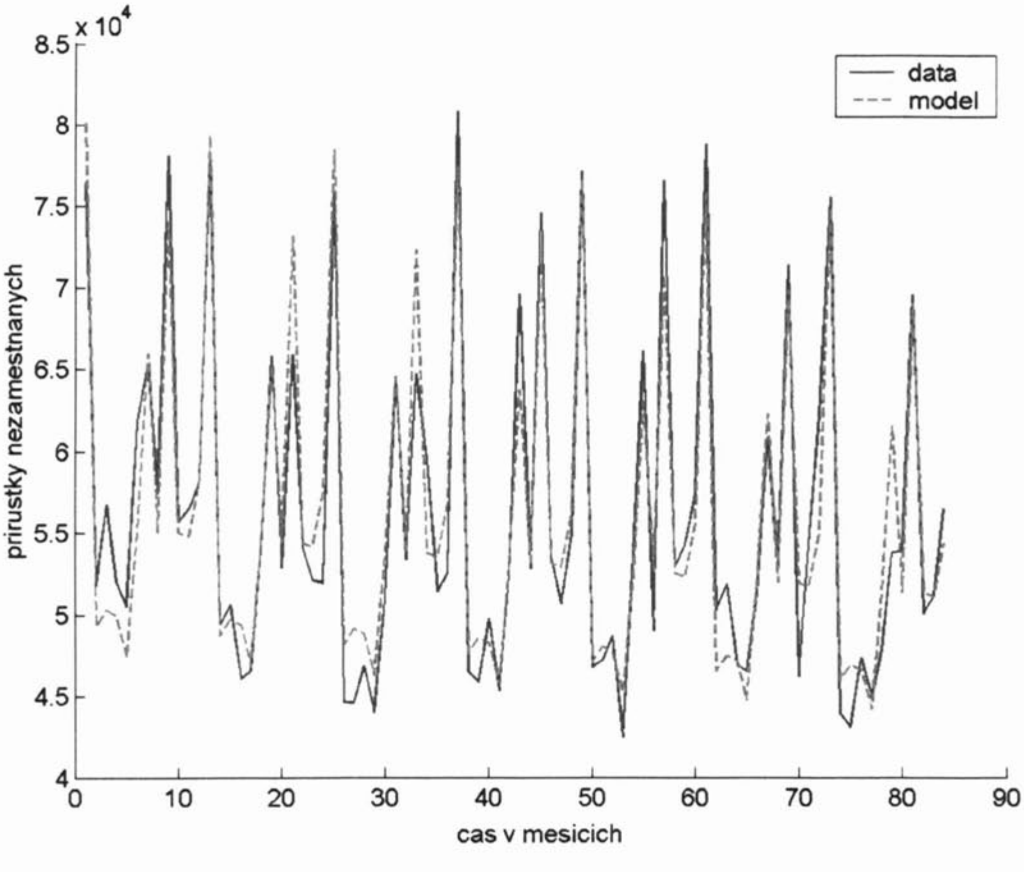
Tabulka 5.1: 95%-ní interval spolehlivosti pro $\boldsymbol{\theta}$.



Obrázek 5.1: Odhad pro parametry sezón.

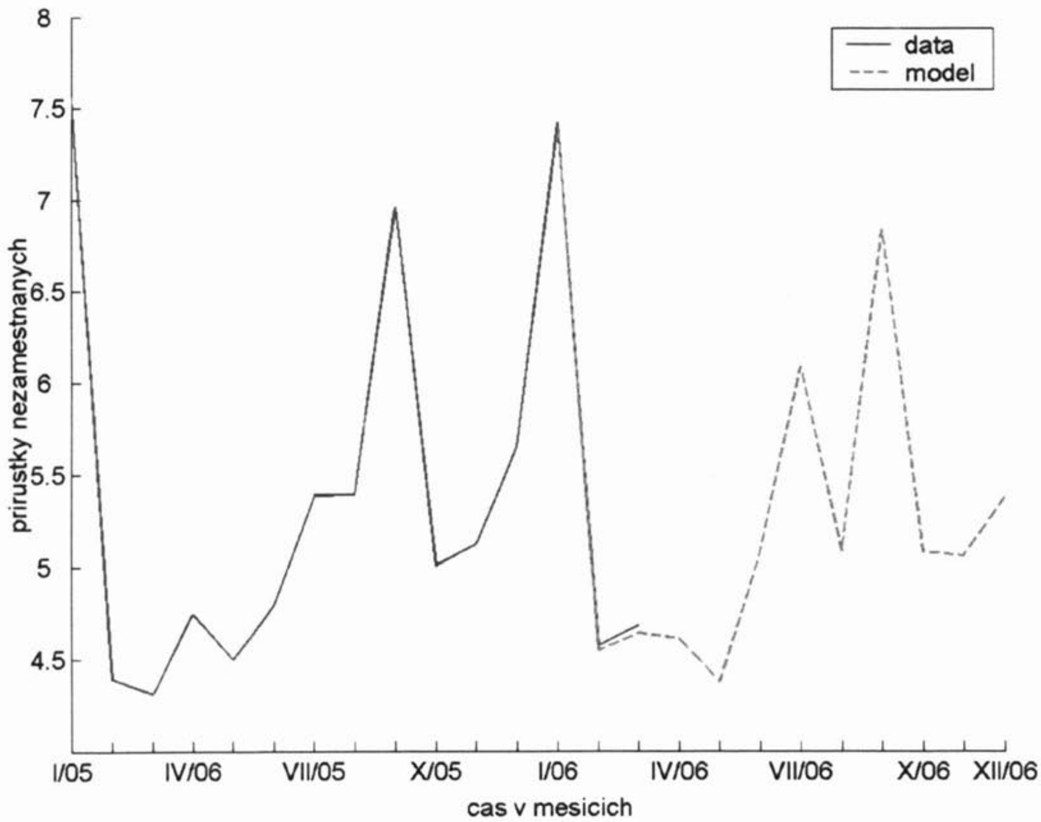
Dále na obrázku 5.2 uvádíme grafické srovnání použitých dat a modelových hodnot, které jsou získány jako střední hodnota při dosazení odhadu $\boldsymbol{\theta}$. Známou vlastností Poissonova rozdělení je, že střední hodnota je rovna parametru intenzity a dostáváme tak:

$$model = de^{\hat{a}+\hat{b}t+\hat{c}_s} \qquad t = 1, \ldots, 7 \quad s = 1, \ldots, 12.$$



Obrázek 5.2: Srovnání modelu a skutečných dat.

Nezaměstnanost se snažíme nejen modelovat, ale také předpovídat. Na obrázku 5.3 uvádíme skutečně pozorované počty nezaměstnaných za rok 2005 a první tři měsíce roku 2006 ve srovnání s hodnotami, které dává model jako předpověď pro rok 2006. Jak můžeme vidět, je shoda modelu a dat za první tři měsíce letošního roku velmi dobrá.



Obrázek 5.3: Predikce.

Model nabízí několik modifikací. Jednou z možností je přidat nekonstantní závislost na letech. Vzrůstá nám tak počet parametrů na osmnáct. Jedenáct parametrů pro sezóny zůstává, ale místo jediného parametru pro roky jich máme šest. První rok volíme jako vztažný a příslušný parametr je tedy roven nule. Dostáváme tak model pro dané t a s ve tvaru

$$\lambda_{t,s} = de^{a+b_t+c_s} \qquad \boldsymbol{\theta} = (a, \boldsymbol{b}, \boldsymbol{c}), \ s = 1, \ldots, 12, \ t = 1, \ldots, 7. \tag{5.1}$$

Maticový zápis může zůstat ve stejném tvaru, pouze se změní tvar matice $\boldsymbol{M}$, a to:

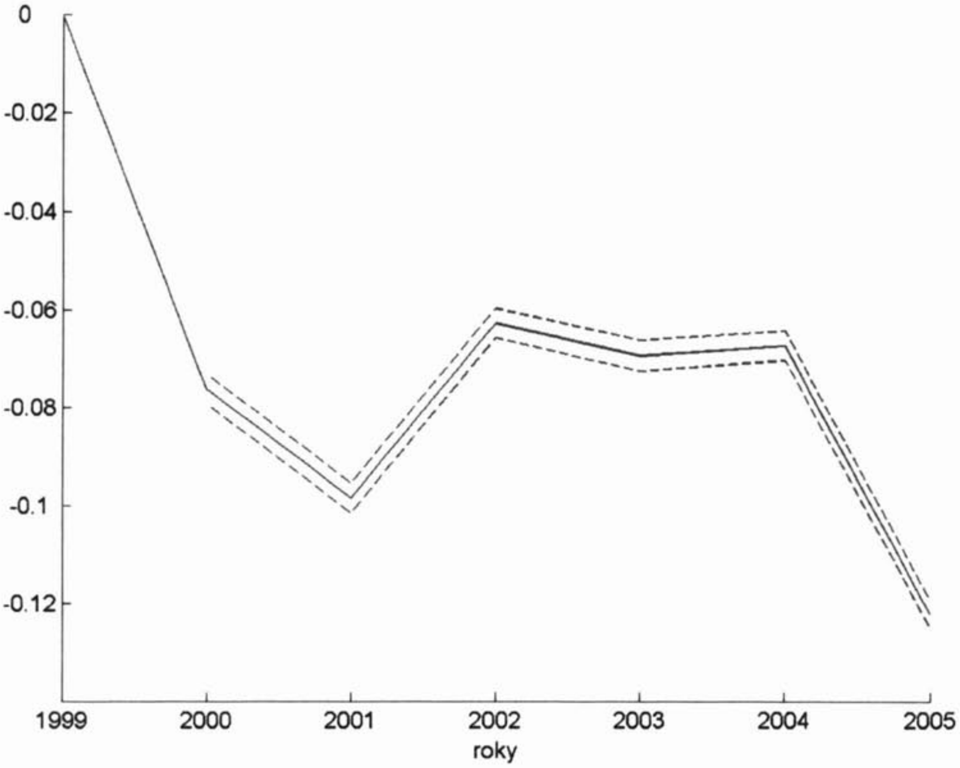
$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \\ \vdots & \vdots & \vdots & \ddots & & \boldsymbol{A} \\ 1 & 0 & 0 & \cdots & 0 & \\ 1 & 1 & 0 & \cdots & 0 & \\ \vdots & \vdots & \vdots & \ddots & & \boldsymbol{A} \\ 1 & 1 & 0 & \cdots & 0 & \\ 1 & 0 & 1 & \cdots & 0 & \\ \vdots & \vdots & \vdots & \ddots & & \boldsymbol{A} \\ 1 & 0 & 1 & \cdots & 0 & \\ \vdots & \vdots & & & \vdots & \\ \vdots & \vdots & & & \vdots & \\ 1 & 0 & 0 & \cdots & 1 & \\ \vdots & \vdots & \vdots & \ddots & & \boldsymbol{A} \\ 1 & 0 & 0 & \cdots & 1 & \end{pmatrix},$$

matice $\boldsymbol{A}$ zůstává stejná.

Odhad parametru $\boldsymbol{\theta}$ byl opět získán metodou maximální věrohodnosti a k nalezení maxima nám pomohla funkce fminunc softwaru MatLab. I nyní vychází všechny složky parametru $\boldsymbol{\theta}$ statisticky významné. Hodnota Waldovy statistiky je $1,48 \cdot 10^5$ a příslušná kritická hodnota rozdělení χ^2 o 18-ti stupních volnosti je rovna $28,87$, a tedy zamítáme hypotézu o nulovosti parametru $\boldsymbol{\theta}$. V tabulce 5.2 je uveden 95%-ní interval spolehlivosti pro parametr $\boldsymbol{\theta}$. Z tabulky vidíme, že parametry příslušné sezónám mají stejné hodnoty jako v prvním modelu, jenom interval spolehlivosti vychází o něco užší. Na obrázku 5.4 následuje grafické znázornění 95%-ního intervalu spolehlivosti pro parametry příslušné rokům. Můžeme vidět, že nezaměstnanost má (s výjimkou přelomu let 2001 a 2002) nekonstantní klesající trend.

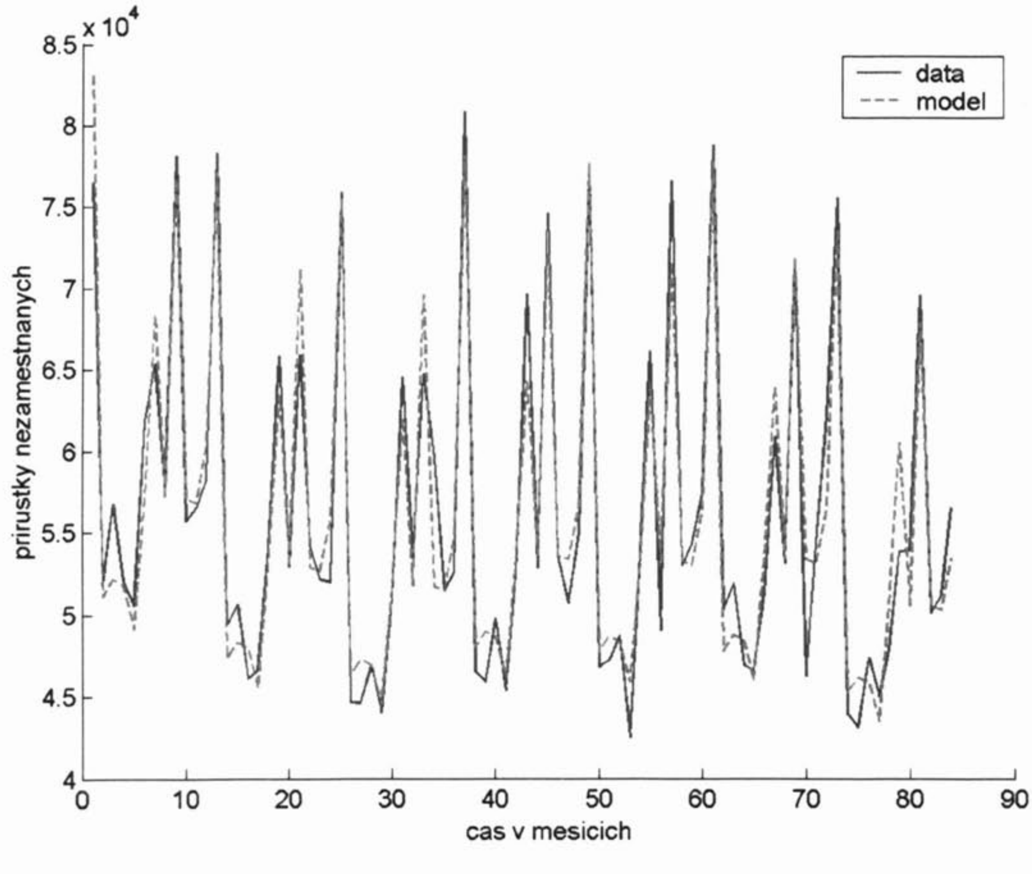
$\boldsymbol{\theta}$	odhad	interval	spolehlivosti
a	$0,0750$	$0,0725$	$0,0775$
b_1	$-0,0763$	$-0,0793$	$-0,0733$
b_2	$-0,0983$	$-0,1014$	$-0,0953$
b_3	$-0,0626$	$-0,0656$	$-0,0596$
b_4	$-0,0691$	$-0,0722$	$-0,0660$
b_5	$-0,0671$	$-0,0701$	$-0,0640$
b_6	$-0,1222$	$-0,1250$	$-0,1193$
c_1	$0,3216$	$0,3177$	$0,3256$
c_2	$-0,1670$	$-0,1712$	$-0,1628$
c_3	$-0,1468$	$-0,1509$	$-0,1427$
c_4	$-0,1535$	$-0,1579$	$-0,1490$
c_5	$-0,2063$	$-0,2109$	$-0,2017$
c_6	$-0,0565$	$-0,0608$	$-0,0523$
c_7	$0,1249$	$0,1210$	$0,1288$
c_8	$-0,0558$	$-0,0595$	$-0,0521$
c_9	$0,2405$	$0,2366$	$0,2444$
c_{10}	$-0,0567$	$-0,0607$	$-0,0528$
c_{11}	$-0,0609$	$-0,0652$	$-0,0566$

Tabulka 5.2: 95%-ní interval spolehlivosti pro $\boldsymbol{\theta}$ v modifikovaném modelu.

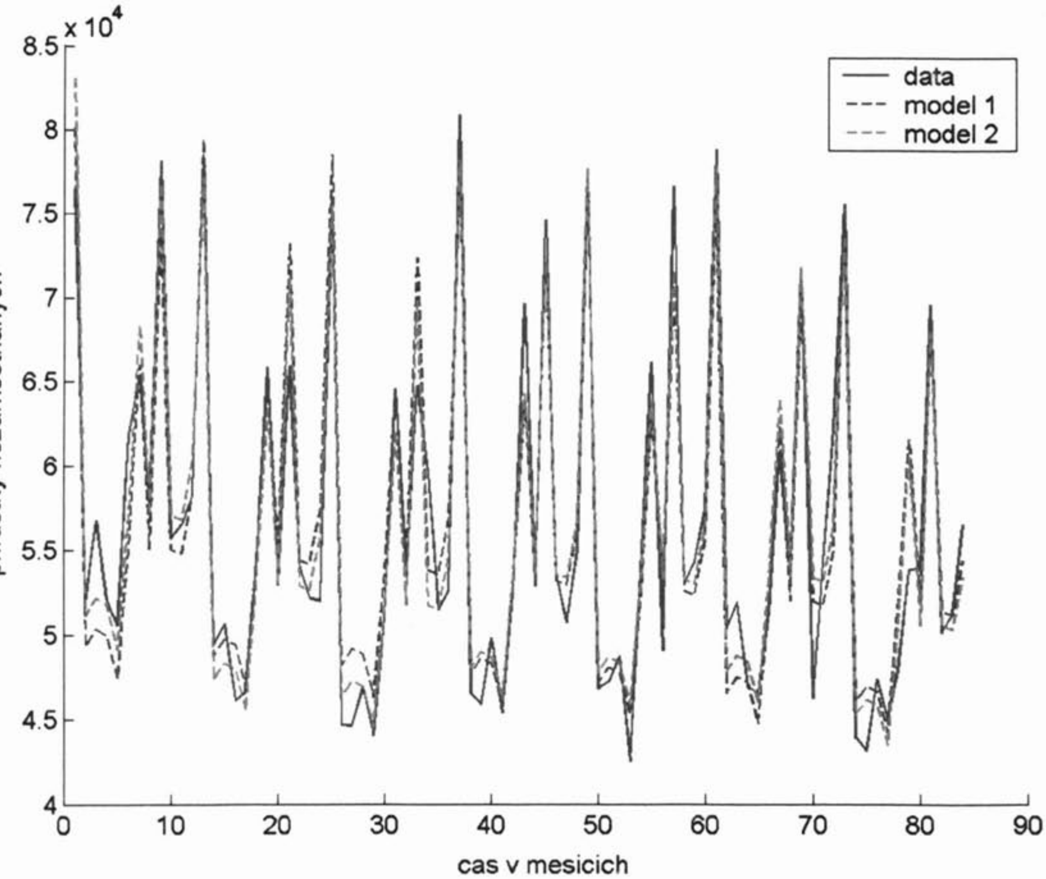


Obrázek 5.4: 95%-ní interval spolehlivosti pro parametry příslušné rokům.

Na obrázku 5.5 uvádíme srovnání modifikovaného modelu a dat, na obrázku 5.6 následuje srovnání výsledků obou modelů. *Model 1* viz. 4.2, *model 2* viz. 5.1, kde modelové hodnoty byly získány stejně jako v případě původního modelu jako střední hodnota při dosazení odhadu $\boldsymbol{\theta}$.



Obrázek 5.5: Srovnání modifikovaného modelu a skutečných dat.



Obrázek 5.6: Srovnání původního a modifikovaného modelu.

Kapitola 6

ZÁVĚR

V této práci jsme se zabývali modelováním nezaměstnanosti pomocí modelů založených na Poissonově procesu. Byly představeny a numericky vyřešeny dva příbuzné modely: původní model viz. (4.2) a modifikovaý model viz. (5.1). Oba dávají podobné výsledky, ale druhý model je přesnější. Přidání parametrů pro roky umožňuje lépe zachytit dlouhodobý vývoj nezaměstnanosti, případné náhlé změny způsobené různými vládními opatřeními a nebo také vliv hospodářské-ho cyklu. Nevýhodou tohoto modelu je jeho vyšší výpočetní náročnost a hlavně to, že ztrácíme možnost předpovídat nezaměstnanost pro následující rok jako tomu bylo učiněno u prvního modelu. Možnost predikce zůstává pouze pro část roku, když už máme odhad příslušného parametru b_i.

Pokusili jsme se také přidat do modelu územní závislost. Výpočty byly provedeny jak pro jednotlivé kraje tak pro různě seskupené. Žádný z těchto pokusů nedával dobré výsledky, pravděpodobně proto, že model tohoto typu je příliš jednoduchý, a proto jsme od dalších pokusů upustili.

Modelování nezaměstnanosti je velmi důležité a na toto téma už bylo zpracováno mnoho sudií. V této práci se nám povedlo ukázat, že i jednoduchý model, kde zohledňujeme pouze vliv času a sezónnost, může dávat poměrně dobré výsledky.

Literatura

[1] Anděl, Jiří *Základy matematické statistiky (Preprint)*; Praha 2002

[2] Buchtová, Božena a kol. *Nezaměstnanost psychologický, ekonomický a sociální problém*; Grada Publishing a.s., Praha 2002

[3] Cipra, Tomáš *Pojistná matematika Teorie a praxe*; EKOPRESS, Praha 1999

[4] Giddens, Anthony *Sociologie*; Argo, Praha 1999

[5] Lehman, E. L.; Casella, George *Theory of Point Estimation* (second edition); Springer-Verlag, New York 1998

[6] Prášková, Zuzana; Lachout, Petr *Základy náhodných procesů*; Karolinum, Praha 1998

Přílohy

původní model

Procedura, kterou byly provedeny výpočty v modelu (4.2). Obsahuje testy významnosti parametrů, výpočet intervalového odhadu, vykresluje uvedené grafy.

```
%modelovani nezamestnanosti pomoci nehomogeniho Poissonova procesu%
%Po(d*exp(a+bt+cA)) metoda maximalni verohodnosti, A sezonost%
%CR%
close all;
clear all; load('cr.txt'); data=cr; d=mean(data);
n=length(data);
k=12; %mesice - pocet radku A%

%%vykresli verohodnost
%x=ones(1,13);
%vysl=[];
%for i=(-6:6)
 %   vysl=[vysl -myfun((i.*x)./100)];
 %end
%plot(vysl)


x0=0.01*ones(1,13);
[xopt,hodnotaf,exitflag,output,grad,hessian]=fminunc(@myfun,x0)


%testy vyznamnosti parametru
'0,95 kvantil chi rozdeleni=0,05 kriticka hodnota 13 stupnu
volnosti' kritChi = chi2inv(0.95,13)
```

```
'Walduv test'
W = ( xopt * (hessian) * xopt' )      % -f => zmizi (-1)*
if W > kritChi disp('zamitame hypotezu o nulovosti parametru')
else disp('nezamitame hypotezu o nulovosti parametru') end

'test pomerem verohodnosti' Lopt=-hodnotaf; L0=-myfun([0 0 0 0 0 0
0 0 0 0 0 0 0]); LR=2*(Lopt - L0) if LR > kritChi disp('zamitame
hypotezu o nulovosti parametru') else disp('nezamitame hypotezu o
nulovosti parametru') end


'intervalovy odhad' io = [];
pom = hessian^(-1);                  % -f => zmizi (-1)*
pomm = diag(pom); for j = 1 : 13
  odhad = [0 0];
  odhad(1) = xopt(j) - (norminv(0.975)*(pomm(j)^(1/2)));
  odhad(2) = xopt(j) + (norminv(0.975)*(pomm(j)^(1/2)));
  io = [io;odhad];
end io


'p hodnoty' pomm.^(1/2); S = xopt'./ (pomm.^(1/2)); pvalues =
2*(ones(13,1)-normcdf(abs(S),0,1))
%p > alfa nezamitam
%p < alfa zamitam

%vytvori M
A=eye(k); %jednotkova
A=A(:,1:k-1); %oriznuty posledni sloupec - vztazeno k prosinci
y=ones(n,1); %pomocna prom. pro M%
M=[]; for i=(1:fix(n/12))
        Mpom=i*ones(12,1);
        M=[M;Mpom]; end
M=[y M];
Apom=A; %pomocna pro tvorbu M%
for i=(1:fix(n/k)-1) Apom=[Apom;A]; end %fix=zaokrouhleni dolu%
Apom=[Apom;Apom(1:rem(n,k),:)]; %rem=zbytek po deleni%
M=[M Apom];


%vykresli fit model data
```

```
model=d*exp(xopt*M');

figure hold on plot(data) plot(model,'r:')

legend('data','model',1) hold off

%predikce pro t=8
load('crp.txt'); crp=[data(73:84);crp];
P=[ones(12,1) 8*ones(12,1) A];      %12x13
modelp=d*exp(xopt*P');

figure hold on plot(crp) plot(modelp,'r:')
legend('data','model',1) hold off
%chyba=zeros(3,1);
%for i=1:3
%    chyba(i)=modelp(i)-crp(i);
%end
%chyba

%vykresli int odhad pro parametry sezon
io1=io(3:13,1); io2=io(3:13,2); figure hold on odhad=xopt(3:13);
odhad=[odhad 0]; plot(odhad) plot(io1,':') plot(io2,':')
```

funkce pro původní model

Zápis (4.4) pro MatLab.

```
function f=myfun(x) %x radkovy vektor 1x13%
%myfun CR, model d*exp(a+bt+cA)%
k=12; %mesice - pocet radku A%
load('cr.txt'); data=cr; n=length(data); d=mean(data);

A=eye(k); %jednotkova
A=A(:,1:k-1); %oriznuty posledni sloupec
y=ones(n,1); %pomocna prom. pro M%
M=[]; for i=(1:fix(n/12))
        Mpom=i*ones(12,1);
        M=[M;Mpom]; end
M=[y M];
Apom=A; %pomocna pro tvorbu M%
for i=(1:fix(n/k)-1) Apom=[Apom;A]; end %fix=zaokrouhleni dolu%
Apom=[Apom;Apom(1:rem(n,k),:)]; %rem=zbytek po deleni%
M=[M Apom];

f=-d*sum(exp(x*M'))+x*M'*data; f=-f;
```

modifikovaný model

Procedura, kterou byly provedeny výpočty v modelu (5.1). Obsahuje testy významnosti parametrů, výpočet intervalového odhadu, vykresluje uvedené grafy.

```
%modelovani nezamestnanosti pomoci nehomogeniho Poissonova procesu%
%Po(d*exp(a+b_t+c_s)) metoda maximalni verohodnosti, A sezonost%
%vztahuje se k:sezony - prosinec, roky - prvni rok
%CR%
close all;
clear all; load('cr.txt'); data=cr; d=mean(data);
n=length(data);
k=12; %mesice - pocet radku A%

%%vykresli verohodnost
%x=ones(1,18);
%vysl=[];
%for i=(-6:6)
%    vysl=[vysl -myfun2((i.*x)./100)];
%end
%plot(vysl)

x0=0.01*ones(1,18);
[xopt,hodnotaf,exitflag,output,grad,hessian]=fminunc(@myfun2,x0)


%testy vyznamnosti parametru
'0,95 kvantil hodnota chi rozdeleni = 0,05 kriticka hodnota, 18
stupnu volnosti' kritChi = chi2inv(0.95,18)

'Walduv test'
W = ( xopt * (hessian) * xopt' )      % -f => zmizi (-1)*
if W > kritChi disp('zamitame hypotezu o nulovosti parametru')
else disp('nezamitame hypotezu o nulovosti parametru') end

'test pomerem verohodnosti' Lopt=-hodnotaf; L0=-myfun([0 0 0 0 0
0 0 0 0 0 0 0]); LR=2*(Lopt - L0) if LR > kritChi disp('zamitame
hypotezu o nulovosti parametru') else disp('nezamitame hypotezu o
nulovosti parametru') end
```

```
'intervalovy odhad' io = [];
pom = hessian^(-1);                % -f => zmizi (-1)*
pomm = diag(pom); for j = 1 : 18
  odhad = [0 0];
  odhad(1) = xopt(j) - (norminv(0.975)*(pomm(j)^(1/2)));
  odhad(2) = xopt(j) + (norminv(0.975)*(pomm(j)^(1/2)));
  io = [io;odhad];
end io


'p hodnoty' pomm.^(1/2); S = xopt'./ (pomm.^(1/2)); pvalues =
2*(ones(18,1)-normcdf(abs(S),0,1))
%p > alfa nezamitam
%p < alfa zamitam

%vytvori M
A=eye(k); %jednotkova
A=A(:,1:k-1); %oriznuty pposledni sloupec - vztazeno k prosinci
y=ones(n,1); %pomocna prom. pro M%
M=zeros(12,6); for i=(1:6)
        pom=zeros(1,6);
        pom(i)=1;
        Mpom=[];
        for j=1:k
            Mpom=[Mpom;pom];
        end
        M=[M;Mpom];
end M=[y M];
Apom=A; %pomocna pro tvorbu M%
for i=(1:fix(n/k)-1) Apom=[Apom;A]; end %fix=zaokrouhleni dolu%
Apom=[Apom;Apom(1:rem(n,k),:)]; %rem=zbytek po deleni%
M=[M Apom]; %vztahuje se k prvnimu roku


%vykresli fit model dat
model=d*exp(xopt*M');


figure hold on plot(data)
%legend('data','model 1','model 2',1)
plot(model,'r:') legend('data','model',1) hold off
```

```
%vykresli int odhad param pro roky
figure hold on roky1=io(2:7,1); roky2=io(2:7,2); plot([0
xopt(2:7)]) plot(roky1,':') plot(roky2,':') hold off

%vykresli int odhad pro parametry sezon

io1=io(8:18,1); io2=io(8:18,2); figure hold on plot([xopt(8:18)
0]) plot(io1,':') plot(io2,':')
```

funkce pro modifikovaný model

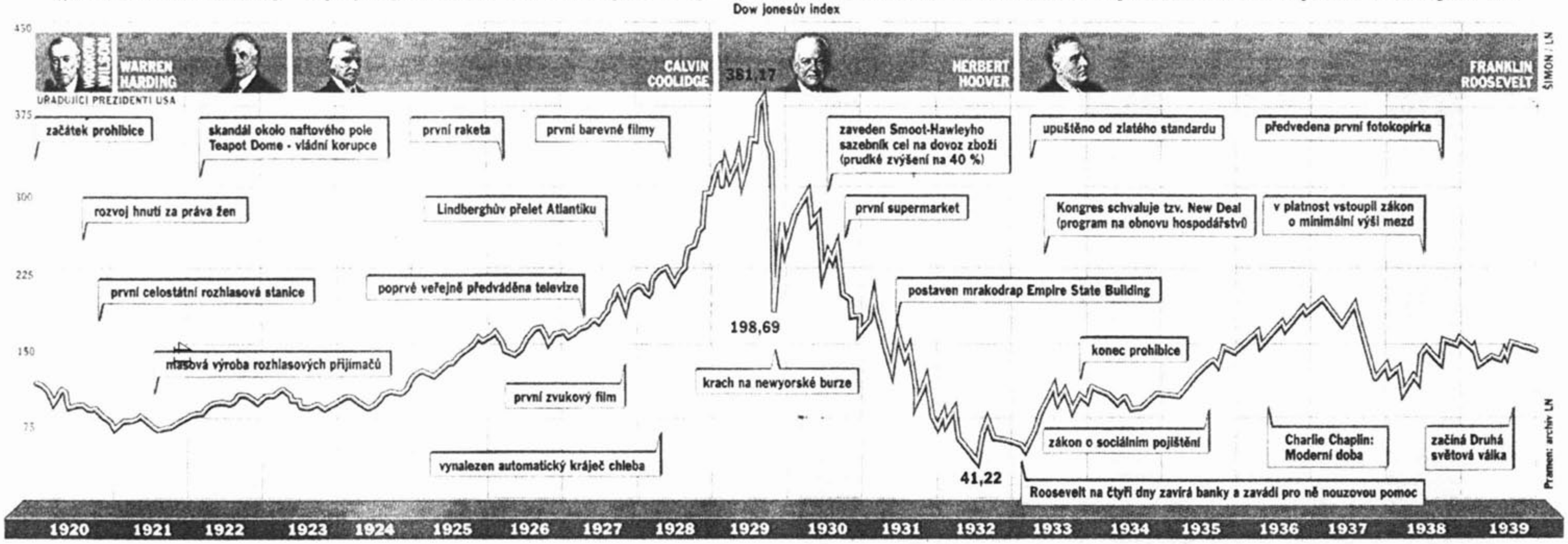
Zápis (4.4) s $\boldsymbol{M}$ příslušnou modelu (5.1).

```
function f=myfun(x) %x radkovy vektor 1x18%
%myfun pro CR, model 2 d*exp(a+b_t+c_s)%


k=12; %mesice - pocet radku A%
load('cr.txt'); data=cr; n=length(data); d=mean(data);

A=eye(k); %jednotkova
A=A(:,1:k-1); %oriznuty posledni sloupec
y=ones(n,1); %pomocna prom. pro M%
M=zeros(12,6); for i=(1:6)
        pom=zeros(1,6);
        pom(i)=1;
        Mpom=[];
        for j=1:k
            Mpom=[Mpom;pom];
        end
        M=[M;Mpom];
end M=[y M];
Apom=A; %pomocna pro tvorbu M%
for i=(1:fix(n/k)-1) Apom=[Apom;A]; end %fix=zaokrouhleni dolu%
Apom=[Apom;Apom(1:rem(n,k),:)]; %rem=zbytek po deleni%
M=[M Apom];

f=-d*sum(exp(x*M'))+x*M'*data; f=-f;
```

Národ směřuje k trvalé a vysoké prosperitě, tvrdili američtí ekonomové pár dní před krachem na newyorské burze

Před sedmdesáti lety začala největší hospodářská krize

„Historické milníky" Spojených států Amerických a vývoj obchodování na Wall Street v průběhu dvacátých a třicátých let



Přehnaná víra v růst cen akcií, nákupy na dluh a příliš vysoké úvěry bank. To byly důvody, které se podílely na hospodářské krizi.

PRAHA - „O budoucnost své země nemám obavy," prohlásil při svém jmenování do funkce v roce 1929 americký prezident Herbert Hoover. Ještě v témž roce, 29. října, však začala krachem na newyorské akciové burze nejhorší ekonomická krize v dějinách Spojených států amerických. „Velká deprese" trvala čtyři roky - až do roku 1933, kdy prezident Franklin Delano Roosevelt při své inauguraci vyhlásil Nový úděl (New Deal), tedy jakýsi plán na obnovu hospodářství. S důsledky ekonomické krize bojovaly Spojené státy až do druhé světové války.

Ekonomická deprese je v USA považována za jednu z temných stránek americké historie, která se nesmí nikdy opakovat. V poslední době se však stále hlasitěji ozývají hlasy, že současná prosperita ekonomicky nejsilnější země světa by mohla v budoucnu vyústit do podobné krize.

Hrozí USA znovu ekonomický kolaps?

Existuje totiž podle nich stále více znaků, kvůli nimž se současný stav americké ekonomiky začíná stále více podobat situaci před krachem na Wall Street v roce 1929. Podle škarohlidů hrozí největší nebezpečí zejména z pádu cen akcií. Ty v současné době dosahují v USA velmi vysokých hodnot a někteří odborníci, včetně šéfa americké centrální banky Alana Greenspana, varují, že může jít pouze o nafouklou bublinu, které hrozí splasknutí. To se ostatně stalo i v říjnu 1929, kdy Dow Jonesův index klesl během několika dnů o několik desítek bodů na pouhý zlomek své původní hodnoty. Lidé tehdy přestali akciím věřit a začali je hromadně prodávat. Akcie se pak staly bezcennými a tisíce lidí se takřka přes noc stalo chudáky. S akciemi úzce souvisí i nálada většiny investorů, která se v současnosti v Americe blíží optimismu z konce dvacátých let. Pět dnů před krachem na Wall Street dokonce ekonom z Yalské univerzity Irving Fischer prohlásil: „Národ směřuje k trvalé a vysoké prosperitě." Tato prohlášení jsou velmi podobná výrokům některých amerických investorů, kteří věří, že akcie budou i nadále růst strmým tempem. Problémem takovéhoto optimismu je skutečnost, že v případě prvních obtíží bude dosavadní víra v ekonomickou prosperitu nahlodána a lidé začnou akcie prodávat. To pak vyústí v začarovaný kruh, kdy prodej způsobuje pokles cen a ten pak další prodej.

Jiným problémem by se mohla stát stále rostoucí předluženost Američanů. Jejich nákupní horečka je totiž tak vysoká, že jí nestačí tempo růstu mezd. Na své utrácení si proto musejí půjčovat peníze od bank. Těm ale hrozí, že v případě problémů by své úvěry nemusely dostat zpátky, což by vedlo k vlně bankrotů.

Marek Bičík

Co způsobilo Velkou depresi

PRAHA - Během let 1929 až 1933 vzrostla nezaměstnanost z 1,6 na 12,8 milionu lidí, osobní příjmy Američanů klesly na polovinu. O padesát procent se díky nadúrodě snížily také ceny zemědělských výrobků, což znamenalo katastrofu pro farmáře. V troskách se ocitl americký akciový trh a zaniklo devět tisíc bank, zavíraly se továrny, doly, tisícovky podnikatelů a farmářů zbankrotovaly.

Velké depresi předcházelo ve Spojených státech období nevídané prosperity. Prudce vzrostla průmyslová výroba a trh byl zaplaven zbožím. Rozvoj rozhlasu, televize a automobilového průmyslu znamenaly převrat v dosud relativně střídmém životním stylu a lidé začali horečnatě utrácet a přestali si své peníze ukládat do bank.

Začaly se však prohlubovat příjmy střední třídy a bohatých podnikatelů. Více než čtyřicet procent celkových příjmů totiž šlo do kapes 0,1 procenta nejbohatších Američanů. Ti zároveň kontrolovali 34 procent všech úspor. Osmdesát procent obyvatel USA naopak nemělo žádné úspory. Poptávka již přestávala stačit stále rostoucí nabídce zboží. Z tohoto důvodu začalo stále více lidí kupovat na úvěr.

bič