

Univerzita Pardubice

Fakulta ekonomicko-správní

Lee-Carterův model mortality

Bc. Jana Šimková

**Diplomová práce
2017**

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Jana Šimková**
Osobní číslo: **E14633**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Pojistné inženýrství: Management finančních rizik**
Název tématu: **Lee-Carterův model mortality**
Zadávací katedra: **Ústav matematiky a kvantitativních metod**

Z á s a d y p r o v y p r a c o v á n í :

Cílem práce je popsat model mortality, který je stochastický, a který umožňuje výpočet trendu úmrtnosti, její predikci a určení očekávané délky života. Parametry pro výpočet odhadneme z dostupných dat platná pro Českou republiku.

Osnova:

- Vývoj úmrtnosti žen a mužů v České republice.
- Výpočet parametrů, trend úmrtnosti a její predikce pomocí Lee-Carterova modelu.
- Určení očekávané délky života.
- Interpretace a vizualizace získaných výsledků a jejich vzájemné porovnání pro muže a ženy.

Rozsah grafických prací: —
Rozsah pracovní zprávy: **cca 50 stran**
Forma zpracování diplomové práce: **tištěná/elektronická**
Seznam odborné literatury:

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org.

CHAN, N. H. (2002). Time series. Applications to Finance, Wiley series in probability and statistics, New York. ISBN 0-471-41117-5.

JINDROVÁ, P., SLAVÍČEK, O. (2012). Life expectancy development and prediction for selected European countries, 6-th International Scientific Conference Managing and Modelling of Financial Risk proceedings, p. 303-312, VŠB-TU Ostrava. ISBN 978-80-248-2835-0.

LEE, R. D., CARTER, L. (1992). Modelling and forecasting the time series of U.S. mortality, Journal of the American Statistical Association, 87, p. 659-671.

RENSHAW, A. E., HABERMAN, S. (2006). A cohort-based extension to Lee-Carter model for mortality reduction factors, Insurance: Mathematics and Economics, 38, p. 556-570.

Vedoucí diplomové práce:


RNDr. Ján Gogola, Ph.D.

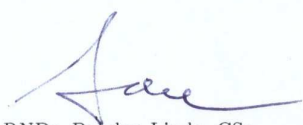
Ústav matematiky a kvantitativních metod

Datum zadání diplomové práce: **4. září 2016**

Termín odevzdání diplomové práce: **28. dubna 2017**


doc. Ing. Romana Provazníková, Ph.D.
děkanka

L.S.


doc. RNDr. Bohdan Linda, CSc.
vedoucí ústavu

V Pardubicích dne 4. září 2016

PROHLÁŠENÍ

Prohlašuji, že jsem tuto práci vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 28. 4. 2017

Jana Šimková

PODĚKOVÁNÍ:

Tímto bych ráda poděkovala svému vedoucímu práce RNDr. Jánů Gogolovi, Ph.D. za jeho odbornou pomoc, cenné rady a poskytnuté materiály, které mi pomohly při zpracování diplomové práce. Dále bych chtěla poděkovat rodině a blízkých přátelům za podporu během celého studia.

ANOTACE

Diplomová práce se zabývá Lee-Carterovým modelem úmrtnosti. První část se zabývá rozdělením a stručnou charakteristikou modelů používaných k modelování úmrtnosti. Dále podrobným popisem Lee-Carterova modelu včetně charakteristiky metody využívané v rámci modelu Lee-Carter k predikci budoucích hodnot. Poslední část je věnována analýze vývoje úmrtnosti na datech pro Českou republiku, zvláště pro muže a ženy, pomocí tohoto modelu a jejich vzájemné porovnání, včetně predikce budoucích hodnot.

KLÍČOVÁ SLOVA

Lee-Carter model, úmrtnost, predikce, ARIMA model, modelování úmrtnosti

TITLE

Lee Carter model mortality

ANNOTATION

The thesis focuses on the Lee-Carter mortality model. The first part deals with the classification and brief characteristics of the models used to model the mortality. A further detailed description of the Lee-Carter model, including the characteristics of the methods used in the model of Lee-Carter to predict future values. The last section is applied to analyze the development of mortality data for the Czech Republic, separately for men and women, using this model and their mutual comparison, including predictions of future values.

KEYWORDS

Lee Carter model, mortality, prediction, ARIMA model, mortality modeling

OBSAH

ÚVOD	9
1 MODEL Y POUŽÍVANÉ K PŘEDPOVÍDÁNÍ MORTALITY	10
1.1 DRUHY MODELŮ	10
1.1.1 Extrapoláční modely.....	11
1.1.2 Specifické extrapolace	11
1.1.3 Projekce pomocí relačních modelů.....	11
1.1.4 Extrapolace s odborným stanoviskem.....	11
1.1.5 Model příčina zpoždění	12
1.1.6 Epidemiologický model.....	12
1.2 DETERMINISTICKÉ A STOCHASTICKÉ MODEL Y ÚMRTNOSTI.....	12
1.2.1 Deterministické modely	12
1.2.2 Stochastické modely úmrtnosti	13
2 LEE-CARTER MODEL ÚMRTNOSTI.....	16
2.1 MODEL	16
2.1.1 Specifická míra úmrtnosti.....	17
2.1.2 Odhady parametrů modelu	18
2.2 ALGORITMUS VÝPOČTU	21
2.3 PREDIKCE BUDOUCÍCH HODNOT.....	22
2.3.1 Boxova-Jenkinsonova metodologie.....	22
2.3.2 Konstrukce modelů Boxovy-Jenkinsonovy metodologie	25
2.3.3 ARIMA Model.....	28
3 APLIKACE LEE-CARTER MODELU NA REÁLNÁ DATA.....	30
3.1 DATA	30
3.2 ODHADY PARAMETRŮ MODELU	31
3.2.1 Parametr α_x	31
3.2.2 Parametr k_t	33
3.2.3 Parametr β_x	34
3.3 PROGNOZOVÁNÍ INDEXU ÚMRTNOSTI k_t	36
3.3.1 Ženy	36
3.3.2 Muži	41
3.3.3 Výsledky a předpovědi parametru k_t	46
3.4 PROJEKCE VÝVOJE ÚMRTNOSTI.....	48
3.4.1 Predikce měř úmrtnosti	48
3.4.2 Predikce střední délky života	48
ZÁVĚR.....	56
POUŽITÁ LITERATURA	57
SEZNAM PŘÍLOH	59

SEZNAM TABULEK

Tabulka 1: Hodnoty odhadnutého parametru α_x pro muže a ženy pro věk 40 – 90	32
Tabulka 2: Hodnoty parametru k_t pro muže a ženy v letech 1965 – 2005	34
Tabulka 3: Hodnoty odhadnutého parametru β_x pro muže a ženy pro věk 1965 - 2005	35
Tabulka 4: Hodnoty informačních kritérií ARIMA modelů pro ČR parametru k_t pro ženy ...	39
Tabulka 5: Hodnoty informačních kritérií ARIMA modelů pro ČR parametru k_t pro muže ..	44
Tabulka 6: Porovnání hodnot konstanty modelu náhodné procházky pro úmrtnostní index k_t podle pohlaví	46
Tabulka 7: Hodnoty předpovězeného parametru k_t pro obě pohlaví pro roky 2006 - 2015 ...	47
Tabulka 8: Odhadnuté a skutečné hodnoty naděje dožití pro určité věky v letech 2006 - 2010	53
Tabulka 9: Odhadnuté a skutečné hodnoty naděje dožití pro určité věky v letech 2011 - 2015	54
Tabulka 10: Odhadnuté hodnoty míry úmrtnosti $m_{x,t}$ pro ženy v období 2006 - 2015	60
Tabulka 11: Odhadnuté hodnoty míry úmrtnosti $m_{x,t}$ pro muže v období 2006 - 2015	61
Tabulka 12: Reálné hodnoty naděje dožití žen v období 1965 – 2015	62
Tabulka 13: Reálné hodnoty naděje dožití mužů v období 1965 - 2015	63
Tabulka 14: Hodnoty standardních chyb se_{kt} pro ženy	64
Tabulka 15: Hodnoty standardních chyb se_{kt} pro muže	64

SEZNAM ILUSTRACÍ

Obrázek 1: Příklady ACF a PACF nestacionární a stacionární řady	25
Obrázek 2: Vývoj střední délky života při narození pro muže a ženy v letech 1965-2005	30
Obrázek 3: Vývoj parametru α_x pro věk 40 - 90 pro ženy a muže	32
Obrázek 4: Vývoj parametru k_t pro muže a ženy v letech 1965 - 2005	33
Obrázek 5: Vývoj parametru β_x pro muže a ženy pro věk 40 - 90	35
Obrázek 6: Vývoj parametru k_t pro ženy v letech 1965 – 2005	36
Obrázek 7: ACF a PACF časové řady parametru k_t pro ženy	37
Obrázek 8: Vývoj diferencovaného parametru k_t pro ženy v letech 1965 – 2005	38
Obrázek 9: ACF a PACF diferencované časové řady k_t pro ženy	38
Obrázek 10: Výsledek $ARIMA(0, 1, 0)$ modelu pro časovou řadu parametru k_t pro ženy	40
Obrázek 11: ACF a PACF reziduí modelu náhodné procházky ČR parametru k_t pro ženy ...	40
Obrázek 12: Histogram rezidui - test normality modelu náhodné procházky ČR parametru k_t pro ženy	41
Obrázek 13: Vývoj parametru k_t pro muže v letech 1965 – 2005	42
Obrázek 14: ACF a PACF časové řady parametru k_t pro muže	42
Obrázek 15: Vývoj diferencovaného parametru k_t pro muže v letech 1965 – 2005	43
Obrázek 16: ACF a PACF diferencované časové řady k_t pro muže	43
Obrázek 17: Výsledek $ARIMA(0, 1, 0)$ modelu pro časovou řadu parametru k_t pro muže	44
Obrázek 18: ACF a PACF reziduí modelu náhodné procházky ČR parametru k_t pro muže ..	45
Obrázek 19: Histogram rezidui - test normality modelu náhodné procházky ČR parametru k_t pro muže	45
Obrázek 20: Prognóza indexu úmrtnosti k_t pro obě pohlaví, 1965 – 2015	47
Obrázek 21: Vývoj střední délky života ženské populace v letech 1965 - 2015	50
Obrázek 22: Vývoj střední délky života ženské populace v letech 1965 - 2015	50
Obrázek 23: Vývoj střední délky života v letech 1965 - 2015 mužské populace	51
Obrázek 24: Vývoj střední délky života v letech 1965 - 2015 mužské populace	51

SEZNAM ZKRATEK A ZNAČEK

ACF	Autokorelační funkce
AIC	Akaike information criterion
AR	Autoregresní proces
ARIMA	Integrovaný smíšený proces
ARMA	Smíšený proces
BIC	Bayes information criterion
ČR	Česká republika
ČŘ	Časová řada
CI	Confidence interval (interval spolehlivosti)
HMD	Human mortality database
MA	Proces klouzavých součtů
OLS	Ordinary least square (metoda nejmenších čtverců)
PACF	Parciální autokorelační funkce
SARIMA	Sezónní integrovaný smíšený proces
SIC	Schwarz information criterion
SVD	Singular value decomposition (singulární rozklad)

ÚVOD

Problematika modelování úmrtnosti patří mezi jedno z aktuálních témat dnešní doby. Jedním z důvodů je stárnutí populace, ke kterému v posledních letech dochází. Tímto vývojem by se dané společnosti měli zabývat, a to právě proto, že ovlivňují nejen ekonomiku, ale také další odvětví, např. školství nebo zdravotnictví. Právě problém stárnutí populace s sebou přináší řadu závazků souvisejících se zdravotní a sociální péčí s důchodovým zabezpečením. Proto, aby společnost mohla tuto problematiku řešit, je vhodné prostřednictvím dostupných modelů sledovat vývoj úmrtnosti a také ho predikovat do budoucna. Tato diplomová práce se zabývá právě jedním z nejpoužívanějších stochastických modelů úmrtnosti, kterým je Lee-Carter model úmrtnosti.

První kapitola teoretické části se zabývá jednotlivými přístupy k modelování úmrtnosti. Obsahuje rozdělení modelů, stručnou charakteristiku a větší zaměření na modely deterministické a stochastické. Následující kapitola je zaměřena na Lee-Carterův model úmrtnosti, kdy je ze začátku model charakterizován obecně a následně popsány metody, pomocí kterých lze odhadnout parametry modelu. Závěr kapitoly je věnován charakteristice Boxovy-Jenkinsonovy metodologie, pomocí které je predikován jediný časově závislý parametr.

Praktická část je zaměřena na vývoj úmrtnosti České republiky zvláště pro muže a ženy věkové kategorie 40 – 90 let v období 1965 – 2005. Rok 2005 jsme vybrali proto, abychom v závěru poslední kapitoly mohli provést predikce pro dané míry úmrtnosti pomocí Lee-Carterova modelu na 10 let dopředu a následně vypočítat střední délku života, kterou jsme mohli porovnat s hodnotami reálnými. Veškerá data pocházejí z *Human Mortality Database* a *Českého statistického úřadu*, výpočty jednotlivých parametrů jsou provedeny v *Microsoft Excel* a v softwaru *Eviews* jsme provedli analýzu indexu úmrtnosti a jeho následnou predikci, díky které jsme poté mohli předpovědět vývoj úmrtnosti pro Českou republiku zvláště pro muže a ženy do roku 2015. V závěru kapitoly prostřednictvím predikovaných měr úmrtnosti jsme vypočítali a předpověděli střední délku života pro vybrané věky.

Cílem této práce je popsat stochastický model úmrtnosti, pomocí kterého analyzujeme a predikujeme vývoj úmrtnosti v ČR ve věku 40-ti až 90-ti let v období 1965 – 2005 a vzájemné porovnání pro muže a ženy, včetně predikcí.

1 MODELY POUŽÍVANÉ K PŘEDPOVÍDÁNÍ MORTALITY

V poslední době se věnuje velká pozornost budoucí úrovni úmrtnosti. Existuje mnoho různých názorů týkající se délky života v budoucnosti. Někteří se domnívají, že délka života bude růst stejně rychle jako v posledních pěti či šesti letech. Jiní zastupují názor, že dojde ke zpomalení růstu délky života a pro určité skupiny může dokonce klesnout. Existují prognózy ohledně poklesu budoucí úmrtnosti, která je závislá na poklesu úmrtnosti z minulosti. Příkladem je USA, kde naděje na dožití v průběhu minulého století vzrostla v důsledku poklesu úmrtnosti lidí v mladším věku. Akademik Jay Olshansky tvrdí, že jakékoliv zvýšení dlouhověkosti musí nastat poklesem úmrtnosti u jiných věkových skupin nebo na základě jiných příčin. [12]

Dalšími spornými body v předvídání úmrtnosti jsou:

- případná existence biologického limitu k životu,
- přístup lidí k tomuto limitu.

James Vaupel navrhuje, že předpovídání přežití je možné rozšířit o různé genetické změny a non-genetické interakce. Existují také diskuze o prognózování úmrtnosti, zda brát v úvahu pokroky v medicíně. Většina prognostiků se shoduje, že medicína má vliv na snižování úmrtnosti, (zvláště ve vztahu k určitým příčinám smrti). Někteří prognostici rovněž zahrnují vývoj technologií, které mohou značně ovlivnit stárnutí a zvýšit délku života. Otázkou zůstává, jak medicínský pokrok zahrnout, aby došlo k přiměřenému odhadu. [12]

1.1 Druhy modelů

S přihlédnutím k významným diskuzím o budoucí úmrtnosti, existuje řada širokých přístupů k předpovídání úmrtnosti. Tyto přístupy mohou být kategorizovány různými způsoby. Jednou z možností je následující klasifikace:

- extrapolační modely,
- kauzální modely zahrnující ekonometrické vztahy,
- modely založené na základních biomedicínských procesech.

V následující části jsme shrnuli některé z nejčastěji používaných metod. Pořadí je podle rostoucí subjektivity znalce při prognózování. [12]

1.1.1 Extrapolační modely

Tyto modely předpokládají, že trend pozorovaný v minulých datech bude přetrvávat i do budoucna. Jejich cílem je popsat úmrtnost matematickou funkcí, jejímž grafem je hladká křivka. Při takovém prognózování je třeba si dát pozor, protože nemusí brát v úvahu různé biologické faktory, pokroky v medicíně nebo změny životního stylu. Tyto změny nebývají významné v kratším časovém horizontu, proto extrapolační modely jsou vhodnější na krátkodobé předpovědi. Přesto se používají i na dlouhodobé předpovědi (např.: Lee-Carter). V kapitole 1.2 se blíže věnujeme rozdělení těchto modelů. [13]

1.1.2 Specifické extrapolace

Specifické modely rozčleňují celkovou úmrtnost podle příčin smrti a předpovídají pravděpodobnost úmrtí pro každou příčinu zvlášť. Prognózy pro každou příčinu se zapojují do extrapolace v minulých trendech. Tato metoda může být obtížnější než prognózování souhrnné úmrtnosti vzhledem k dostupnosti údajů podle příčin, které jsou méně spolehlivé než data celková. Výhodou tohoto postupu je, že může poskytnout lepší pochopení faktorů celkových změn v úmrtnosti. Přesto přesnost výsledných prognóz úmrtnosti není vždy nejlepší. Nevyřešeným problémem zůstává korelace mezi příčinami smrti, které jsou obecně nevyčíslitelné. [12]

1.1.3 Projekce pomocí relačních modelů

Jedním ze způsobů projekce pomocí relačních modelů může být odkaz na „pokročilou“ populaci. V tomto přístupu se předpokládá, že budoucí míra úmrtnosti je následována dynamikou pozorované úmrtnosti pro samostatné „unikátní“ populace. Jinými slovy, úmrtnostní profil prognózované populace je považován za úmrtnostní profil, kterého bude dosaženo v nějakém časovém horizontu. Tento přístup je vhodnější pro rozvojové země než pro rozvinuté země. [12]

1.1.4 Extrapolace s odborným stanoviskem

Tento přístup je založený na extrapolačním modelu, který přímo zahrnuje silné předpoklady prognostika pro vývoj budoucí úmrtnosti. Modelovat pomocí matematického modelu, aniž by došlo k posouzení a pochopení kontextu, je nebezpečné. Znalecký posudek může být stejně jednoduchý jako předpoklad o nejvyšší úrovni úmrtnosti nebo, že zlepšení v úmrtnosti následuje specifickou cestu v čase. Přístup můžeme použít pro souhrnnou úmrtnost i pro konkrétní příčinu. [12]

1.1.5 Model příčina zpoždění

Předpoklad tohoto přístupu vychází ze zlepšení úmrtnosti pro konkrétní příčinu, na kterou má vliv zpoždění úmrtí pro danou příčinu ve vyšším věku. K poklesu úmrtnosti z důvodu některé nemoci dochází v důsledku lékařského pokroku nebo změny životního stylu, které mohou smrt oddálit. Tyto účinky opožděné úmrtnosti podle příčin se mohou odhadnout za podmínky, že se příčina přesunula na mladší věkové kategorie. To bude vysvětleno na příkladu onemocnění srdce. Pokud dojde ke zpoždění úmrtnosti o pět let, přesune se pozorovaná úmrtnost do věkové kategorie o pět let mladší. Tedy pro věk 65 by se konkrétní úmrtnost použila jako odhad pro 60-ti leté. Filozofie tohoto přístupu je podobná projekcím s odkazem na více „pokročilé“ populace. Rozdílem je, že se zde úmrtnost aplikuje na stejný počet obyvatel, nikoliv na úmrtnost odlišné populace. [12]

1.1.6 Epidemiologický model

Model analyzuje vztah mezi specifickými rizikovými faktory a jejich vlivem na úmrtnost. Do rizikových faktorů se zahrnuje kouření, obezita, sociálně-ekonomické postavení, specifické nemoci. Metoda je podobná přístupu příčina zpoždění, ale zde je kladen důraz na dopad specifických rizikových faktorů nikoliv příčinám smrti. Nejdříve se odhadne dopad na míru úmrtnosti každého rizikového faktoru, který se potom může promítnout do budoucnosti s rizikovými faktory. Podmínkou je, že se rozdělení může měnit v čase. Přestože byl v poslední době zaznamenán vývoj v modelování demografických a epidemiologických procesů úmrtnosti, pro přesnější odhad vztahu mezi rizikovými faktory a úmrtností je zapotřebí zvolit správný model. [12]

1.2 Deterministické a stochastické modely úmrtnosti

V následující podkapitole se budeme věnovat v praxi nejrozšířenějším modelům, a to jsou extrapoláčnı modely, které byly popsány výše. Podle přístupu k pozorovaným hodnotám můžeme modely rozdělit na deterministické a stochastické. [13]

1.2.1 Deterministické modely

Tyto modely předpokládají, že hodnoty pozorované v minulosti (v množině roků T) mohou být proložené nějakým trendem, nejčastěji pomocí exponenciální funkce, a tento trend bude pokračovat i v budoucnosti. Abychom dostali požadované trendy, potom stačí extrapolovat tento trend do budoucnosti.

Důležitá je především volba množiny T . V případech, kdy máme k dispozici pozorování daleko do minulosti, nemusí být vhodné zahrnutí všech dat, ale omezit se na kratší časový úsek. Efekt pozorování dávno v minulosti by měl mít na předpověď jen malý vliv, avšak při jejich zahrnutí do množiny T se může stát, že budoucí hodnoty ovlivní výrazně. [13]

- ***Exponenciální formule***

Velmi často se používá jednoduchý přístup, který předpokládá budoucí vývoj logaritmu pravděpodobnosti úmrtí q_x ve věku x jako lineární. To znamená, že pro $h = 1, 2, \dots, n - 1$ existuje hodnota δ_x pro kterou platí:

$$\ln q_x(t_{h+1}) - \ln q_x(t_h) \approx -\delta_x(t_{h+1} - t_h), \quad (1)$$

nebo

$$\frac{q_x(t_{h+1})}{q_x(t_h)} \approx \exp\{-\delta_x(t_{h+1} - t_h)\}, \quad (2)$$

kde pro každý věk x je parametr δ_x možné odhadnout například pomocí metody nejmenších čtverců.

Tento přístup zastupuje tzv. *horizontální přístup*, který spočívá v analýze úmrtnostního profilu v čase, vztahující se k věku x . [13]

- ***Zákony úmrtnosti***

Problematika zákonů úmrtnosti je motivována snahou modelovat lidskou úmrtnost pomocí matematických vzorců. Jednotlivé zákony jsou potom charakterizovány typickým tvarem funkce popisující intenzity úmrtnosti nebo pravděpodobnosti úmrtí. Mezi nejznámější úmrtnostní zákony patří Moivrův zákon, Gompertzův zákon nebo Makehamův zákon. [3]

Tento způsob představuje tzv. *vertikální přístup* k předpovědi úmrtnosti, neboť parametry jsou odhadované pro každý rok zvlášť. [13]

1.2.2 Stochastické modely úmrtnosti

Projekce popsané výše, nebraly v úvahu stochastickou povahu úmrtnosti, byly založeny na deterministických scénářích a směsi analýz smrti podle příčin a znaleckého posudku. Deterministické metody v posledních desetiletích systematicky podhodnocovaly úmrtnost. Z tohoto důvodu se motivuje potřeba použít stochastické modely úmrtnosti. [12]

Při stochastickém modelování úmrtnosti považujeme pozorované míry úmrtnosti za realizace náhodných veličin reprezentující minulou úmrtnost. Stochastický přístup nám

umožňuje určit kromě bodových odhadů úmrtnosti i odhady intervalové, na druhou stranu však vyžaduje splnění určitých statistických předpokladů (rozdělení počtu úmrtí). Některé modely modelují logaritmus věkově specifické míry úmrtnosti $m_{x,t}$, jiné logitovou transformací pravděpodobnosti úmrtí osoby ve věku x , označovanou $\text{logit}(q_x(t))$. Některé modely představují lepší spolehlivost pro určitý typ dat. Ale volba modelu by neměla být podřízena pouze samotnou vhodností dat, ale pozornost by se měla zaměřit i na účely prognózy a dynamiky úmrtnosti. Mezi další faktory k uvažování patří robustnost, kvalitativní kritéria (např.: transparentnost, šetrnost). [12], [13]

Tyto modely je možné rozdělit ještě do tří skupin, a to *age-period modely*, které nevyužívají složku kohortního efektu. V opačném případě, při zahrnutí kohortního efektu se jedná o takzvané *age-period-cohort modely*. Poslední skupinou jsou *p-spliny*, které jsou funkce definované po částech polynomy. V následující části bude stručně představeno několik nejčastěji používaných modelů, ve kterých se používá následující zápis:

$\beta_x^{(i)}$ funkce vyjadřující efekty stárnutí obyvatelstva, kde x označuje věk,

$\kappa_t^{(i)}$ funkce vyjadřující období související s efekty, kde t označuje období nebo čas,

$\gamma_c^{(i)}$ funkce kohorty v souvislosti s efekty, kde $c = t - x$ označuje kohortu narození. [12]

- **Lee-Carter model**

V roce 1992 Ronald Lee a Lawrence Carter navrhli model pro předpovídání budoucí úmrtnosti, kde parametry modelu byly kalibrovány na úmrtnost z historické zkušenosti.

Jde o nejjednodušší stochastický prognostický model z řady modelů, které zahrnují různé stupně složitosti vnímání věku, období a kohorty v datech. Model můžeme zapsat následovně:

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)}. \quad (3)$$

Nevýhodou složitých modelů je vyžadování pečlivé kalibrace a větší počet parametrů ztěžuje intuitivní pocit z toho, jak model reaguje na změny v parametrech. Jelikož se tato práce zabývá tímto modelem, budeme se mu v následující kapitole věnovat podrobněji.

- **Renwhaw-Haberman model**

Tento model je zobecněním verze Lee-Carterova modelu, který zahrnuje kohortní efekt. Vyjádřit jej můžeme ve formě:

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}. \quad (4)$$

K rovnici Lee-Carterova modelu se přidává dynamický účinek kohorty v závislosti na roku narození. Vlivem této kohorty se efekt může měnit až podle věku $\beta_x^{(3)}$.

- ***Currie Age-Period-Cohort model***

Currie navrhuje zjednodušenou verzi Renshaw-Haberman modelu. Věk, období a kohortní efekt ovlivňují úmrtnost samostatně, nezávisle. Model se formuluje:

$$\ln m_{x,t} = \beta_x^{(1)} + \kappa_t^{(2)} + \gamma_{t-x}^{(3)}. \quad (5)$$

- ***Cairns, Blake and Dowd model***

Modeluje se přímo na původní úmrtnosti, nikoliv na centrální míře úmrtnosti. Nejjednodušší dvoufaktorový model má tvar:

$$\text{logit } q_x(t) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}), \quad (6)$$

kde \bar{x} je středním věkem pro věky použité v analýze.

Model existuje v celé řadě rozšířených verzí s různými dostatečnými podmínkami k pokrytí skupinového efektu s lineární, kvadratickou nebo časovou závislostí variability. Tento model se liší od předchozích stochastických modelů předpokladem funkčního vztahu úmrtnosti a věku. [12]

2 LEE-CARTER MODEL ÚMRTNOSTI

Jak bylo zmíněno výše, model navrhli a popsali v roce 1992 Ronald Lee a Lawrence Carter v článku „*Modelling and forecasting US mortality*“. Jedná se o metodu, která je extrapoláčnı́ a nemá žádnou snahu začlenit poznatky zdravotnı́ch, behaviorálnı́ch nebo sociálnı́ch vlivů na změny úmrtnosti. Většinou se používá na dlouhodobé předpovědi agregovaných veličin úmrtnosti, např.: očekávané délky života. Dnes patří Lee-Carter model k nejčastěji používaným modelům.

Jeho předností je, že kombinuje bohatý, přesto šetrný demografický model s metodami statistických časových řad, což je pevně založeno na trvalých dlouhodobých historických trendech. Poskytuje pravděpodobnostní oblasti spolehlivosti pro své předpovědi. Mnoho metod předpokládá hornı́ hranici časového intervalu lidského života, nebo odůvodněnı́ zpomalenı́ ve výsledcích délky života nějakým jiným způsobem. Tento model však umožňuje, aby úmrtnost podle věku klesala exponenciálně bez nějakých omezenı́. Naopak model není určený na spolehlivou projekci úmrtnostnı́ch měř pro konkrétnı́ věky. Pro konkrétnı́ věky, nebo pokud nás zajímají změny v kratším časovém úseku, je potom vhodné použít jiný model. [11]

2.1 Model

Navržený model charakterizuje specifické míry úmrtnosti podle věku v závislosti na časovém indexu. Předpokládá se, že přirozený logaritmus specifické míry úmrtnosti je možné rozložit na 3 složky. Na první složku, která je na čase nezávislá, druhou, která popisuje celkovou míru úmrtnosti a je závislá na čase, a třetí složku, která charakterizuje vliv určitého věku na specifickou míru úmrtnosti v závislosti na změně celkové míry úmrtnosti a je zase závislá na čase. Matematické vyjádření tohoto modelu má tvar:

$$\ln(m_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad (7)$$

kde:

$m_{x,t}$ je specifická míra úmrtnosti ve věku x a čase t ,

α_x jsou aritmetické průměry jednotlivých řádků zlogaritmované matice $m_{x,t}$,

β_x představuje vliv jednotlivých věků na změnu specifické míry úmrtnosti,

k_t reprezentuje časový trend,

$\varepsilon_{x,t}$ je náhodná chyba (tzv. bílý šum) s nulovou střední hodnotou a konstantním rozptylem σ^2 . [7]

Důležitou vlastností $\varepsilon_{x,t}$ je tedy homoskedasticita, stejný rozptyl $\varepsilon_{x,t}$ pro všechny věky, to je často také zmiňováno jako největší slabina modelu.

Parametry modelu není možné určit jednoznačně, a proto je vhodné zavést určitá omezení. Nejčastěji se setkáme s normalizací, kterou původně navrhli Lee a Carter: [11]

$$\sum_{t=\ell_1}^{t_n} k_t = 0 \quad \text{a} \quad \sum_{x=x_1}^{x_n} \beta_x = 1. \quad (8)$$

2.1.1 Specifická míra úmrtnosti

Je zřejmé, že úmrtnost se výrazně liší podle věku, ale také podle profese, rodinného stavu, pohlaví a dalších charakteristik. A proto se místo obecné míry úmrtnosti nejčastěji pro výpočet úmrtnosti dělí populace do kategorií podle věku a pohlaví, protože zejména podle věku jsou v úmrtnosti největší rozdíly a i podle pohlaví se úmrtnost výrazně liší.

Pro takovéto kategorie se pak počítané míry úmrtnosti nazývají *specifické míry úmrtnosti*, přesněji věkově a pohlavně specifické míry úmrtnosti. Pro specifické míry úmrtnosti podle jiných charakteristik se používá termín *diferenční míry úmrtnosti*, které se počítají zřídka. [8]

Způsob určení věkově a pohlavně specifických měr úmrtnosti je zřejmý – počet zemřelých dané kategorie je vztažený k době, pro kterou jsou členové populace vystaveni riziku úmrtí (*doba expozice*). Teoreticky lze dobu expozice určit jakou součet dob, po něž jsou jednotliví členové vystaveni riziku úmrtí. Prakticky je velmi obtížné tyto údaje pro jednotlivé členy populace získat, i když individuální údaje o narození, úmrtí, přistěhování nebo vystěhování k dispozici jsou. Proto se místo skutečné doby expozice používá její odhad. Odhadne se buď jako průměr počtu členů na začátku a na konci období – průměr *počátečního a koncového stavu*, nebo jako počet členů populace ke středu období – *střední stav*, což můžeme zapsat jako: [9]

$$m_{x,t} = \frac{D_{x,t}}{\bar{S}_{x,t}}, \quad (9)$$

kde:

$m_{x,t}$ je specifická míra úmrtnosti ve věku x a čase t ,

$D_{x,t}$ je počet zemřelých v dokončeném věku x v roce t ,

$\bar{S}_{x,t}$ je střední stav x -letých osob v roce t . [7]

2.1.2 Odhady parametrů modelu

K odhadu parametrů existuje několik postupů, které se liší zejména v postupu používaném pro odhad parametru β_x . Lee a Carter upřednostňují metodu nejmenších čtverců, SVD (Singular value decomposition). Další možností je např. metoda maximální věrohodnosti. Postup SVD je početně náročný, a tak v roce 2005 přišli Haberman a Russolillo v článku „*Lee-Carter mortality forecasting: application to the Italian population*“ s metodou, která využívá pro odhad parametru β_x lineární regresi. Nyní si blíže popíšeme metodu nejmenších čtverců a metodu maximální věrohodnosti. [7]

- **Metoda nejmenších čtverců**

Metoda nejmenších čtverců slouží k odhadu parametrů v rovnici (7) tak, jak byly popsány v kapitole 2.1 v předchozím textu. Označme $\hat{m}_{x,t}$ pozorovanou intenzitu úmrtnosti ve věku x během roku t , kde $x \in \{x_1, x_2, \dots, x_m\}$ a $t \in \{t_1, t_2, \dots, t_n\}$.

Lee a Carter v původním článku používali k řešení metodu nejmenších čtverců singulární rozklad (SVD), který aplikují na matici $[\ln \hat{m}_{x,t} - \alpha_x]$. Úlohou této metody je najít odhady parametrů $\hat{\alpha}_x, \hat{\beta}_x, \hat{k}_t$ které minimalizují funkci ve tvaru:

$$O_{LS}(\alpha, \beta, k) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} (\ln \hat{m}_{x,t} - \alpha_x - \beta_x k_t)^2. \quad (10)$$

Položením $\frac{\partial}{\partial \alpha_x} O_{LS} = 0$ získáme vztah:

$$\sum_{t=t_1}^{t_n} \ln \hat{m}_{x,t} = (t_n - t_1 + 1)\alpha_x + \beta_x \sum_{t=t_1}^{t_n} k_t \quad (11)$$

a využijeme omezení (8) pro k_t , získáme odhad parametru α_x :

$$\hat{\alpha}_x = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \ln \hat{m}_{x,t}, \quad (12)$$

který vyjadřuje průměr logaritmů pozorovaných hodnot v čase. [11], [13]

Pro určení odhadů β_x a k_t použijeme postup metody *singulární rozkladu SVD*:

Věta 1: *Nechť A je libovolná čtvercová matice. Pak existují ortogonální matice U a V a diagonální matice*

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix} \quad (13)$$

na jejíž diagonále jsou vlastní čísla matice $\sqrt{A^T A}$ tak, že

$$A = U\Sigma V^T. \quad (14)$$

Rozklad (13) se nazývá singulární rozklad a vlastní čísla matice $\sqrt{A^T A}$ se nazývají singulární čísla matice A . [10]

Pro odhad zbylých parametrů β_x, k_t využijme rozklad SVD. Položme $a_{x,t} = \ln \hat{m}_{x,t} - \hat{\alpha}_x$, které jsou prvky matice A z věty 1. Hledáme pak odhady minimalizující:

$$\tilde{O}_{LS}(\beta, k) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} (a_{x,t} - \beta_x k_t)^2. \quad (15)$$

Řešení je dané singulárním rozkladem matice A . Označme λ_1 jako největší vlastní číslo matice $A^T A$, u_1 je vlastní vektor odpovídající největšímu vlastnímu číslu λ_1 čtvercové matice $A^T A$ dimenze $(t_n - t_1 + 1)$ a necht' v_1 je odpovídající vlastní vektor čtvercové matice AA^T dimenze $(x_m - x_1 + 1)$. Nejlepší aproximace prvního řádu A ve smyslu metody nejmenších čtverců je potom dána vztahem:

$$A \approx A^* = \sqrt{\lambda_1} v_1 u_1^T, \quad (16)$$

odkud lze odvodit odhady parametrů:

$$\hat{\beta}_x = \frac{v_1}{\sum_{j=1}^{x_n - x_1 + 1} v_{1j}} \quad \text{a} \quad \hat{k}_t = \sqrt{\lambda_1} \left(\sum_{j=1}^{x_n - x_1 + 1} v_{1j} \right) u_1 \quad (17)$$

za podmínky $\sum_{j=1}^{x_n - x_1 + 1} v_{1j} \neq 0$.

Alternativou k odhadu pomocí SVD rozkladu je numerický výpočet jednotlivých parametrů na základě *Newton-Raphsonova algoritmu*. (viz. [13]).

Index úmrtnosti \hat{k}_t je odhadnutý tak, aby minimalizoval chyby logaritmů úmrtnostních měr, namísto přímo chyb úmrtnostních měr. Proto Lee a Carter navrhli, že je vhodné po získání odhadů α_x, β_x , metodou nejmenších čtverců ještě jednou přepočítat odhady \hat{k}_t ,

ale tentokrát na základě hodnot již získaných metodou SVD nebo Newton-Raphsonovým algoritmem. Lee a Carter zvolili úpravu \hat{k}_t podle reálně pozorovaných počtů úmrtí v roce t , tak aby reálně pozorované počty úmrtí odpovídaly vypočteným hodnotám. Znova odhadnuté hodnoty \hat{k}_t jsou pro každé t řešením následující rovnice:

$$\sum_{x=x_1}^{x_m} D_{x,t} = \sum_{x=x_1}^{x_m} E_{x,t} e^{\hat{\alpha}_x + \hat{\beta}_x \hat{k}_t}, \quad (18)$$

kde $E_{x,t}$ je centrální expozice úmrtnosti, tedy velikost populace čelící riziku smrti ve věku x v čase t a $D_{x,t}$ je počet mrtvých ve věku x v čase t .

Rovnice (18) nemá explicitní řešení, a proto je potřeba postupovat numericky, např.: opět pomocí metody Newton-Raphsonův algoritmus. Tyto znovu odhadnuté hodnoty \hat{k}_t se mohou relativně významně lišit od původních odhadů tohoto parametru. Důvodem je stejná váha, kterou přisuzuje SVD nízkým úmrtnostním mírám mladých lidí stejně jako vysokým hodnotám starších. Mladí se však logicky podílejí o hodně méně na celkové úmrtnosti, než je tomu u starších generací.

Na závěr se výsledné odhady upravují, tak aby splňovaly podmínky zmíněné v (8):

$$\begin{aligned} \hat{\alpha}_x &= \hat{\alpha}_x + \hat{\beta}_x \bar{k}, \\ \hat{\beta}_x &= \frac{\hat{\beta}_x}{\hat{\beta}_\bullet}, \\ \hat{k}_t &= (\hat{k}_t - \bar{k}) \hat{\beta}_\bullet, \end{aligned} \quad (19)$$

kde \bar{k} je průměr původních k_t a $\hat{\beta}_\bullet$ je součet původních β_x , tedy: [11], [13]

$$\bar{k} = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \hat{k}_t \quad \hat{\beta}_\bullet = \sum_{x=x_1}^{x_m} \hat{\beta}_x. \quad (20)$$

- **Metoda maximální věrohodnosti**

Další možností, jak odhadnout parametry, je pomocí metody maximální věrohodnosti. Tato metoda na rozdíl od metody nejmenších čtverců nepracuje s věkově specifickými mírami úmrtnosti, ale s počty úmrtí osob ve věku x v roce t , $D_{x,t}$ a s příslušnými expozicemi $E_{x,t}$. Dále předpokládáme, že $D_{x,t}$ má Poissonovo rozdělení s parametrem $E_{x,t} m_{x,t}$, zapsané jako:

$$D_{x,t} \sim Po(E_{x,t} \exp(\alpha_x + \beta_x k_t)). \quad (21)$$

Věrohodnostní funkce má potom tvar:

$$L(\alpha, \beta, k) = \prod_{x=x_1}^{x_m} \prod_{t=t_1}^{t_n} \frac{\left(E_{x,t} \exp(\alpha_x + \beta_x k_t) \right)^{D_{x,t}}}{D_{x,t}!} \exp\left(-E_{x,t} \exp(\alpha_x + \beta_x k_t)\right), \quad (22)$$

a následně sestavíme logaritmickou věrohodnostní funkci, kterou použijeme pro výpočty:

$$l(\alpha, \beta, k) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} \left(D_{x,t} (\alpha_x + \beta_x k_t) - E_{x,t} \exp(\alpha_x + \beta_x k_t) \right) + c, \quad (23)$$

kde c je konstanta, která nezávisí na odhadovaných parametrech.

Odhady parametrů $\hat{\alpha}_x, \hat{\beta}_x, \hat{k}_t$ poté získáme maximalizací této funkce, kdy položíme parciální derivace dle parametrů α_x, β_x, k_t rovné nule a poté můžeme k nalezení řešení použít opět jednorozměrný Newton-Raphsonův algoritmus. Vztahy pro jednotlivé parametry jsou potom následující:

$$\begin{aligned} \hat{\alpha}_x^{(k+1)} &= \hat{\alpha}_x^{(k)} - \frac{\sum_{t=t_1}^{t_n} \left(D_{x,t} - E_{x,t} \exp\left(\hat{\alpha}_x^{(k)} + \hat{\beta}_x^{(k)} \hat{k}_t^{(k)}\right) \right)}{\sum_{t=t_1}^{t_n} \left(-E_{x,t} \exp\left(\hat{\alpha}_x^{(k)} + \hat{\beta}_x^{(k)} \hat{k}_t^{(k)}\right) \right)}, \\ \hat{k}_t^{(k+1)} &= \hat{k}_t^{(k)} - \frac{\sum_{x=x_1}^{x_m} \left(D_{x,t} - E_{x,t} \exp\left(\hat{\alpha}_x^{(k+1)} + \hat{\beta}_x^{(k)} \hat{k}_t^{(k)}\right) \right) \hat{\beta}_x^{(k)}}{\sum_{x=x_1}^{x_m} \left(-E_{x,t} \exp\left(\hat{\alpha}_x^{(k+1)} + \hat{\beta}_x^{(k)} \hat{k}_t^{(k)}\right) \right) \left(\hat{\beta}_x^{(k)} \right)^2}, \\ \hat{\beta}_x^{(k+1)} &= \hat{\beta}_x^{(k)} - \frac{\sum_{t=t_1}^{t_n} \left(D_{x,t} - E_{x,t} \exp\left(\hat{\alpha}_x^{(k+1)} + \hat{\beta}_x^{(k)} \hat{k}_t^{(k+1)}\right) \right) \hat{k}_t^{(k+1)}}{\sum_{t=t_1}^{t_n} \left(-E_{x,t} \exp\left(\hat{\alpha}_x^{(k+1)} + \hat{\beta}_x^{(k)} \hat{k}_t^{(k+1)}\right) \right) \left(\hat{k}_t^{(k+1)} \right)^2}. \end{aligned} \quad (24)$$

V tomto případě na rozdíl od předešlé metody nejmenších čtverců není potřeba upravovat parametr k_t , protože používáme přímo počty úmrtí. Odhady je ale nutné zase upravit tak, aby vyhovovali podmínkám zmíněných v (8). [1] [13]

2.2 Algoritmus výpočtu

Prvním krokem algoritmu je odhad parametru α_x , který vypočítáme pro všechna x jako aritmetický průměr z hodnot $\ln(m_{x,t})$ pro všechna t , tedy:

$$\alpha_x = \frac{1}{T} \sum_{t=t_1}^{t_n} \ln(m_{x,t}) \quad T = t_n - t_1 + 1, \quad (25)$$

kde:

T je rovno počtu prvků posloupnosti t_1, \dots, t_n . Jedná se tedy o aritmetické průměry řádků matice $[\ln(m_{x,t})]$.

n představuje rozsah analyzovaného souboru.

Druhým krokem je odhadnutí hodnot k_t . Ty odhadneme jako součty hodnot v jednotlivých sloupcích matice $[\ln(m_{x,t}) - \alpha_x]$, tedy:

$$k_t = \sum_{x=x_1}^{x_n} (\ln(m_{x,t}) - \alpha_x). \quad (26)$$

Posledním krokem je výpočet parametru β_x . K tomu využijeme lineární regresi bez absolutního členu. Zvolíme hodnoty $(\ln(m_{x,t}) - \alpha_x)$ za závisle proměnnou a za nezávisle proměnnou volíme hodnoty k_t . [14]

2.3 Predikce budoucích hodnot

Po odhadnutí všech parametrů modelu, můžeme přistoupit k predikci hodnot $m_{x,t}$ pro $t > t_n$. K tomu potřebujeme predikovat hodnoty k_t pro $t > t_n$, kdy parametry α_x a β_x jsou neměnné v čase. Cílem je tedy nalézt vhodný model vystihující chování úmrtnostního indexu k_t , který je jedinou časově závislou proměnnou v rovnici (7). Lee a Carter ve svém článku navrhuje použít standardní Boxovy-Jenkinsonovu metodologii vedoucí k nalezení nejvhodnějšího ARIMA modelu, který si popíšeme později. Nyní si charakterizujeme Boxovu-Jenkinsonovu metodologii. [7], [11]

2.3.1 Boxova-Jenkinsonova metodologie

Tato metodologie je nazvána podle autorů známé monografie Box a Jenkins (1970), která shrnula ve své době všechny tehdejší poznatky o této problematice a hlavně jim dala algoritmickou podobu.

Přistupuje k analýze časových řad na základě speciálních stochastických modelů a je tak schopna uspokojivě modelovat časové řady velmi obecných průběhů, které jsou v rámci dekompozičního přístupu nevladatelné. Charakterizuje se nejen stochastickým modelováním trendu a sezónnosti, ale hlavně velkým důrazem na (auto)korelační analýzu,

kteřá zastupuje jádro jejich postupů. Předmětem analýzy potom mohou být i časové řady se silně závislými pozorováními.

Mezi některé modely patří lineární proces, proces klouzavých součtů MA, autoregresní proces AR, smíšený proces ARMA, model ARIMA nebo SARIMA. Nyní si uvedeme některé výhody a nevýhody dané metodologie.

- **Výhody:**

- stochastické modely ARMA jsou velice přizpůsobivé, takže jsou použitelné i pro časové řady velmi obecných průběhů;
- je možno doložit velké množství úspěšných aplikací;
- softwarová podoba metodologie je dnes obvykle dostupná ve většině ekonometrických a statistických programových systémů;
- zatím pro analýzu časově závislých pozorování neexistuje „lepší“ rutinní nástroj.

- **Nevýhody:**

- metodologie vyžaduje delší časové řady (jako minimální délka se doporučuje 40 až 50 pozorování);
- je v podstatě nerealizovatelná bez počítače vybaveného příslušným softwarem a bez určité instruktáže;
- praktické vysvětlení zkonstruovaných modelů bývá obtížná; argumentem zde často bývají numerické výstupy (např. předpovědi) získané pomocí těchto modelů. [2]

Důležitou úlohou při modelování časových řad hraje autokorelační a parciální autokorelační funkce, protože časové řady se vyznačují silnou korelovaností v čase a stacionarita. Tyto charakteristiky si blíže charakterizujeme.

- **Stacionarita**

Stacionarita časové řady znamená, že chování této řady je v určitém smyslu stochasticky ustálené. Stochastický proces *striktně stacionární* znamená, že pravděpodobnostní chování příslušného stochastického procesu je invariantní vůči posunům v čase. Vzhledem k problematickému ověřování striktní stacionarity v praxi, byl v analýze zaveden méně omezující pojem *slabá stacionarita*, kdy stačí, aby stochastický proces byl invariantní vůči

posunům v čase pouze v rámci momentů do druhého řádu. Při analýze časových řad pomocí Boxovy-Jenkinsonovy metodologie pracujeme se slabou stacionaritou označovanou jako *stacionarita*. Úroveň a rozptyl časové řady jsou konstantní v čase (trend, sezónnost či proměnný rozptyl nelze se stacionaritou sloučit, a proto musí být přednostně z řady odstraněny).

- **Autokorelační funkce (ACF)**

ACF podává informaci o síle lineární závislosti mezi veličinami y_t a y_{t-k} . Jelikož v praxi obvykle neznáme hodnotu ACF dané stacionární řady, je potřeba tuto funkci odhadnout. Její odhad ACF lze zapsat jako:

$$r_k = \frac{c_k}{c_0} \quad \text{pro } k = 0, 1, \dots, n-1, \quad (27)$$

kde c_k je odhad *autokovarianční funkce*, vyjádřené:

$$c_k = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}) \quad \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad \text{pro } k = 0, 1, \dots, n-1, \quad (28)$$

kde \bar{y} je odhad *střední hodnoty*.

- **Parciální autokorelační funkce (PACF)**

Jelikož korelace mezi náhodnými veličinami y_t a y_{t-k} může být způsobena jejich korelací s veličinami ležícími mezi nimi a proto se zavádí funkce PACF, která podává informaci o korelaci veličin y_t a y_{t-k} očištěnou o vliv veličin ležících mezi nimi. Ani tady v praxi hodnoty PACF nejsou k dispozici, tak se používá opět odhad, definovaný ve tvaru:

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_j} \quad \text{pro } k > 1, \quad (29)$$

kde

$$r_{kj} = r_{k-1,j} - r_{kk} \cdot r_{k-1,k-j} \quad \text{pro } j = 1, \dots, k-1. \quad (30)$$

Chování ACF a PACF v rámci Boxovy-Jenkinsonovy metodologie je důležitým ukazatelem toho, jaký typ modelu pro danou časovou řadu použít. [2]

Na obrázku 1 jsou zobrazeny příklady korelogramů ACF a PACF. V levém korelogramu, kde je vidět, že ACF má pozvolně klesající trend, a naopak PACF je useknutá hned v prvním

bodě funkce. Tento průběh ACF a PACF je jasným znakem, že řada je nestacionární. Nestacionarita dané časové řady byla odstraněna diferencováním, což můžeme vidět na korelogramu ACF a PACF vpravo. U ACF vidíme mírně exponenciální pokles a tak v rámci části identifikace modelu, která je součástí následující podkapitoly konstrukce modelů, byl zvolen model ARIMA (0, 1, 1) s konstantou.

Autocorrelation		Partial Correlation		AC	PAC	Autocorrelation		Partial Correlation		AC	PAC
		1	0.908	0.908				1	0.389	0.389	
		2	0.819	-0.033				2	-0.074	-0.265	
		3	0.731	-0.044				3	-0.241	-0.129	
		4	0.650	-0.005				4	-0.315	-0.218	
		5	0.582	0.022				5	-0.132	0.036	
		6	0.519	-0.013				6	-0.108	-0.221	
		7	0.450	-0.078				7	-0.112	-0.120	
		8	0.370	-0.101				8	-0.088	-0.163	
		9	0.292	-0.044				9	0.066	0.079	
		10	0.215	-0.053				10	0.278	0.121	
		11	0.143	-0.038				11	0.247	0.021	
		12	0.080	-0.019				12	0.069	-0.044	
		13	0.025	-0.014				13	-0.160	-0.111	
		14	-0.015	0.040				14	-0.075	0.204	
		15	-0.055	-0.041				15	0.025	0.018	
		16	-0.092	-0.018				16	0.006	0.011	

Obrázek 1: Příklady ACF a PACF nestacionární a stacionární řady.

Zdroj: Eviews - vlastní zpracování

2.3.2 Konstrukce modelů Boxovy-Jenkinsonovy metodologie

Konstrukci modelu pro danou časovou řadu pomocí této metodologie se doporučuje provádět ve třech krocích, a to identifikace modelu, odhad modelu a diagnostika modelu. Tyto kroky si nyní popíšeme.

- **Identifikace modelu**

Rozhodnutí, jaký typ modelu zvolit, je jednou z nejtěžších úloh při výstavbě těchto modelů. Nejprve je vhodné prozkoumat **graf časové řady**, i když se jedná hlavně o subjektivní zhodnocení situace, ale většinou je možné rozpoznat přítomnost trendu a lze danou časovou řadu případně stacionarizovat.

Dalším krokem je identifikace pomocí **ACF** a **PACF**, díky kterým můžeme potvrdit, že řadu je třeba stacionarizovat. Po případné stacionarizaci na základě prohlídky grafických záznamů korelogramů ACF a PACF časové řady, kdy se této řadě snažíme přiřadit nejvhodnější typ modelu, tedy nalezení hodnot řádu p autoregresního procesu AR a řádu q procesu klouzavých součtů MA. V případě pochybností testujeme potenciální bod useknutí k_0 , za kterou ACF nebo PACF začíná být nulová, buď pomocí *Bartlettovy aproximace* s přibližným kritickým oborem (na hladině významnosti pět procent), pro určitou ACF:

$$|r_k| \geq 2 \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)} \quad \text{pro některé } k > k_0, \quad (31)$$

nebo prostřednictvím *Quenouilleovy aproximace* s kritickým oborem (na hladině významnosti pět procent), pro určitou PACF:

$$|r_{kk}| \geq 2 \sqrt{\frac{1}{n}} \quad \text{pro některé } k > k_0. \quad (32)$$

Poslední možností je identifikace pomocí **informačních kritérií**. Jde o modernější přístup, který (teoreticky) umožňuje identifikaci daného modelu plně zautomatizovat s vyloučením subjektivního zásahu analytika. K problému identifikace modelu $ARMA(p, q)$ se zde přistupuje jako k problému odhadu parametrů p a q prostřednictvím optimalizace

$$(\hat{p}, \hat{q}) = \arg \min_{(k, l)} A(k, l), \quad (33)$$

kde $A(k, l)$ je vhodné kritérium, k jehož sestavení musíme pro danou řadu odhadnout model $ARMA(k, l)$, a minimalizaci realizujeme přes zvolenou síť hodnot $k = 0, 1, \dots, K$ a $l = 0, 1, \dots, L$. Nejčastěji se využívají kritéria, která s využitím teorie informace penalizují zbytečně velké řády k, l . Jsou to například kritérium *AIC (Akaike information criterion)*:

$$AIC(k, l) = \ln \hat{\sigma}_{k, l}^2 + \frac{2(k + l + 1)}{n}, \quad (34)$$

nebo kritérium *BIC (Bayes information criterion)*:

$$BIC(k, l) = \ln \hat{\sigma}_{k, l}^2 + \frac{(k + l + 1) \ln n}{n}, \quad (35)$$

kde $\hat{\sigma}_{k, l}^2$ je odhadnutý rozptyl bílého šumu procesu $ARMA(k, l)$ a n je délka dané řady.

- **Odhad modelu**

Dalším krokem konstrukce je odhad modelu, kdy je především možné využít vztahy mezi parametry daného modelu a jeho autokorelacemi. Tyto momentové odhady se obvykle berou jen za předběžné a slouží jako počáteční hodnoty pro vlastní odhadové postupy prováděné většinou iteračně. Procedury pro konstrukci finálních odhadů v uvažovaných modelech jsou zejména softwarovou záležitostí. V případě autoregresního procesu $AR(p)$ je možné použít klasický OLS odhad, který je za předpokladu stacionarity konzistentní. Nebo například u stacionárního a invertibilního modelu $ARMA(p, q)$ se používají nejčastěji NLS odhady pomocí algoritmů typu Gauss-Newton.

- **Diagnostika modelu**

Posledním krokem konstrukce modelů v rámci Boxovy-Jenkinsonovy metodologie je diagnostika modelu, která je velmi propracovaná. Jedná se o to, že pomocí různých diagnostických nástrojů kontrolujeme, zda je skutečně kompatibilní s analyzovanými daty.

První na co se zaměřujeme, je **kontrola stacionarity**, u které se především kontroluje, zda kořeny odhadnutého autoregresního polynomu leží vně jednotkového kruhu v komplexní rovině (nebo pokud ekvivalentně jejich převrácené hodnoty, které jsou kořeny autoregresního polynomu, zapsaného s opačným pořadím mocnin $z^p - \varphi_1 z^{p-1} - \dots - \varphi_p$, leží uvnitř tohoto kruhu). Možné je také rozdělit řadu do několika úseků a testovat shodnost odhadnutých úrovní, rozptylů a autokorelací (či vyšších momentů) mezi jednotlivými úseky. Nebo je možné testovat tzv. *odezvu na impuls*, která spočívá v analýze, jakou odezvu má v odhadnutém modelu impuls, který nastal v jediném časovém okamžiku nebo opakovaně od daného časového okamžiku a stanovuje následné hodnoty procesu.

Dále se provádí **kontrola struktury ARMA**, kterou se rozumí shoda korelační struktury odhadnuté z dat (tj. funkcí r_k a r_{kk}) s korelační strukturou vypočítanou z odhadnutého modelu, který ověřujeme.

Předposlední kontrolou je **grafická prohlídka vypočteného bílého šumu**, kdy jeho grafický záznam, odhadnutý korelogram, histogram apod. mohou naznačovat eventuální vady modelu. Ve standardní situaci předpokládáme nulovou střední hodnotu, konstantní rozptyl, nekorelovanost a normalitu.

Posledním krokem diagnostiky modelu je **testování nekorelovanosti pro vypočtený bílý šum**, který můžeme za předpokladu normality testovat pomocí testu založeného na *Bartlettově aproximaci*, kde ale nyní počítáme s odhadnutými autokorelacemi vypočteného bílého šumu $r_k(\hat{\varepsilon}_t)$. Nulová hypotéza nekorelovanosti má potom kritický obor (na hladině významnosti pět procent)

$$|r_k(\hat{\varepsilon}_t)| \geq 2 \sqrt{\frac{1}{n}} \quad \text{pro } k = 1, 2, \dots \quad (36)$$

Také se velmi často používají tzv. *Q-testy*, které testují souhrnně významnost prvních K autokorelací odhadnutého bílého šumu (konstantu K je důležité zvolit předem s doporučenou výší $K \approx \sqrt{n}$, n je délka dané řady). Současně také dochází k ověření použité struktury $ARMA(p, q)$, neboť testová *Q-statistika* za platnosti nulové hypotézy, že původní řada se řídí modelem $ARMA(p, q)$, má pak asymptoticky rozdělení $\chi^2(K - p - q)$.

Mezi nejčastěji používané Q-statistiky se používá *Boxova-Pierceova statistika* s kritickým oborem (na hladině významnosti α)

$$Q = n \sum_{k=1}^K (r_k(\hat{\varepsilon}_t))^2 \geq \chi_{\alpha}^2(K - p - q), \quad (37)$$

a silnější testově *Ljungova-Boxova statistika* s kritickým oborem (na hladině významnosti α)

$$Q = n(n+2) \sum_{k=1}^K \frac{1}{n-k} (r_k(\hat{\varepsilon}_t))^2 \geq \chi_{\alpha}^2(K - p - q) \cdot [2] \quad (38)$$

2.3.3 ARIMA Model

Jedná se o modely určené pro časové řady se stochastickým trendem, které jsou nestacionární, ale můžeme je stacionarizovat diferencováním. Jedná se o integrovaný smíšený proces řádu p, d, q , značený jako *ARIMA*(p, d, q), který má tvar:

$$\varphi(B) \Delta^d y_t = \alpha + \theta(B)\varepsilon_t \quad (39)$$

kde

α je konstanta;

$\Delta^d y_t$ je d -tá diference modelované časové řady y_t ;

$\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ je autoregresní operátor, který vzniká „useknutím“ invertovaného tvaru lineárního procesu v bodě, který odpovídá zpoždění p ;

$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ je operátor klouzavých součtů, který vzniká „useknutím“ lineárního procesu v bodě, který odpovídá zpoždění q . [2]

Jedná se tedy o to, že v takovém modelu ARIMA nejprve provedeme stacionarizaci pomocí vhodné diference modelované řady a vzniklou stacionární řadu následně modelujeme pomocí smíšeného modelu ARMA. V praxi řád diferencování d nepřekročí 2, rutinní časové řady ekonomického a finančního charakteru většinou mívají $d = 1$ a speciálně řady spotřebitelských indexů či nominálních mezd někdy mohou mít $d = 2$. Jak odhadnout řád diferencování d pro pozorovanou řadu, je možné několika možnostmi:

- testy na jednotkový kořen (Dickeyův-Fullerův test, rozšířený Dickeyův-Fullerův test, KPSS-test, ...),
- subjektivní prohlídka průběhu řad $y_t, \Delta y_t, \Delta^2 y_t, \dots$ a jejich odhadnutých korelogramů ACF a PACF,

- porovnání výběrových směrodatných odchylek řad $y_t, \Delta y_t, \Delta^2 y_t, \dots$,
- aplikace informačních kritérií, modifikovaných pro modely ARIMA.

V praxi se nejčastěji používá metoda, která spočívá v prohlídce průběhu časových řad $y_t, \Delta y_t, \Delta^2 y_t, \dots$ a jejich odhadnutých korelogramů ACF a PACF. Speciální lineární pokles odhadnutých autokorelací je znamením pro další diferencování řady. [2]

3 APLIKACE LEE-CARTER MODELU NA REÁLNÁ DATA

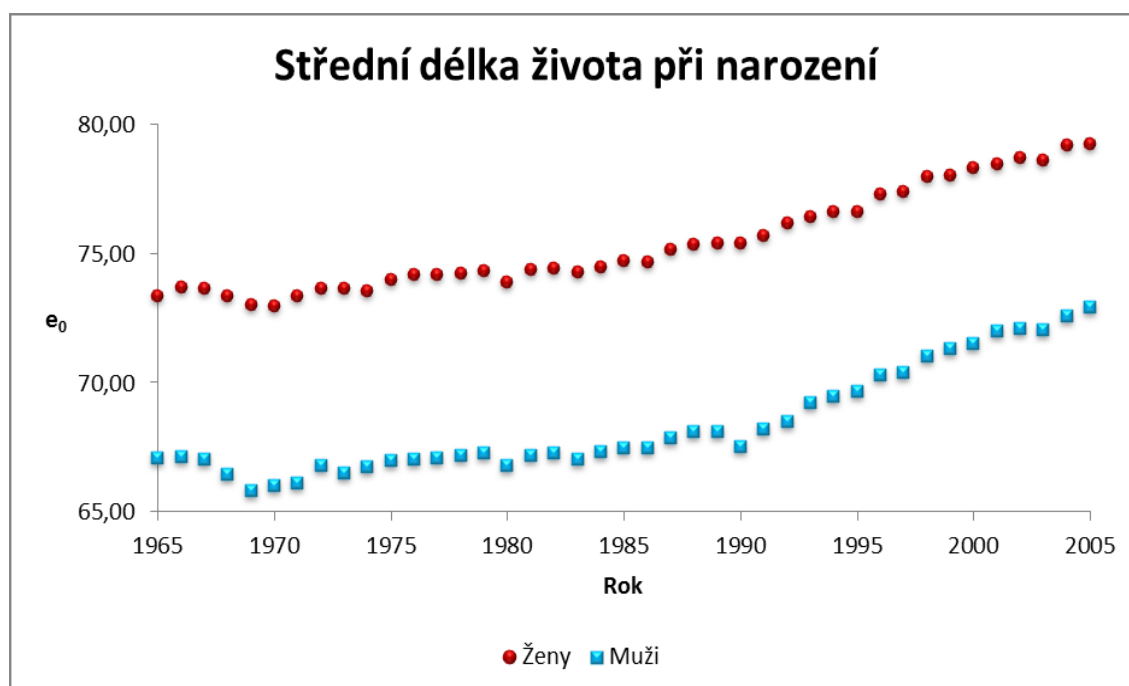
V následující kapitole budeme aplikovat výše popsany Lee-Carterův model mortality na data pro Českou republiku. V závěru budou vypočteny předpovědi měr úmrtnosti a očekávané střední délky života, které budou porovnány s aktuálními daty.

K výpočtu parametrů používáme *Microsoft Excel 2010* a pro modelování parametru k_t , pomocí ARIMA modelu, využíváme software *Eviews*. Pro všechny testy a odhady byla zvolena hladina významnosti 5 % ($\alpha = 0,05$).

3.1 Data

Model bude aplikován na data pro ČR v letech 1965 - 2005 pro věkovou kategorii 40 – 90 let, zvláště pro muže a ženy. Údaje jsme získali z veřejně dostupné databáze *Human mortality database*, ve které lze nalézt původní, ale i upravená data. My budeme využívat očekávanou délku života (*life expectancy*), počty úmrtí (*death rates*) a expozici k riziku (*exposure to risk*). Míru úmrtnosti $m_{x,t}$, lze potom vyjádřit jako podíl počtu úmrtí a expozice k riziku.

Nejdříve se podíváme na vývoj střední délky života (naděje dožití), která je jedním z měřítek vyspělosti dané společnosti a s jejím prodlužováním jsou spojené určité nároky na společnost, např. růst náročnosti na důchodové zabezpečení, sociální a zdravotní péči. [7] Podrobný vývoj pro ČR zvláště pro muže a ženy daného období je znázorněný na obrázku 2.



Obrázek 2: Vývoj střední délky života při narození pro muže a ženy v letech 1965-2005

Zdroj: vlastní zpracování podle [6]

Když se podíváme na graf, zjistíme, že i pro českou populaci dochází k prodlužování života (tzv. riziko dlouhověkosti). A proto modelování a následná predikce Lee-Carterovým modelem dává smysl, protože riziko dlouhověkosti je největší motivací pro stochastické modelování úmrtnosti. Během sledovaného období se střední délka života při narození zvýšila o necelých 7 let u žen a 8 let u mužů. Z grafu je vidět, že naděje dožití při narození pro ženy je vyšší než pro muže a to v průměru o 7 let. Jak už bylo zmíněno i v rámci ČR dochází ke stárnutí populace, což je v grafu znázorněno rostoucím trendem, který v některých letech narušili mírné odchylky, kdy v jednotlivých letech, např.: v roce 1980 nebo 1990, došlo k malému snížení očekávané délky života při narození, jak pro muže tak i ženy.

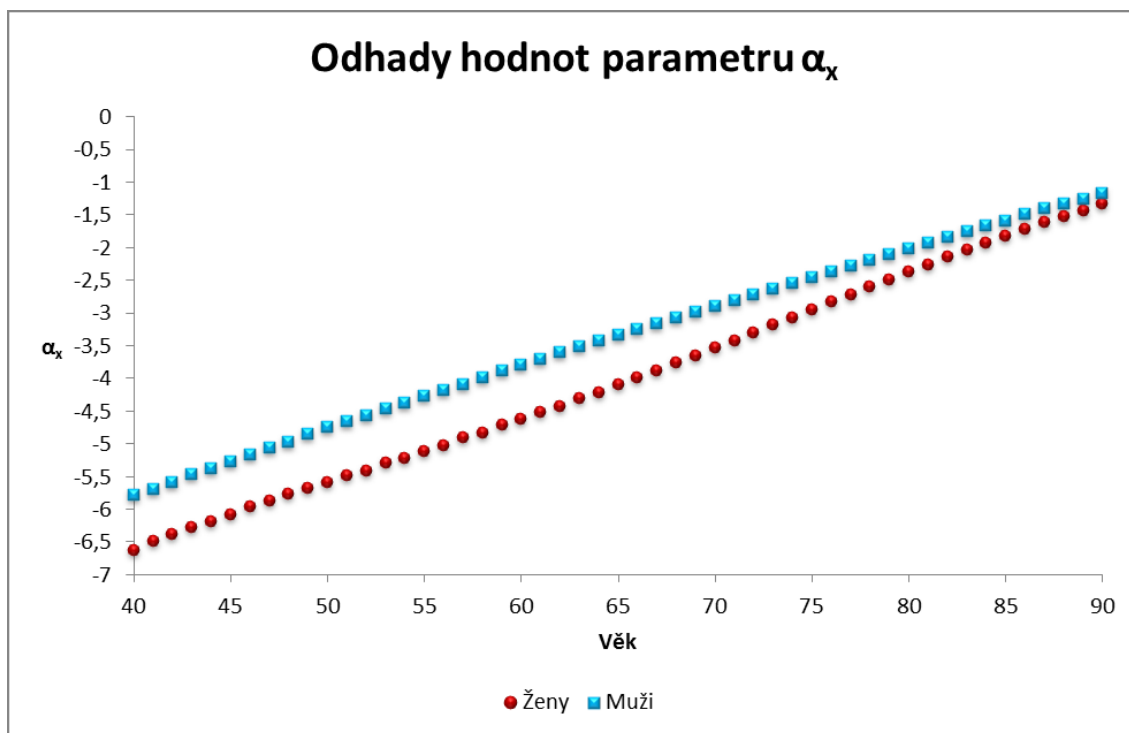
3.2 Odhady parametrů modelu

V této podkapitole se budeme věnovat odhadu parametrů, které byly popsány výše ve 2. kapitole. Nejdříve musíme sestavit matici měr úmrtnosti pro muže a ženy, které vypočítáme podle vzorce (9), tedy podílem počtu zemřelých a doby expozice. Jak je vidět ze vzorce (7), následně je potřeba, před vstupem do modelu míry úmrtnosti logaritmicky transformovat. Vznikne nám matice $\ln(m_{x,t})$, která má počet řádků roven počtu věkových kategorií a počet sloupců roven počtu let sledovaného období.

K odhadnutí modelu pro danou matici měr úmrtnosti potom hledáme řešení rovnice (7) při splnění podmínek (8). Ve 2. kapitole byly popsány dvě metody odhadů, zejména pro parametr β_x a k_t . Jak již bylo zmíněno, v původním článku Lee a Carter použili metodu *SVD*, ta je ovšem početně náročná a proto jsme pro výpočet našeho modelu použili metodu, kterou popsali v [5] Haberman a Russolilo.

3.2.1 Parametr α_x

Odhad parametru α_x vypočítáme podle rovnice (25) jako aritmetický průměr z hodnot $\ln(m_{x,t})$. Parametr znázorňuje obecný věkový profil úmrtnosti nezávislý na čase t a jeho hlavním rysem je rostoucí trend. Tento rostoucí vývoj je možné vidět i na obrázku 3, lze si i všimnout, že hodnoty parametru pro mužskou populaci jsou vyšší, což znamená jistou nadúmrtnost. S vyšším věkem parametr pro ženskou populaci začal vykazovat strmější trend vývoje a tak se od 70-ti let α_x pro muže začala přibližovat k hodnotám parametru pro ženy. Konkrétní hodnoty odhadnutého parametru α_x jsou zobrazeny v tabulce 1.



Obrázek 3: Vývoj parametru α_x pro věk 40 - 90 pro ženy a muže

Zdroj: vlastní zpracování

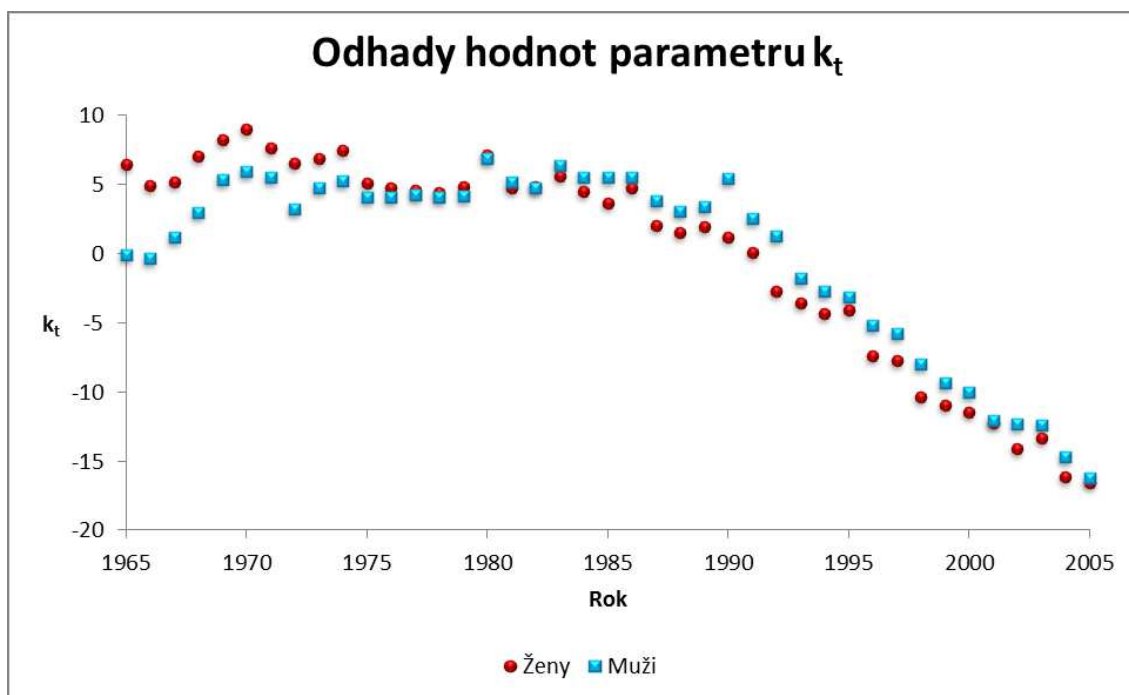
Tabulka 1: Hodnoty odhadnutého parametru α_x pro muže a ženy pro věk 40 – 90

Věk	α_x		Věk	α_x		Věk	α_x	
	Ženy	Muži		Ženy	Muži		Ženy	Muži
40	-6,6157	-5,7715	57	-4,9001	-4,0753	74	-3,0545	-2,5396
41	-6,4836	-5,6767	58	-4,8131	-3,9738	75	-2,9315	-2,4539
42	-6,3712	-5,5754	59	-4,7066	-3,8797	76	-2,8231	-2,3611
43	-6,2715	-5,4561	60	-4,6107	-3,7769	77	-2,7006	-2,2707
44	-6,1830	-5,3675	61	-4,5001	-3,6884	78	-2,5808	-2,1870
45	-6,0656	-5,2680	62	-4,4185	-3,5966	79	-2,4723	-2,0970
46	-5,9479	-5,1493	63	-4,2969	-3,5034	80	-2,3507	-2,0056
47	-5,8668	-5,0439	64	-4,2035	-3,4199	81	-2,2430	-1,9240
48	-5,7503	-4,9593	65	-4,0829	-3,3233	82	-2,1257	-1,8276
49	-5,6734	-4,8421	66	-3,9720	-3,2388	83	-2,0209	-1,7494
50	-5,5759	-4,7360	67	-3,8675	-3,1529	84	-1,9148	-1,6577
51	-5,4727	-4,6435	68	-3,7548	-3,0571	85	-1,8065	-1,5764
52	-5,4027	-4,5578	69	-3,6370	-2,9671	86	-1,7099	-1,4854
53	-5,2878	-4,4464	70	-3,5174	-2,8825	87	-1,6030	-1,3961
54	-5,2019	-4,3554	71	-3,4066	-2,8007	88	-1,5086	-1,3257
55	-5,1091	-4,2552	72	-3,2898	-2,7129	89	-1,4163	-1,2562
56	-5,0124	-4,1726	73	-3,1699	-2,6200	90	-1,3270	-1,1656

Zdroj: vlastní zpracování

3.2.2 Parametr k_t

Základní časový trend úmrtnosti pro ČR zvlášť pro muže a ženy v letech 1965 – 2005 zachycuje parametr k_t . Hodnoty vypočítané na základě rovnice (26) jsou zobrazeny v tabulce 2. Z grafu na obrázku 4 je možné vidět, že vývoj hodnot v letech 1965 – 2005 je pro obě pohlaví skoro stejný, i když ze začátku sledovaného období parametr pro ženy dosahoval vyšších hodnot. Z grafu vidíme a hodnoty v tabulce to jen potvrzují, že od roku 1986 hodnoty vykazují postupně se snižující trend a tentokrát parametr k_t pro mužskou populaci nabývá vyšších hodnot než pro ženy, ale v posledním roce sledovaného období se hodnoty parametrů téměř shodují.



Obrázek 4: Vývoj parametru k_t pro muže a ženy v letech 1965 - 2005

Zdroj: vlastní zpracování

Tabulka 2: Hodnoty parametru k_t pro muže a ženy v letech 1965 – 2005

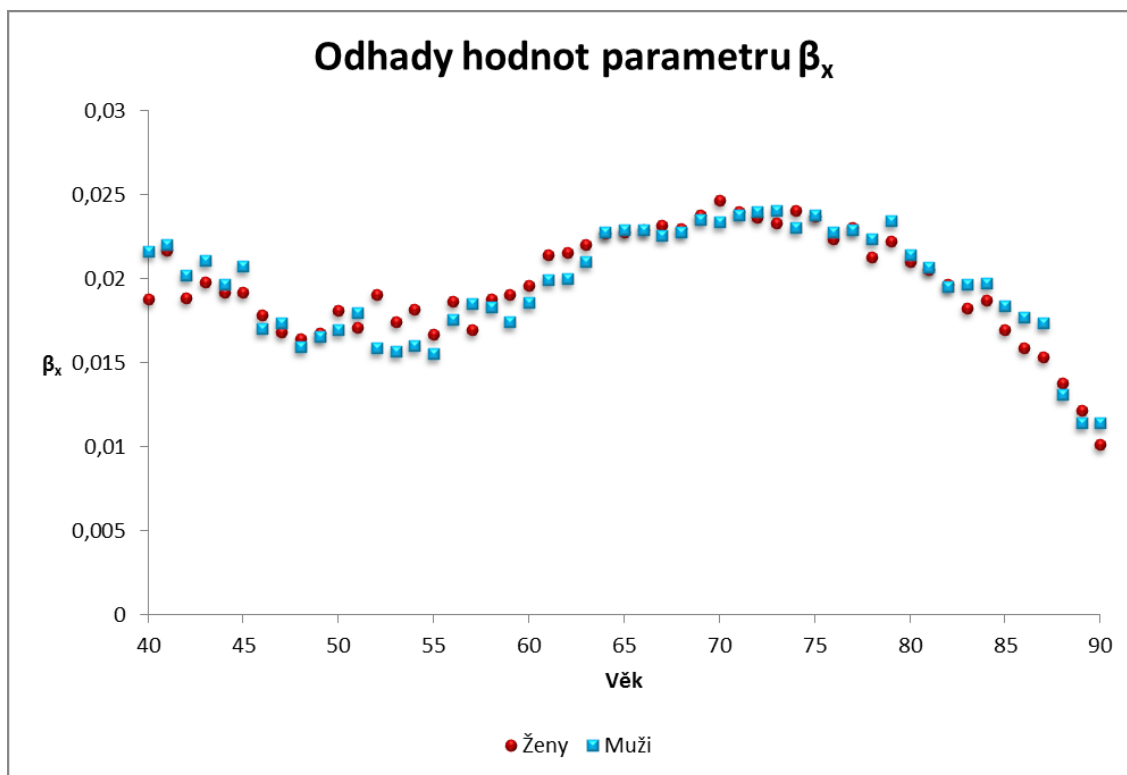
Rok	k_t		Rok	k_t		Rok	k_t	
	Ženy	Muži		Ženy	Muži		Ženy	Muži
1965	6,41161	-0,05832	1979	4,8041	4,1740	1993	-3,5323	-1,7981
1966	4,93997	-0,35282	1980	7,1583	6,8910	1994	-4,3019	-2,7064
1967	5,16179	1,18492	1981	4,7339	5,1493	1995	-4,1225	-3,1468
1968	7,05465	2,97911	1982	4,8026	4,7729	1996	-7,3806	-5,1563
1969	8,24309	5,35195	1983	5,6284	6,3762	1997	-7,7245	-5,7994
1970	9,00947	5,95282	1984	4,5240	5,4756	1998	-10,3758	-8,0214
1971	7,65391	5,55496	1985	3,6793	5,4950	1999	-10,9505	-9,3495
1972	6,56103	3,23608	1986	4,7678	5,5507	2000	-11,5025	-10,0100
1973	6,88352	4,78451	1987	1,9941	3,7833	2001	-12,2424	-12,0471
1974	7,48763	5,29273	1988	1,4850	3,0371	2002	-14,1011	-12,3492
1975	5,05070	4,05774	1989	1,9822	3,3966	2003	-13,3117	-12,4358
1976	4,75494	4,07772	1990	1,1842	5,4398	2004	-16,1570	-14,7006
1977	4,61788	4,23224	1991	0,0450	2,5389	2005	-16,6047	-16,1937
1978	4,38268	4,06218	1992	-2,6942	1,2781			

Zdroj: vlastní zpracování

3.2.3 Parametr β_x

Naopak parametr β_x vysvětluje rozdíly vlivu času na úmrtnost mezi věkovými skupinami. Pro výpočet parametru jsme vycházeli na základě článku „*Lee-Carter mortality forecasting: application to the Italian population*“ v [5]. Jak bylo zmíněno výše, autoři pro výpočet použili lineární regresi bez absolutního členu, kde za závisle proměnnou byly zvolené hodnoty $(\ln(m_{x,t}) - \alpha_x)$ a za nezávisle proměnnou hodnoty parametru k_t .

β_x popisuje tendenci úmrtnosti, ve věku x , se měnit v závislosti na parametru k_t . To znamená, že když je parametr v některých věkových kategoriích vyšší než v jiných, poté i míry úmrtnosti kolísají mnohem více. Vývoj parametru β_x je možné vidět na obrázku 5 a konkrétní vypočtené hodnoty v tabulce 3. Z grafu je vidět, že odhadnutý parametr β_x proti α_x více kolísá a tudíž nevykazuje v případě žen i mužů stálý trend. Zejména do věku 60-ti let je možné vidět, že parametr kolísá hlavně pro ženy, ale i pro muže, i když rozdíly nejsou tak velké. Po věku 60-ti let hodnoty měly tendenci růst a kolísání se ustálilo. Mezi roky 65 až 75 hodnoty parametru pro obě pohlaví byly skoro stejné, což potvrzují i hodnoty v tabulce. Naopak s vyšším věkem hodnoty parametru β_x začaly pro obě pohlaví opět klesat.



Obrázek 5: Vývoj parametru β_x pro muže a ženy pro věk 40 - 90

Zdroj: vlastní zpracování

Tabulka 3: Hodnoty odhadnutého parametru β_x pro muže a ženy pro věk 1965 - 2005

Věk	β_x		Věk	β_x		Věk	β_x	
	Ženy	Muži		Ženy	Muži		Ženy	Muži
40	0,0188	0,0216	57	0,0169	0,0184	74	0,0240	0,0230
41	0,0216	0,0220	58	0,0187	0,0183	75	0,0237	0,0237
42	0,0188	0,0202	59	0,0190	0,0174	76	0,0223	0,0227
43	0,0197	0,0210	60	0,0195	0,0185	77	0,0230	0,0228
44	0,0191	0,0196	61	0,0213	0,0199	78	0,0212	0,0223
45	0,0191	0,0207	62	0,0215	0,0199	79	0,0222	0,0234
46	0,0178	0,0170	63	0,0220	0,0210	80	0,0210	0,0214
47	0,0167	0,0173	64	0,0227	0,0227	81	0,0205	0,0206
48	0,0163	0,0159	65	0,0227	0,0229	82	0,0196	0,0195
49	0,0167	0,0165	66	0,0229	0,0228	83	0,0182	0,0196
50	0,0181	0,0169	67	0,0231	0,0225	84	0,0187	0,0197
51	0,0171	0,0179	68	0,0229	0,0227	85	0,0169	0,0183
52	0,0190	0,0158	69	0,0237	0,0235	86	0,0158	0,0176
53	0,0174	0,0156	70	0,0246	0,0233	87	0,0153	0,0173
54	0,0181	0,0159	71	0,0239	0,0237	88	0,0137	0,0131
55	0,0166	0,0155	72	0,0236	0,0239	89	0,0121	0,0114
56	0,0186	0,0175	73	0,0232	0,0240	90	0,0101	0,0114

Zdroj: vlastní zpracování

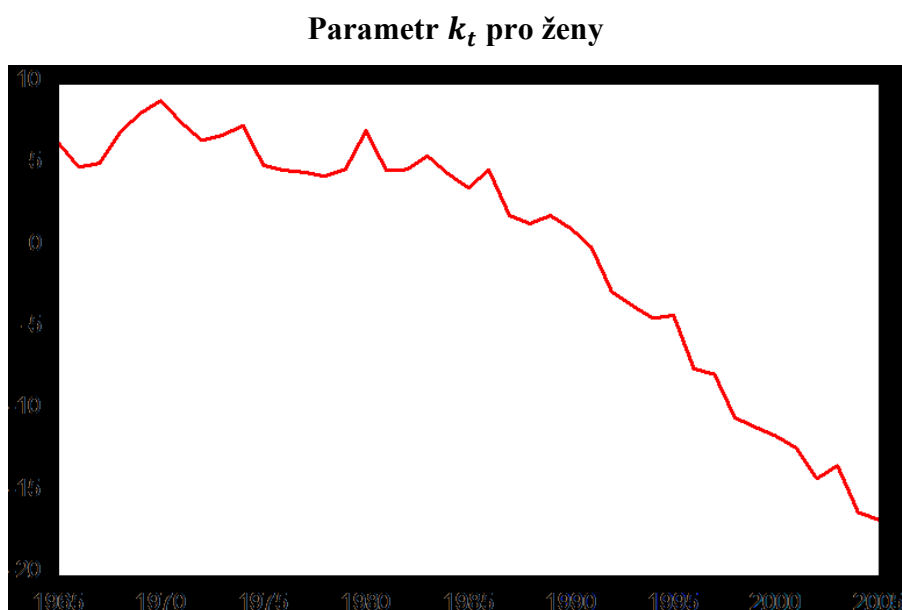
3.3 Prognózování indexu úmrtnosti k_t

Abychom mohli provést predikci měr úmrtnosti, je potřeba vytvořit odhad pouze budoucího parametru k_t , protože jak bylo zmíněno výše, zbylé dva parametry α_x a β_x zůstávají v čase konstantní. Předpovědi budeme počítat do roku 2015, abychom mohli předpovědi porovnat s hodnotami aktuálními. Prognózování indexu úmrtnosti k_t se provádí modelováním řady s použitím Boxovy-Jenkinsonovy metodologie pomocí integrovaného smíšeného procesu $ARIMA(p, d, q)$, který je s metodologií charakterizovaný v kapitole 2.3. Nyní se budeme zabývat podrobnější analýzou parametru k_t pro muže a ženy zvláště a na závěr provedeme předpovědi až do roku 2015.

3.3.1 Ženy

Prvním krokem konstrukce modelů prostřednictvím Boxovy-Jenkinsonovy metodologie je identifikace modelu, která se provede např.: prohlédnutím grafu nebo korelogramů ACF a PACF. V této části, se zjišťuje, zda je daná řada stacionární, tudíž je invariantní v čase (tzn., že nepodléhá změnám v průměru nebo variabilitě). V případě, že je daná řada nestacionární je potřeba ji vhodnou transformací stacionarizovat.

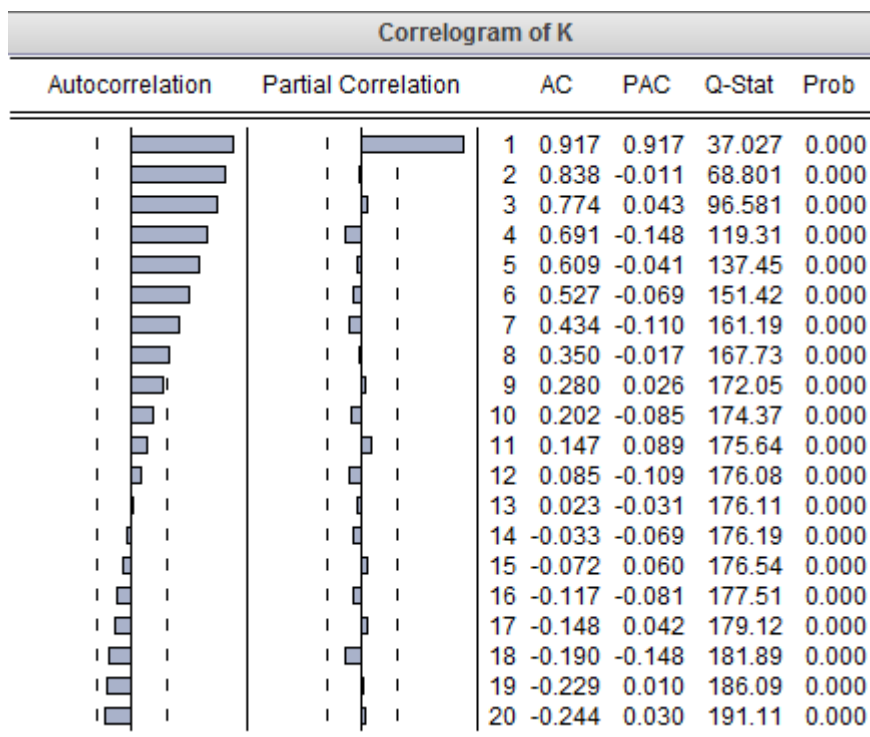
Nyní se podíváme na průběh časové řady, který je možné vidět na obrázku 6. Z grafu je patrné, že řada charakterizující vývoj úrovně úmrtnosti pro ženy v letech 1965 – 2005 vykazuje trend, který je ovšem různorodý. Přičemž nejvýraznější a nejvýznamnější část této křivky tvoří prudce klesající tendence od počátku 90. let.



Obrázek 6: Vývoj parametru k_t pro ženy v letech 1965 – 2005

Zdroj: Eviews – vlastní zpracování

Už při pohledu na zřetelný trend v průběhu analyzované časové řady pro ženy, můžeme tvrdit, že se jedná o nestacionární řadu, kterou je potřeba případným způsobem transformovat. Při nejistotě, zda je daná časová řada stacionární, je možné využít statistický software, který nám vykreslí autokorelační funkci a parciální autokorelační funkci nebo testy stacionarity. V našem případě pro modelování časových řad pomocí Boxovy-Jenkinsonovy metodologie byl využit software *Eviews*. Průběh ACF a PACF časové řady parametru k_t pro ženy je zobrazen na obrázku 7.



Obrázek 7: ACF a PACF časové řady parametru k_t pro ženy

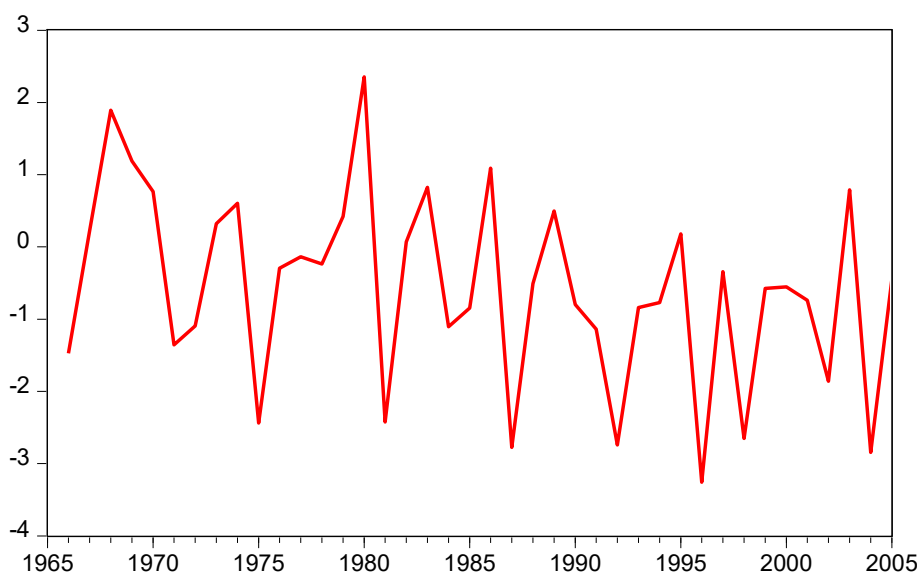
Zdroj: Eviews – vlastní zpracování

Jak je možné vidět na obrázku výše, i průběh těchto funkcí nám jednoznačně potvrzuje, že se jedná o nestacionární časovou řadu. Protože hodnoty autokorelační funkce klesají pomalým lineárním tempem a naopak parciální autokorelační funkce má první hodnoty vysoké, blíží se k jedničce a zbylé hodnoty jsou malé, což je znakem nestacionarity. A proto pro další pokračování v analýze je nutné data stacionarizovat.

Samotná transformace nestacionární ČŘ se provádí vhodným *diferencováním*, což znamená odstranění trendu. Po diferencování by měla být nestacionarita řady odstraněna. Nyní se znovu podíváme na graf vývoje diferencované časové řady (viz. obrázek 8), abychom zjistili, zda se nám podařilo odstranit trend. Z grafu je patrné, že trendu jsme se zbavili. Pro úplnost uvedeme ještě průběh ACF a PACF diferencované řady, zobrazené na obrázku 9,

kteře nam opet potvrzuj, e se nam danou asovou řadu podařilo stacionarizovat. A tudíž můžeme přistoupit k dalšímu kroku konstrukce a to je odhad modelu.

Parametr k_t pro ženy po diferencování



Obrázek 8: Vývoj diferencovaného parametru k_t pro ženy v letech 1965 – 2005

Zdroj: Eviews – vlastní zpracování

Correlogram of D(K)						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.118	-0.118	0.5974	0.440
		2	0.028	0.015	0.6328	0.729
		3	0.088	0.094	0.9818	0.806
		4	0.104	0.127	1.4848	0.829
		5	-0.086	-0.065	1.8392	0.871
		6	0.439	0.422	11.344	0.078
		7	-0.223	-0.191	13.884	0.053
		8	-0.056	-0.113	14.046	0.081
		9	0.149	0.114	15.253	0.084
		10	0.085	0.042	15.656	0.110
		11	0.072	0.232	15.959	0.143
		12	0.105	-0.118	16.627	0.164
		13	-0.099	0.029	17.236	0.189
		14	-0.095	-0.135	17.823	0.215
		15	0.130	-0.052	18.965	0.215
		16	-0.148	-0.173	20.502	0.198
		17	0.141	0.137	21.962	0.186
		18	-0.188	-0.169	24.675	0.134
		19	-0.122	-0.148	25.867	0.134
		20	-0.008	0.072	25.872	0.170

Obrázek 9: ACF a PACF diferencované časové řady k_t pro ženy

Zdroj: Eviews – vlastní zpracování

Po diferencování dané ČŘ je zřejmé, že hledaný $ARIMA(p, d, q)$ model bude prozatím vypadat takto: $(p, 1, q)$, protože d označuje řád difference. Dalším krokem je nalezení vhodných parametrů p a q . Jak bylo řečeno výše v $ARIMA$ modelech dochází ke kombinaci autoregresivních procesů $AR(p)$ a procesů klouzavých součtů $MA(q)$. Lee a Carter v původním článku použili model, kdy oba tyto parametry jsou nulové, tedy: $ARIMA(0, 1, 0)$, který nazýváme *náhodnou procházkou*.

Pro odhadování vhodných hodnot parametrů existuje několik metod, v první řadě prozkoumáním průběhu dané ACF a PACF a nebo v případě nejasností použít informační kritérium *Akaike information criterion (AIC)* a nebo *Bayes information criterion (BIC)*, často také označované jako *Schwarz information criterion*, které byly popsány výše. Jako vodítko pro výběr modelu použijeme právě tato dvě kritéria, kdy budeme vybírat model s nejnižší hodnotou daného kritéria. Provedli jsme odhady devíti modelů, kdy jsme kombinovali hodnoty řádů p a q od 0 do 2. V následující tabulce 4 jsou zobrazené hodnoty informačních kritérií AIC a BIC pro jednotlivé modely.

Tabulka 4: Hodnoty informačních kritérií $ARIMA$ modelů pro ČŘ parametru k_t pro ženy

Model	SIC	AIC
(0,1,0)	3,4874	3,4452
(0,1,1)	3,6590	3,5323
(1,1,0)	3,6581	3,5314
(1,1,1)	3,7502	3,5813
(2,1,0)	3,7500	3,5812
(0,1,2)	3,7486	3,5797
(2,1,1)	3,8103	3,5914
(1,1,2)	3,7928	3,5817
(2,1,2)	3,8746	3,6212

Zdroj: Eviews – vlastní zpracování

Z hodnot v tabulce je jasné, že i pro naši analyzovanou časovou řadu bude nejlepší model $ARIMA(0, 1, 0)$, tedy *náhodná procházka (random walk)*. Tento model může i nemusí obsahovat konstantu, nazývanou *drift*. Většina studií tuto konstantu v modelu používá, protože jinak by byl model tvořen pouze kumulativním součtem bílého šumu a v našem případě by prognóza dané časové řady byla tvořena konstantní přímkou. Do budoucna tento průběh ale nepředpokládáme, a proto se snažíme konstantu do modelu započítat. Výsledek dané náhodné procházky ze softwaru *Eviews* je znázorněn na obrázku 10, který odhadl konstantu na $-0,5754$. Bližší popis výsledků bude obsažen v závěru této kapitoly pro obě pohlaví společně.

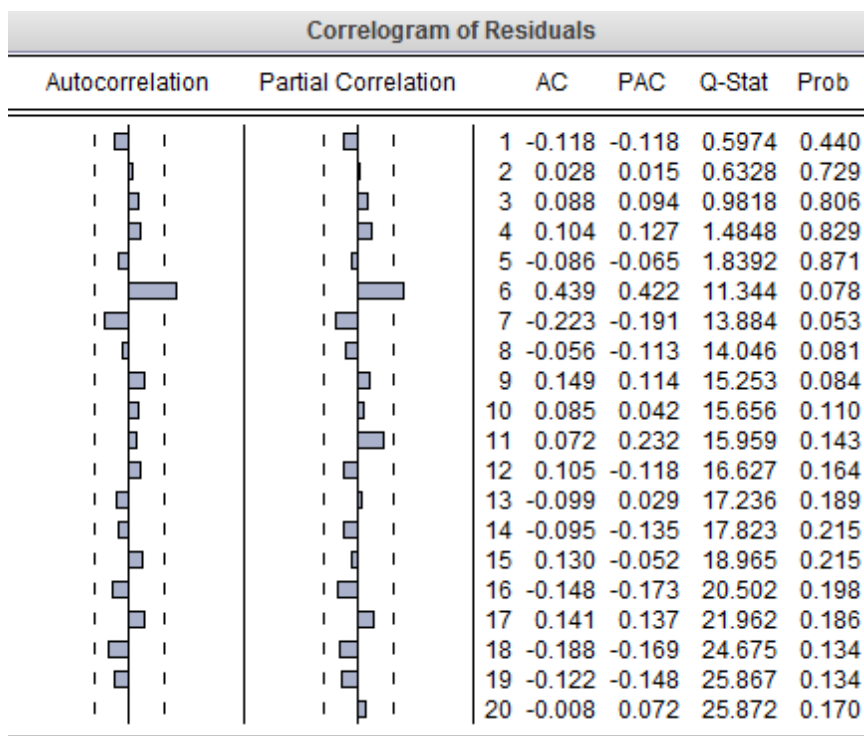
Dependent Variable: DK
 Method: Least Squares
 Date: 04/19/17 Time: 19:29
 Sample (adjusted): 1966 2005
 Included observations: 40 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.575407	0.211588	-2.719469	0.0097
R-squared	0.000000	Mean dependent var		-0.575407
Adjusted R-squared	0.000000	S.D. dependent var		1.338199
S.E. of regression	1.338199	Akaike info criterion		3.445209
Sum squared resid	69.84034	Schwarz criterion		3.487431
Log likelihood	-67.90419	Hannan-Quinn criter.		3.460475
Durbin-Watson stat	2.223788			

Obrázek 10: Výsledek $ARIMA(0, 1, 0)$ modelu pro časovou řadu parametru k_t pro ženy

Zdroj: Eviews – vlastní zpracování

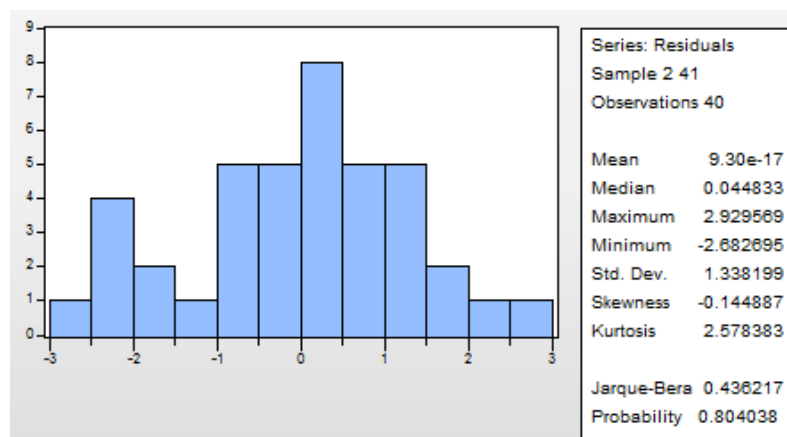
Po odhadnutí modelu můžeme přejít k poslednímu kroku konstrukce modelů a tím je diagnostika modelu. Jinak řečeno, pomocí různých nástrojů kontrolujeme, zda model dostatečně vysvětluje analyzovanou časovou řadou indexu úmrtnosti. My budeme testovat nekorelovanost a normalitu pro vypočtený bílý šum. Testování nekorelovanosti vypočteného šumu zkontrolujeme pomocí korelogramů ACF a PACF zobrazených na obrázku 11. Jak je možné vidět, hodnoty ACF i PACF jsou uvnitř intervalu spolehlivosti, což znamená, že rezidua již nejsou korelovaná. To potvrzují i hodnoty ve sloupci *Prob*, které jsou větší, než zvolená hladina významnosti 0,05.



Obrázek 11: ACF a PACF reziduí modelu náhodné procházky ČŘ parametru k_t pro ženy

Zdroj: Eviews – vlastní zpracování

Dále budeme testovat normalitu reziduí (vypočteného bílého šumu), pomocí histogramu, který je zobrazený na obrázku 12. Testujeme nulovou hypotézu: H_0 = rezidua mají normální rozdělení, oproti alternativní hypotéze: H_1 = rezidua nemají normální rozdělení. Hodnota *Probability* je 0,804038, což je větší než hladina významnosti 0,05, tudíž nulovou hypotézu nezamítáme a můžeme tvrdit, že rezidua odhadnutého modelu mají normální rozdělení.



Obrázek 12: Histogram rezidui - test normality modelu náhodné procházky ČŘ parametru k_t pro ženy

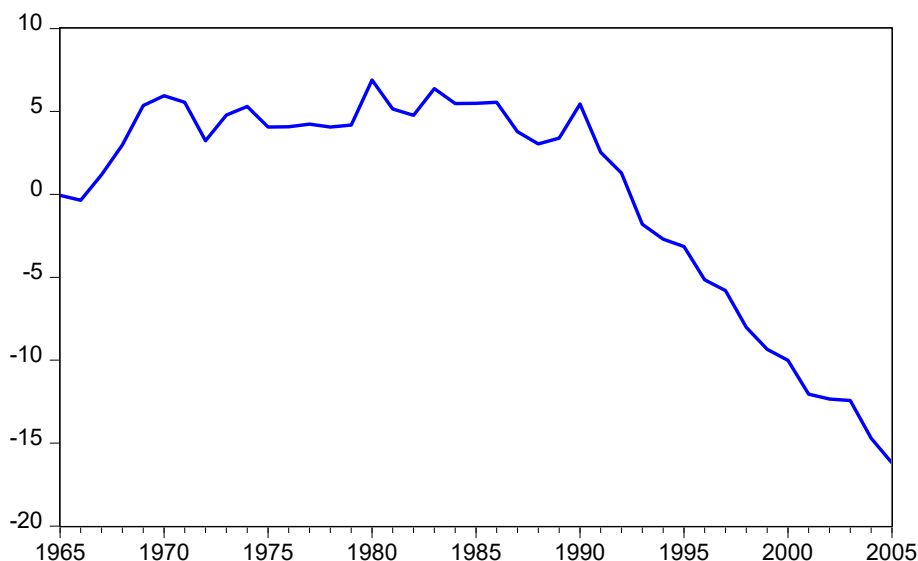
Zdroj: Eviews – vlastní zpracování

Z výsledků diagnostiky námi zvoleného modelu náhodné procházky s driftem můžeme tvrdit, že popisuje danou časovou řadu indexu úmrtnosti žen, vzhledem k jejímu průběhu celkem dobře. Tudíž můžeme přistoupit k vytvoření předpovědi daného parametru k_t pro ženy od roku 2006 do roku 2015. Výsledné hodnoty jsou uvedeny v tabulce 7 a znázorněny na obrázku 20 společně s hodnotami k_t pro muže na konci této kapitoly.

3.3.2 Muži

Nyní provedeme analýzu časové řady parametru k_t pomocí těchto tří kroků konstrukce modelu také pro mužskou populaci. V rámci identifikace modelu si prohlédneme graf časové řady parametru k_t , znázorněný na obrázku 13. Z grafu je opět patrné, že se bude jednat o nestacionární řadu. Vývoj úmrtnostního indexu k_t mužské populace v letech 1965 – 2005 je přibližně stejný jako u žen. Tedy vykazuje klesající, nesourodý trend, který zejména od 90. let nabral strmější průběh. Přítomnost tohoto trendu znamená, že daná časová řada je nestacionární.

Parametr k_t pro muže



Obrázek 13: Vývoj parametru k_t pro muže v letech 1965 – 2005

Zdroj: Eviews – vlastní zpracování

To, že je potřeba danou řadu stacionarizovat, potvrzují i korelogramy ACF a PACF znázorněné na obrázku 14. Opět můžeme vidět u ACF pozvolné klesání lineárním tempem a u PACF jednu významnou hodnotu blíží se k jedné a zbylé hodnoty blízké nule. Lze tedy říci, že se jedná o nestacionární řadu, kterou diferencujeme.

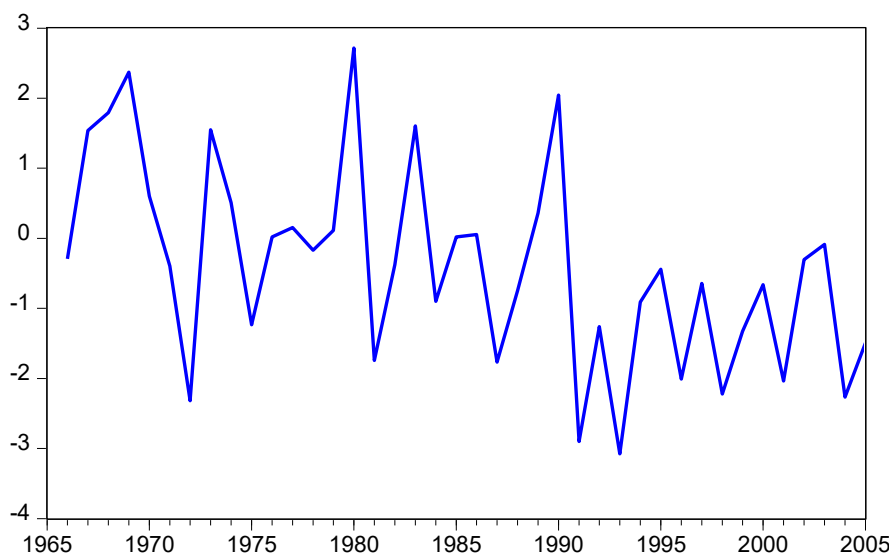
Correlogram of K						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.904	0.904	36.006	0.000
		2	0.811	-0.034	65.709	0.000
		3	0.731	0.023	90.494	0.000
		4	0.640	-0.106	110.00	0.000
		5	0.544	-0.081	124.48	0.000
		6	0.454	-0.035	134.86	0.000
		7	0.355	-0.113	141.40	0.000
		8	0.257	-0.065	144.92	0.000
		9	0.181	0.043	146.72	0.000
		10	0.104	-0.070	147.34	0.000
		11	0.032	-0.020	147.40	0.000
		12	-0.038	-0.078	147.49	0.000
		13	-0.110	-0.079	148.24	0.000
		14	-0.155	0.066	149.82	0.000
		15	-0.189	-0.008	152.25	0.000
		16	-0.196	0.119	154.95	0.000
		17	-0.223	-0.155	158.61	0.000
		18	-0.258	-0.103	163.72	0.000
		19	-0.279	-0.004	169.96	0.000
		20	-0.282	0.029	176.62	0.000

Obrázek 14: ACF a PACF časové řady parametru k_t pro muže

Zdroj: Eviews – vlastní zpracování

Poté, co diferencujeme řadu, prohlédneme její průběh v grafu (viz. obrázek 15) spolu s korelogramy ACF a PACF, které můžeme vidět na obrázku 16. Z obrázku 15 i 16, je tedy patrné, že se nám řadu podařilo stacionarizovat, tzn., že se jedná o ČŘ s v čase konstantním průměrem, rozptylem a autokorelací. A můžeme přistoupit k odhadu modelu.

Parametr k_t pro muže po diferencování



Obrázek 15: Vývoj diferencovaného parametru k_t pro muže v letech 1965 – 2005

Zdroj: Eviews – vlastní zpracování

Correlogram of DK						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.160	0.160	1.1008	0.294
		2	0.079	0.055	1.3778	0.502
		3	0.191	0.176	3.0423	0.385
		4	0.084	0.028	3.3717	0.498
		5	0.172	0.146	4.7950	0.441
		6	0.253	0.189	7.9646	0.241
		7	0.187	0.122	9.7531	0.203
		8	-0.143	-0.267	10.827	0.212
		9	0.104	0.073	11.418	0.248
		10	0.221	0.161	14.164	0.166
		11	-0.001	-0.070	14.164	0.224
		12	0.147	0.043	15.462	0.217
		13	0.005	-0.083	15.463	0.279
		14	-0.057	-0.008	15.674	0.334
		15	-0.060	-0.114	15.913	0.388
		16	0.014	-0.100	15.928	0.458
		17	0.058	0.073	16.171	0.512
		18	-0.226	-0.204	20.075	0.329
		19	-0.085	-0.096	20.657	0.356
		20	-0.025	0.090	20.709	0.414

Obrázek 16: ACF a PACF diferencované časové řady k_t pro muže

Zdroj: Eviews – vlastní zpracování

Díky těmto výsledkům, budeme zase sestavovat model $ARIMA(p, 1, q)$, kde hodnoty parametrů autoregresního procesu $AR(p)$ a procesu klouzavých průměrů $MA(q)$ zvolíme na základě porovnání kritérií AIC a BIC , kdy nejhodnější model bude ten s nejnižším kritériem. Opět jsme sestavili devět modelů, kdy jsme volili hodnoty řádů p a q od 0 do 2. Výsledné hodnoty kritérií jednotlivých modelů znázorňuje tabulka 5.

Tabulka 5: Hodnoty informačních kritérií ARIMA modelů pro ČR parametru k_t pro muže

Model	SIC	AIC
(0,1,0)	3,6300	3,5878
(0,1,1)	3,7909	3,6642
(1,1,0)	3,7888	3,6622
(1,1,1)	3,8182	3,6493
(2,1,0)	3,8774	3,7085
(0,1,2)	3,8823	3,7134
(2,1,1)	3,9076	3,6965
(1,1,2)	3,9060	3,6949
(2,1,2)	3,9832	3,7298

Zdroj: Eviews – vlastní zpracování

Na základě porovnání kritérií jsme vybrali, jako v případě ženské populace, náhodnou procházku s driftem, tedy model $ARIMA(0, 1, 0)$, který dosahoval ze všech vytvořených modelů nejnižších hodnot, jak pro kritérium SIC i AIC . Konstanta byla odhadnutá na hodnotu $-0,4034$, což je možné vidět na výsledku ze softwaru *Eviews* na obrázku 17.

Dependent Variable: DK
Method: Least Squares
Date: 04/20/17 Time: 10:31
Sample (adjusted): 1966 2005
Included observations: 40 after adjustments

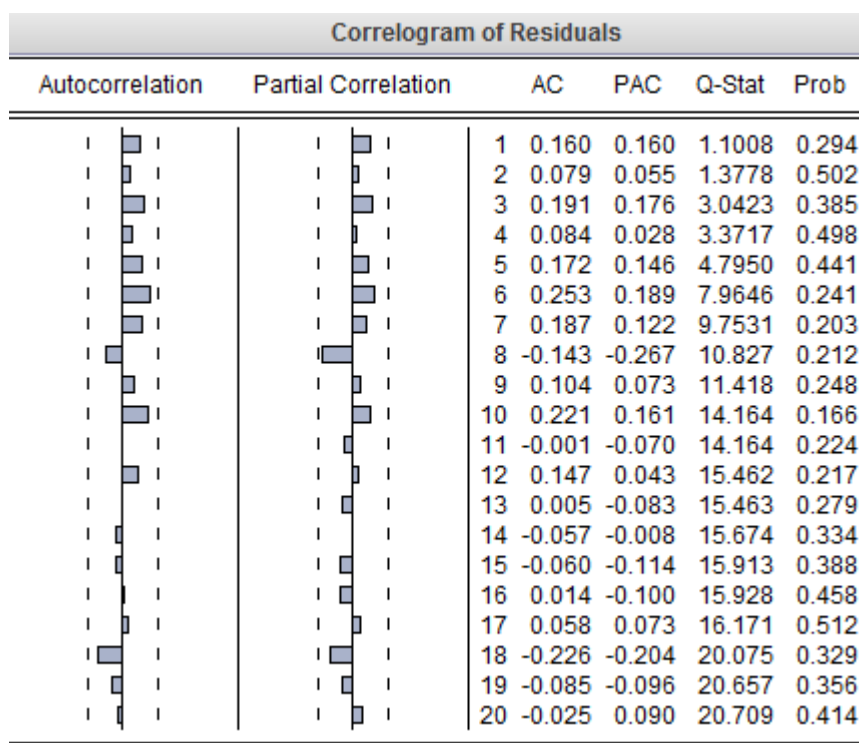
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.403384	0.227226	-1.775254	0.0837

R-squared	0.000000	Mean dependent var	-0.403384
Adjusted R-squared	0.000000	S.D. dependent var	1.437103
S.E. of regression	1.437103	Akaike info criterion	3.587818
Sum squared resid	80.54537	Schwarz criterion	3.630040
Log likelihood	-70.75637	Hannan-Quinn criter.	3.603084
Durbin-Watson stat	1.665390		

Obrázek 17: Výsledek $ARIMA(0, 1, 0)$ modelu pro časovou řadu parametru k_t pro muže

Zdroj: Eviews – vlastní zpracování

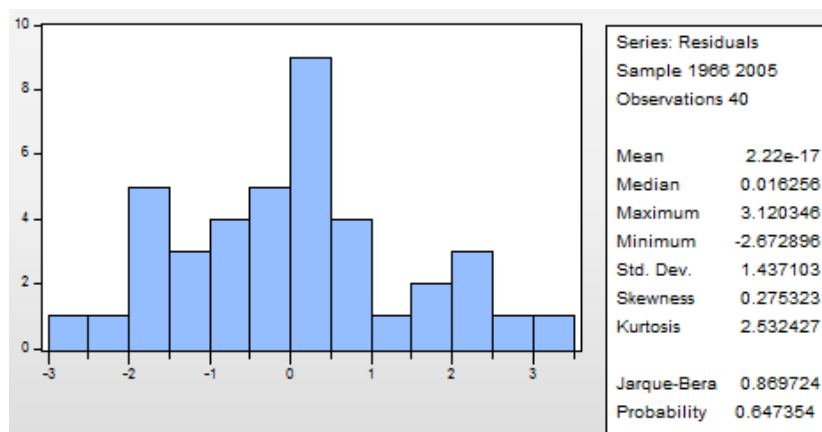
Po odhadnutí modelu můžeme přejít k poslední části, a to k diagnostice daného modelu, jako v případě ženské populace. Na obrázku 18 je znázorněný korelogram ACF a PACF vypočteného bílého šumu, kterým testujeme jeho nekorelovanost.



Obrázek 18: ACF a PACF reziduí modelu náhodné procházky ČR parametru k_t pro muže

Zdroj: Eviews – vlastní zpracování

Jak je patrné z obrázku, vypočtený bílý šum už není zatížen autokorelací, protože hodnoty ACF a PACF jsou uvnitř intervalu spolehlivosti a všechny hodnoty ve sloupci *Prob* jsou větší než naše hladina významnosti 0,05. Následně budeme testovat normalitu vypočteného bílého šumu, kdy opět testujeme nulovou hypotézu H_0 = rezidua mají normální rozdělení, oproti alternativní: H_1 = rezidua nemají normální rozdělení. Daný histogram s hodnotou *Probability* je znázorněný na obrázku 19, kdy hodnota *Probability* je 0,6474, což je také větší jak hladina významnosti 0,05. Tudíž nulovou hypotézu zamítáme stejně jako v případě ženské populace a můžeme tvrdit, že vypočtený bílý šum má normální rozdělení.



Obrázek 19: Histogram reziduí - test normality modelu náhodné procházky ČR parametru k_t pro muže

Zdroj: Eviews – vlastní zpracování

Na základě výsledků diagnostiky tohoto modelu, můžeme tvrdit, že náhodná procházka s driftem celkem dobře popisuje průběh indexu úmrtnosti k_t pro muže stejně jako v případě ženské populace. Můžeme tedy začít s předpovídáním parametru k_t pro muže i ženy. Těmto předpovědím se budeme věnovat v následující podkapitole.

3.3.3 Výsledky a předpovědi parametru k_t

Pro případ vývoje indexu úmrtnosti k_t pro obě pohlaví ve sledovaném období 1965 - 2005 jsme zvolili jako nejvhodnější model ARIMA (0, 1, 0). Tabulka 6 přibližuje hodnoty odhadnuté konstanty pro obě pohlaví zvlášť a další charakteristiky. Otestujeme statistickou významnost konstant pro obě pohlaví porovnáním p hodnoty s hodnotou zvolené hladiny významnosti 0,05. Jak můžeme vidět, statisticky významná je pouze konstanta v případě ženské populace, naopak u mužů tuto konstantu za statisticky významnou považovat nemůžeme. I přes tento výsledek konstantu do modelu zahrneme, protože do budoucna předpokládáme, že parametr k_t pro muže bude pokračovat v klesajícím trendu. V případě nezahrnutí konstanty by index úmrtnosti pro muže nabýval stejných hodnot a vývoj by tedy měl konstantní průběh.

Tabulka 6: Porovnání hodnot konstanty modelu náhodné procházky pro úmrtnostní index k_t podle pohlaví

Pohlaví	Odhad	Stadardní chyba	hodnota t	hodnota p
Muži	-0,4034	0,2272	-1,7753	0,0837
Ženy	-0,5754	0,2116	-2,7195	0,0097

Zdroj: Eviews – vlastní zpracování

Na základě výsledků námi zvolených ARIMA modelů, můžeme model zapsat následující rovnicí, pro muže:

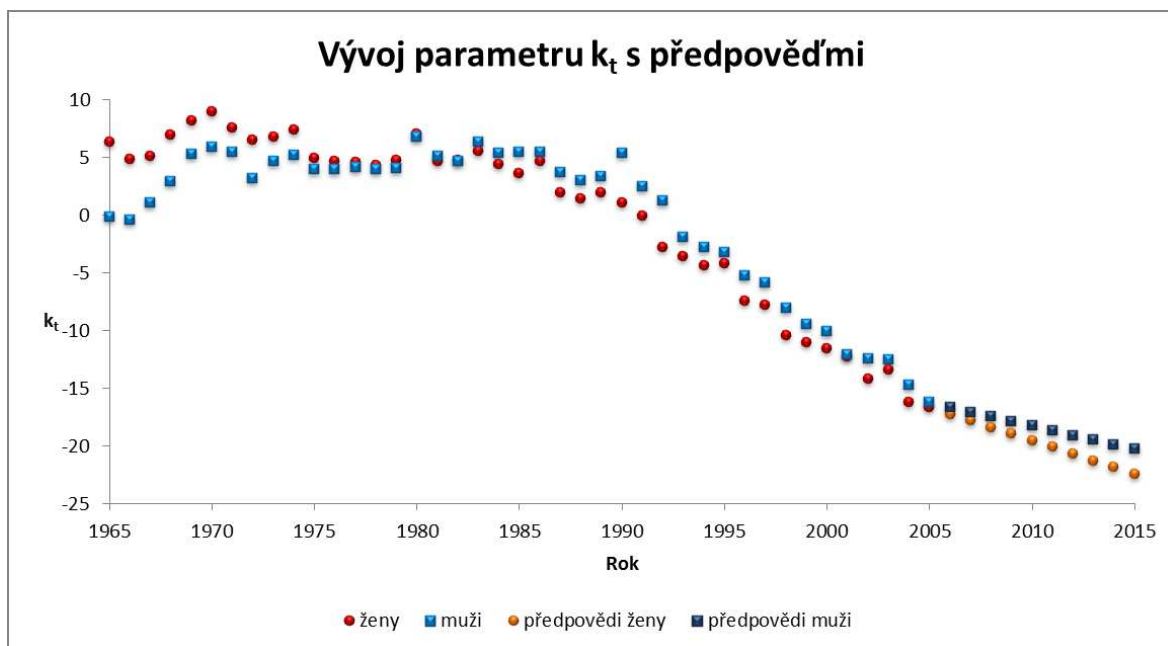
$$\hat{k}_t = \hat{k}_{t-1} - 0,4034 + \varepsilon_t \quad (40)$$

a pro ženy:

$$\hat{k}_t = \hat{k}_{t-1} - 0,5754 + \varepsilon_t, \quad (41)$$

kde ε_t označuje chybovou (zbytkovou) složku (bílý šum).

Z těchto rovnic potom můžeme vypočítat předpovědi parametru k_t pro obě pohlaví od roku 2006 do roku 2015.



Obrázek 20: Prognóza indexu úmrtnosti k_t pro obě pohlaví, 1965 – 2015

Zdroj: vlastní zpracování

Výše uvedený obrázek 20 vyjadřuje odhadovaný vývoj budoucích hodnot úmrtnostního indexu k_t zvláště pro muže a ženy pomocí modelu náhodné procházky s driftem. U obou časových řad tedy můžeme předpokládat, že budou i nadále klesat a v případě ženské populace s větší intenzitou než u mužské. Hodnoty předpovězeného parametru k_t jak pro ženskou i mužskou populaci přibližuje tabulka 7. Nyní už máme všechny potřebné výpočty a můžeme přistoupit k predikci měr úmrtnosti, kterou rozebíráme dále.

Tabulka 7: Hodnoty předpovězeného parametru k_t pro obě pohlaví pro roky 2006 - 2015

Rok	k_t	
	Ženy	Muži
2006	-17,1801	-16,5971
2007	-17,7555	-17,0004
2008	-18,3309	-17,4038
2009	-18,9063	-17,8072
2010	-19,4817	-18,2106
2011	-20,0571	-18,6140
2012	-20,6325	-19,0174
2013	-21,2079	-19,4207
2014	-21,7833	-19,8241
2015	-22,3587	-20,2275

Zdroj: vlastní zpracování

3.4 Projekce vývoje úmrtnosti

Poté co jsme na základě odpovídajícího ARIMA modelu získali prognózu časové řady indexu úmrtnosti k_t zvlášť pro muže a ženy, můžeme jednoduše odhadnout budoucí vývoj úmrtnosti pro ženy a muže v období 2006 – 2015. Následně prostřednictvím odhadnutých měř úmrtnosti $m_{x,t}$ vypočítáme očekávanou délku života pro vybrané věkové kategorie, které na konci této kapitoly budou porovnány s hodnotami z úmrtnostních tabulek získaných z *Human mortality database* a *Českého statistického úřadu*.

3.4.1 Predikce měř úmrtnosti

Odhad budoucích hodnot měř úmrtnosti pro muže a ženy zvlášť dostaneme upravením vzorce (7), který potom vypadá následovně:

$$\hat{m}_{x,t+s} = e^{(\hat{\alpha}_x + \hat{\beta}_x \hat{k}_{t+s})}, \quad (42)$$

kde $\hat{m}_{x,t}$ označuje míry úmrtnosti ve věku x a s periodu prognózování.

Vypočtené hodnoty úmrtnostních měř $m_{x,t}$ od 40-ti do 90-ti let pro období 2006 – 2015 zvlášť pro muže a ženy jsou zobrazeny v tabulkách 10 a 11 v příloze 1. Pomocí těchto odhadnutých hodnot je možné následně vypočítat jednu z důležitých charakteristik, kterou sledujeme vývoj úmrtnosti pro danou věkovou kategorii.

3.4.2 Predikce střední délky života

Jak již bylo řečeno střední délka života, neboli naděje dožití je jedním z měřítek, kterým se sleduje vyspělost dané společnosti. Pomocí námi získaných měř úmrtnosti $m_{x,t}$ lze vypočítat pravděpodobnost úmrtí ve věku x , označené q_x , pro jednotlivé věky zvlášť pro muže a ženy a z toho vyplývající další důležité charakteristiky uveřejněné v úmrtnostních tabulkách. Základem při výpočtech těchto jednotlivých charakteristik je po odhadnutí daných q_x získání hodnot pravděpodobnosti dožití se věku x , značené p_x , vypočítané ze vztahu $q_x + p_x = 1$. [14]

Poté na základě těchto dvou hodnot lze vypočítat další charakteristiky, které jsou uvedeny v úmrtnostních tabulkách (např. l_x – počet osob dožívajících se věku x , nebo d_x – počet zemřelých ve věku x) a pomocí kterých odhadneme budoucí hodnoty střední délky života pro období 2006 – 2015. Hodnota q_x se potom vypočítá na základě námi odhadnutých měř úmrtnosti $m_{x,t}$ zvlášť pro muže a ženy pro jednotlivé roky predikovaného období podle vztahu:

$$\hat{q}_x = 1 - e^{(-m_{x,t})}, \quad (43)$$

kde:

\hat{q}_x je pravděpodobnost úmrtí ve věku x ,

$m_{x,t}$ námi odhadnutá míra úmrtnosti věku x a času t .

Dalšími několika výpočty jednotlivých charakteristik se dostaneme k odhadu hodnot střední délky života ve věku x , kterou značíme e_x , získané následujícím vztahem:

$$e_x = \frac{T_x}{L_x}, \quad (44)$$

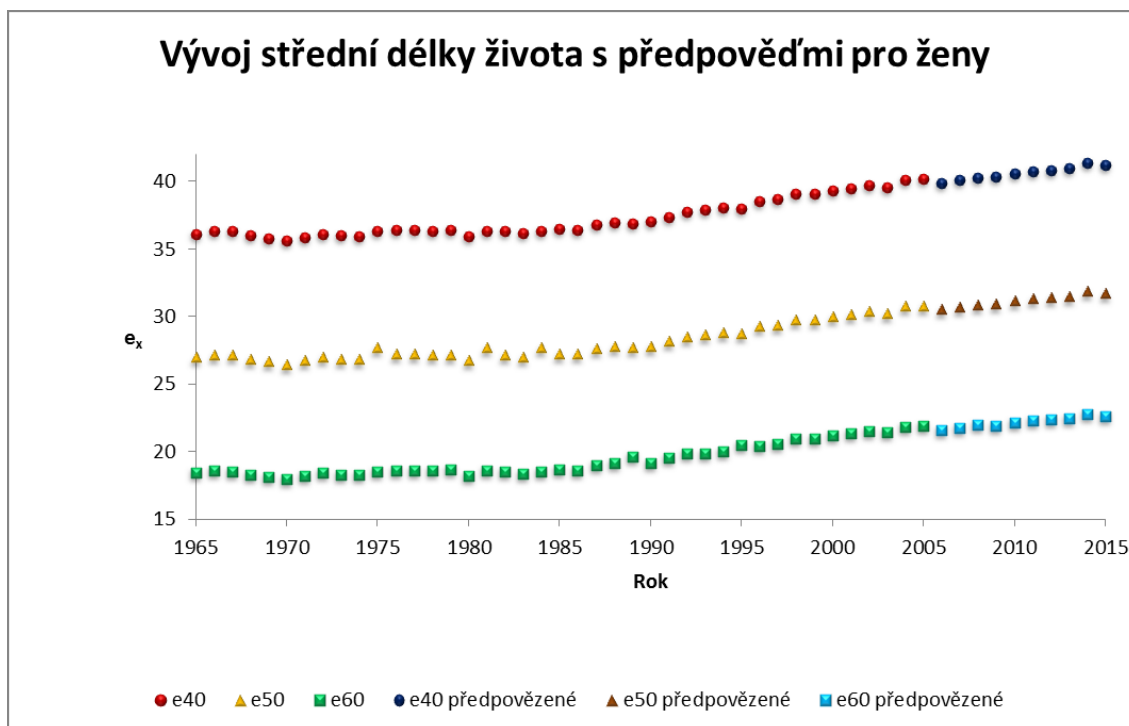
kde:

T_x vyjadřuje počet let života, které má tabulková generace (nikoliv jednotlivec) v daném věku ještě před sebou,

L_x je hypotetický průměrný počet žijících v dokončeném věku x let. [3]

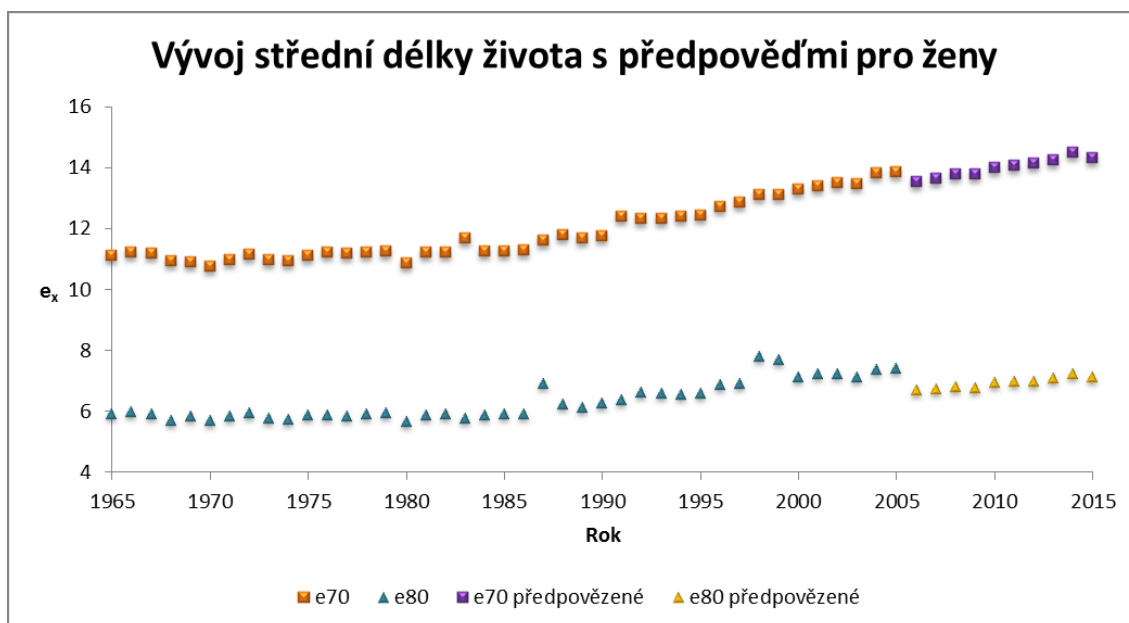
Pro naše sledované období jsme vybrali střední délku života zvláště pro muže a ženy věkové kategorie 40-ti, 50-ti, 60-ti, 70-ti a 80-ti let. Reálné hodnoty jsme získali z *Human mortality database* a z *Českého statistického úřadu*, protože pro rok 2015 nejsou hodnoty na stránkách *HMD* k dispozici. Tyto převzaté hodnoty pro období 1965 – 2015 jsou v tabulkách 12 a 13 v příloze 2. Hodnoty pro roky 2006 – 2015 jsme odhadli pomocí námi vypočítaných měr úmrtnosti $m_{x,t}$, které jsme dosadili do vzorce (43) a dalšími výpočty dostali potřebné charakteristiky pro výpočet dané střední délky života pomocí (44).

Na následujících obrázcích 21 a 22 vidíme střední délky života pro již zmíněné vybrané věky ženské populace. Na obou obrázcích, střední délka života pro všechny vybrané věkové kategorie vykazuje pozvolně rostoucí trend. Také je možné si všimnout, že došlo k mírnému podhodnocení námi odhadnutých hodnot střední délky života pro všechny věky. Na obrázku 22 je vidět, že střední délka života 80-ti letých vykazuje v některých letech největší odchylky, zejména v roce 1987, 1998, 1999. Podobné odchylky, i když menšího rozsahu, vykazuje i naděje dožití pro 70-ti leté ženy v letech 1983 a 1991. Z obrázku 21 je patrné, že téměř žádné odchylky v trendu nemá střední délka života 40-ti letých žen, k větším odchylkám došlo v případě naděje dožití 50-ti letých žen, a to v letech 1975, 1981 a 1984. U všech středních délek života vybraných věkových kategorií zaznamenáme pozvolnější růst do roku 1990, od kterého lze poté vidět růst trochu strmější. Toto potvrzují i hodnoty, které najdeme v tabulce 12 v příloze 2.



Obrázek 21: Vývoj střední délky života ženské populace v letech 1965 - 2015

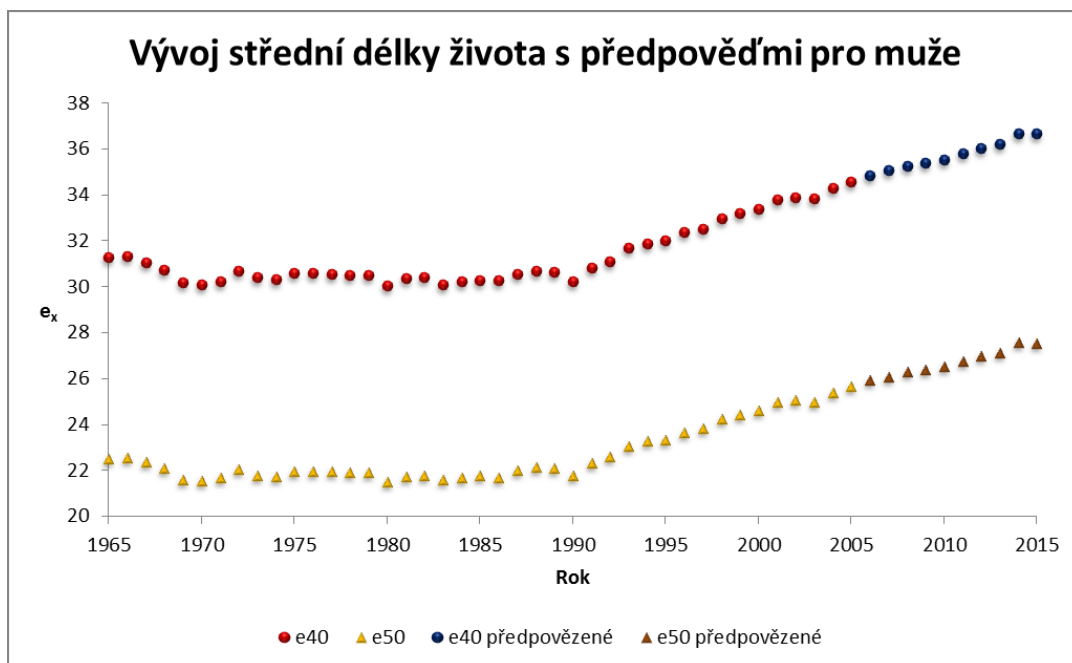
Zdroj: vlastní zpracování



Obrázek 22: Vývoj střední délky života ženské populace v letech 1965 - 2015

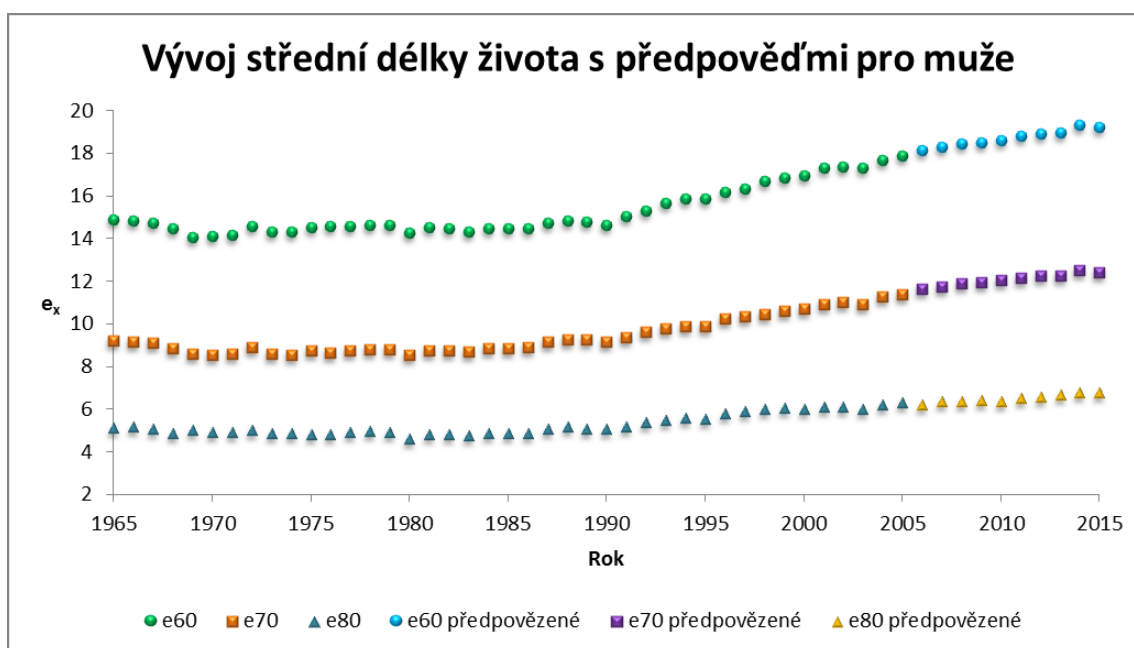
Zdroj: vlastní zpracování

Na obrázcích 23 a 24 jsou znázorněny grafy pro vývoj střední délky života vybraných věkových kategorií v letech 1965 – 2015 mužské populace.



Obrázek 23: Vývoj střední délky života v letech 1965 - 2015 mužské populace

Zdroj: vlastní zpracování



Obrázek 24: Vývoj střední délky života v letech 1965 - 2015 mužské populace

Zdroj: vlastní zpracování

Z obou grafů je patrné, že vývoj naděje dožití mužské populace je jiný než v případě žen. Zatímco u žen střední délka života vykazovala pro všechny věkové kategorie pozvolný růst, v případě mužů to tak jednoznačné není. Zejména pro věky 40-ti a 50-ti let ze začátku sledovaného období hodnoty klesaly, v roce 1970 se hodnoty ustálily a s mírnými odchylkami lze říci, že tvoří přímku a rostoucí tendenci zaznamenaly hodnoty až od roku 1991. Oproti

ženské populaci v těchto letech nabrala podstatně strměji rostoucí tendenci. Podobný průběh vykazuje i naděje dožití v případě 60-ti letých mužů, což vidíme na obrázku 24. Z tohoto grafu vidíme, že nejmenší nárůst zaznamenala střední délka života pro 80-ti leté muže a lze říci, že i vykazovala nejméně odchylek. Ale u všech kategorií si můžeme všimnout, že ze začátku sledovaného období několik let hodnoty klesaly. Také oproti ženské populaci, námi odhadnuté hodnoty od roku 2006 nepodhodnocují vývoj střední délky života pro všechny věkové kategorie mužské populace. Dané hodnoty je možné opět vidět v příloze 2 v tabulce 13.

Jelikož jsme hodnoty odhadovali do roku 2015, můžeme je porovnat s hodnotami, které jsme přezvali ze stránek *Human mortality database* a z *Českého statistického úřadu*. Na následujících dvou tabulkách 8 a 9 můžeme vidět námi odhadnuté střední délky života pro vybrané věkové kategorie za období 2006 – 2015 spolu s horní a dolní hranicí 95% intervalu spolehlivosti a také hodnoty skutečné, zvláště pro muže a ženy. Horní a dolní hranici 95% intervalu spolehlivosti jsme vypočítali, stejně jako autoři v původním článku, a to pro horní mez:

$$\text{horní mez} = e^{(2 \cdot \beta_x \cdot se_{k_t})}, \quad (45)$$

a pro dolní mez:

$$\text{dolní mez} = e^{(-2 \cdot \beta_x \cdot se_{k_t})}, \quad (46)$$

kde:

β_x je parametr odhadnutý Lee-Carterovým modelem,

se_{k_t} jsou standardní chyby při odhadu parametru k_t modelu ARIMA(0, 1, 0).

Hodnoty se_{k_t} jsou zobrazeny v tabulkách 14 a 15 v příloze 3. Standardních chyby rostou s druhou odmocninou vzdálenosti k předpovědnímu horizontu. [11]

V obou tabulkách 8 a 9 je možné si všimnout, že naše odhadnuté hodnoty ve většině případů nedosahují velikých rozdílů od hodnot reálných. Nyní se podíváme blíže na jednotlivé tabulky. V tabulce 8 vidíme hodnoty pro mužskou a ženskou populaci v letech 2006 – 2010. Bližším porovnáním hodnot skutečných a predikovaných jsme zjistili, že u mužské populace se skutečné hodnoty vešly do námi vytvořeného 95% intervalu spolehlivosti. U žen to ve většině případů bylo stejné jako u mužů, ale reálné hodnoty byly velmi blízké k horní hranici CI 95%. Skutečná hodnota střední délky života 80-ti let v roce 2008 se do odhadnutého 95% intervalu spolehlivosti nevešla, ale pouze o 0,06. I v letech 2009

a 2010 naděje dožití 80-ti let pro ženy přesahovala horní hranici intervalu spolehlivosti opět pouze s malým rozdílem o 0,02.

Tabulka 8: Odhadnuté a skutečné hodnoty naděje dožití pro určité věky v letech 2006 - 2010

Střední délka života	2006							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	40,64	39,91	40,92	38,92	35	34,80	35,81	33,81
50	31,22	30,50	31,50	29,50	26,04	25,89	26,90	24,90
60	22,34	21,62	22,63	20,63	18,19	18,10	19,11	17,11
70	14,29	13,55	14,56	12,56	11,66	11,57	12,58	10,58
80	7,68	6,68	7,69	5,69	6,41	6,17	7,18	5,18
e_x	2007							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	40,83	40,07	41,08	39,08	35,3	35,04	36,05	34,05
50	31,4	30,64	31,65	29,65	26,25	26,05	27,07	25,07
60	22,51	21,76	22,77	20,77	18,41	18,25	19,27	17,27
70	14,41	13,64	14,65	12,65	11,84	11,70	12,71	10,71
80	7,76	6,73	7,74	5,74	6,62	6,30	7,32	5,32
e_x	2008							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	41,07	40,27	41,29	39,29	35,44	35,21	36,22	34,22
50	31,62	30,83	31,84	29,84	26,44	26,24	27,26	25,26
60	22,76	21,96	22,98	20,98	18,57	18,41	19,43	17,43
70	14,64	13,81	14,82	12,82	12,01	11,86	12,88	10,88
80	7,87	6,80	7,81	5,81	6,68	6,33	7,35	5,35
e_x	2009							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	41,1	40,33	41,35	39,35	35,56	35,34	36,36	34,36
50	31,64	30,88	31,90	29,90	26,52	26,34	27,36	25,36
60	22,71	21,95	22,96	20,96	18,58	18,45	19,47	17,47
70	14,58	13,79	14,80	12,80	12,02	11,90	12,92	10,92
80	7,81	6,78	7,79	5,79	6,62	6,36	7,38	5,38
e_x	2010							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	41,39	40,59	41,61	39,61	35,8	35,51	36,53	34,53
50	31,92	31,13	32,14	30,14	26,74	26,49	27,51	25,51
60	22,96	22,17	23,19	21,19	18,77	18,57	19,59	17,59
70	14,82	14,01	15,02	13,02	12,18	11,99	13,01	11,01
80	7,98	6,94	7,96	5,96	6,73	6,35	7,37	5,37

Zdroj: vlastní zpracování podle [4], [6]

V tabulce 9 jsou zobrazeny hodnoty pro muže a ženy v letech 2011 – 2015. Z této tabulky můžeme vyčíst, že skutečné hodnoty naděje dožití pro všechny věkové kategorie mužů se opět vešly do odhadnutého 95% intervalu spolehlivosti. U žen už přesahovalo horní hranici

intervalu spolehlivosti více skutečných hodnot. Jednalo se zejména zase o střední délku života 80-ti letých ve všech letech, tedy od roku 2011 do 2015. V roce 2012 a 2014 se do odhadnutého intervalu spolehlivosti nevešla reálná hodnota pro střední délku života 70-ti let. Ovšem rozdíly skutečných hodnot od hodnot horní hranice 95% intervalu spolehlivosti už byly větší než v případě naděje dožití 80-ti let v roce 2008, 2009 a 2010.

Tabulka 9: Odhadnuté a skutečné hodnoty naděje dožití pro určité věky v letech 2011 - 2015

Střední délka života	2011							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	41,6	40,73	41,75	39,75	36,07	35,79	36,81	34,81
50	32,11	31,25	32,27	30,27	26,96	26,72	27,75	25,75
60	23,14	22,28	23,30	21,30	18,96	18,77	19,79	17,79
70	15	14,09	15,11	13,11	12,34	12,13	13,15	11,15
80	8,1	6,97	7,99	5,99	6,87	6,47	7,50	5,50
e_x	2012							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	41,7	40,82	41,84	39,84	36,29	36,01	37,04	35,04
50	32,21	31,33	32,36	30,36	27,16	26,94	27,97	25,97
60	23,22	22,35	23,38	21,37	19,05	18,88	19,90	17,90
70	15,5	14,15	15,17	13,17	12,40	12,21	13,24	11,24
80	8,14	6,99	8,01	6,01	6,90	6,53	7,56	5,56
e_x	2013							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	41,82	40,94	41,96	39,96	36,47	36,19	37,22	35,22
50	32,29	31,43	32,45	30,45	27,34	27,10	28,13	26,13
60	23,3	22,43	23,45	21,45	19,11	18,94	19,96	17,96
70	15,17	14,25	15,28	13,28	12,41	12,23	13,25	11,25
80	8,21	7,08	8,10	6,10	6,97	6,63	7,66	5,66
e_x	2014							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	42,39	41,37	42,39	40,39	36,95	36,63	37,66	35,66
50	32,86	31,85	32,88	30,88	27,80	27,52	28,55	26,55
60	23,8	22,80	23,83	21,83	19,52	19,29	20,32	18,32
70	15,56	14,52	15,54	13,54	12,74	12,49	13,52	11,52
80	8,52	7,21	8,24	6,24	7,17	6,73	7,76	5,75
e_x	2015							
	Ženy				Muži			
	Realita	Predikce	95 % CI		Realita	Predikce	95 % CI	
40	42,12	41,22	42,25	40,25	37,08	36,63	37,66	35,66
50	32,57	31,69	32,71	30,71	27,90	27,49	28,52	26,52
60	23,51	22,62	23,65	21,65	19,58	19,21	20,24	18,24
70	15,26	14,34	15,37	13,37	12,81	12,38	13,41	11,41
80	8,24	7,10	8,13	6,12	7,49	6,74	7,77	5,77

Zdroj: vlastní zpracování podle [4], [6]

Shrnutím lze říci, že námi vypočítané hodnoty pomocí Lee-Carter modelu úmrtnosti a predikcí parametru k_t ARIMA modelem byly odhadnuté v rámci možností ve většině případů dobře. Díky těmto hodnotám jsme vypočítali predikci hodnot pro naději dožití vybraných věkových kategorií v letech 2006 – 2015 a následně porovnali se skutečnými hodnotami, které se většinou vešly mezi horní a dolní hranici 95% intervalu spolehlivosti. Pouze v případě ženské populace u věkové kategorie 80 let nám většina hodnot přesahovala horní hranici intervalu a ve dvou letech i pro střední délku života 70-ti let.

ZÁVĚR

Cílem diplomové práce bylo analyzovat vývoj úmrtnosti pro Českou republiku zvlášť pro muže a ženy ve věkové kategorii 40 – 90 let v období 1965 – 2005 pomocí Lee-Carterova modelu úmrtnosti.

V první kapitole byly charakterizovány modely, díky kterým je možné sledovat vývoj úmrtnosti. V následující kapitole byl podrobně popsán Lee-Carter model úmrtnosti, metody odhadů parametrů a Boxova-Jenkinsonova metodologie využitá pro modelování indexu úmrtnosti.

Kapitola 3 se zaměřila na modelování úmrtnosti prostřednictvím Lee-Carter modelu úmrtnosti v ČR pro vybrané věkové kategorie a období. Ze začátku kapitoly došlo k charakteristice vybraných dat pomocí střední délky života při narození, odhadnutí jednotlivých parametrů metodou nejmenších čtverců. Z veškerých hodnot byly zobrazeny grafy i tabulky, které byly vzájemně porovnány pro ženy a muže zvlášť ve věkové kategorii 40-ti až 90-ti let ve sledovaném období 1965 – 2005. Následně byla provedena analýza časově závislého parametru, indexu úmrtnosti, Boxovou-Jenkinsonovou metodologií, kdy nejvhodnějším modelem byla zvolena náhodná procházka s driftem jak pro muže a ženy. Tento parametr byl prognózován do roku 2015.

Na konci kapitoly byly pomocí predikovaného parametru vypočítány nové míry úmrtnosti do roku 2015, díky kterým se vypočítala střední délka života pro věky 40-ti, 50-ti, 60-ti, 70-ti a 80-ti let, zvlášť pro muže a ženy. Pro tyto hodnoty byly opět sestaveny grafy a tabulky, které jsme porovnali pro mužskou a ženskou populaci. V závěru jsou tyto predikované hodnoty porovnány s reálnými hodnotami získanými z *Human mortality database* a *Českého statistického úřadu*. Během porovnání těchto hodnot bylo zjištěno, že odhadnuté hodnoty se tolik nelišily od hodnot skutečných a v nejvíce případech se skutečné hodnoty vešly mezi horní a dolní hranici odhadnutého 95% intervalu spolehlivosti. Pouze s výjimkou střední délky života 80-ti let pro ženy, kdy většina hodnot přesahovala horní hranici intervalu spolehlivosti, a to v letech 2008 – 2010, 2011 – 2015. V roce 2012 a 2014 hodnota střední délky života 70-ti let pro ženy také přesáhla horní hranici 95% intervalu spolehlivosti.

Lze tedy říci, že stárnutí populace, nebo-li riziko dlouhověkosti je problémem i České republiky, který by měla vláda řešit a to nejen proto, že to je měřítkem vyspělosti dané společnosti, ale také že dochází k ovlivnění některých odvětví, např. ekonomika, školství, zdravotnictví.

POUŽITÁ LITERATURA

- [1] BROUHNS, N., DENUIT, M., VERMUNT, J. K., *A Poisson Log- Bilinear Regression Approach to the Construction of Protect Life Tables*, Insurance: Mathematics and Economics 31, 2002, 373-393.
- [2] CIPRA, T. *Finanční ekonometrie*. 2. upravené vydání. Praha: Ekopress, 2013. 538 s. ISBN 978-80-86929-93-4.
- [3] CIPRA, T. *Pojistná matematika: teorie a praxe*. 2. aktualiz. vydání. Praha: Ekopress, 2006. 411 s. ISBN 80-869-2911-6.
- [4] ČESKÝ STATISTICKÝ ÚŘAD. *Úmrtnostní tabulky 2015* [online]. [cit. 2017-04-21]. Dostupné z: <<https://www.czso.cz/csu/czso/umrtnostni-tabulky-za-cr-regiony-soudrznosti-a-kraje>>
- [5] HABERMAN, S., RUSSOLILLO, M. *Lee-Carter mortality forecasting: application to the Italian population*. Cass Business School London, Faculty of Actuarial Science and Statistics, 2005. 28 s. Actuarial Research Paper No. 167. ISBN 1-901615-93-6.
- [6] HUMAN MORTALITY DATABASE. *Data series for Czech republic*. University of California, Berkeley (USA), Max Planck Institute for Demographic Research.[online] [cit. 2017-04-15] Dostupné z: <<http://www.mortality.org/cgi-bin/hmd/country.php?cntr=CZE&level=1>>
- [7] JINDROVÁ P., SLAVÍČEK, O. *Vývoj a predikce střední délky života ve vybraných evropských zemích*. 6-th International Scientific Conference Managing and Modelling of Financial Risk proceedings. VŠB – TU Ostrava, 2012. 303 – 312. ISBN: 978-80-248-2835-0.
- [8] KOSCHIN F. *Aktuárská demografie (úmrtnost a životní pojištění)*. Praha: Vysoká škola ekonomická, 1997. 124 s. ISBN 80-7079-112-8.
- [9] KOSCHIN, F. *Vybrané demografické modely*. Praha: Vysoká škola ekonomická, 1995. 49 s. ISBN 80-7079-761-4.
- [10] KRÁTKÝ, M. *Využití SVD pro indexování latentní sémantiky*. [online] 2002. Technická zpráva. VŠB - TU Ostrava, Katedra informatiky. [cit. 2017-03-17] Dostupné z: <http://www.cs.vsb.cz/arg/techreports/lsi-svd_ma.pdf>
- [11] LEE, R. D., CARTER, L. *Modeling and Forecasting U. S. Mortality*. Journal of the American Statistical Association, vol. 87. 1992. 659 – 671.

- [12] MORGAN, J. P. *LifeMetrics: A toolkit for measuring and managing longevity and mortality risks*. 2007 [online]. [cit. 2017-03-13]. Dostupné z:
<https://www.jpmorgan.com/cm/BlobServer/lifemetrics_technical.pdf?blobkey=id&blobwhere=1158472448701&blobheader=application/pdf&blobheadername1=Cache-Control&blobheadervalue1=private&blobcol=urldata&blobtable=MungoBlobs>
- [13] PITACCO, E., DENUIT, M., HABERMAN, S., OLIVIERI, A.M. *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press, 2009. 395 s. ISBN 978-0-19-954727-2.
- [14] SLAVÍČEK O., JINDROVÁ P. *Využití Lee-Carterova modelu pro predikci Střední délky života*. Univerzita Pardubice – FES. 2013.

SEZNAM PŘÍLOH

Příloha 1: Odhadnuté míry úmrtnosti $m_{x,t}$ pro období 2006 -2015

Příloha 2: Skutečné hodnoty střední délky života pro vybrané věky a období 1965 – 2015

Příloha 3: Hodnoty standardních chyb při odhadu parametru k_t ARIMA modelem

PŘÍLOHA 1: Odhadnuté míry úmrtnosti $m_{x,t}$ pro období 2006 -2015

Tabulka 10: Odhadnuté hodnoty míry úmrtnosti $m_{x,t}$ pro ženy v období 2006 - 2015

věk/Rok	$m_{x,t}$									
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
40	0,0010	0,0010	0,0009	0,0009	0,0009	0,0009	0,0009	0,0009	0,0009	0,0009
41	0,0011	0,0010	0,0010	0,0010	0,0010	0,0010	0,0010	0,0010	0,0010	0,0009
42	0,0012	0,0012	0,0012	0,0012	0,0012	0,0012	0,0012	0,0011	0,0011	0,0011
43	0,0013	0,0013	0,0013	0,0013	0,0013	0,0013	0,0013	0,0012	0,0012	0,0012
44	0,0015	0,0015	0,0015	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0013
45	0,0017	0,0017	0,0016	0,0016	0,0016	0,0016	0,0016	0,0015	0,0015	0,0015
46	0,0019	0,0019	0,0019	0,0019	0,0018	0,0018	0,0018	0,0018	0,0018	0,0018
47	0,0021	0,0021	0,0021	0,0021	0,0020	0,0020	0,0020	0,0020	0,0020	0,0019
48	0,0024	0,0024	0,0024	0,0023	0,0023	0,0023	0,0023	0,0023	0,0022	0,0022
49	0,0026	0,0026	0,0025	0,0025	0,0025	0,0025	0,0024	0,0024	0,0024	0,0024
50	0,0028	0,0027	0,0027	0,0027	0,0027	0,0026	0,0026	0,0026	0,0026	0,0025
51	0,0031	0,0031	0,0031	0,0030	0,0030	0,0030	0,0030	0,0029	0,0029	0,0029
52	0,0032	0,0032	0,0032	0,0031	0,0031	0,0031	0,0030	0,0030	0,0030	0,0029
53	0,0038	0,0037	0,0037	0,0036	0,0036	0,0036	0,0035	0,0035	0,0035	0,0034
54	0,0040	0,0040	0,0040	0,0039	0,0039	0,0038	0,0038	0,0038	0,0037	0,0037
55	0,0045	0,0045	0,0045	0,0044	0,0044	0,0043	0,0043	0,0042	0,0042	0,0042
56	0,0048	0,0048	0,0047	0,0047	0,0046	0,0046	0,0045	0,0045	0,0044	0,0044
57	0,0056	0,0055	0,0055	0,0054	0,0054	0,0053	0,0053	0,0052	0,0052	0,0051
58	0,0059	0,0058	0,0058	0,0057	0,0056	0,0056	0,0055	0,0055	0,0054	0,0053
59	0,0065	0,0064	0,0064	0,0063	0,0062	0,0062	0,0061	0,0060	0,0060	0,0059
60	0,0071	0,0070	0,0070	0,0069	0,0068	0,0067	0,0066	0,0066	0,0065	0,0064
61	0,0077	0,0076	0,0075	0,0074	0,0073	0,0072	0,0072	0,0071	0,0070	0,0069
62	0,0083	0,0082	0,0081	0,0080	0,0079	0,0078	0,0077	0,0076	0,0075	0,0075
63	0,0093	0,0092	0,0091	0,0090	0,0089	0,0088	0,0086	0,0085	0,0084	0,0083
64	0,0101	0,0100	0,0099	0,0097	0,0096	0,0095	0,0094	0,0092	0,0091	0,0090
65	0,0114	0,0113	0,0111	0,0110	0,0108	0,0107	0,0106	0,0104	0,0103	0,0101
66	0,0127	0,0126	0,0124	0,0122	0,0121	0,0119	0,0118	0,0116	0,0114	0,0113
67	0,0141	0,0139	0,0137	0,0135	0,0133	0,0132	0,0130	0,0128	0,0126	0,0125
68	0,0158	0,0156	0,0154	0,0152	0,0150	0,0148	0,0146	0,0144	0,0142	0,0140
69	0,0175	0,0173	0,0170	0,0168	0,0166	0,0164	0,0161	0,0159	0,0157	0,0155
70	0,0194	0,0192	0,0189	0,0186	0,0184	0,0181	0,0179	0,0176	0,0174	0,0171
71	0,0220	0,0217	0,0214	0,0211	0,0208	0,0205	0,0202	0,0200	0,0197	0,0194
72	0,0248	0,0245	0,0242	0,0239	0,0235	0,0232	0,0229	0,0226	0,0223	0,0220
73	0,0282	0,0278	0,0274	0,0271	0,0267	0,0264	0,0260	0,0257	0,0253	0,0250
74	0,0312	0,0308	0,0304	0,0300	0,0296	0,0291	0,0287	0,0284	0,0280	0,0276
75	0,0355	0,0350	0,0346	0,0341	0,0336	0,0332	0,0327	0,0323	0,0319	0,0314
76	0,0405	0,0400	0,0395	0,0390	0,0385	0,0380	0,0375	0,0370	0,0365	0,0361
77	0,0452	0,0446	0,0441	0,0435	0,0429	0,0423	0,0418	0,0412	0,0407	0,0402
78	0,0526	0,0519	0,0513	0,0507	0,0501	0,0495	0,0489	0,0483	0,0477	0,0471
79	0,0577	0,0569	0,0562	0,0555	0,0548	0,0541	0,0534	0,0527	0,0521	0,0514
80	0,0665	0,0657	0,0649	0,0641	0,0633	0,0626	0,0618	0,0611	0,0603	0,0596
81	0,0747	0,0738	0,0729	0,0721	0,0712	0,0704	0,0696	0,0688	0,0680	0,0672
82	0,0852	0,0843	0,0833	0,0824	0,0815	0,0805	0,0796	0,0787	0,0779	0,0770
83	0,0970	0,0960	0,0950	0,0940	0,0930	0,0920	0,0911	0,0901	0,0892	0,0882
84	0,1069	0,1058	0,1047	0,1036	0,1024	0,1014	0,1003	0,0992	0,0981	0,0971
85	0,1229	0,1217	0,1205	0,1194	0,1182	0,1171	0,1159	0,1148	0,1137	0,1126
86	0,1378	0,1366	0,1354	0,1341	0,1329	0,1317	0,1305	0,1293	0,1282	0,1270
87	0,1548	0,1535	0,1522	0,1508	0,1495	0,1482	0,1469	0,1456	0,1443	0,1431
88	0,1747	0,1733	0,1720	0,1706	0,1693	0,1680	0,1666	0,1653	0,1640	0,1627
89	0,1970	0,1956	0,1943	0,1929	0,1916	0,1902	0,1889	0,1876	0,1863	0,1850
90	0,2231	0,2218	0,2205	0,2192	0,2180	0,2167	0,2155	0,2142	0,2130	0,2117

Zdroj: vlastní zpracování

Tabulka 11: Odhadnuté hodnoty míry úmrtnosti $m_{x,t}$ pro muže v období 2006 - 2015

věk/Rok	$m_{x,t}$									
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
40	0,0022	0,0022	0,0021	0,0021	0,0021	0,0021	0,0021	0,0020	0,0020	0,0020
41	0,0024	0,0024	0,0023	0,0023	0,0023	0,0023	0,0023	0,0022	0,0022	0,0022
42	0,0027	0,0027	0,0027	0,0026	0,0026	0,0026	0,0026	0,0026	0,0025	0,0025
43	0,0030	0,0030	0,0030	0,0029	0,0029	0,0029	0,0029	0,0028	0,0028	0,0028
44	0,0034	0,0033	0,0033	0,0033	0,0033	0,0032	0,0032	0,0032	0,0032	0,0031
45	0,0037	0,0036	0,0036	0,0036	0,0035	0,0035	0,0035	0,0035	0,0034	0,0034
46	0,0044	0,0043	0,0043	0,0043	0,0043	0,0042	0,0042	0,0042	0,0041	0,0041
47	0,0048	0,0048	0,0048	0,0047	0,0047	0,0047	0,0046	0,0046	0,0046	0,0045
48	0,0054	0,0054	0,0053	0,0053	0,0053	0,0052	0,0052	0,0052	0,0051	0,0051
49	0,0060	0,0060	0,0059	0,0059	0,0058	0,0058	0,0058	0,0057	0,0057	0,0057
50	0,0066	0,0066	0,0065	0,0065	0,0065	0,0064	0,0064	0,0063	0,0063	0,0062
51	0,0072	0,0071	0,0070	0,0070	0,0069	0,0069	0,0068	0,0068	0,0067	0,0067
52	0,0081	0,0080	0,0080	0,0079	0,0079	0,0078	0,0078	0,0077	0,0077	0,0076
53	0,0090	0,0090	0,0089	0,0089	0,0088	0,0088	0,0087	0,0087	0,0086	0,0085
54	0,0099	0,0098	0,0097	0,0097	0,0096	0,0095	0,0095	0,0094	0,0094	0,0093
55	0,0110	0,0109	0,0108	0,0108	0,0107	0,0106	0,0106	0,0105	0,0104	0,0104
56	0,0115	0,0114	0,0114	0,0113	0,0112	0,0111	0,0110	0,0110	0,0109	0,0108
57	0,0125	0,0124	0,0123	0,0122	0,0121	0,0120	0,0120	0,0119	0,0118	0,0117
58	0,0139	0,0138	0,0137	0,0136	0,0135	0,0134	0,0133	0,0132	0,0131	0,0130
59	0,0155	0,0154	0,0153	0,0152	0,0151	0,0150	0,0149	0,0147	0,0146	0,0145
60	0,0168	0,0167	0,0166	0,0165	0,0163	0,0162	0,0161	0,0160	0,0159	0,0157
61	0,0180	0,0178	0,0177	0,0175	0,0174	0,0173	0,0171	0,0170	0,0169	0,0167
62	0,0197	0,0195	0,0194	0,0192	0,0191	0,0189	0,0188	0,0186	0,0185	0,0183
63	0,0212	0,0211	0,0209	0,0207	0,0205	0,0204	0,0202	0,0200	0,0199	0,0197
64	0,0224	0,0222	0,0220	0,0218	0,0216	0,0214	0,0212	0,0210	0,0208	0,0207
65	0,0247	0,0244	0,0242	0,0240	0,0238	0,0235	0,0233	0,0231	0,0229	0,0227
66	0,0269	0,0266	0,0264	0,0261	0,0259	0,0256	0,0254	0,0252	0,0249	0,0247
67	0,0294	0,0291	0,0289	0,0286	0,0284	0,0281	0,0279	0,0276	0,0273	0,0271
68	0,0323	0,0320	0,0317	0,0314	0,0311	0,0308	0,0305	0,0303	0,0300	0,0297
69	0,0349	0,0345	0,0342	0,0339	0,0336	0,0332	0,0329	0,0326	0,0323	0,0320
70	0,0380	0,0377	0,0373	0,0370	0,0366	0,0363	0,0360	0,0356	0,0353	0,0350
71	0,0410	0,0406	0,0402	0,0398	0,0394	0,0391	0,0387	0,0383	0,0380	0,0376
72	0,0446	0,0442	0,0438	0,0434	0,0429	0,0425	0,0421	0,0417	0,0413	0,0409
73	0,0489	0,0484	0,0480	0,0475	0,0470	0,0466	0,0461	0,0457	0,0453	0,0448
74	0,0539	0,0534	0,0529	0,0524	0,0519	0,0514	0,0509	0,0505	0,0500	0,0495
75	0,0580	0,0574	0,0569	0,0563	0,0558	0,0553	0,0547	0,0542	0,0537	0,0532
76	0,0647	0,0641	0,0635	0,0629	0,0624	0,0618	0,0612	0,0607	0,0601	0,0596
77	0,0707	0,0700	0,0694	0,0688	0,0681	0,0675	0,0669	0,0663	0,0657	0,0651
78	0,0775	0,0768	0,0761	0,0754	0,0748	0,0741	0,0734	0,0728	0,0721	0,0715
79	0,0833	0,0825	0,0818	0,0810	0,0802	0,0795	0,0787	0,0780	0,0773	0,0766
80	0,0944	0,0936	0,0928	0,0920	0,0912	0,0904	0,0896	0,0888	0,0881	0,0873
81	0,1037	0,1029	0,1020	0,1012	0,1003	0,0995	0,0987	0,0979	0,0971	0,0963
82	0,1164	0,1155	0,1146	0,1137	0,1128	0,1119	0,1110	0,1102	0,1093	0,1085
83	0,1256	0,1247	0,1237	0,1227	0,1217	0,1208	0,1198	0,1189	0,1180	0,1170
84	0,1374	0,1363	0,1353	0,1342	0,1331	0,1321	0,1310	0,1300	0,1290	0,1279
85	0,1525	0,1514	0,1503	0,1492	0,1481	0,1470	0,1459	0,1448	0,1437	0,1427
86	0,1689	0,1677	0,1665	0,1654	0,1642	0,1630	0,1619	0,1607	0,1596	0,1584
87	0,1858	0,1845	0,1832	0,1819	0,1806	0,1794	0,1781	0,1769	0,1757	0,1744
88	0,2138	0,2127	0,2116	0,2104	0,2093	0,2082	0,2071	0,2060	0,2050	0,2039
89	0,2358	0,2347	0,2337	0,2326	0,2315	0,2305	0,2294	0,2284	0,2273	0,2263
90	0,2580	0,2569	0,2557	0,2545	0,2533	0,2522	0,2510	0,2499	0,2487	0,2476

Zdroj: vlastní zpracování

PŘÍLOHA 2: Skutečné hodnoty střední délky života pro vybrané věky a období 1965 – 2015

Tabulka 12: Reálné hodnoty naděje dožití žen v období 1965 – 2015

Rok	e_x				
	40	50	60	70	80
1965	36,13	26,95	18,42	11,1	5,89
1966	36,34	27,12	18,58	11,21	5,98
1967	36,31	27,11	18,54	11,2	5,91
1968	35,99	26,82	18,27	10,92	5,69
1969	35,78	26,63	18,13	10,91	5,83
1970	35,61	26,43	17,97	10,74	5,67
1971	35,87	26,71	18,22	10,97	5,84
1972	36,12	26,96	18,47	11,16	5,94
1973	36,03	26,83	18,32	10,98	5,75
1974	35,95	26,78	18,27	10,93	5,71
1975	36,31	27,7	18,54	11,13	5,86
1976	36,39	27,15	18,61	11,22	5,85
1977	36,4	27,16	18,63	11,19	5,83
1978	36,36	27,1	18,6	11,23	5,91
1979	36,37	27,14	18,64	11,27	5,93
1980	35,92	26,68	18,18	10,88	5,65
1981	36,32	27,7	18,56	11,21	5,88
1982	36,34	27,12	18,53	11,22	5,9
1983	36,17	26,92	18,38	11,7	5,77
1984	36,35	27,7	18,53	11,26	5,86
1985	36,49	27,2	18,64	11,27	5,91
1986	36,38	27,15	18,58	11,28	5,9
1987	36,83	27,55	18,99	11,62	6,9
1988	36,99	27,73	19,15	11,79	6,24
1989	36,89	27,65	19,6	11,68	6,13
1990	37,02	27,76	19,16	11,77	6,26
1991	37,34	28,11	19,51	12,4	6,37
1992	37,75	28,47	19,85	12,34	6,6
1993	37,87	28,58	19,88	12,32	6,59
1994	38,03	28,73	20	12,4	6,56
1995	38,02	28,71	20,5	12,43	6,59
1996	38,54	29,2	20,44	12,72	6,88
1997	38,67	29,35	20,6	12,85	6,91
1998	39,06	29,69	20,93	13,11	7,8
1999	39,12	29,74	20,95	13,11	7,7
2000	39,33	29,98	21,21	13,3	7,12
2001	39,48	30,12	21,35	13,4	7,23
2002	39,71	30,31	21,49	13,51	7,24
2003	39,59	30,2	21,42	13,46	7,11
2004	40,1	30,7	21,83	13,82	7,38
2005	40,16	30,75	21,91	13,87	7,4
2006	40,64	31,22	22,34	14,29	7,68
2007	40,83	31,4	22,51	14,41	7,76
2008	41,07	31,62	22,76	14,64	7,87
2009	41,1	31,64	22,71	14,58	7,81
2010	41,39	31,92	22,96	14,82	7,98
2011	41,6	32,11	23,14	15	8,1
2012	41,7	32,21	23,22	15,5	8,14
2013	41,82	32,29	23,3	15,17	8,21
2014	42,39	32,86	23,8	15,56	8,52
2015	42,12	32,57	23,51	15,26	8,24

Zdroj: vlastní zpracování podle [4], [6]

Tabulka 13: Reálné hodnoty naděje dožití mužů v období 1965 - 2015

Rok	e_x				
	40	50	60	70	80
1965	31,24	22,48	14,83	9,16	5,08
1966	31,27	22,51	14,8	9,12	5,16
1967	31,02	22,32	14,68	9,04	5,05
1968	30,7	22,04	14,43	8,81	4,85
1969	30,13	21,58	14,03	8,55	4,96
1970	30,07	21,5	14,05	8,48	4,87
1971	30,18	21,65	14,14	8,54	4,9
1972	30,64	22,01	14,54	8,85	5
1973	30,37	21,73	14,29	8,54	4,83
1974	30,3	21,7	14,26	8,51	4,81
1975	30,57	21,94	14,48	8,68	4,79
1976	30,56	21,92	14,52	8,62	4,79
1977	30,51	21,92	14,56	8,72	4,88
1978	30,49	21,89	14,61	8,75	4,92
1979	30,49	21,87	14,57	8,76	4,86
1980	30,01	21,46	14,25	8,47	4,59
1981	30,32	21,71	14,47	8,71	4,8
1982	30,37	21,74	14,43	8,71	4,77
1983	30,07	21,54	14,29	8,67	4,75
1984	30,21	21,67	14,43	8,78	4,82
1985	30,24	21,72	14,44	8,79	4,85
1986	30,23	21,67	14,41	8,85	4,82
1987	30,53	21,98	14,69	9,11	5,05
1988	30,66	22,12	14,79	9,22	5,14
1989	30,62	22,06	14,73	9,22	5,02
1990	30,19	21,76	14,58	9,1	5,05
1991	30,81	22,3	14,99	9,33	5,16
1992	31,07	22,57	15,25	9,59	5,37
1993	31,67	23,04	15,64	9,72	5,44
1994	31,85	23,23	15,81	9,85	5,56
1995	31,97	23,31	15,81	9,82	5,48
1996	32,32	23,62	16,15	10,18	5,78
1997	32,47	23,81	16,31	10,3	5,85
1998	32,93	24,2	16,63	10,43	5,94
1999	33,18	24,4	16,79	10,56	6,01
2000	33,34	24,58	16,93	10,68	5,96
2001	33,76	24,95	17,27	10,88	6,09
2002	33,84	25,01	17,3	10,95	6,05
2003	33,8	24,93	17,25	10,86	5,98
2004	34,27	25,37	17,64	11,21	6,19
2005	34,55	25,62	17,83	11,32	6,26
2006	35	26,04	18,19	11,66	6,41
2007	35,3	26,25	18,41	11,84	6,62
2008	35,44	26,44	18,57	12,01	6,68
2009	35,56	26,52	18,58	12,02	6,62
2010	35,8	26,74	18,77	12,18	6,73
2011	36,07	26,96	18,96	12,34	6,87
2012	36,29	27,16	19,05	12,4	6,9
2013	36,47	27,34	19,11	12,41	6,97
2014	36,95	27,8	19,52	12,74	7,17
2015	37,08	27,90	19,58	12,81	7,49

Zdroj: vlastní zpracování podle [4], [6]

PŘÍLOHA 3: Hodnoty standardních chyb při odhadu parametru k_t ARIMA modelem

Tabulka 14: Hodnoty standardních chyb se_{k_t} pro ženy

Rok	Standardní chyba
2006	0,21159
2007	0,29923
2008	0,36648
2009	0,42318
2010	0,47313
2011	0,51828
2012	0,55981
2013	0,59846
2014	0,63476
2015	0,66910

Zdroj: vlastní zpracování

Tabulka 15: Hodnoty standardních chyb se_{k_t} pro muže

Rok	Standardní chyba
2006	0,22723
2007	0,32135
2008	0,39357
2009	0,45445
2010	0,50809
2011	0,55659
2012	0,60118
2013	0,64269
2014	0,68168
2015	0,71855

Zdroj: vlastní zpracování