

Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Demografie
Studijní obor: Demografie



Ing. Petr Mazouch, Ph.D.

GENERAČNÍ ÚMRTNOST A JEJÍ MODELOVÁNÍ
Cohort mortality pattern and its modelling

Diplomová práce

Vedoucí závěrečné práce: RNDr. Klára Hulíková Tesárková, Ph.D.

Praha, 2013

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 19. 8. 2013

.....

Poděkování

Na tomto místě bych velmi rád poděkoval RNDr. Kláře Hulíkové Tesárkové, Ph.D., protože předkládaná práce by nemohla nikdy vzniknout bez jejího obětavého přístupu a obrovské dávky entuziasmu, kterou dávala do každého setkání nebo konzultace, na kterých jsme společně diskutovali možná úskalí a možnosti rozvoje tématu, které se nakonec podařilo přetavit v předkládanou diplomovou práci. Tyto konstruktivní rozhovory byly završeny úžasně pečlivým a profesionálním přístupem při čtení jednotlivých verzí práce. Děkuji.

Práce vznikla v rámci projektu „Generační úmrtnostní tabulky České republiky: data, biometrické funkce a trendy“ Grantové agentury České republiky, č. GAP404/12/0883.

Generační úmrtnost a její modelování

Abstrakt

Cílem práce je představit jednoduchý model predikující míry úmrtnosti založený na modelování generační úmrtnosti, kde je základním předpokladem konstantní úroveň poměrů úmrtnosti mezi dvěma po sobě jdoucími věkovými skupinami pro všechny kohorty. V práci je odvozen základní model, který předpokládá tyto vlastnosti vývoje úmrtnosti a je ověřen na časových řadách Švédska, které jsou dostatečně dlouhé. Výsledky jsou srovnány s empirickými hodnotami již vymřelých kohort a je proveden odhad také pro kohorty dosud žijící. Model při aplikaci na reálných datech vykazuje velmi dobrou schopnost reagovat na změny vývoje. V poslední části práce jsou zmíněny možné změny a modifikace prezentovaného modelu, které by vedly k více stabilním výsledkům.

Klíčová slova: úmrtnost, modelování úmrtnosti, generační úmrtnost, vývoj úmrtnosti, odhad úmrtnosti

Cohort mortality pattern and its modelling

Abstract

The aim of the thesis is to introduce a simple alternative approach of mortality modeling (future estimates of mortality rates) which is based on cohort mortality patterns. The assumption of constant change of the force of mortality between two following ages across cohorts is the fundamental base of the proposed model. Theoretical aspects of the model are discussed and its assumptions are verified on time series of Swedish data which are long enough for this purpose. Results of the model application are compared with empirical results of already extinct cohorts and some estimates for still living cohorts are made too. Applications on real data showed good results which respected changes in the mortality development. In the last part of the thesis some possible modification are discussed which could cause results less variable.

Keywords: mortality, mortality modelling, cohort mortality, mortality pattern, mortality estimate

Obsah

Přehled použitých zkratk	7
Seznam tabulek	8
Seznam obrázků	9
1. Úvod	11
1.1 Cíle práce	12
1.2 Struktura práce	12
1.3 Diskuse s literaturou	13
2. Zdrojová data a diskuse metodologie modelování úmrtnosti	17
2.1 Zdrojová data	17
2.2 Metodologie modelování úmrtnosti	18
2.2.1 Diskuse vybraných přístupů modelování úmrtnosti	18
2.2.2 Popis základních principů navržené alternativní metody predikce	24
2.2.3 Konstrukce úmrtnostních tabulek	31
2.2.4 Vybrané vlastnosti časových řad	32
2.2.5 Shrnutí metodických přístupů	34
3. Analýza dat a výsledky aplikace navržené metodologie predikce úmrtnosti	35
3.1 Analýza vývoje měř úmrtnosti České republiky a Švédska	35
3.2 Analýza vývoje poměrů a rozdílů měř úmrtnosti	39
3.3 Výsledky aplikace navrhované metodologie predikce úmrtnosti	45
3.4 Shrnutí analýzy dat a výsledků aplikace navržené metodologie predikce úmrtnosti	56
4. Diskuse silných a slabých stránek navrhovaného modelu a možnosti jeho dalšího rozvoje	57
4.1 Variabilita výsledků navrhovaného modelu	57
4.1.1 Alternativy odhadu základní úmrtnosti $bm_{x,z}$	58

4.1.2 Vyrovnávání odhadovaných měr úmrtnosti	60
4.2 Další možné modifikace navrhovaného modelu	62
4.3 Celkové zhodnocení navrhovaného modelu	63
5. Závěr	64
Přílohy.....	71

Přehled použitých zkratk

ACF	Autokorelační funkce
APC	Age-period-cohort model
HMD	Human Mortality Database
AP	Age-period model
AC	Age-cohort model
ČSÚ	Český statistický úřad
GM	Gompertz-Makeham model
LC	Lee-Carter model
MNČ	Metoda nejmenších čtverců
PACF	Parciální autokorelační funkce
PC	Period-cohort model
SVD	Singular Value Decomposition

Seznam tabulek

Tab. 1. Odhad budoucích specifických měr úmrtnosti a střední délky života, srovnání empirickými hodnotami, kohorta 1900, muži, Švédsko	46
Tab. 2. Odhad budoucích specifických měr úmrtnosti a střední délky života, kohorta 1946, Česká republika	55

Seznam obrázků

Obr. 1: Lexisův diagram, červeně označen 1. hlavní soubor událostí $D_{x,z}$ a zeleně exponovaný počet obyvatel $P_{x,z}$	18
Obr. 2: Ilustrace LC modelu – odhad budoucích hodnot měř úmrtnosti v roce 2020 na základě vývoje měř úmrtnosti do roku 2013	21
Obr. 3: Ilustrace principu metodického postupu odhadu kohortních měř úmrtnosti (kohorta 1953).....	26
Obr. 4: Predikce měř úmrtnosti od věku 60 let – ilustrativní příklad.....	30
Obr. 5: Kompletní schéma modelu predikce kohortních měř úmrtnosti.....	30
Obr. 6: Korelogram funkce ACF a PACF.....	33
Obr. 7: Míry úmrtnosti, transversální pohled, 3. hlavní soubory událostí, 1950–2008, 30–95 let, Česká republika, ženy	36
Obr. 8: Míry úmrtnosti, transversální pohled, 3. hlavní soubory událostí, 1950–2008, 30–95 let, Česká republika, ženy – jiná rotace.....	37
Obr. 9: Míry úmrtnosti, kohortní pohled, kohorty 1780, 1850, 1914 a 1930, 30–95 let, 1. hlavní soubory událostí, Švédsko, muži	37
Obr. 10: Vývoj specifických měř úmrtnosti ve věku 50, 60, 70, 80 a 90 let, 1. hlavní soubory událostí, kohorty 1780–1959, Švédsko, muži.....	38
Obr. 11: Rozdíly měř úmrtnosti $d_{x,z}$, kohorty 1871–1977, 30–95 let, 1. hlavní soubory událostí, Česká republika, ženy.....	39
Obr. 12: Rozdíly měř úmrtnosti $d_{x,z}$, kohorty 1780, 1850, 1914 a 1930, 30–95 let, 1. hlavní soubory událostí Švédsko, muži	40
Obr. 13: Rozdíly měř úmrtnosti $d_{x,z}$, pro $x = 50, 60, 70, 80$ a 90 let, 1. hlavní soubory událostí, kohorty 1780–1959, Švédsko, muži	40
Obr. 14: Poměry měř úmrtnosti $ar_{x,z}$, kohorty 1871–1977, 30–95 let, 1. hlavní soubory událostí, Česká republika, ženy.....	41
Obr. 15: Poměry měř úmrtnosti $ar_{x,z}$, kohorty 1780, 1850, 1914 a 1930, 30–95 let, 1. hlavní soubory událostí, Švédsko, muži	42
Obr. 16: Poměry měř úmrtnosti $ar_{x,z}$, pro $x = 50, 60, 70, 80$ a 90 let, 1. hlavní soubory událostí, kohorty 1780–1949, 1. hlavní soubory událostí, Švédsko, muži.....	42
Obr. 17: Odhad ACF a PACF pro časovou řadu $ar_{x,z}$, pro $x = 60$, kohorty 1780–1949, Švédsko, muži	43

Obr. 18: Průměrné hodnoty $ar_{x,z}$ a směrodatné odchylky $ar_{x,z}$, věk 30–95 let, kohorty 1780–1979, Švédsko	44
Obr. 19: Srovnání empirických a odhadovaných hodnot $m_{x,z}$, kohorta 1780, Švédsko, ženy	47
Obr. 20: Srovnání empirických a odhadovaných hodnot $\ln(m_{x,z})$ kohorta 1880, Švédsko, muži	48
Obr. 21: Srovnání empirických a odhadovaných hodnot střední délky života v přesném věku 60 let, kohorty 1761–1945, Švédsko, muži	49
Obr. 22: Srovnání empirických a odhadovaných hodnot střední délky života 60leté osoby, kohorty 1761–1945, Švédsko, ženy	50
Obr. 23: Relativní odchylka mezi empirickou a odhadovanou hodnotou kohortní střední délky života 60leté osoby, kohorty 1761–1912, Švédsko, v %	52
Obr. 24: Vývoj specifických měr úmrtnosti podle věku, 1. hlavní soubory událostí, empirické a odhadované hodnoty, kohorta 1946, Česká republika, muži	53
Obr. 25: Vývoj specifických měr úmrtnosti podle věku, 1. hlavní soubory událostí, empirické a odhadované hodnoty, kohorta 1946, Česká republika, ženy	54
Obr. 26: Srovnání vývoje kohortní střední délky života 60leté osoby a odhadu základní míry úmrtnosti $bm_{x,z}$, kohorty 1761–1945, Švédsko, ženy	58
Obr. 27: Vývoj empirických hodnot $\ln(m_{x,z})$ pro $x = 65$ a 66 , kohorty 1761–1943, Švédsko, ženy, data proložena regresní funkcí (polynomem), odhad budoucích hodnot do roku 1953	61

Kapitola 1

Úvod

Zkoumání úmrtnostních zákonů je základním demografickým tématem od samotného počátku formování demografie jako vědní disciplíny. Již v 17. století, v jednom z prvních děl, které se dají označit jako analytické, představil John Graunt model úmrtnosti londýnského obyvatelstva (Roubíček, 1997).

V dnešní době je zkoumání úmrtnosti a jejích pravidel (či nepravidelností) ještě více akcentováno díky napojení na různé pojistné produkty nebo na sociální politiku státu. Jen těžko si dovedeme představit pojistný produkt (pojištění smrti nebo dožití), který neuvažuje s vývojem úmrtnosti a neanalyzuje současný a neodhaduje budoucí vývoj úmrtnostních poměrů. Také formování sociální politiky je např. v případě systému výplaty penzí (tedy jakési doživotní anuity) závislé na vývoji úmrtnosti. Oblastí, kde je nutné uvažovat úroveň úmrtnosti (ať již současnou nebo budoucí) je však mnohem více.

Možnost důkladné analytické práce s ukazateli úmrtnosti je (na rozdíl od minulosti) možná díky dvěma skutečnostem. První je existence, v rozvinutých zemích dlouhodobě, dobré evidence událostí (úmrť) a stavu obyvatelstva. Ta umožňuje sledovat ukazatele v dlouhodobém horizontu a rozšiřuje tak možnosti analýzy např. o longitudinální (kohortní) přístupy, které by v případě krátkodobého počtu pozorování nebylo možné v praxi použít.

Druhým faktorem, který výrazně zvyšuje možnosti analýzy je rozvoj výpočetní techniky. Některé metody jsou matematicky velmi složité a nalezení řešení vyžaduje dobré zázemí z hlediska výpočetní techniky, což v minulosti nebylo vždy samozřejmostí (a v historii vzdálenější než 50 let to bylo prakticky dokonce nemožné).

Výše zmíněné důvody vedly k tomu, že se v problematice zkoumání vývoje úmrtnosti v posledních letech objevuje celá řada přístupů, více či méně matematicky složitých (a uvažujících různé časové přístupy), jejichž pouhý výčet by vystačil na obsáhlou publikaci. Matematická složitost však často vede k obtížím při aplikaci, předkládaná práce se proto snaží představit alternativní přístup k analyzování (a predikování) úmrtnosti v kohortním pohledu s využitím co nejjednoduššího výpočetního aparátu, tak aby byla metoda přístupná co nejširší odborné veřejnosti.

1.1 Cíle práce

Základním cílem předkládané diplomové práce je navržení takové metody odhadu budoucího vývoje úrovně úmrtnosti ve vyšších věcích, která by byla založená na kohortním pohledu a mohla představovat jistou alternativu k dnes nejčastěji používaným metodám predikce úmrtnosti. S tímto základním (obecným) cílem je spojeno několik (konkrétních) dílčích cílů, které je nutné před dosažením cíle základního naplnit.

Pro naplnění hlavního cíle, tedy nalezení jednoduché metody predikce úmrtnosti v kohortním pohledu, je prvotním dílčím cílem především analýza minulých trendů úmrtnosti kohort, kde jsou dostupná data. Uvažovány budou jen věky nad 60 let, čímž je částečně eliminován problém nedostupnosti dat a zároveň respektována potřeba predikce především v těchto věcích v souvislosti např. s udržitelností sociálních systémů. Analýze budou podrobeny nejen specifické míry úmrtnosti, ale i jejich podíly (koeficienty růstu), neboť na základě studia dosud známých a užívaných postupů je předpokládáno, že právě koeficienty růstu by mohly sloužit jako vhodný základ navrhovaného postupu.

Nutnou podmínkou pro využití koeficientů růstu měr úmrtnosti mezi jednotlivými věky v rámci projekčního modelu je nalezení určitého trendu nebo jinak predikovatelného vývoje v minulých letech. Nalezení a ověření této vlastnosti je tedy dalším dílčím cílem práce. Za tímto účelem budou využita data za Švédsko, protože se jedná o zemi, za kterou existují nejdelší časové řady dat, které je možné pro analýzu využít. Pro ilustraci budou použita i data za ČR disponující však mnohem kratší časovou řadou vstupních dat.

Pokud se podaří na základě naplnění výše zmíněných podmínek formulovat alternativní přístup k odhadu kohortní úmrtnosti, bude dílčím cílem také ověření jeho vhodnosti na empirických datech (na datech za již vymřelé kohorty).

Vedlejším (nikoliv dílčím) efektem celé práce je možnost využití diskutovaných metod obecně pro analýzu a ověření na jiných než švédských datech a využití navrhovaných modifikací a dalších teoretických možností rozvoje v práci navržené metody, které jsou diskutovány ve čtvrté kapitole.

1.2 Struktura práce

V rámci úvodní části bude ještě provedena stručná diskuse s literaturou, která se modelováním úmrtnosti zabývá, s důrazem na literaturu, která se věnuje kohortnímu přístupu. Zmíněny budou však také v současnosti populární metody predikce, které se využívají v transverzálním přístupu.

Druhá kapitola obsahuje popis datových zdrojů, jejichž data jsou analyzována ve třetí části. Stěžejní částí druhé kapitoly je metodický popis navrženého postupu odhadu kohortní úmrtnosti včetně zdůvodnění jednotlivých kroků. Součástí je také srovnání s přístupy zmíněnými v diskusi s literaturou v úvodní kapitole. Z důvodu analýzy časových řad je součástí také stručný popis metodiky pro testování přítomnosti některých trendů časových řad.

Třetí kapitola obsahuje aplikaci a popis výsledků modelu, které byly spočteny na základě metodiky z předchozí kapitoly, aplikované na data Švédska (a z části České republiky). Kapitola

obsahuje ověření teoretických předpokladů modelu z předchozí kapitoly. Výsledky jsou ilustrovány na výstupech z kohortních úmrtnostních tabulek pro věky nad 60 let, které je možné analyzovat v čase. V závěru kapitoly je srovnání výsledků modelu s empirickými výsledky již vymřelých kohort, opět za využití funkcí z úmrtnostních tabulek.

Čtvrtá kapitola se věnuje diskusi výsledků a možných způsobů rozvoje navržené metodiky. V jednotlivých krocích jsou rozebrány silné a slabé stránky metodiky, možné modifikace a další možnosti využití přístupů z ostatních metod.

1.3 Diskuse s literaturou

V oblasti problematiky modelování ukazatelů úmrtnosti existuje celá řada různých přístupů. Nejprve je dobré, pro potřeby této práce, rozdělit problematiku do dílčích skupin, se kterými se můžeme při modelování setkat. První skupinou, kterou je možné označit jako historicky nejstarší, jsou metody, které se věnují zkoumání zákonů úmrtnosti na transverzálních datech, zejména prostřednictvím modelu úmrtnostní tabulky, kde se snaží o odhad vývoje úmrtnosti v závislosti na věku a v případě vyšších věků o její vyrovnání (podrobnější popis i s odkazy na relevantní zdroje následuje dále).

Druhou skupinu metod lze vymezit jako modely zabývající se predikcí funkcí úmrtnostní tabulky (zejména střední délky života nebo specifických pravděpodobností úmrtí), které uvažují opět pouze transverzální časové hledisko. Třetí skupina pak zahrnuje metody, které do analýzy zahrnují také vliv kohorty.

Protože by některé přístupy bylo dobré podrobněji rozebrat i z matematického hlediska, budou tyto zahrnuty také do následující kapitoly zabývající se metodikou. V následujícím textu se tedy omezíme pouze na slovní hodnocení modelů, matematické vyjádření některých důležitých modelů bude součástí následující kapitoly.

Řada modelů vznikla v době, kdy jejich ověření (či použití) bylo téměř nemožné. Když v roce 1825 publikoval Gompertz (1825) svůj článek, kde vysvětloval vývoj úmrtnosti jako přirozený exponenciální růst, kdy je funkce úrovně úmrtnosti závislá pouze na věku, bylo možné ověřit tuto úvahu pouze na transverzálních datech, a to ještě velmi omezeně z důvodu absence jakéhokoliv technického nástroje (reálně ověřil Gompertz svou teorii i kohortně – na populaci hmyzu – viz publikaci Gavrilova a Gavrilovové (1991)).

Podobně, když v roce 1860 Gompertzovu teorii, která byla založena na přímé souvislosti mezi úmrtností a věkem, rozvinul Makeham, jenž do vztahu přidal konstantu, která vyjadřovala část úmrtnosti nezávislou na věku (Makeham, 1860), šlo spíše o teoretické úvahy než o jasně kvantifikované modely s odhady parametrů (nalezení parametrů funkce prostřednictvím King-Hardyho metody – podrobněji odhad parametrů diskutuje Fiala (2005) – se v dnešní době může jevit jako jednoduché, ale bez výpočetní techniky by bylo téměř nemožné nebo určitě velmi nesnadné).

Kromě deficitu v oblasti technického zázemí, je nutné uvažovat také o dalších limitacích, které aplikaci zmíněných (ale i nezmíněných) modelů znesnadňovaly. Jde zejména o chyby nebo nepřesnosti v evidenci událostí nebo obyvatelstva. Tato skutečnost může, i přes sebelepší

technické zajištění vést k nepřesnostem v modelech. Blíže se tomuto problému věnuje např. Koschin (1981). Význam kvality dat diskutuje také Pastor (2007).

S rozvojem výpočetní techniky a se zvětšujícím se rozsahem disponibilních dat, vhodných k analýze, roste také počet a složitost modelů, které se odborníci snaží využít k odhadu vývoje úmrtnosti. Důraz je kladen zejména na odhad vývoje v nejvyšších věcích, které se jeví jako období, jež úplně nerespektuje původní Gompertzovu teorii o exponenciálním růstu úmrtnosti s věkem (více k diskusi o platnosti Gompertzova modelu v následující kapitole). Z těchto důvodů byly odvozeny další metody – např. modifikovaná Gompertz-Makehamova funkce popsaná Koschinem (1999) nebo dnes velmi populární Kannistova metoda (Thatcher a kol., 1998) používaná např. pro vyrovnávání úmrtností v Human Mortality Database¹ (Wilmoth a kol., 2007).

Přehled vybraných metod, které je možné pro modelování úmrtností využít, byl publikován v článku Burcina a kol. (2010) a pracovníky Katedry demografie a geodemografie Přírodovědecké fakulty Univerzity Karlovy v Praze byla dokonce vyvinuta softwarová aplikace DeRaS – nástroj pro modelování intenzit úmrtnosti a konstrukci úmrtnostních tabulek (Burcin a kol., 2011). Aplikaci vybraných metod vyrovnávajících úmrtnost na datech a další využití např. grafických výstupů lze nalézt také v disertační práci Hulíkové Tesárkové (2012).

Všechny doposud zmíněné metody, které se váží k úmrtnostním tabulkám, mají za úkol spíše interpolaci hodnot (nebo proložení empirických dat nejlépe hladkou funkcí). Výjimkou jsou extrapolace ukazatelů ve vyšších věcích. Predikcí vývoje úmrtnosti v budoucnu se však nevěnuje z těchto metod žádná (tedy k tomu není ani uzpůsobena). Navíc je většina aplikací prováděna na transversálních úmrtnostních tabulkách.

S rozvojem datových zdrojů se rozšiřují také přístupy k analýze úmrtnosti. Rozšiřuje se uplatnění analýz, které využívají také kohortní přístup, tedy longitudinální. Vznikají kohortní úmrtnostní tabulky, což umožňuje hodnotit vývoj úmrtnosti z další perspektivy.

I přes absenci kompletních kohortních úmrtnostních tabulek pro Českou republiku zde byly v literatuře diskutovány některé fragmenty kohortních analýz z teoretické, ale i aplikační stránky. Teoretické modely konstrukce, které uvádí ve své knize např. Pavlík a kol. (1986) jsou zde doplněny také jednou kompletní úmrtnostní tabulkou pro kohortu žen narozenou v roce 1875. Rychtaříková a kol. (1994) analyzuje dostupná kohortní data za období 1950–1990. Koschin (1981) nebo Hulíková Tesárková a Mazouch (2013) diskutují také rozdíly mezi funkcemi úmrtnostních tabulek transversálních a longitudinálních. V posledním jmenovaném je také v příloze seznam zemí, které mají veřejně dostupné kohortní úmrtnostní tabulky.

Kromě klasických úmrtnostních tabulek také vznikají modely, které do faktorů ovlivňujících úroveň úmrtnosti zahrnují kromě věku a období ještě faktor kohorty. Nejznámějším a asi nejpoužívanějším případem jsou takzvané Age-period-cohort modely (také APC modely). Vycházejí z předpokladu, že úroveň úmrtnosti je dána nejenom věkem (úmrtnost se během života bezesporu mění), ale také období (existují více či méně exponovaná období, kdy je úmrtnost vyšší či nižší – viz např. zhoršení úmrtnostních poměrů v 60. letech v ČR – více viz v článku Hulíkové Tesárkové a Mazoucha (2013)) a v neposlední řadě také kohortou. Vliv

¹ Human Mortality Database (HMD) – dostupná z <http://www.mortality.org>

kohorty se dá také definovat jako vliv nepříznivé (nebo příznivé) události, který se projeví v úmrtnostních poměrech v pozdějším věku, což ve svém článku podrobněji rozebrali např. Elo a Preston (1992).

Skutečný vliv období, kterým kohorta v určitém věku prochází, se může projevit různými způsoby. Jedna z teorií předpokládá, že osoby, které v určitém období projdou nepříznivými podmínkami, jsou těmito podmínkami ovlivněni (oslabeni), a to se může negativně promítnout na jejich pozdější úrovni úmrtnosti. Druhá teorie předpokládá tzv. selekční faktor, tedy že se v nepříznivé době vyselektují slabší jedinci, což může opět ovlivnit (tentokrát pozitivně) budoucí úroveň úmrtnosti kohorty v následujících obdobích (Wilmoth, 2006). Prostřednictvím APC modelů (nebo jejich částí AP, AC nebo PC) se dají konstruovat předpovědi budoucího vývoje úrovně úmrtnosti, praktickou aplikaci lze nalézt např. v diplomové práci Čadyho (2012).

Mezi další metody patří predikce vývoje některých funkcí z úmrtnostních tabulek (nejčastěji střední délky života). Takové predikce můžeme nalézt např. v publikaci Českého statistického úřadu, kde se v rámci prognózy obyvatelstva uvažuje také vývoj střední délky života s následnou dekompozicí – tedy určením hlavních věkových skupin, kde se předpokládá změna úmrtnosti (Český statistický úřad, 2013).

V současnosti velmi oblíbenou metodou modelování a predikování úrovně úmrtnosti je také metoda Lee-Cartera (Lee, Carter, 1992). Jedná se o prediktivní metodu, která odhaduje budoucí vývoj specifických měr úmrtnosti. Ve své základní podobě kombinuje faktor průměrné úrovně úmrtnosti pro určitý věk (nezávislý na čase), faktor obecné úrovně úmrtnosti (jeho vývoj v čase) a věkově specifickou konstantu, která vyjadřuje, jak rychle nebo pomalu kolísá úmrtnost v závislosti na věku, když se mění obecná úroveň úmrtnosti (Lee, Carter, 1992). Matematické vyjádření modelu bude podrobněji popsáno v následující kapitole (i s diskusí výhod a nevýhod). Obecně je nutné mít na paměti, že tato metoda (ve své základní podobě) je založena na predikci transverzálních dat do teoreticky libovolně vzdáleného okamžiku, můžeme tedy prostřednictvím této metody odhadnout střední délku života v libovolném roce (samozřejmě základem metody je predikce vývoje specifických měr úmrtnosti). Aplikací Lee-Carterovy metody na projekci vývoje úmrtnosti na nemoci oběhové soustavy v České republice se věnovala Habartová (2009). Součástí článku je také stručný popis obecné formy Lee-Carterovy metody.

Model založený na kombinaci pouze věku a času se jevil jako nedostatečně specifikovaný, což vedlo k tomu, že různé další studie rozšiřovaly Lee-Carterovu metodu zahrnutím kohortního faktoru do modelu. Mezi nejznámější patří práce Renshawa a Habermana, která rozšiřuje základní Lee-Carterovu metodu právě o zahrnutí kohortního faktoru (Renshaw, Haberman, 2006). Srovnání s některými dalšími modely (AP, AC) lze nalézt také v práci Alaie a Sherrise (2012).

Modifikacemi Lee-Carterovy metody se zabývají mnozí další odborníci, např. v článku Andeeva a Vaupela můžeme nalézt modely rozvíjející Lee-Carterův model, založené na různých (alternativních) předpokladech budoucího vývoje úmrtnosti (Andeev, Vaupel, 2006).

Většina takových přístupů, které jsou definovány jako regresní model, však často vyžaduje náročný statistický software. Rozšiřování modelů o další faktory vede k nárůstu počtu proměnných a výsledný model tak může mít i několik set proměnných (např. Alai, Sherris, 2012). Řešení takové úlohy je pak pochopitelně náročné na technické vybavení.

Zvláštní skupinou modelů, které se věnují analýze úmrtnosti, jsou modely, které nemodelují přímo hodnoty konkrétních specifických úmrtnostních měr, ale jejich vzájemné vztahy. Jde zejména o poměry mezi jednotlivými měrami a jejich vývoj v čase. V práci Habermana a Renshawa je např. modelovaný vývoj rozdílu měr úmrtností konkrétní věkové skupiny mezi dvěma časovými obdobími (dvěma roky). Autoři tento vývoj označují jako „mortality improvement“ a odhadují tak tedy „rychlost zlepšování“ úmrtnosti pro konkrétní věkovou skupinu v budoucnu (Haberman, Renshaw, 2013).

Další možností je analyzovat vývoj poměrů (nebo rozdílů) mezi specifickými měrami úmrtnosti dvou věkových skupin v čase. Analýza rozdílů byla publikována (Mazouch, 2010), ale tento přístup se nejevil jako optimální. Modelování poměrů se věnuje např. práce Börgera a Aleksice (2013), Mazoucha (2009), Mazoucha a Hulíkové Tesárkové (2010) nebo Mazoucha a Hulíkové Tesárkové (2012). Podrobnější popis zmíněných metod bude v následující kapitole.

Protože zmíněný přehled možných přístupů k modelování úmrtnosti není v žádném případě úplný a má za úkol pouze naznačit směry, kterými se lze ubírat, bylo by vhodné zmínit některé přehledové publikace, které se predikovaním úmrtností zabývají. Není překvapivé, že velká část publikací, obsahujících takové modely, vzniká na zakázku komerčních společností, tedy pojišťoven nebo penzijních fondů. Takové obsáhlé přehledy i s aplikačními příklady můžeme nalézt například v publikaci Cairnse a kol. (2009) nebo Pitacca a kol. (2009).

Další diskuse technických parametrů vybraných modelů bude součástí následující kapitoly, věnující se metodickému vymezení problému.

Kapitola 2

Zdrojová data a diskuse metodologie modelování úmrtnosti

V této kapitole budou představeny datové zdroje, na kterých budou následně v kapitole 3 ilustrovány výsledky aplikace metodických postupů. Součástí metodické části kapitoly, jejíž stěžejní část je popis vlastní metodiky predikce kohortních měř úmrtnosti, je také detailnější popis některých již zmíněných postupů z předchozí kapitoly, které tak rozšiřují diskusi s literaturou. V závěru je zmíněna stručná metodika konstrukce úmrtnostních tabulek a přístupy pro hodnocení vlastností časových řad.

2.1 Zdrojová data

Protože je kvalita a konzistence dat pro analýzu velmi důležitá, byla jako zdroj dat v praktické části této práce zvolena Human Mortality Database (Human Mortality Database, 2012). Základní zemí, za kterou budou provedeny vybrané analýzy, bylo zvoleno Švédsko. Data z HMD jsou pro Švédsko dostupná již od roku 1751 (v případě sledování počtu zemřelých podle kalendářního roku). Kohortní data, jsou z HMD dostupná již od kohort s rokem narození 1676 (do roku 1981). Použitá data jsou ve formě specifických měř úmrtnosti a ve struktuře 1×1 , což symbolizuje údaje za jednotlivé roky kohort i věku. Specifické míry úmrtnosti $m_{x,z}$ jsou pak spočteny jako

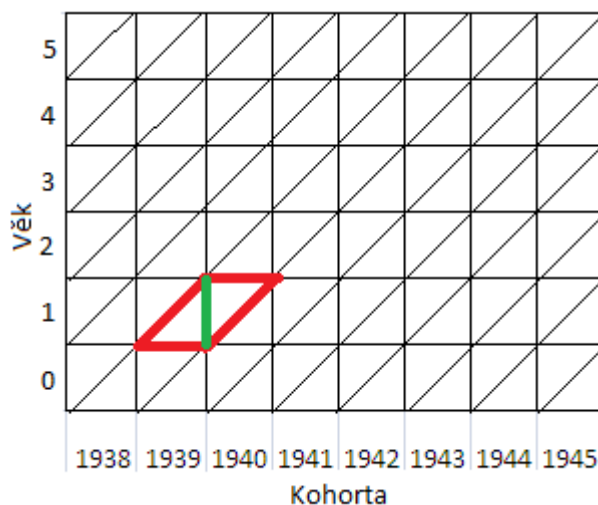
$$m_{x,z} = \frac{D_{x,z}}{P_{x,z}}, \quad (1)$$

kde x je věk při posledních narozeninách (dokončený věk), z je rok narození příslušné kohorty, $D_{x,z}$ počet zemřelých ve věku x z kohorty z a $P_{x,z}$ je exponovaný počet osob ve věku x z kohorty z .

Pro lepší ilustraci viz následující Obr. 1, kde je červeně vyznačen 1. hlavní soubor událostí kohorty narozené např. v roce 1938 ve věku 1 rok a zeleně vyznačen odhadovaný počet exponovaných osob, zde uvažovaný jako střední stav – počet jednoletých osob k 31. 12. 1939 (koncový stav roku 1939) resp. k 1. 1. 1940 (počáteční stav roku 1940).

V případě České republiky jsou data v rámci Human Mortality Database dostupná od roku 1950 (transverzálně) resp. od kohorty narozené 1871 (do roku 1981). Uspořádání dat je stejné jako v případě Švédska.

Obr. 1: Lexisův diagram, červeně označen 1. hlavní soubor událostí $D_{x,z}$ a zeleně exponovaný počet obyvatel $P_{x,z}$



2.2 Metodologie modelování úmrtnosti

Predikci úmrtnosti, jak bylo částečně popsáno v úvodní části, je možné založit na různých přístupech. V následující části bude pozornost věnována diskusi různých přístupů, které modelují vývoj specifických měr úmrtnosti s důrazem na metody, které modelují jejich vztahy. Jsou zde diskutovány také přístupy, jejichž aplikace nebudou součástí kapitoly s výsledky, ale jejich popis umožní porovnání těchto přístupů s hlavní popisovanou metodou. Cílem je diskuse základních principů běžně užívaných metod, vybrané aspekty jednotlivých z nich pak budou uvažovány pro využití v dále navrhovaném alternativním přístupu.

V další části metodologie modelování úmrtnosti bude navržen postup odhadu kohortních měr úmrtnosti na základě vlastní metodologie. Poslední část této podkapitoly je věnována metodickému popisu konstrukce úmrtnostních tabulek a přístupů k modelování časových řad, což je nezbytné pro analýzu výsledků v následující kapitole.

2.2.1 Diskuse vybraných přístupů modelování úmrtnosti

V úvodní kapitole v části diskuse s literaturou byla zmíněna celá řada přístupů, které se při modelování úmrtnosti v současné době využívají. Základní principy některých z nich budou popsány v následující části.

Gompertz-Makehamova funkce (dále značena jako **GM**), která se využívá zejména pro vyrovnávání měr úmrtností ve vyšších věcích při konstrukci úmrtnostních tabulek, má tvar

$$m_x = A + B * c^x, \quad (2)$$

kde m_x je míra úmrtnosti ve věku x , člen A je konstanta vyjadřující úroveň úmrtnosti nezávislé na věku a člen $B * c^x$ vyjadřuje část úmrtnosti závislou na věku (Fiala, 2005).

Zmíněná funkce se zdá být teoreticky jednoduchá, ale její řešení, odhad všech tří parametrů, úplně snadné není. Pro podrobnější popis odhadu parametrů je možné postupovat např. podle Fialy (2005). Při hledání řešení je nutné učinit několik subjektivních rozhodnutí a samotné

řešení je pak tedy možné částečně ovlivnit. Více k této problematice nalezneme v práci Hulíkové Tesárkové (2012).

Protože Gompertz-Makehamovu funkci lze aplikovat na kohortní i na transverzální data, je vhodné definovat také specifické míry úmrtnosti periodické, tedy kombinující věk a kalendářní rok události (třetí variantu, která kombinuje kalendářní rok a rok narození, není pro potřeby této práce nutné definovat – nebude dále nijak analyzována), tedy tzv. 3. hlavní soubory událostí (tj. kombinace věku a kalendářního roku) lze psát:

$$m_{x,t} = \frac{D_{x,t}}{P_{x,t}}, \quad (3)$$

kde x je dokončený věk, t je kalendářní rok, ve kterém sledujeme počet zemřelých, $D_{x,t}$ počet zemřelých ve věku x v kalendářním roce t a $P_{x,t}$ je exponovaný počet osob ve věku x v kalendářním roce t .

V přístupu Gompertz-Makehama (možná ještě lépe v původním vyjádření Gompertze) lze nalézt předpoklad konstantního koeficientu růstu úmrtnosti s věkem, který je ve výrazu (2) vyjádřen jako c . V případě transverzálního přístupu by bylo možné tento koeficient vyjádřit jako

$$ar_{x,t} = \frac{m_{x,t}}{m_{x-1,t}}, \quad (4)$$

kde $ar_{x,t}$ je **poměr periodických specifických** měr podle věku, x je dokončený věk, t je kalendářní rok ve kterém sledujeme příslušnou specifickou míru úmrtnosti, $m_{x,t}$ je spočteno na základě (3).

V případě kohortního přístupu by pak koeficient růstu bylo možné vyjádřit jako

$$ar_{x,z} = \frac{m_{x,z}}{m_{x-1,z}}, \quad (5)$$

kde $ar_{x,z}$ je **poměr kohortních specifických** měr podle věku, x je dokončený věk, z je rok narození příslušné kohorty, $m_{x,z}$ je spočteno na základě (1).

Pro Gompertzovo c by pak platil předpoklad, že:

$$ar_{x,t} = ar_{x+1,t}, \quad (6)$$

resp.

$$ar_{x,z} = ar_{x+1,z}, \quad (7)$$

v případě kohortního přístupu, neboli, předpokládá, že hodnota c je konstantní pro všechny hodnoty věku. Pochybnosti o tomto předpokladu můžeme nalézt v článku Koschina (1999) nebo Thatcher a kol. (1998), kde se v nejvyšších věcích nepředpokládá exponenciální růst úmrtnosti, ale spíše zpomalení rychlosti jejího růstu.

Protože předpoklad o konstantním koeficientu růstu, definovaném jako (6) nebo (7), byl z počátku také jedním z předpokladů vlastní metody výpočtu kohortní prognózy (popsané dále),

ověření tohoto přístupu bude na empirických datech provedeno v následující kapitole, kde bude zřetelné, že koeficient růstu není konstantní ani v mladších věcích.

Lee-Carterova metoda (LC) ve své základní podobě, definovaná Lee a Carterem (1992) uvažuje, že

$$\ln m_{x,t} = \alpha_x + \beta_x * \kappa_t + \varepsilon_{x,t}, \quad (8)$$

kde $m_{x,t}$ je míra úmrtnosti ve věku x a v čase t , α_x je průměrný věkový profil úmrtnosti, β_x je věkově specifická konstanta, κ_t je index úmrtnosti v čase t a $\varepsilon_{x,t}$ je reziduum, tedy vlivy, které nejsou vysvětleny modelem.

Při bližším popisu jednotlivých komponent výrazu (8) lze říci, že α_x je průměrný věkový profil úmrtnosti (podrobněji viz dále), který není závislý na čase. Parametr β_x , věkově specifická konstanta, ukazuje, jak citlivě reaguje úmrtnost ve věku x na změnu κ_t . Zde je možné nalézt určitou podobnost s parametrem B z GM.

Jediný člen, který je závislý na čase t (a naopak nezávislý na věku x), je parametr κ_t , index celkové úrovně úmrtnosti v čase t . Je tedy zřejmé, že základem predikce bude odhad vývoje parametru κ_t v čase, tedy trend vývoje úrovně úmrtnosti. Parametr b_x pak tento trend umocňuje (nebo lépe multiplikuje) podle toho, zda je v určitém věku úmrtnost vyšší nebo nižší než trend (Pitacco a kol., 2007).

Odhad vývoje κ_t je založen na Box-Jenkinsově metodologii (více o této metodě např. Arlt, Arltová, 2005). Jakmile jsou odhadnuty budoucí hodnoty obecného indexu úmrtnosti κ_t , pak je teprve možné odhadnout také budoucí hodnoty specifických měr úmrtnosti (Habartová, 2009).

Jak píše Habartová (2009) nebo Pitacco a kol. (2007), odhad regrese nelze provést klasickou metodou nejmenších čtverců, ale je nutné použít metodu rozkladu matice – Singular Value Decomposition (SVD). Z důvodu nalezení pouze jednoho řešení, je nutné přijmout další podmínky, bez nichž by nebylo jednoznačné řešení možné najít (Pitacco a kol., 2007):

$${}_x \beta_x = 1, \quad (9)$$

a

$${}_t \kappa_t = 0. \quad (10)$$

Při dodržení těchto podmínek je zřejmé, že hodnoty α_x jsou rovny průměrným hodnotám $\ln(m_{x,t})$ v čase (Habartová, 2009).

Odhad $m_{x,t+1}$ pak vypadá následovně:

$$m_{x,t+1} = \exp \alpha_x + \beta_x * \kappa_t = m_{x,t} * \exp \beta_x * \kappa_t - \kappa_{t-1}, \quad (11)$$

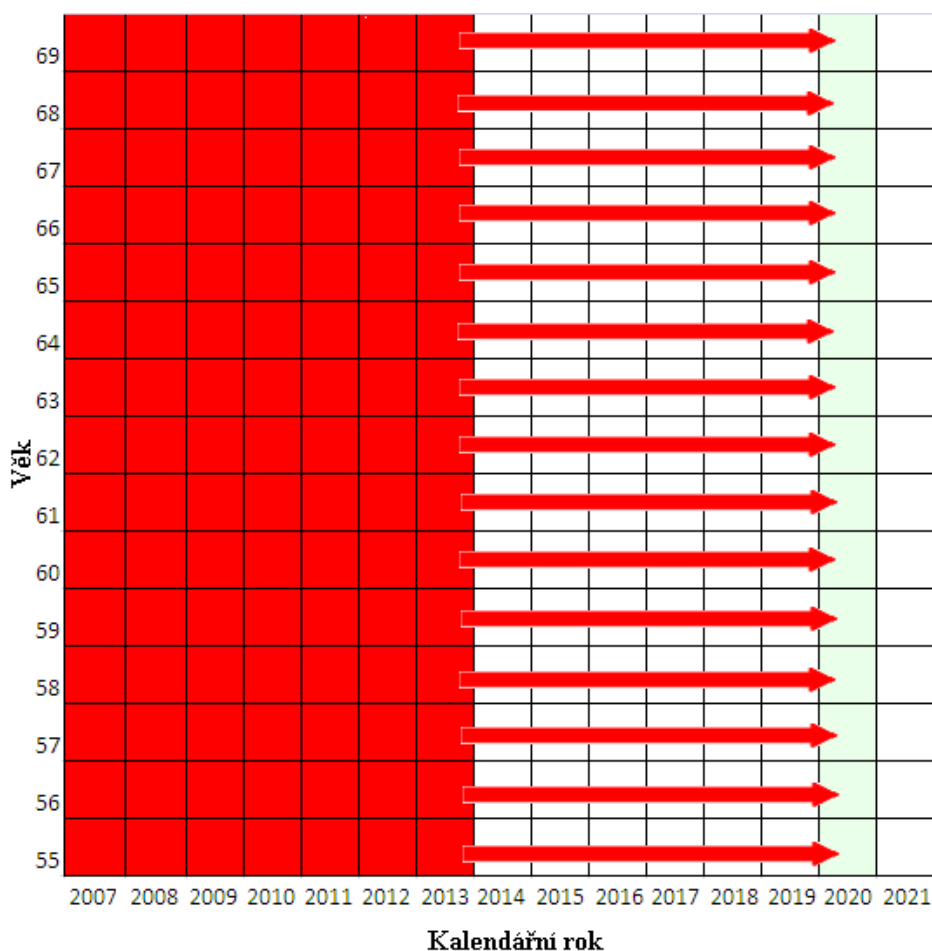
kde $m_{x,t}$ je míra úmrtnosti ve věku x a v čase t , α_x je průměrný věkový profil úmrtnosti, β_x je věkově specifická konstanta, κ_t je index úmrtnosti v čase t (Pitacco a kol., 2007).

Obr. 2 ilustruje predikci budoucích měr úmrtnosti, založenou na transverzálních hodnotách specifických měr úmrtnosti, které se na základě modelu predikují do budoucnosti.

Jedna z nevýhod LC metody se týká náhodné složky $\varepsilon_{x,t}$ ze základní rovnice (8) LC modelu. U té se předpokládá její homoskedasticita², která je základním předpokladem pro řešení úlohy prostřednictvím metody nejmenších čtverců (MNC). Požadavky na vlastnosti náhodné složky popisuje např. Arlt a Arltová (2005). Kromě nulové střední hodnoty $\varepsilon_{x,t}$ (který je zřejmě v LC splněn) je druhým požadavkem konstantní rozptyl náhodné složky. Splnění druhé podmínky je v případě úmrtnosti téměř nemožné, protože rozptyl (variabilita) měr úmrtnosti roste s věkem (Pitacco a kol., 2009).

Brouhns a kol. publikovali modifikaci odhadu měr úmrtnosti na základě LC metody, kde není nutné dodržet zmíněné podmínky homoskedasticity. Místo odhadu budoucí míry úmrtnosti se pokusili odhadnout počet zemřelých při použití Poissonova rozdělení (Brouhns a kol., 2002).

Obr. 2: Ilustrace LC modelu – odhad budoucích hodnot měr úmrtnosti v roce 2020 na základě vývoje měr úmrtnosti do roku 2013



Pozn. Červeně jsou označeny 3. hlavní soubory událostí, na základě kterých se odhadují specifické míry úmrtnosti v budoucnu (konkrétně v r. 2020). Světle zeleně je vyznačena plocha, ke které se odhad vztahuje.

² Náhodné chyby ε_x mají konstantní a konečný rozptyl $E \varepsilon_x^2 = \sigma^2$ (Meloun, Militký, 2002)

Předpoklad Poissonova rozdělení říká, že počet zemřelých ve věku x v kalendářním roce t je náhodná veličina popsaná

$$D_{x,t} \sim Po P_{x,t} * \mu_{x,t} , \quad (12)$$

kde $D_{x,t}$ počet zemřelých ve věku x v roce t , $P_{x,t}$ je exponovaný počet osob ve věku x v roce t , $\mu_{x,t}$ je intenzita úmrtnosti a výraz $Po P_{x,t} * \mu_{x,t}$ znamená, že náhodná veličina má Poissonovo rozdělení s parametrem $P_{x,t} * \mu_{x,t}$.

Brouhns a kol. uvažovali log-bilineární intenzitu úmrtnosti, což je možné vyjádřit jako

$$\ln \mu_{x,t} = \alpha_x + \beta_x * \kappa_t, \quad (13)$$

kde α_x je průměrný věkový profil úmrtnosti, β_x je věkově specifická konstanta, κ_t je index úmrtnosti v čase t . Tedy výraz je téměř totožný s (8) s výjimkou reziduí, která v tomto vyjádření nejsou. Odhady parametrů se v tomto případě neprovádí pomocí MNČ, ale jsou odhadovány na základě metody maximální věrohodnosti (Brouhns a kol., 2002).

Z výše uvedeného vyplývá, že základem metody LC je modelování parametru κ_t , pro nějž se využívá metody Box-Jenkinse (i Brouhns a kol., 2002 využili této metody). Zajímavou možností nahrazení tohoto parametru nabízí ve své práci Buettner (2002). Místo parametru κ_t využívá konstrukce modelů úmrtnostních tabulek (k metodice jejich konstrukce viz dále) a funkce e_0 , tedy střední délky života při narození. Do LC modelu tak byl implementován místo indexu úmrtnosti index reflektující vývoj úrovně střední délky života (což je také jeden z ukazatelů úrovně úmrtnosti). Pak model vypadá následovně:

$$\ln m_{x,e} = \alpha_x + \beta_x * \kappa_e + \varepsilon_{x,e}, \quad (14)$$

kde $m_{x,e}$ je míra úmrtnosti ve věku x při konkrétní úrovni e_0 , α_x je průměrný věkový profil úmrtnosti, β_x je věkově specifická konstanta, κ_e je index trendu střední délky života při narození a $\varepsilon_{x,e}$ je reziduum, tedy vlivy, které nejsou vysvětleny modelem. Model touto úpravou ztrácí informaci o čase (není zde index t) a je vázaný na hodnotu e_0 (Pitacco a kol., 2009).

Modifikace LC modelu, kterou provedli Renshaw a Haberman (2006) byla svým způsobem revoluční, protože zavedli do modelu další faktor, kterým byl vliv kohorty. Tím se pochopitelně stal model složitějším:

$$\ln m_{x,t} = \alpha_x + \ln F_{x,t} , \quad (15)$$

kde $m_{x,t}$ je míra úmrtnosti ve věku x a v čase t , α_x je průměrný věkový profil úmrtnosti a $F_{x,t}$ je redukční faktor úmrtnosti, který lze vyjádřit jako

$$F_{x,t} = \exp \beta_x^{(0)} * t_{t-x} + \beta_x^{(1)} * \kappa_t , \quad (16)$$

kde $\beta_x^{(0)} * t_{t-x}$ znázorňuje kohortní efekt a $\beta_x^{(1)} * \kappa_t$ je efekt periody. Je zřejmé, že původní LC model je jen specifický druh (16), kdy $\beta_x^{(0)} = 0$. Podobně jestliže $\beta_x^{(1)} = 0$, pak se jedná o model AC (Renshaw, Haberman, 2006).

Rozšíření modelu o efekt kohorty způsobuje určité komplikace při hledání řešení. Základní problém tkví ve vztahu

$$\text{kohorta} + \text{věk} = \text{perioda}, \quad (17)$$

tedy v situaci, kdy kombinace dvou proměnných je třetí proměnná. Více o řešení tohoto problému můžeme najít v článku zabývajícím se aplikací APC modelů v demografii, který publikoval Wilmoth (2006). Praktické použití tohoto přístupu také využil Čady (2012) ve své diplomové práci, kde diskutuje možnosti řešení zmíněného problému závislosti kohorty, věku a periody.

Řešení problému v zahrnutí kohortního efektu do LC modelu lze nalézt v práci Rensawa a Habermana (2006). Na stejném místě můžeme nalézt také postup pro predikci měr úmrtnosti:

$$m_{x,t_n+s} = m_{x,t_n} * F_{x,t_n+s} = m_{x,t_n} * \exp \left(\beta_x^{(0)} * (t_{t_n-x+s} - t_{t_n-x}) + \beta_x^{(1)} * (\kappa_{t_n+s} - \kappa_{t_n}) \right), \quad (18)$$

kde m_{x,t_n+s} je odhad míry úmrtnosti ve věku x , v čase $t_n + s$, t_n je počáteční rok predikce, s je délka predikce, m_{x,t_n} je míra úmrtnosti ve věku x v počátečním roce predikce, F_{x,t_n+s} je redukční faktor úmrtnosti, $\beta_x^{(0)}$ je věkově specifická konstanta kohortního efektu, t_{t_n-x+s} a t_{t_n-x} je index kohorty $t_n - x + s$, resp. $t_n - x$ a κ_{t_n+s} a κ_{t_n} je index úmrtnosti v období $t_n + s$, resp. t_n .

Modelování změn (rozdílů) měr úmrtnosti je poslední oblast modelování úmrtnosti, které bude v metodické části věnována pozornost. Na ni pak navazuje podkapitola věnující se popisu vlastní metody predikce úmrtnosti, jejíž princip je také založen na modelování změn úmrtnosti.

V článku Habermana a Renshawa (2013) můžeme nalézt projekční přístup, založený na analýze diferencí měr úmrtnosti. Haberman a Renshaw v článku uvažují, že spočteme-li rozdíly měr úmrtnosti podle věku ($m_{x-1,t-1} - m_{x,t-1}$), pak jsou (s výjimkou možného výkyvu způsobeného nehodami) většinou negativní. Jestliže pak ještě nahradíme $m_{x,t-1}$ hodnotou $m_{x,t}$ (vývoj mezi dvěma roky je zanedbatelný), můžeme stále očekávat záporné hodnoty, které se vyvíjejí rostoucím trendem respektujícím efekt stárnutí (Haberman, Renshaw, 2013).

Haberman a Renshaw se rozhodli modelovat veličinu $z_{x,t}$, které je dána vztahem

$$z_{x,t} = 2 * \frac{1 - \frac{m_{x,t}}{m_{x,t-1}}}{1 + \frac{m_{x,t}}{m_{x,t-1}}}, \quad (19)$$

kde $m_{x,t}$ je míra úmrtnosti ve věku x a v čase t . Odhad míry úmrtnosti pak bude dle (19) vypadat následovně

$$m_{x+j,t_n+j} = m_{x+j-1,t_n+j-1} * \frac{2-z_{x+j,t_n+j}}{2+z_{x+j,t_n+j}}, \quad (20)$$

kde $m_{x,t}$ je míra úmrtnosti ve věku x a v čase t , t_n je počáteční rok predikce, j je počet let predikce (Haberman, Renshaw, 2013).

Kombinaci modelů využívajících změn měr úmrtnosti a APC modelů se zabývali autoři Börger a Aleksis (2013), kteří zvolili jako ukazatel vývoje úmrtnosti vztah definovaný jako

$$v_{x,t} = \frac{q_{x,t-1} - q_{x,t}}{q_{x,t-1}} = 1 - \frac{q_{x,t}}{q_{x,t-1}}, \quad (21)$$

kde $q_{x,t}$ je pravděpodobnost úmrtí ve věku x a v čase t .

Veličinu $v_{x,t}$ pak následně vyjádřili prostřednictvím APC modelů:

$$v_{x,t} = a_x + p_t + c_{t-x} + \varepsilon_{x,t}, \quad (22)$$

kde x je věk, t je období, a_x je věkově závislá komponenta, p_t je komponenta období, c_{t-x} je kohortní komponenta a $\varepsilon_{x,t}$ je reziduum (Börger, Aleksis, 2013).

Autoři pak přechází, z důvodů lepších vlastností pro modelování, na ukazatel redukce úmrtnosti

$$r_{x,t} = 1 - v_{x,t} = \frac{q_{x,t}}{q_{x,t-1}}, \quad (23)$$

kde $q_{x,t}$ je pravděpodobnost úmrtí ve věku x a v čase t (Börger, Aleksis, 2013).

2.2.2 Popis základních principů navržené alternativní metody predikce

Projekční model, který je detailně vysvětlen a popsán v následující části a stal se základem této diplomové práce, byl v jeho první navržené podobě publikován již v roce 2009 jako výsledek jednoduchého modelování úmrtnosti založeného na měřích úmrtnosti a jejich vzájemném vztahu. V článku Mazoucha (2009) byly prezentovány hodnoty podílů měr úmrtností, založené na vztahu (4)

$$ar_{x,t} = \frac{m_{x,t}}{m_{x-1,t}}, \quad (4)$$

kde $ar_{x,t}$ je **poměr transverzálních specifických** měr podle věku, x je dokončený věk, t je kalendářní rok ve kterém sledujeme příslušnou specifickou míru úmrtnosti, $m_{x,t}$ je spočteno na základě (3).

Vysoká variabilita měr poměrů $ar_{x,t}$ vedla k tomu, že hodnoty byly vyrovnány klouzavým průměrem. I přes to nebyly výsledky příliš uspokojivé, což bylo přisuzováno transverzální povaze dat (Mazouch, 2009).

V dalších člancích pak byla následně využita již data kohortní (Mazouch, 2010; Mazouch, Hulíková Tesárková, 2010; Mazouch, Hulíková Tesárková, 2012), přesto lze pozorovat u jiných

autorů, že transverzální způsob považují za vhodný (Haberman, Renshaw, 2013, Börger, Aleksis, 2013).

Mazouch (2010) a Mazouch a Hulíková Tesárková (2010) analyzovali kohortní data, kdy byly hodnoceny vztahy (5) a nový vztah (24), tedy

$$ar_{x,z} = \frac{m_{x,z}}{m_{x-1,z}}, \quad (5)$$

kde $ar_{x,z}$ je **poměr kohortních specifických** měr podle **věku**, x je dokončený věk, z je rok narození příslušné kohorty, $m_{x,z}$ je spočteno na základě (1) a

$$ad_{x,z} = m_{x,z} - m_{x-1,z}, \quad (24)$$

kde $ad_{x,z}$ je **rozdíl kohortních specifických** měr podle **věku**, x je dokončený věk, z je rok narození příslušné kohorty, $m_{x,z}$ je spočteno na základě (1).

Mazouch a Hulíková Tesárková (2012) pak využili předchozí analýzy kohortních dat ke konstrukci vlastní metodiky odhadu budoucího vývoje kohortních měr úmrtnosti, jejíž základní principy budou popsány v této části.

Vlastní model odhadu budoucích specifických měr úmrtnosti je založen na vztahu (5), důvodem jsou předpoklady o vývoji tohoto ukazatele v čase (viz dále).

Jestliže platí (což předpokládá ve svých úvahách také Gompertz, 1825 aj.), že od určitého věku (např. 30 let)

$$m_{x,z} > m_{x-1,z}, \quad (25)$$

pak musí být hodnota $ar_{x,z} > 1$ a také $ad_{x,z} > 0$ (Mazouch, 2010).

Pokud by se podařilo z vývoje $ar_{x,z}$ zjistit určitou pravidelnost (nejlépe monotónnost vývoje), pak by nebylo nutné modelovat všechny míry úmrtnosti, ale stačilo by modelovat míru jednu a ostatní od ní díky znalosti vývoje $ar_{x,z}$ konkrétní kohorty, dopočítat. Tento fakt lze považovat za základní myšlenku popsaného způsobu modelování kohortní úmrtnosti a možnosti její predikce.

Situaci (a základní předpoklad metody) ilustruje Obr. 3, kde je vyznačena kohorta narozená v roce 1953 a jejíž informace o vývoji úmrtnosti končí v roce 2013 (informace o úrovni úmrtnosti v dalších letech nejsou k dispozici). Další vývoj úmrtnosti sledované kohorty je tedy nutné odhadnout. Pokud by jednotlivé hodnoty $ar_{x,z}$ pro věky 60 a více let vykazovaly systematický (nebo jinak odhadnutelný) průběh v čase (tedy mezi kohortami), pak na základě odhadů $ar_{x,1953}$, lze odhadnout i zbývající průběh měr úmrtnosti kohorty narozené 1953, tedy obecně $ar_{m,1953}$, kde $m = x + 1, x + 2, \dots, \omega$.

Tento postup vyžaduje kvalitní analýzu empirických hodnot $ar_{x,z}$. Vývoj by měl být v čase stabilní, aby hodnoty mohly být využity v modelu predikce měr úmrtnosti. V této části práce je možné stanovit určité teoretické předpoklady, založené na výsledcích z článku Mazoucha a Hulíkové Tesárkové (2012), které by vedly k nepříliš složitému modelu predikce a jejichž naplnění bude ověřeno v následující kapitole na reálných hodnotách.

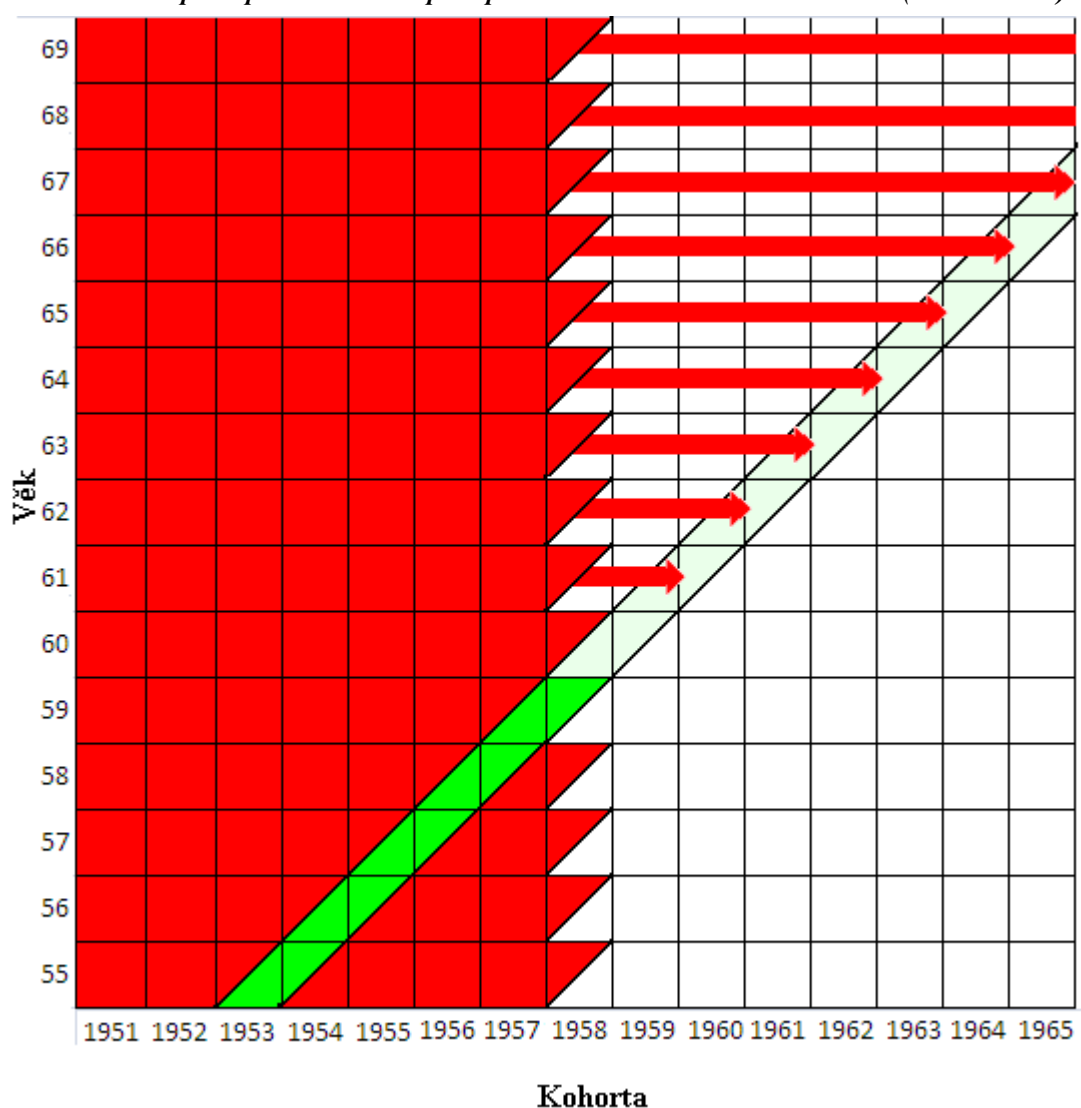
Tímto předpokladem je konstantní hodnota $ar_{x,z}$ pro všechny uvažované kohorty,

$$ar_{x,z} = ar_{x,z+1} , \quad (26)$$

kde $ar_{x,z}$ je poměr kohortních specifických měr podle věku, x je dokončený věk, z je rok narození příslušné kohorty (Mazouch, Hulíková Tesárková, 2012).

Tento předpoklad se zdá být velmi striktní a v praxi těžko dosažitelný (jak se ukáže v kapitole výsledků, není však tak vzdálen od reality). Znamenal by, že koeficient růstu úmrtnosti mezi dvěma konkrétními věky je pro jakoukoliv kohortu stále stejný, tedy konstantní (v kapitole diskuse bude rozebrán i přístup, který nevyžaduje toto relativně striktní pravidlo konstantní úrovně) a jeho budoucí vývoj (nebo v tomto případě úroveň) tak lze odhadnout na základě minulé hodnoty (Mazouch, Hulíková Tesárková, 2012).

Obr. 3: Ilustrace principu metodického postupu odhadu kohortních měr úmrtnosti (kohorta 1953)



Pozn. Červeně (a tmavě zeleně) jsou označeny známé hodnoty, na základě kterých se odhadují koeficienty $ar_{x,z}$. Tmavě zeleně je označena plocha známých hodnot kohorty 1953. Šípky znázorňují směr aplikace hodnot $ar_{x,z}$ na specifické míry úmrtnosti. Světle zeleně je vyznačena plocha, ke které se odhad vztahuje.

Protože data budou jistě vykazovat určitou variabilitu (tedy vztah (26) nebude platit doslovně, protože koeficienty růstu pro věk x budou v čase kolísat), pak by bylo dobré, aby dané časové řady koeficientů $ar_{x,z}$ pro konkrétní věk x byly mezi jednotlivými kohortami alespoň **stacionární**³ (o testech a vlastnostech stacionárních časových řad viz oddíl na konci této kapitoly), tedy $ar_{x,z} \approx ar_{x,z+1} \approx ar_{x,z+2} \approx \dots \approx k$, kde k je konstanta.

Z důvodu variability $ar_{x,z}$ v čase (tedy mezi jednotlivými kohortami), je pro konkrétní **modelovanou** kohortu z nutné určit hodnoty $ar_{x,z}$, pro jednotlivá x , prostřednictvím např. průměrné hodnoty, tedy pro každé x spočteme $ar_{x,z,n,s}$ z n předchozích kohort. Pak můžeme hodnoty $ar_{x,z,n,s}$ využít pro odhad budoucích hodnot měř úmrtnosti kohorty z do budoucna.

Pro odhad $ar_{x,z,n,s}$ je možné využít několika způsobů, např.

$$ar_{x,z,n,s} = \frac{\sum_{k=1}^n \alpha^{k-1} * ar_{x,z-k-x+s}}{\sum_{k=0}^n \alpha^{k-1}}, \text{ pro } x = s, \dots, \omega - 1 \quad (27)$$

kde $ar_{x,z,n,s}$ je průměrná hodnota **poměru kohortních specifických** měr podle věku (pro konkrétní věk x v rámci všech uvažovaných kohort, tj. pro všechny hodnoty z), x je věk (jehož maximem je poslední věk, pro který lze spočítat $ar_{x,z}$), z je rok narození kohorty, α je váha nabývající hodnot (0;1), kterou lze předem subjektivně zvolit podle požadované paměti modelu (viz dále), n je počet kohort, které se do výpočtu započítávají a s je nejnižší věk, pro který začínáme počítat $ar_{x,z,n,s}$.

Jedná se tedy o vážený průměr $ar_{x,z}$ z n předchozích kohort pro konkrétní věk x , jehož váhy α^{k-1} , pro $k = 1, \dots, n$ exponenciálně klesají směrem do minulosti s růstem k (čím vzdálenější hodnota, tím menší váha). Tímto je zajištěna určitá „omezená“ paměť modelu, kdy nejsou do výpočtu započítávány všechny historické kohorty (jestliže je jejich počet nižší než n , pak pochopitelně ano) a navíc jejich váha se směrem do minulosti zmenšuje.

Složitost vzorce (zejména indexu u ar) je dána tím, že je nutné zajistit, aby byly do výpočtu zahrnovány relevantní kohorty. Níže je uveden praktický příklad výpočtu pro kohortu narozenou v roce 1950, kdy je odhad budoucích měř úmrtnosti realizován pro věky 60 a více let a kdy je do výpočtu započítáváno 40 předchozích kohort ($n = 40$). Pro příklad uvažujme, že o kohortě 1950 je v období konstrukce odhadu budoucího vývoje úmrtnosti známa poslední hodnota míry úmrtnosti pro věk 59 let. Pro odhad $ar_{x,z,n,s}$ ve věku 60 je tedy nutné mít informaci o úrovni $ar_{60,z,n,s}$ za 40 předchozích kohort (tedy $z = 1910, \dots, 1949$). Pro odhad $ar_{61,z,n,s}$ je však nutné počítat s kohortami opět o jeden rok staršími, tedy z období 1909–1948, protože za mladší kohorty v tomto věku ještě empirická data nejsou známá, atd. Posun o náležitý počet kohort zpět je ve vzorci (27) zajištěn právě indexem ($z - k - x + s$) u členu ar .

Příklad praktické aplikace vzorce (27) pro kohortu narozenou 1950 ($z = 1950$), pro věky 60 (tj. výpočet první průměrné hodnoty bude pro věk $s = 60$) a více let a pro $n = 40$ (bude využito informace z předchozích 40 let) vypadá následovně:

³ Vykazovaly konstantní trend – viz podkapitulu 2.2.4

Pro věk 60 let:

$$\begin{aligned} ar_{60,1950,40,60} &= \frac{\alpha^0 * ar_{60,1950-1-60+60}}{\sum_{k=0}^{40} \alpha^{k-1}} + \frac{\alpha^1 * ar_{60,1950-2-60+60}}{\sum_{k=0}^{40} \alpha^{k-1}} + \dots \\ &+ \frac{\alpha^{39} * ar_{60,1950-40-60+60}}{\sum_{k=0}^{40} \alpha^{k-1}} \\ &= \frac{\alpha^0 * ar_{60,1949}}{\sum_{k=0}^{40} \alpha^{k-1}} + \frac{\alpha^1 * ar_{60,1948}}{\sum_{k=0}^{40} \alpha^{k-1}} + \dots + \frac{\alpha^{39} * ar_{60,1910}}{\sum_{k=0}^{40} \alpha^{k-1}} \end{aligned}$$

Pro věk 61 let:

$$\begin{aligned} ar_{61,1950,40,60} &= \frac{\alpha^0 * ar_{61,1950-1-61+60}}{\sum_{k=0}^{40} \alpha^{k-1}} + \frac{\alpha^1 * ar_{61,1950-2-61+60}}{\sum_{k=0}^{40} \alpha^{k-1}} + \dots \\ &+ \frac{\alpha^{39} * ar_{61,1950-40-61+60}}{\sum_{k=0}^{40} \alpha^{k-1}} \\ &= \frac{\alpha^0 * ar_{61,1948}}{\sum_{k=0}^{40} \alpha^{k-1}} + \frac{\alpha^1 * ar_{61,1947}}{\sum_{k=0}^{40} \alpha^{k-1}} + \dots + \frac{\alpha^{39} * ar_{61,1909}}{\sum_{k=0}^{40} \alpha^{k-1}} \end{aligned}$$

...

Pro věk 100 let:

$$\begin{aligned} ar_{100,1950,40,60} &= \frac{\alpha^0 * ar_{100,1950-1-100+60}}{\sum_{k=0}^{40} \alpha^{k-1}} + \frac{\alpha^1 * ar_{100,1950-2-100+60}}{\sum_{k=0}^{40} \alpha^{k-1}} + \dots \\ &+ \frac{\alpha^{39} * ar_{100,1950-40-100+60}}{\sum_{k=0}^{40} \alpha^{k-1}} \\ &= \frac{\alpha^0 * ar_{100,1909}}{\sum_{k=0}^{40} \alpha^{k-1}} + \frac{\alpha^1 * ar_{100,1908}}{\sum_{k=0}^{40} \alpha^{k-1}} + \dots + \frac{\alpha^{39} * ar_{100,1870}}{\sum_{k=0}^{40} \alpha^{k-1}} \end{aligned}$$

...

Další možností odhadu $ar_{x,z,n,s}$ je využití např. způsobu

$$ar_{x,z,n,s} = \frac{25ar_{x,z-x+s,n} + 75ar_{x,z-x+s,n}}{2}, \quad (28)$$

kde $ar_{x,z,n,s}$ je průměrná hodnota **poměru kohortních specifických** měr podle **věku**, x je věk, z je rok narození kohorty, $25ar_{x,z-x+s,n}$ je hodnota dolního kvartilu (25. percentilu) určeného z n posledních hodnot $ar_{x,z}$ předcházejících kohortě $z - x + s$, $75ar_{x,z-x+s,n}$ je hodnota horního kvartilu (75. percentilu) určeného z n posledních hodnot $ar_{x,z}$ předcházejících kohortě $z - x + s$ a s je nejnižší věk, pro který začínáme počítat $ar_{x,z}$ a n je počet kohort které se do výpočtu započítávají.

Opět je zde (identicky jako v případě vzorce (27)) nutné zajistit posun zahrnovaných kohort směrem do minulosti, aby byly do výpočtu zahrnuty pouze kohorty, za které jsou v době výpočtu již známá empirická data.

Tento přístup může, vzhledem k možným extrémním hodnotám $ar_{x,z}$, vykazovat vyšší robustnost výsledků. Na druhou stranu je zde opravdu nutné dodržení podmínky předpokladu stacionarity hodnot $ar_{x,z}$ v čase (viz kapitolu 2.2.4).

Důležitým krokem je volba věku, který bude základem pro následný odhad měr úmrtnosti, tedy o věk v souladu s výše popsanou symbolikou označený jako $s - 1$ (teoreticky se jedná o věk, pro který je známa poslední hodnota úmrtnosti pro danou kohortu). Je tedy patrné, že výsledný odhadovaný vývoj úmrtnosti konkrétní kohorty je závislý na prvním zvoleném věku, od kterého bude úmrtnost predikována, resp. na hodnotě úmrtnosti v tomto věku (dále značena jako $bm_{x,z}$).

Protože je však volatilita jednotlivých specifických měr úmrtnosti velmi vysoká i zde je nutné přistoupit ke stanovení „základní“ úrovně úmrtnosti (tj. výchozí úrovně úmrtnosti $bm_{x,z}$, kde $x = s - 1$) jinak než prostým dosazením empirické hodnoty (která může být vlivem náhodných odchylek vychýlena směrem k vyšším i nižším hodnotám a ovlivnit tak celý výsledek). Jednou z možností je spočtení „průměrné“ míry úmrtnosti, která by případnou volatilitu (extrémní odchylku) rozložila do více věkových skupin v rámci kohorty. Zmiňovaný odhad základní hodnoty může být tedy např. následující:

$$bm_{x,z} = \text{geomean } m_{x-4,z}, m_{x+4,z} = \\ =^9 \frac{m_{x-4,z} * m_{x-3,z} * m_{x-2,z} * m_{x-1,z} * m_{x,z} * m_{x+1,z} * m_{x+2,z} * m_{x+3,z} * m_{x+4,z}}{9}, \quad (29)$$

kde $bm_{x,z}$ je základní míra úmrtnosti, x je věk a z je rok narození kohorty. Volba počtu zahrnovaných měr úmrtnosti do výpočtu není závazná, ale při počtu 9 by již měl dávat dostatečně robustní výsledek, který eliminuje náhodné výkyvy základní míry úmrtnosti.

Po určení všech $ar_{x,z,n,s}$ pro všechna x pro které bude probíhat odhad úmrtnosti u konkrétní kohorty, je možné modelovat úmrtnosti dle následujícího vzorce, kde nejprve základní úroveň úmrtnosti $bm_{x,z}$ vynásobíme příslušným koeficientem $ar_{x,z,n,s}$:

$$m_{x+1,z} = bm_{x,z} * ar_{x+1,z,n,s} \quad (30)$$

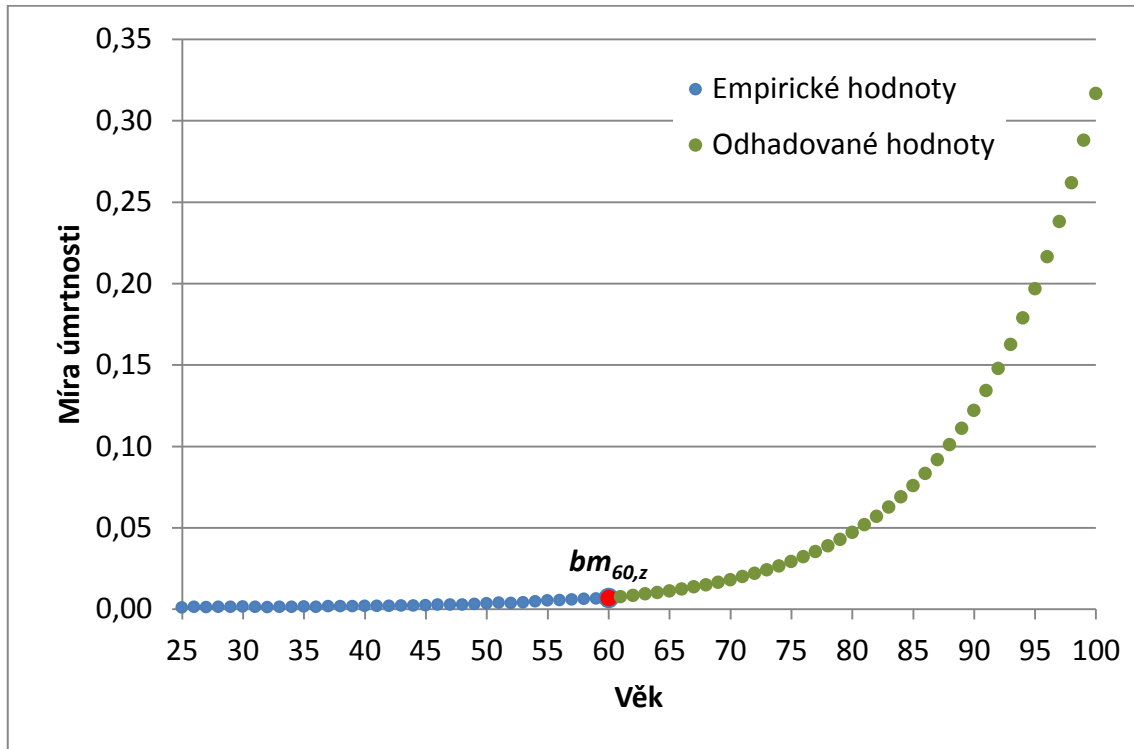
kde $m_{x+1,z}$ je odhad $m_{x+1,z}$, x je věk, z je rok narození kohorty, $bm_{x,z}$ je základní míra úmrtnosti, $ar_{x,z,n,s}$ je průměrná hodnota poměru kohortních specifických měr úmrtnosti a následně další hodnoty $m_{a,z}$, pro $a = x+2, \dots, \omega$ postupně vypočteme dle

$$m_{a,z} = m_{a-1,z} * ar_{a,z}, \quad (31)$$

kde $m_{a,z}$ je odhad $m_{a,z}$, a je věk, z je rok narození kohorty, $ar_{x,z,n,s}$ je průměrná hodnota poměru kohortních specifických měr úmrtnosti odpovídající danému věku.

Názorně je výsledek aplikace vidět na Obr. 4. Hodnoty měr úmrtnosti jsou nad 60 let věku (což je základní věk) odhadovány, červeně je vyznačena základní míra úmrtnosti $bm_{x,z}$.

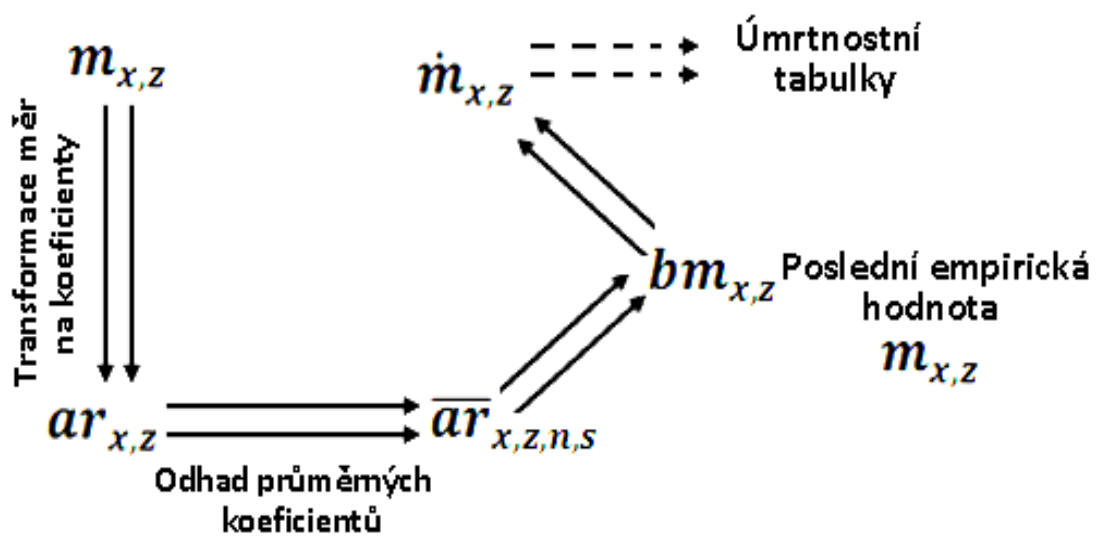
Obr. 4: Predikce měr úmrtnosti od věku 60 let – ilustrativní příklad



Pozn.: do věku 60 let (včetně) jsou úmrtnosti známy, nad 60 let jsou odhadovány dle vztahu (30) resp. (31).

Kompletní postup výše zmíněného modelu je znázorněn na Obr. 5. Základem je převedení měr úmrtnosti $m_{x,z}$ na poměry měr $ar_{x,z}$, následně odhad průměrných koeficientů růstu $ar_{x,z,n,s}$ a přes odhad poslední empirické hodnoty $bm_{x,z}$ až po konečný odhad modelových měr úmrtnosti $\hat{m}_{x,z}$. Praktická aplikace bude předvedena v následující kapitole.

Obr. 5: Kompletní schéma modelu predikce kohortních měr úmrtnosti



2.2.3 Konstrukce úmrtnostních tabulek

Pro hodnocení některých výsledků bude využito ilustrace na hodnotách funkcí úmrtnostních tabulek, proto je vhodné, definovat vztahy mezi jednotlivými funkcemi. Značení bylo převzato z knihy Roubíčka (1997) s jediným rozdílem, kdy do indexu specifické míry úmrtnosti byl přidán symbol z , aby bylo zřejmé, že se jedná o konstrukci kohortních tabulek a mírnou úpravou formulace počátečního věku, který v tomto případě nebude 0, ale 60 let a posledního, otevřeného věkového intervalu, který Roubíček nedefinoval.

Úmrtnostní tabulky budou sestavovány pro vyšší věky, tedy není nutné uvažovat problematiku 0letých, pro všechny věkové skupiny jsou vypočteny z kohortních specifických měr úmrtnosti pravděpodobnosti úmrtí q_x mezi přesnými věky x a $x+1$ (Roubíček, 1997):

$$q_x = \frac{2 \cdot m_{x,z}}{2 + m_{x,z}}. \quad (32)$$

Pravděpodobnost dožití p_x pro všechna x je spočtena jako jednotkový doplněk hodnoty q_x (Roubíček, 1997):

$$p_x = 1 - q_x. \quad (33)$$

Hodnota počtů dožívajících se přesného věku l_x pro všechna x je spočtena postupně od prvního věku pomocí hodnoty p_x (tamtéž):

$$l_{x+1} = p_x \cdot l_x, \quad (34)$$

přičemž pro první věk, který je v tabulce uvažován (kořen tabulky) je hodnota l_{60} rovna 100000, pak hodnota l_{61} je spočtena jako

$$l_{61} = p_{60} \cdot 100000 \text{ (tamtéž)}. \quad (35)$$

Tabulkové počty zemřelých d_x jsou spočteny jako rozdíl dvou sousedních hodnot l_x v úmrtnostní tabulce (Roubíček, 1997):

$$d_x = l_x - l_{x+1}. \quad (36)$$

Hodnoty počtu žijících v dokončeném věku L_x jsou spočteny pomocí vzorce (tamtéž):

$$L_x = (l_x + l_{x+1})/2, \quad (37)$$

s výjimkou posledního otevřeného věkového intervalu (značeno jako $x+$; tento interval končí věkem $\omega - 1$, kde věk ω je první věk, kterého se nikdo nedožije), který se spočte jako podíl poslední hodnoty l_{x+} a m_{x+} :

$$L_{x+} = \frac{l_{x+}}{m_{x+}}. \quad (38)$$

Počet zbývajících let života populace žijící v dokončeném věku x T_x je spočten jako kumulace souborů žijících L_x od nejvyššího věku tabulky $\omega - 1$ postupně až po nejnižší věk tabulky (60 let), tedy

$$T_x = T_{x+1} + L_x \quad (\text{Roubíček, 1997}). \quad (39)$$

Střední délka života osoby ve věku x se vypočítá jako poměr počtu zbývajících let života T_x a počtu dožívajících se přesného věku l_x :

$$e_x = \frac{T_x}{l_x} \quad (\text{tamtéž}). \quad (40)$$

V úmrtnostních tabulkách není aplikována žádná metoda vyrovnávání či vyhlazování, uvedené vztahy (32)–(40) jsou tedy kompletní metodikou, která je v této práci pro konstrukci úmrtnostních tabulek využita.

2.2.4 Vybrané vlastnosti časových řad

Výraz (26) předpokládá určitou neměnnost vývoje hodnot $ar_{x,z}$ v čase (mezi kohortami). Tato neměnnost (konstantnost) je samozřejmě ve vztahu (26) definována pouze jako teoretický (a velmi striktní) předpoklad, který je nutné ověřit. Je proto nutné nejprve popsat vlastnosti stochastického procesu (neboli časové řady). Ty se dají popsat celou řadou nástrojů. Jejich obsáhlý výčet lze nalézt např. v knize autorů Arlt a Arltové (2005). V této části bude zmíněna základní stručná teorie vlastností procesu bílého šumu, který je definován jako proces stacionární (Arlt, Arltová, 2005).

„Stochastický proces lze definovat jako v čase uspořádanou řadu náhodných veličin $X(s, t)$, kde $s \in S$ a $t \in T$ a kde pro každé $t \in T$ je $X(\cdot, t)$ náhodná veličina definovaná na výběrovém prostoru S . Pro každé $s \in S$ je $X(s, \cdot)$ realizace stochastického procesu definovaná na indexní řadě T , tj. uspořádaná řada čísel, z níž každé odpovídá jedné hodnotě indexní řady“ (Arlt, Arltová, 2005, s. 30–31).

„Jestliže je stochastický proces a_t řadou nekorelovaných náhodných veličin jednoho pravděpodobnostního rozdělení s konstantní střední hodnotou $E a_t = \mu_a$, konstantním rozptylem $D a_t = \sigma_a^2$ a kovariancí $\gamma_k = C a_t, a_{t-k} = 0$ pro všechna $k \neq 0$ (kde k značí zpoždění – minulé období, předcházející období t), potom se takový proces nazývá jako **proces bílého šumu**“ (Arlt, Arltová, 2005, s. 34). Taková časová řada nevykazuje žádný trend a neexistuje v ní ani žádná opakující se část (např. sezónnost nebo cyklus).

Pak je pro proces bílého šumu autokorelační funkce $\rho_k = 1$ pro $k = 0$ a $\rho_k = 0$ pro $k \neq 0$, kde autokorelační funkcí (ACF) rozumíme

$$\rho_k = \frac{\gamma_k}{D(a_t) \cdot D(a_{t-k})}, \quad (41)$$

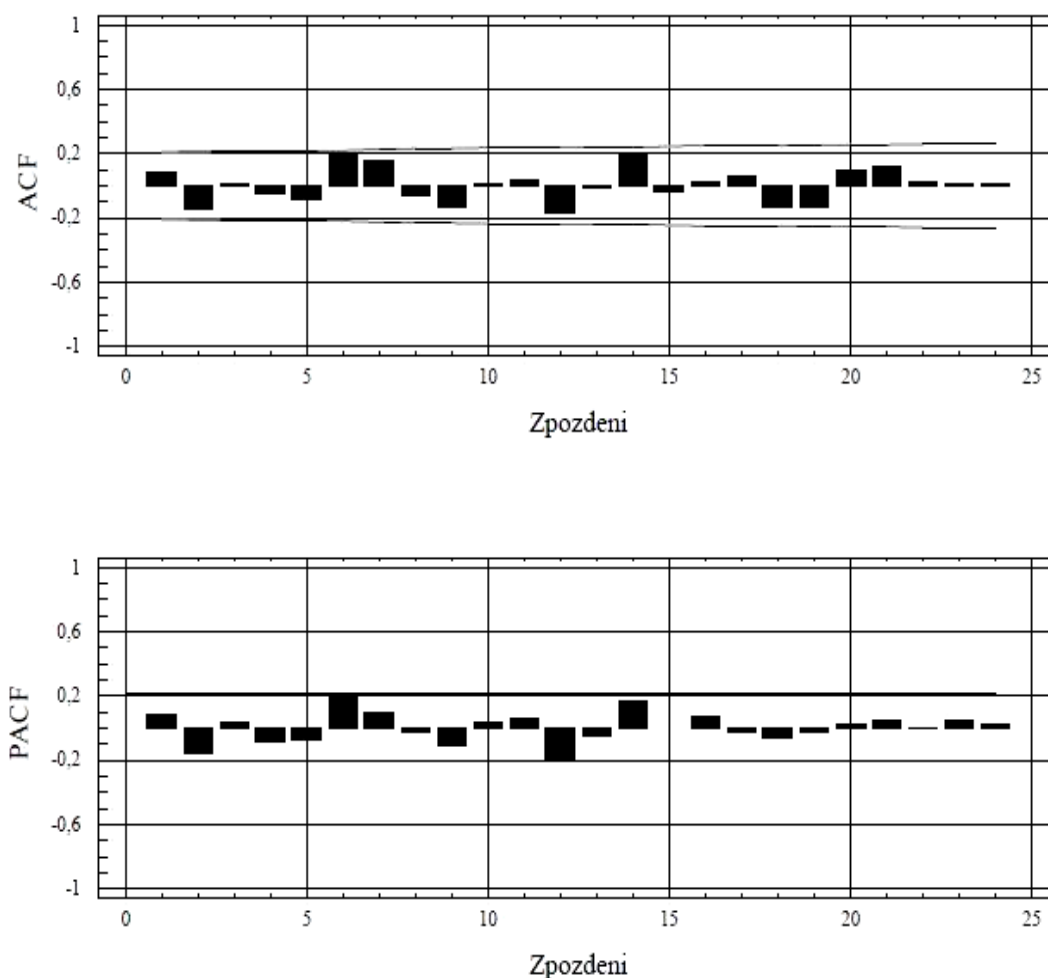
pro niž platí v případě stacionárního procesu $D a_t = D a_{t-k} = \gamma_0$. Obecné vlastnosti ACF jsou uvedeny v knize autorů Arlt a Arltová, 2005, s. 32.

Kromě vlastností ACF, které vyjadřují vztah mezi dvěma náhodnými veličinami lze konstruovat ještě parciální autokorelační funkci (PACF), která podává informaci o korelaci veličin a_t a a_{t-k} očištěnou o vliv veličin ležících mezi nimi. Značí se ϕ_{kk} a pro její vyjádření viz knihu Arlta a Arltové (2005), s. 33.

Pro proces bílého šumu je nutné, aby i PACF měla stejné vlastnosti, jako ACF, tedy $\phi_{kk} = 1$ pro $k = 0$ a $\phi_{kk} = 0$ pro $k \neq 0$.

Pro využití v této práci je však nutné nalezení určitého klíče k určení, zda je možné daný proces (jinak řečeno vývoj časové řady) považovat za bílý šum či ne (nulové hodnoty ACF a PACF pro $k \neq 0$ jsou pouze teoretickým předpokladem). V tomto případě bude vhodným nástrojem grafické znázornění vývoje ACF, popř. PACF, které je možné prostřednictvím tzv. korelogramu.

Obr. 6: Korelogram funkce ACF a PACF



Pozn.: Zpracováno v software Statgraphics for Windows 3.1. Pokud jednotlivé sloupce nepřekročí vyznačené intervaly (přibližně hodnoty -0,2, resp. 0,2), pak jsou splněny podmínky pro definici procesu bílého šumu.

Jak ve svém článku píše Arltová a Arlt, korelogramy autokorelační (ACF) a parciální autokorelační funkce (PACF) se používají většinou pro posouzení, zda řada reziduí má charakter tzv. bílého šumu. Je účelné použít jejich grafy při modelování časových řad. Jednotlivé sloupce autokorelační funkce vyjadřují sílu lineární závislosti mezi hodnotami

časové řady. První sloupec vyjadřuje sílu lineární závislosti mezi řadou a_t a řadou a_{t-1} , $t = 2, 3, \dots, T$, druhý sloupec vyjadřuje sílu lineární závislosti mezi řadou a_t a řadou a_{t-2} , $t = 3, 4, \dots, T$ atd. Hodnota autokorelační funkce je statisticky významná tehdy, překoná-li určitou mez stanovenou na zvolené hladině významnosti, což lze vyčíst např. na Obr. 6. vyznačenými intervaly.

První sloupec parciální autokorelační funkce je totožný s prvním sloupcem autokorelační funkce, vyjadřuje rovněž sílu lineární závislosti mezi řadou a_t a řadou a_{t-1} , $t = 2, 3, \dots, T$, druhý sloupec vyjadřuje sílu lineární závislosti mezi řadou a_t a řadou a_{t-2} , $t = 3, 4, \dots, T$, zohledňuje přitom přítomnost řady a_{t-1} , třetí sloupec vyjadřuje sílu lineární závislosti mezi řadou a_t a řadou a_{t-3} , $t = 4, 5, \dots, T$, zohledňuje přitom přítomnost řady a_{t-1} a a_{t-2} atd. (Arltová, Arlt, 1995).

Tento stručný návod jak hodnotit proces bílého šumu není v žádném případě vyčerpávající, cílem bylo představit jednoduché řešení, vhodné pro rychlé hodnocení časové řady. Pro hlubší analýzu časových řad, její nástroje a postupy je nutné využít další nástroje zmíněné v uvedených zdrojích (Arlt, Arltová, 2005; Arltová, Arlt, 1995). Pro účel této práce by však popis dalších metod byl spíše přebytečný.

2.2.5 Shrnutí metodických přístupů

V této podkapitole bylo základním cílem stručně popsat vybrané metodické přístupy, které je možné při modelování úmrtnosti využít. Stěžejní částí kapitoly je konstrukce vlastního modelu odhadu budoucího vývoje kohortních měr úmrtnosti, jejíž výsledky, po aplikaci na datech, budou diskutovány v následujících kapitolách této práce. Základem metody je predikce měr úmrtnosti založená na analýze poměrů měr úmrtnosti a jejich vývoji v čase.

Kromě zmíněné metody byly diskutovány také další přístupy, zmíněné již v diskusi s literaturou. Mezi základní patří Gompertz-Makehamův přístup exponenciálního růstu úmrtnosti s věkem, Lee-Carterova metoda, předpokládající (v čase neměnnou) závislost věkově specifických měr úmrtnosti na celkovém vývoji úmrtnosti, nebo různé modifikace navazující na Lee-Carterovu metodu, které jsou založené např. na rozšíření modelu o kohortní faktor nebo nahrazení predikce vývoje celkové úrovně úmrtnosti jinými charakteristikami, jako je třeba střední délka života při narození.

Mezi nejnovější přístupy patří metody modelování úmrtnosti na základě diferencí nebo poměrů měr úmrtnosti, ať již v kohortním nebo transverzálním přístupu. To jen potvrzuje skutečnost, že předpoklady v práci navrženého modelu nejsou ojedinělým názorem mezi odbornou veřejností a její rozvoj by mohl najít uplatnění např. v aktuáрске praxi.

Kapitola 3

Analýza dat a výsledky aplikace navržené metodologie predikce úmrtnosti

Struktura této kapitoly je do značné míry podřízena posloupnosti metodologie odhadu měř úmrtnosti, popsané v kapitole předchozí. Na úvod je provedena stručná analýza dat specifických měř úmrtnosti s ilustrací vlivu kohorty a období na úroveň úmrtnosti. Data pochází z HMD (viz podkapitola 2.1 Zdrojová data) za dvě země, Českou republiku a Švédsko.

V další části jsou diskutovány výsledky rozdílů a poměrů měř úmrtností v souvislosti s jejich vlastnostmi vhodnými pro odhady budoucích měř úmrtnosti, zejména jde o hodnocení variability a trendu vývoje hodnot rozdílů a poměrů měř úmrtnosti v čase, který je pro následné odhady budoucích měř úmrtnosti zásadní.

V závěrečné části kapitoly jsou prezentovány výsledky aplikace navržené metodiky predikce úmrtnosti v podobě odhadů specifických měř úmrtnosti a tabulkových funkcí. Výsledky jsou porovnány s empirickými hodnotami již vymřelých kohort, čímž je ověřena přesnost modelu. Pro dosud žijící kohorty je proveden odhad tabulkových funkcí do budoucna.

V kapitole je pro vizualizaci výsledků využito softwaru SAS (verze 9.2) pro 3D grafy, softwaru Statgraphics for Windows (verze 3.1) pro korelogramy a balíčku Microsoft Office Excel (verze 2007) pro ostatní grafické výstupy.

3.1 Analýza vývoje měř úmrtnosti České republiky a Švédska

Vývoj úmrtnosti ve vyšším věku (např. nad 30 let) se teoreticky řídí zákonem, který ve své teorii uvažuje i Gompertz (1825) aj. V předchozí kapitole byl tento zákon vyjádřen vztahem (25), tedy $m_{x,z} > m_{x-1,z}$. Pochopitelně lze tento vztah přepsat i do podoby transverzální, tedy

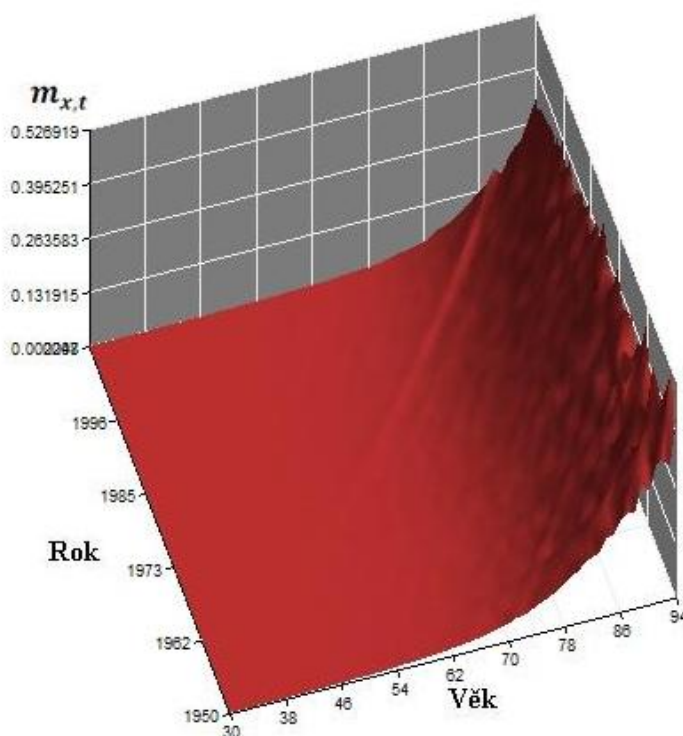
$$m_{x,t} > m_{x-1,t} , \quad (42)$$

což se dá interpretovat tak, že míry úmrtnosti s věkem rostou (ať již v kohortním nebo transverzálním vyjádření).

Jestliže se pokusíme tuto teorii ověřit na empirických datech, zjistíme, že průběh úmrtnosti podle věku je (v závislosti na pohledu) narušován buď kohortními vlivy nebo vlivy transverzálními (vlivy období).

Graficky je tento fakt patrný z obr. 7, kde jsou zobrazeny míry úmrtnosti z transverzálního pohledu, tedy podle 3. hlavních souborů událostí. Z průběhů měř úmrtnosti lze pozorovat, že předpoklad o rostoucích hodnotách $m_{x,t}$ s věkem je obhajitelný s výjimkou výkyvů, které jsou způsobeny nejspíše právě kohortním efektem. Ten lze pozorovat v podobě vln, které vedou přibližně pod úhlem 45 stupňů mezi osami roku a věku. Patrně nejvýraznější vlna je viditelná od věku přibližně 60 let a roku 1975. Je způsobena kohortou narozenou v období 1. světové války ($1975 - 60 = 1915$). Další výraznější kohortní vlivy jsou patrné pro ještě starší kohorty, ale podobně jako efekt kohorty narozené v období 1. světové války, jsou viditelné až od vyšších věků, do přibližně 50 let věku není žádný vliv na úmrtnost z tohoto grafického pohledu patrný.

Obr. 7: Míry úmrtnosti, transverzální pohled, 3. hlavní soubory událostí, 1950–2008, 30–95 let, Česká republika, ženy



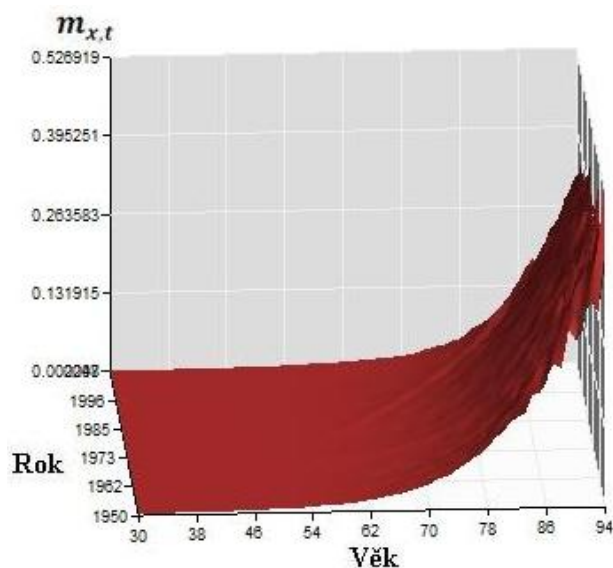
Pozn.: Zpracováno v softwaru SAS 9.2

Zdroj: Human Mortality Database (2012)

Proti tomu lze pozorovat výraznější vlivy efektu období, tedy roku. Jsou to všechny vlny, které jsou rovnoběžné s osou věku. Tyto vlny mohou být způsobeny změnami úmrtnostních faktorů v konkrétních letech (nebo i delších obdobích). Známa je např. zhoršená úroveň úmrtnosti v České republice v 60. a 70. letech minulého století (Srb, 1972 nebo Hulíková Tesárková, Mazouch 2013).

Obr. 8 pak na stejných datech z jiného pohledu (rotace obrázku) ukazuje, že úroveň úmrtnosti se celkově mění, dochází k postupnému „naklání plochy“ znázorňující úroveň úmrtnosti, což svědčí o poklesu celkové úrovně úmrtnosti v čase. Podobné výsledky jsou patrné také z kohortního pohledu, který bude však demonstrován, z důvodu delších časových řad, na datech Švédska.

Obr. 8: Míry úmrtnosti, transverzální pohled, 3. hlavní soubory událostí, 1950–2008, 30–95 let, Česká republika, ženy – jiná rotace

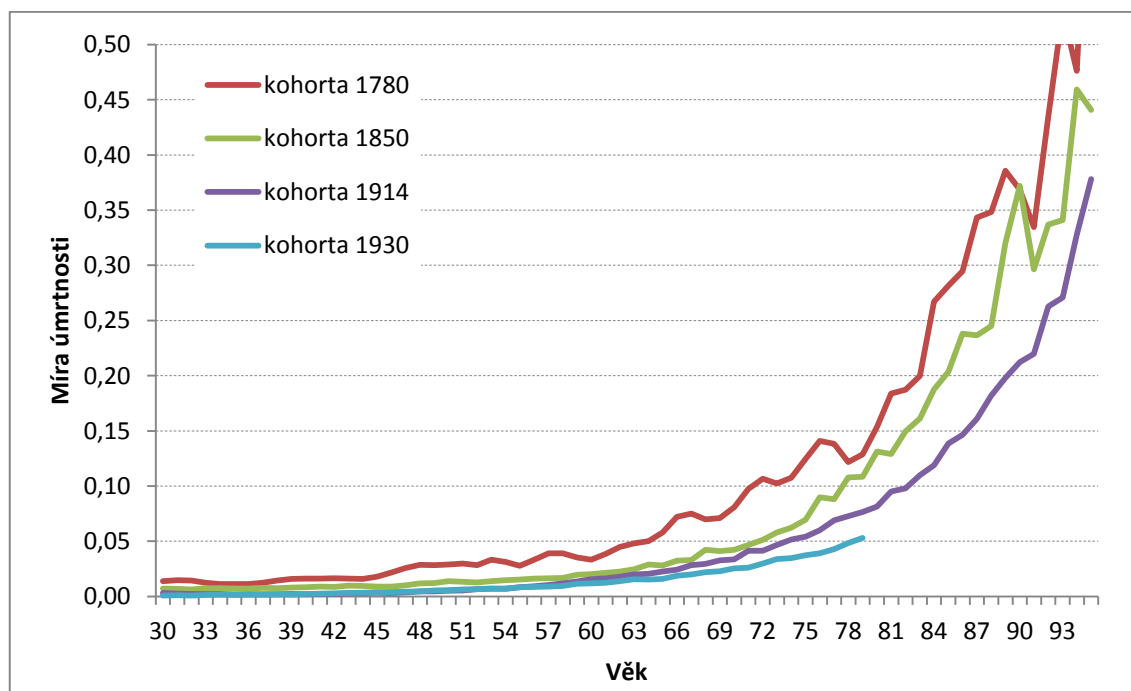


Pozn.: Zpracováno v softwaru SAS 9.2

Zdroj: Human Mortality Database (2012)

Na Obr. 9 je vidět, že předpoklad rostoucí míry úmrtnosti s věkem platí i v případě kohortního přístupu. Křivky mladších kohort jsou také hladší, což znamená, že zde existuje menší vliv transverzálního faktoru. Variabilita ve vyšších věcích nemusí být nutně zapříčiněna efektem období, ale může být způsobena také nízkým počtem událostí a exponovaných osob, kdy i malé výkyvy v počtech zemřelých mohou ovlivnit specifickou míru úmrtnosti.

Obr. 9: Míry úmrtnosti, kohortní pohled, kohorty 1780, 1850, 1914 a 1930, 30–95 let, 1. hlavní soubory událostí, Švédsko, muži



Zdroj: Human Mortality Database (2012)

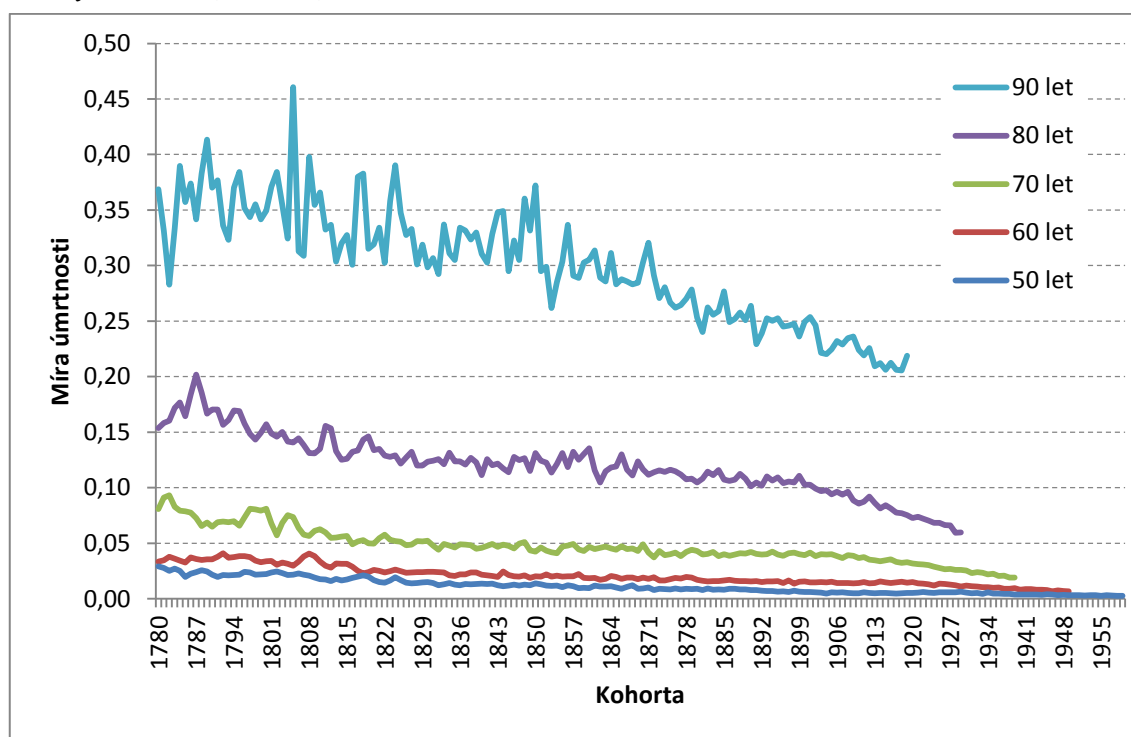
Lze také pozorovat, že hodnoty úmrtnosti pro jednotlivé kohorty postupně klesají, tedy teoreticky platí, že v dlouhém období

$$m_{x,z} > m_{x,z+1}. \quad (43)$$

Z pohledu celé kohorty se dá hovořit o procesu rektangularizace, který popisují ve svém příspěvku např. Burcin a kol. (2009). Z průběhu části křivky kohorty narozené v roce 1930 je patrné, že její hodnoty měr úmrtnosti, pokud nenastane žádná výrazná změna úmrtnostních poměrů, což by však nejspíš zasáhlo i ostatní kohorty, budou i v dalších věcích pod úrovní úmrtnosti kohorty narozené v roce 1914.

Obr. 10 znázorňuje vývoj specifických měr úmrtnosti pro vybrané věky. Je zde patrný výrazný pokles všech sledovaných měr v čase. V případě nejstarších kohort je zřetelná vysoká variabilita, která však s postupem času klesá. Přesto lze pozorovat, že variabilita s věkem roste, tedy starší věkové skupiny mají vyšší variabilitu úmrtnosti (důvody viz výše).

Obr. 10: Vývoj specifických měr úmrtnosti ve věku 50, 60, 70, 80 a 90 let, 1. hlavní soubory událostí, kohorty 1780–1959, Švédsko, muži



Zdroj: Human Mortality Database (2012)

Z hlediska modelování měr úmrtnosti jsou tyto informace podstatné zejména z několika důvodů. Prvním je skutečnost, že i empirická data potvrzují, že platí teoretické předpoklady jako je růst měr úmrtnosti s věkem, tedy platnost vztahu (25), resp. i (42) v případě transverzálního přístupu a toto pravidlo se nemění v čase (je patrné pro všechny roky v případě transverzálního přístupu i pro všechny kohorty v případě longitudinálního přístupu).

Přesto lze pozorovat skutečnost, že intenzita úmrtnosti se v čase mění. Dochází ke snižování úmrtnosti ve všech věkových skupinách, což znamená, že zde existuje trend, v tomto případě klesající. Tuto skutečnost lze pozorovat v dlouhém období u všech věkových skupin.

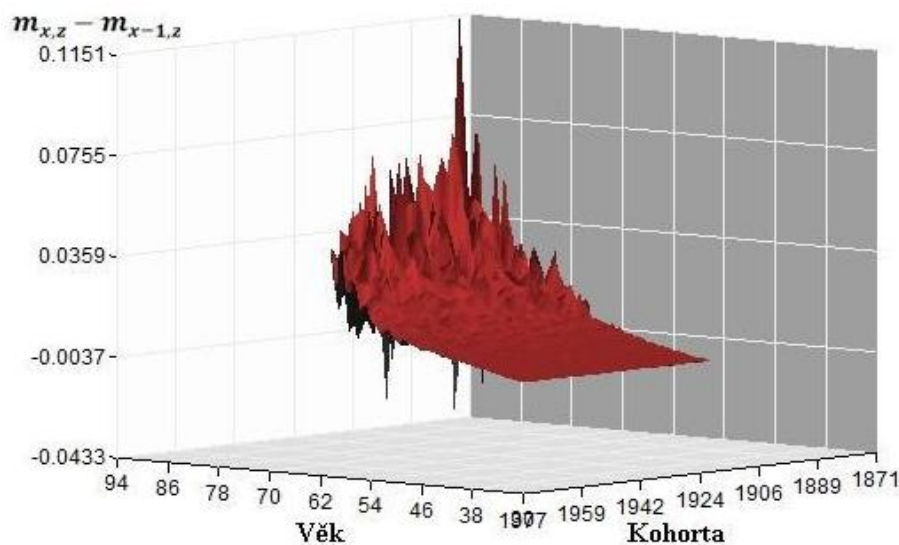
Protože je navrhovaná metoda predikce úmrtnosti založená na modelování vztahů mezi specifickými měrami úmrtnosti, je vhodné, aby byla provedena i analýza těchto ukazatelů, zejména jejich vývoje v čase.

3.2 Analýza vývoje poměrů a rozdílů měr úmrtnosti

V této části práce bude již pozornost soustředěna výhradně na ukazatele založené na kohortním přístupu. Na Obr. 11 je znázorněn vývoj rozdílů kohortních specifických měr úmrtnosti, jak byl definován ve výrazu (24). V nízkých věcích jsou hodnoty diferencí velmi malé, blízké nule. S růstem věku se jejich hodnoty zvyšují a maxima dosahují v nejvyšších věcích.

Tato skutečnost není nijak překvapivá, protože absolutní difference specifických měr úmrtnosti jsou závislé na úrovni hodnot samotných specifických měr, které (jak bylo ukázáno v předchozí části) s věkem rostou, a rychlosti jejich nárůstu s věkem. Musí tak růst také absolutní difference. Z obrázku je již méně patrné, že v čase (mezi kohortami) hodnoty absolutních diferencí pro konkrétní věky také klesají. Opět je to způsobeno vývojem absolutních hodnot specifických měr úmrtnosti, které v čase klesají (opět viz předchozí část). Tato skutečnost bude demonstrována ještě na dalších grafech.

Obr. 11: Rozdíly měr úmrtnosti $d_{x,z}$ kohorty 1871–1977, 30–95 let, 1. hlavní soubory událostí, Česká republika, ženy

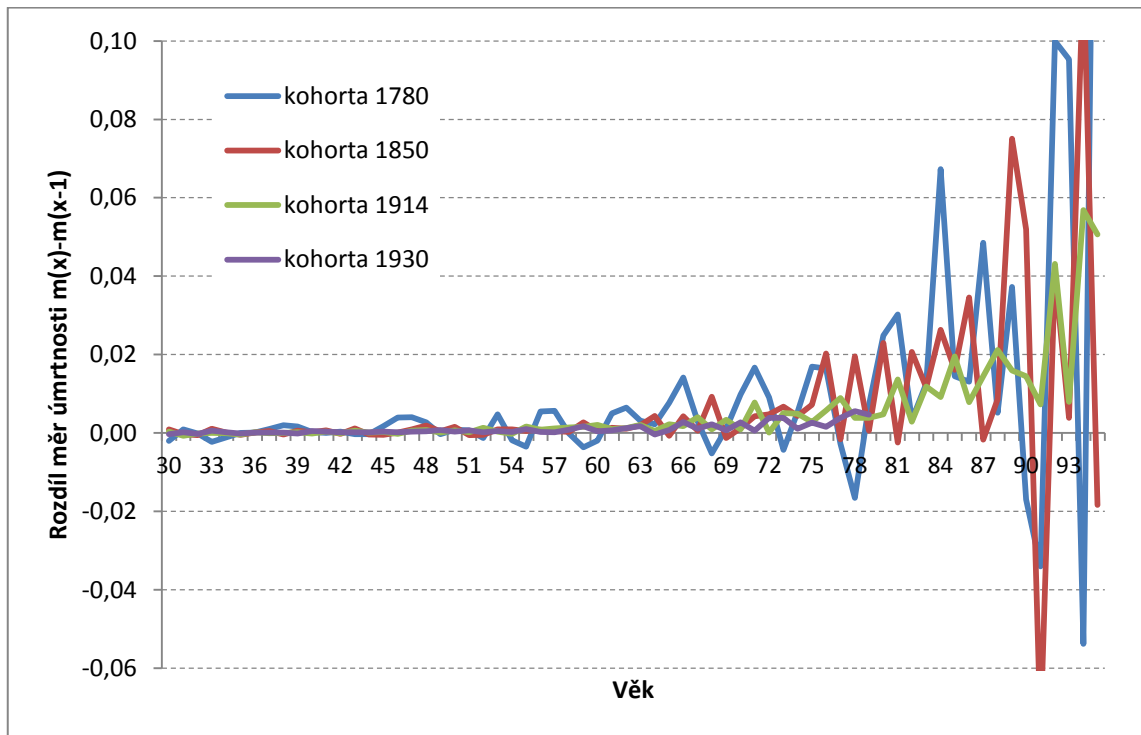


Pozn.: Zpracováno v softwaru SAS 9.2. Hodnota $d_{x,z} = m_{x,z} - m_{x-1,z}$.

Zdroj: Human Mortality Database (2012)

Pro vybrané švédské kohorty na Obr. 12 a 13 se potvrzují předchozí závěry. Diference specifických měr nabývají v nízkých věcích malých hodnot a s růstem absolutních hodnot (viz Obr. 9) rostou i hodnoty absolutních diferencí. Také je patrné, že průběh diferencí úmrtnosti se liší mezi jednotlivými kohortami, nejnižších hodnot dosahují kohorty nejmladší. Opět je to zapříčiněno vlivem klesajících hodnot měr úmrtnosti v čase (mezi kohortami). Tato skutečnost je zřejmá také z Obr. 13, kde difference měr pro vybrané věky v čase klesá (a klesá také její variabilita – opět díky klesající absolutní úrovni hodnot).

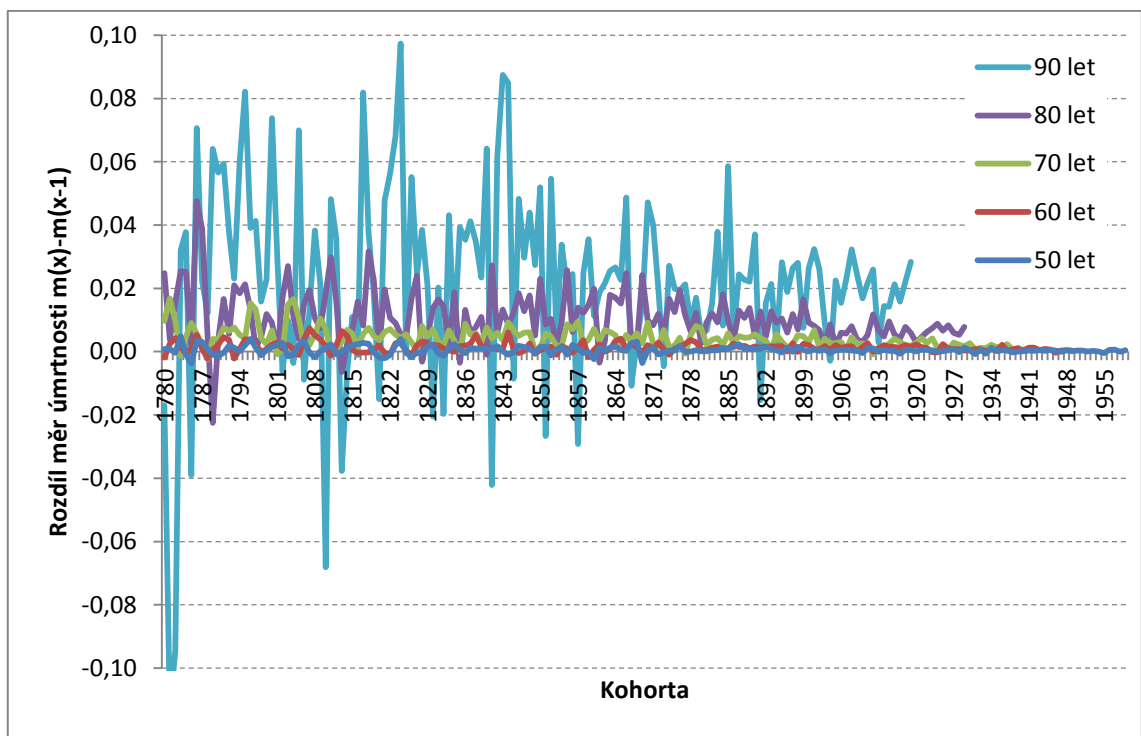
Obr. 12: Rozdíly měr úmrtnosti $d_{x,z}$ kohorty 1780, 1850, 1914 a 1930, 30–95 let, 1. hlavní soubory událostí Švédsko, muži



Pozn.: Hodnota $d_{x,z} = m_{x,z} - m_{x-1,z}$.

Zdroj: Human Mortality Database (2012)

Obr. 13: Rozdíly měr úmrtnosti $d_{x,z}$ pro $x = 50, 60, 70, 80$ a 90 let, 1. hlavní soubory událostí, kohorty 1780–1959, Švédsko, muži

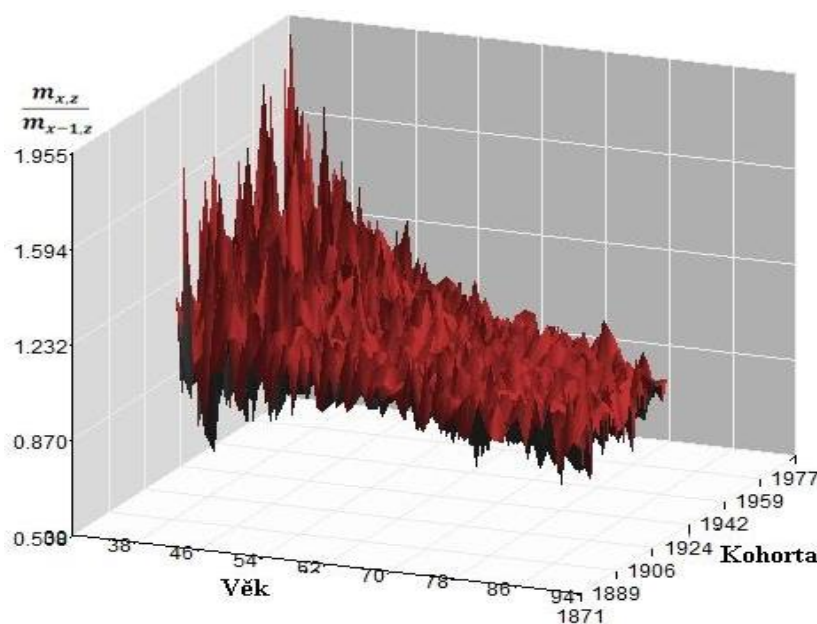


Pozn.: Hodnota $d_{x,z} = m_{x,z} - m_{x-1,z}$.

Zdroj: Human Mortality Database (2012)

V případě poměrů kohortních specifických měr úmrtnosti $ar_{x,z}$, spočtených na základě vztahu (5), které jsou zobrazeny na Obr. 14, můžeme vidět na první pohled velmi nestabilní vývoj ukazatele v čase. Tento nestabilní vývoj však nevykazuje, na rozdíl od diferencí specifických měr, žádný trend. Lze vidět, že se s věkem mění variabilita ukazatele, která je vyšší v nejmladších a nejstarších věkových skupinách. Opět se jedná o věkové skupiny, ve kterých i malé počty událostí mohou způsobit velké výkyvy. Ve věcích, které můžeme označit za čtenější z hlediska počtu událostí, je variabilita nejnižší.

Obr. 14: Poměry měr úmrtnosti $ar_{x,z}$ kohorty 1871–1977, 30–95 let, 1. hlavní soubory událostí, Česká republika, ženy



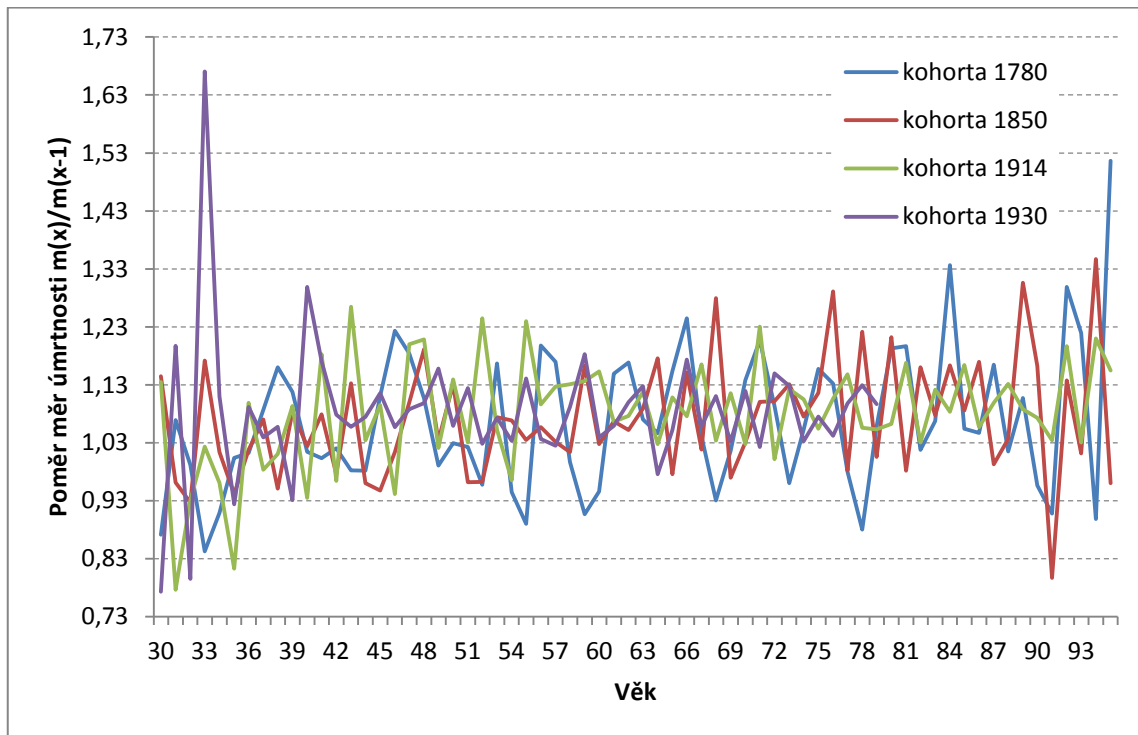
Pozn.: Zpracováno v softwaru SAS 9.2. Hodnoty $ar_{x,z} = \frac{m_{x,z}}{m_{x-1,z}}$.

Zdroj: Human Mortality Database (2012)

Při detailnějším pohledu na vybrané kohorty Švédska na Obr. 15 lze pozorovat, že časové řady koeficientů růstu (poměr měr úmrtnosti mezi po sobě jdoucími věky) nevykazuje žádný výrazný růst nebo pokles (přesto se v rámci další části této kapitoly přesvědčíme, že velmi mírný růst se v těchto datech skrývá). Výjimkou může být nejnižší věk do 35 let, kdy je velmi vysoká variabilita a hodnoty tudíž velmi kolísají. Mezi kohortami je patrný velmi mírný pokles variability.

Srovnáme-li vývoj poměrů měr úmrtnosti $ar_{x,z}$ pro konkrétní věky v čase na Obr. 16, zjistíme, že tyto hodnoty jsou velmi variabilní. Jejich variabilita roste s věkem, tedy největší variabilitu vykazují nejstarší věkové skupiny (a podobně tomu bude u nejmladších věkových skupin), což již bylo zdůvodněno. Z hlediska trendu však můžeme říci, že ani jedna z řad (ani jeden věk) nevykazuje žádný systematický růst či pokles. Tato skutečnost vede k závěru, že pro odhad budoucích hodnot úmrtnosti je nejlepší použít právě poměr specifických měr úmrtnosti, protože vykazuje sice vysokou variabilitu (srovnáme-li to např. s diferencemi), ale tato je v čase, pro konkrétní hodnotu věku x , přibližně konstantní a hlavně konstantní je také trend vývoje hodnot poměrů měr úmrtnosti v čase.

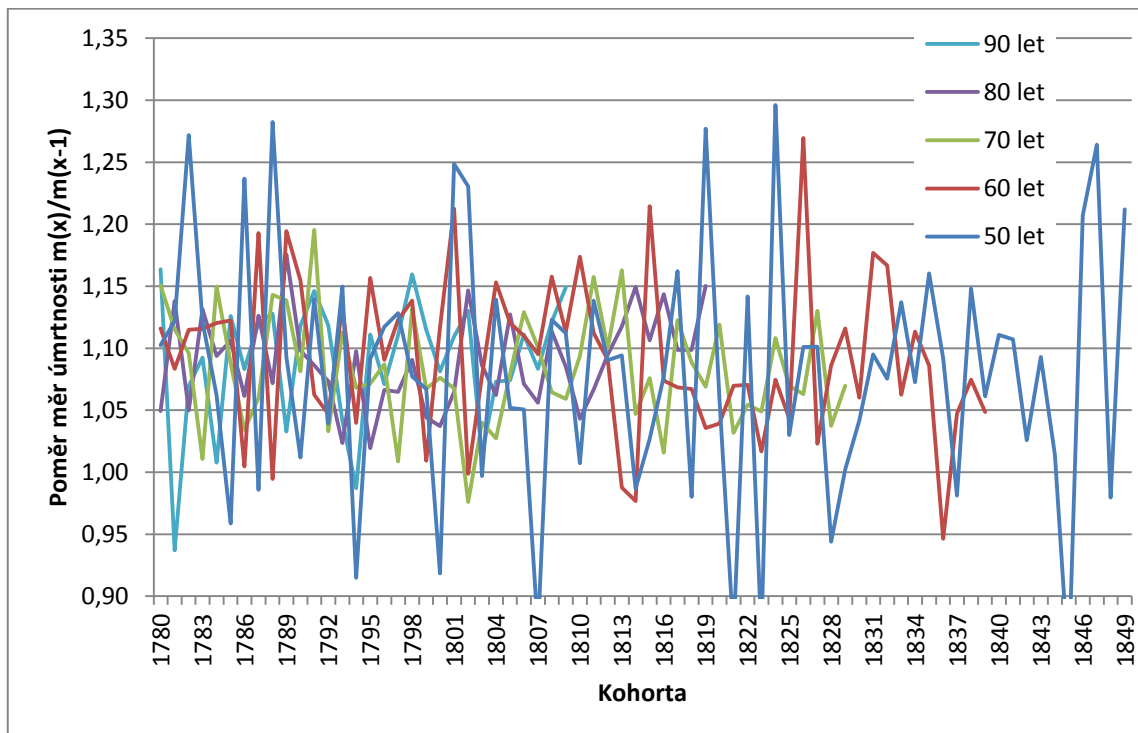
Obr. 15: Poměry měr úmrtnosti $ar_{x,z}$ kohorty 1780, 1850, 1914 a 1930, 30–95 let, 1. hlavní soubory událostí, Švédsko, muži



Pozn.: Hodnoty $ar_{x,z} = \frac{m_{x,z}}{m_{x-1,z}}$.

Zdroj: Human Mortality Database (2012)

Obr. 16: Poměry měr úmrtnosti $ar_{x,z}$ pro $x = 50, 60, 70, 80$ a 90 let, 1. hlavní soubory událostí, kohorty 1780–1949, 1. hlavní soubory událostí, Švédsko, muži

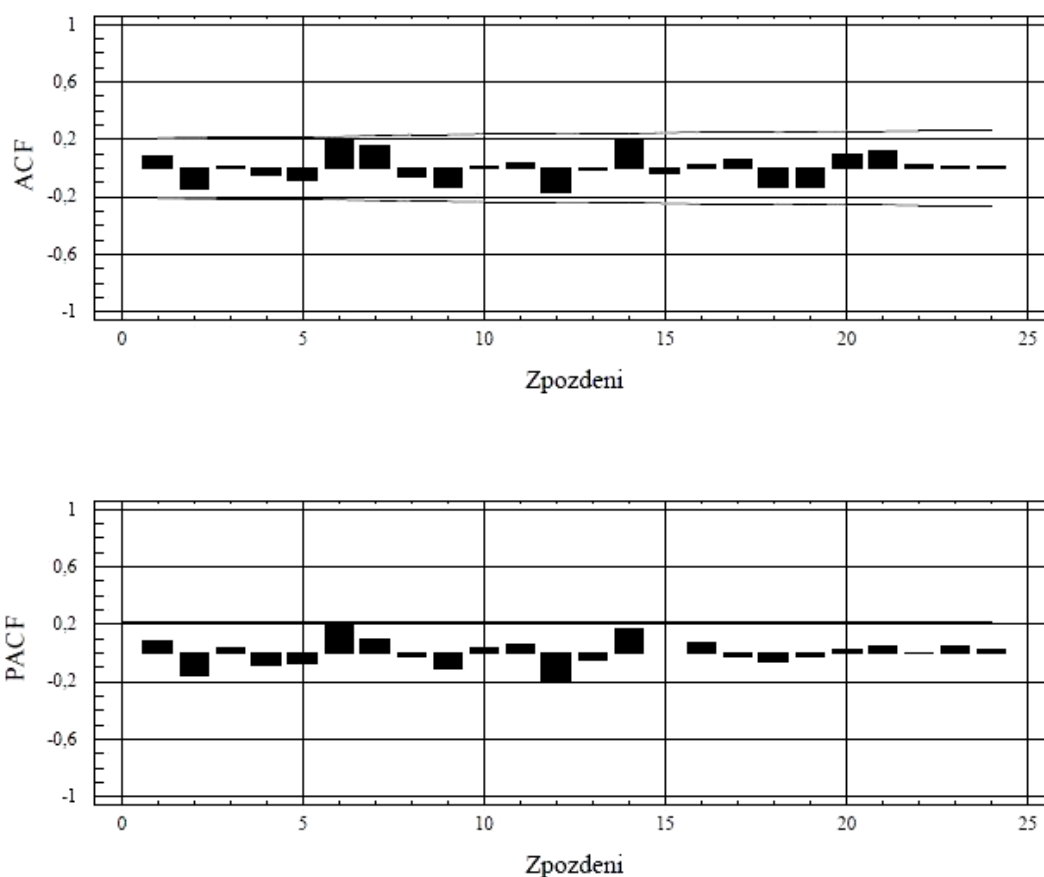


Pozn.: Hodnoty $ar_{x,z} = \frac{m_{x,z}}{m_{x-1,z}}$.

Zdroj: Human Mortality Database (2012)

Důkaz stacionarity pouhým vizuálním posouzením na vybraných věcích není pochopitelně dostačující. Je proto nutné podrobit tyto časové řady důkladnějšímu zkoumání v podobě analýzy popsané v kapitole 2.2.4. Pro každý věk x je nutné analyzovat časovou řadu $ar_{x,z}$ a zjistit hodnoty její ACF a PACF. Nejjednodušší způsob je (jak bylo popsáno v 2.2.4) využít korelogram funkcí ACF a PACF.

Obr. 17: Odhad ACF a PACF pro časovou řadu $ar_{x,z}$ pro $x = 60$, kohorty 1780–1949, Švédsko, muži



Pozn.: Zpracováno v softwaru Statgraphics for Windows 3.1

Zdroj: Human Mortality Database (2012)

Z obou korelogramů je patrné, že žádná hodnota ACF ani PACF nepřekračuje výrazně horní nebo dolní interval spolehlivosti, tedy je možné uvažovat, že hodnoty ACF i PACF se neliší od nuly (na hladině významnosti $p = 0,05$). Hodnoty $ar_{60,z}$ jsou tedy v čase stacionární a jejich vývoj lze charakterizovat jako proces bílého šumu. V tom případě je jejich nejlepší odhad budoucí úrovně střední hodnota hodnot $ar_{60,z}$. Pro ostatní hodnoty x jsou výsledky korelogramů $ar_{x,z}$ v příloze č. 1. Je z nich patrné, že hodnoty žádné časové řady nepřekračují intervaly spolehlivosti a že tedy všechny časové řady $ar_{x,z}$ pro věky 60 a více let, pro muže i ženy ve Švédsku můžeme považovat za procesy bílého šumu a predikovat je na základě střední hodnoty.

Pokud budeme hodnotit vývoj průměrných hodnot $ar_{x,z}$, tedy $ar_{x,z,n,s}$, pak zjistíme, že hodnoty se mezi jednotlivými věky x mění. Hodnoty $ar_{x,z,n,s}$ byly spočteny na základě vzorce

(28) a jejich výsledky jsou v Obr. 18. Spolu s hodnotami $ar_{x,z,n,s}$ jsou zde zobrazeny také směrodatné odchylky hodnot $ar_{x,z}$ pro jednotlivé věkové skupiny x spočtené jako

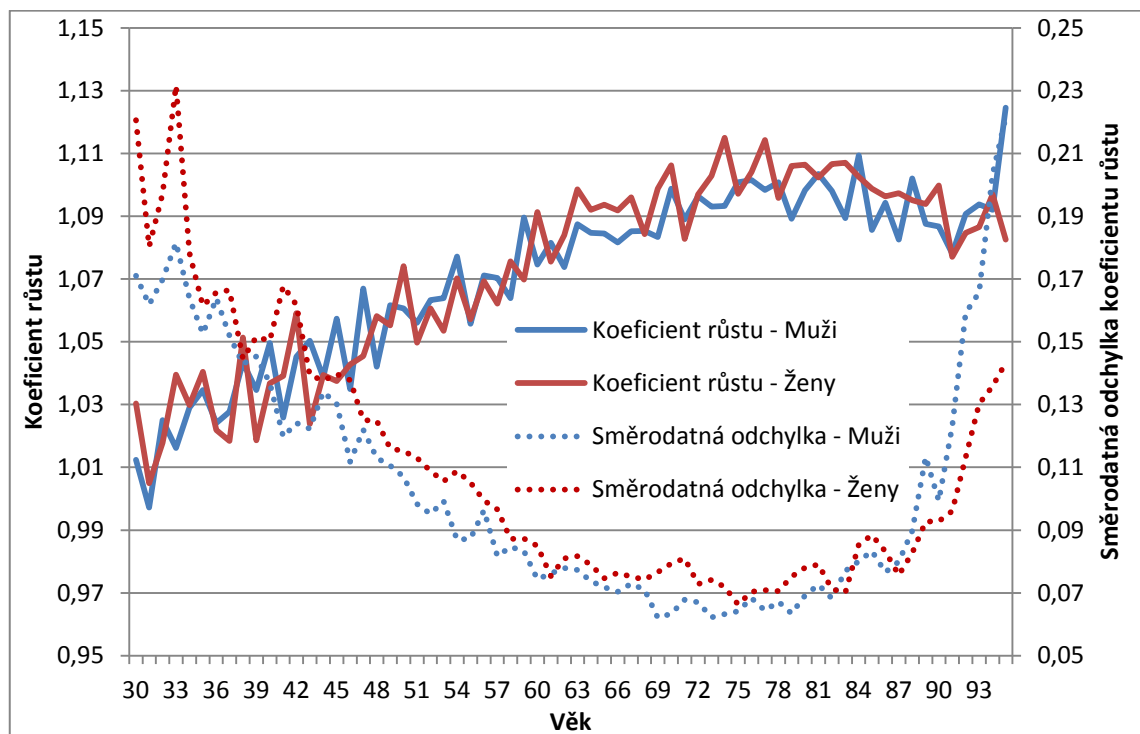
$$\frac{\frac{b}{z=a} ar_{x,z} - ar_{x,z}^2}{a-b}, \text{ pro } x = s, \dots, \omega-1, \quad (44)$$

kde $ar_{x,z}$ je poměr měr úmrtnosti, x je věk, z je kohorta, $ar_{x,z}$ je průměr vypočtený z hodnot $ar_{x,z}$ pro kohorty narozené mezi roky a a b , a a je rok narození první kohorty zahrnované do výpočtu a b je rok narození poslední kohorty zahrnované do výpočtu.

Z obrázku je patrný růst hodnot $ar_{x,z,n,s}$ až do věku přibližně 80 let, následuje konstantní úroveň hodnoty $ar_{x,z,n,s}$ a po ní pokles hodnot $ar_{x,z,n,s}$. Tyto hodnoty lze interpretovat jako koeficient c z vyjádření Gompertzovy funkce (Gompertz, 1825). V tomto případě je však zřejmé, že c není konstantní, tedy bylo by lepší vyjádřit jej jako c_x , tedy věkově specifický koeficient růstu.

Předpoklad, že je nutné spočítat pro každý věk x individuální hodnotu $ar_{x,z,n,s}$ se podařilo prokázat. Vývoj směrodatné odchylky i v tomto případě odpovídá variabilitě specifických měr úmrtnosti, která dosahuje nejnižších hodnot v nízkých a nejvyšších věcích (důvody opět viz výše).

Obr. 18: Průměrné hodnoty $ar_{x,z}$ a směrodatné odchylky $ar_{x,z}$ věk 30–95 let, kohorty 1780–1979, Švédsko



Pozn.: Průměrné hodnoty $ar_{x,z}$ spočteny na základě vzorce (28) z důvodu porovnání s ukazateli variability. Hodnoty koeficientů růstu jsou na levé svislé ose a hodnoty směrodatných odchylek koeficientů růstu na pravé svislé ose.

Zdroj: Human Mortality Database (2012)

3.3 Výsledky aplikace navrhované metodologie predikce úmrtnosti

Dříve než budou hodnoceny samotné výsledky aplikace metodologie predikce úmrtnosti, bude dobré na konkrétním příkladu demonstrovat jednotlivé kroky metody. Po analýze dat, která ukázala, že předpoklady navrhovaného modelu predikce jsou stanoveny správně lze přistoupit k odhadu budoucích měr úmrtnosti.

V první fázi je tedy nutné spočítat hodnoty $ar_{x,z}$ z hodnot jednotlivých měr úmrtnosti $m_{x,z}$ na základě vzorce (5)

$$ar_{x,z} = \frac{m_{x,z}}{m_{x-1,z}}$$

Tyto hodnoty je dobré spočítat za co nejdelší časové období. Praktická demonstrace odhadu budoucího vývoje úmrtnosti je provedena na švédské kohortě mužů, narozené v roce 1900 (viz Tab. 1) pro věky vyšší než 60 let.

Průměrné hodnoty poměrů měr úmrtnosti $ar_{x,z,n,s}$ budou počítány z předchozích 40 hodnot, tedy ve výrazu (27) $ar_{x,z,n,s}$ jsou parametry $z = 1900$, $n = 40$ a $s = 61$ (první věk pro který odhadujeme míru úmrtnosti $m_{x,z}$). Hodnoty věku, pro který bude proveden odhad nabývají hodnot $x = 61, 62, \dots, 96$.

Přestože byl výpočet $ar_{x,z,n,s}$ popsán v metodické části práce, bude dobré ještě důkladně okomentovat určitá úskalí celého výpočtu. Je třeba mít na paměti, že počet kohort, které se do výpočtu zahrnují, je velký a model má tak velmi dlouhou „paměť“.

Např. pro výpočet prvního průměrného poměru měr úmrtnosti $ar_{61,1900,40,61}$ dle vzorce (27) musíme znát hodnoty $ar_{61,z}$ pro kohorty narozené od roku 1860 do 1899. Pro hodnotu $ar_{62,1900,40,61}$ pak analogicky hodnoty $ar_{62,z}$ pro kohorty narozené 1859 až 1898, atd. Pokud budeme predikci měr úmrtnosti konstruovat např. do věku 100 let, pak pro hodnotu $ar_{100,1900,40,61}$ potřebujeme informace o $ar_{100,z}$ u kohort narozených 1821 až 1860.

Kromě skutečnosti, že model využívá informace o kohortách, které již mohou být (podle velikosti zvoleného n) v době konstrukce predikce úmrtnosti dávno vymřelé, je nutné disponovat informacemi o mírách úmrtnosti za poměrně dlouhé období (opět v závislosti na zvoleném n). Je-li $n = 40$, pak musíme znát míry úmrtnosti za 40 předchozích let (ve skutečnosti za 41, protože uvažujeme 1. hlavní soubory událostí).

Po výpočtu hodnot $ar_{x,z,n,s}$ je nutné odhadnout první hodnotu $bm_{x,z}$, která je základní hodnotou, od které se pak odhadují následující hodnoty $m_{x+1,z}$. Hodnota $bm_{x,z}$ je spočtena na základě vzorce (29) z empirických měr úmrtnosti $m_{x,z}$.

V Tab. 1 je tato hodnota v prvním řádku ve sloupci odhadovaných hodnot $m_{x,z}$, konkrétně hodnota 0,01513. Od této hodnoty budou následně odhadovány další hodnoty $m_{x+1,z}$, a to na základě vzorce (30) a (31). Nejprve je tedy hodnota 0,015 vynásobena koeficientem růstu $ar_{61,1900,40,61}$, který má hodnotu 1,091, čímž získáme hodnotu 0,016⁴. V dalším kroku je nutné vynásobit získanou hodnotu 0,016 hodnotou příslušného koeficientu růstu, která je 1,084, čímž se získá další hodnota $m_{x+1,1900}$, tentokrát již pro $x = 62$.

⁴ Hodnoty v tabulce jsou zaokrouhlené, výsledky tedy nejsou úplně přesné.

Tab. 1. Odhad budoucích specifických měr úmrtnosti a střední délky života, srovnání s empirickými hodnotami, kohorta 1900, muži, Švédsko

Věk	Koeficient růstu	Odhadovaná hodnota	Empirická hodnota	Odhadovaná hodnota	Empirická hodnota
x	$ar_{x,z,n,s}$	$m_{x,z}$	$m_{x,z}$	$e_{x,z}$	$e_{x,z}$
60	–	0,015130	0,015697	17,63	17,66
61	1,091	0,016513	0,015693	16,89	16,93
62	1,084	0,017898	0,018221	16,16	16,19
63	1,111	0,019878	0,021290	15,45	15,48
64	1,112	0,022094	0,022685	14,75	14,80
65	1,092	0,024126	0,023636	14,07	14,13
66	1,095	0,026410	0,027542	13,40	13,46
67	1,114	0,029420	0,031063	12,74	12,82
68	1,110	0,032670	0,034781	12,11	12,21
69	1,086	0,035472	0,036318	11,49	11,62
70	1,119	0,039703	0,039269	10,89	11,03
71	1,102	0,043737	0,047730	10,31	10,45
72	1,104	0,048288	0,047966	9,75	9,94
73	1,087	0,052509	0,054624	9,21	9,41
74	1,117	0,058644	0,059977	8,68	8,91
75	1,110	0,065092	0,066522	8,17	8,43
76	1,097	0,071379	0,071362	7,69	7,97
77	1,130	0,080651	0,079357	7,22	7,52
78	1,098	0,088518	0,088010	6,78	7,10
79	1,099	0,097323	0,093426	6,37	6,71
80	1,110	0,107981	0,102589	5,97	6,32
81	1,112	0,120076	0,112602	5,59	5,95
82	1,100	0,132092	0,123378	5,24	5,60
83	1,086	0,143416	0,131689	4,91	5,27
84	1,099	0,157576	0,145626	4,59	4,94
85	1,100	0,173307	0,155398	4,29	4,64
86	1,103	0,191177	0,169212	4,01	4,34
87	1,096	0,209523	0,191000	3,75	4,05
88	1,094	0,229185	0,203012	3,52	3,80
89	1,083	0,248136	0,223078	3,30	3,54
90	1,086	0,269427	0,249482	3,09	3,31
91	1,106	0,297955	0,272690	2,90	3,11
92	1,057	0,315087	0,278450	2,73	2,93
93	1,077	0,339196	0,338008	2,57	2,72
94	1,125	0,381710	0,316532	2,41	2,62
95	1,110	0,423705	0,361019	2,32	2,42
96	1,028	0,435691	0,433076	2,30	2,26

Zdroj: Human Mortality Database (2012)

Pozn.: Hodnoty $ar_{x,z,n,s}$ spočteny dle vzorce (27) pro $z = 1900$, $n = 40$, $s = 61$ a $\alpha = 0,9$.

Analogicky je možné postupovat dále až po věk nejvyšší (tedy 96 let). Odhadnuté hodnoty měř úmrtnosti je pak možné použít jako základ konstrukce úmrtnostní tabulky a pro hodnocení úmrtnostních poměrů využít funkcí úmrtnostní tabulky.

Z výsledků prezentovaných v Tab. 1 je vidět, že hodnoty odhadnutých měř úmrtnosti pro kohortu narozenou v roce 1900 se od empirických hodnot příliš neliší. V tomto případě tedy model velmi dobře vystihuje realitu, což se projevuje také v hodnotách střední délky života, jejíž rozdíl je ve věku 60 let téměř nepatrný (s vědomím, že i empirické hodnoty mohou vykazovat určitou míru „nepřesnosti“ – viz článek Pastora (2007) nebo Koschina (1981)) a lze jej tedy považovat za nevýznamný.

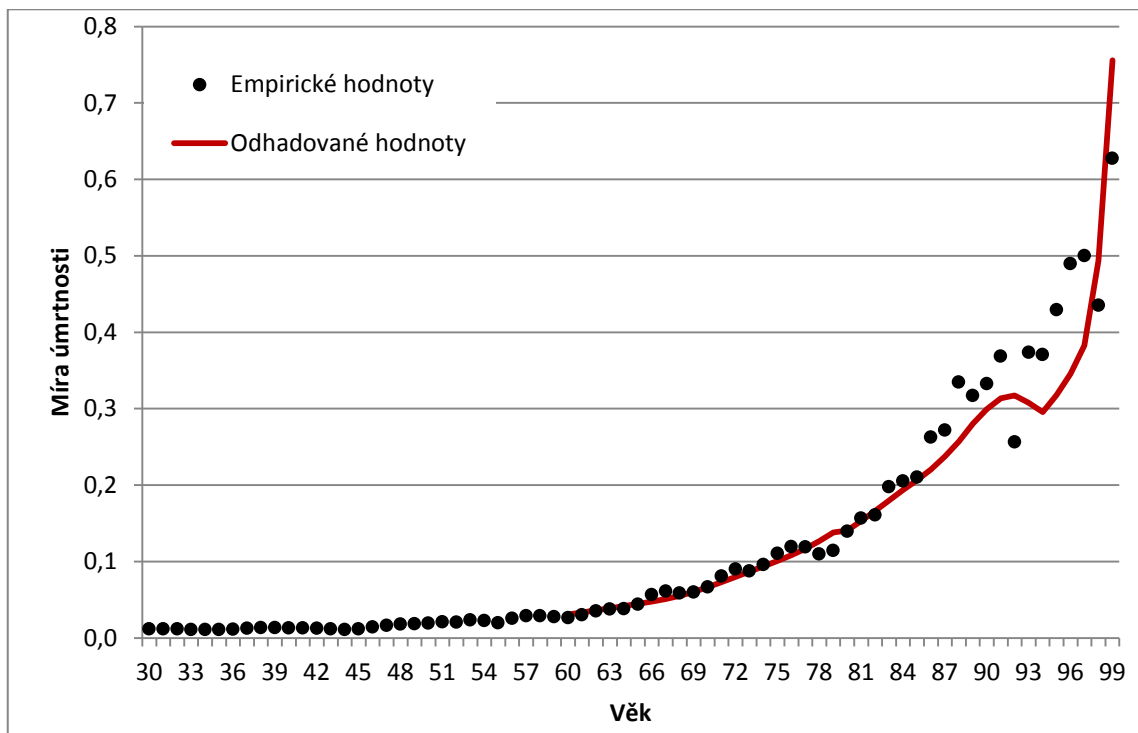
Prezentovaný postup navrhované metodiky, demonstrováný na kohortě mužů narozených v roce 1900, byl zopakován pro všechny kohorty od roku 1761 až do roku 1945 pro švédské muže i ženy. V následující části budou vybrané výsledky prezentovány formou různých metod.

Pro analýzu výsledků aplikace navrženého modelu není nutné využít vždy jen tabulkových funkcí. Pro detailnější informaci lze empirické a odhadované hodnoty vynést do grafu v podobě specifických měř úmrtnosti.

Na Obr. 19 jsou porovnány hodnoty empirických a odhadovaných hodnot měř úmrtnosti pro kohortu švédských žen narozenou v roce 1780. Zde je patrné, že v určitých věcích jsou empirické hodnoty vyšší než odhadované a v některých věcích naopak nižší. Celkově pro kohortu 1780 je rozdíl mezi empirickou a odhadovanou střední délkou života 60leté osoby přibližně 0,15 roku ve prospěch odhadované míry.

Na grafu je patrné, že od přibližně 85 let věku model empirickou úmrtnost podhodnocuje, tedy odhaduje lepší úmrtnostní poměry, což zvyšuje pochopitelně také ukazatel střední délky života.

Obr. 19: Srovnání empirických a odhadovaných hodnot $m_{x,20}$ kohorta 1780, Švédsko, ženy



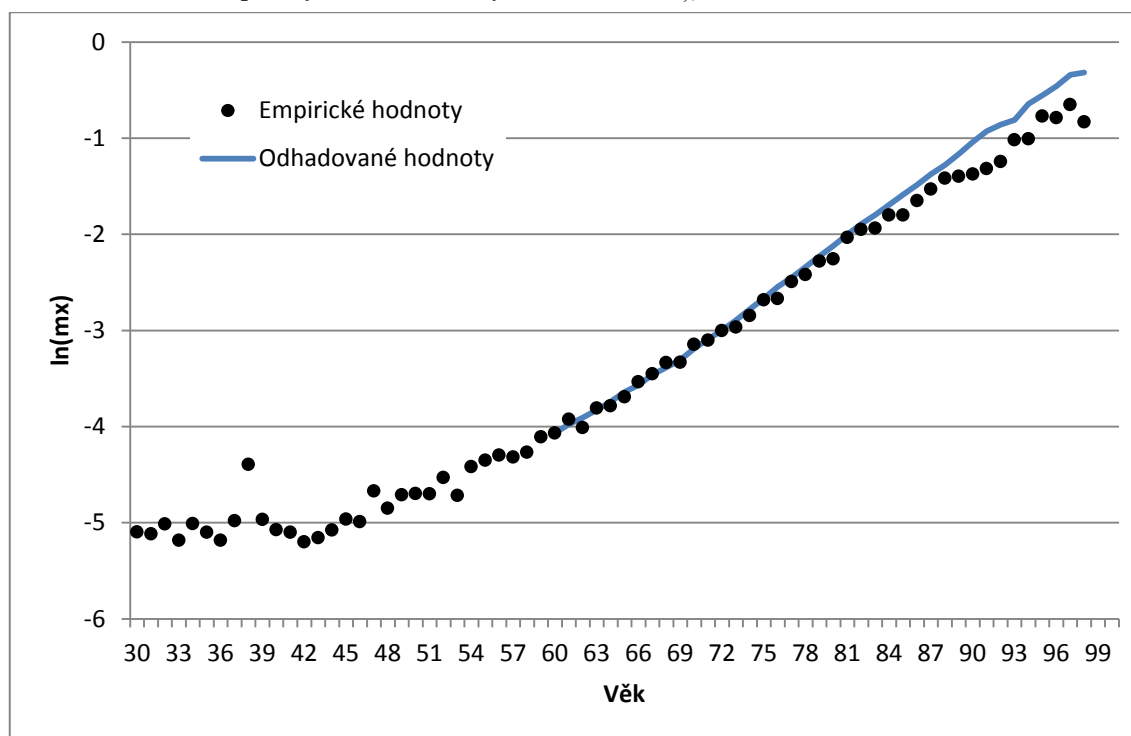
Zdroj: Human Mortality Database (2012)

Další možností, jak znázornit a porovnat výsledky můžeme vidět na Obr. 20. Zde jsou porovnány přirozené logaritmy empirických a odhadovaných specifických měr úmrtnosti kohorty švédských mužů, narozené v roce 1880.

Informace logaritmovaných hodnot měr úmrtnosti je ve skutečnosti stejná, jako kdyby byly hodnoty ponechány ve svém původním měřítku. Při vynesení do grafu je však průběh zlogaritmované exponenciální funkce lineární (v případě Obr. 20 jde vidět, že průběh hodnot měr úmrtnosti má exponenciální tvar přibližně od 55 let věku). Modelování zlogaritmovaných měr úmrtnosti využívají také některé modely uvedené v předchozí kapitole (např. Lee-Carterův model, viz základní rovnici modelu (8)).

Výsledek kohorty švédských mužů narozených v roce 1880 ukazuje, že model v tomto případě hodnoty proti empirickým pozorováním nadhodnocoval, tedy předpovídal horší úmrtnostní poměry, než v realitě byly. Rozdíl mezi empirickou a odhadovanou střední délkou života 60leté osoby byl v tomto případě přibližně 0,35 roku. Podobně jako u kohorty žen 1780 začíná výraznější odchylka odhadovaných hodnot míry úmrtnosti od empirických přibližně kolem věku 85 let, v tomto případě ale obráceně, empirické hodnoty měr úmrtnosti byly modelem nadhodnoceny a model tak předpokládal horší úmrtnostní poměry, než v realitě byly.

Obr. 20: Srovnání empirických a odhadovaných hodnot $\ln(m_{x,z})$ kohorta 1880, Švédsko, muži



Zdroj: Human Mortality Database (2012)

Z výše uvedených výsledků je zřejmé, že navrhovaný model odhadu úmrtnosti v některých případech empirické hodnoty specifických měr úmrtnosti nadhodnotil a v některých případech podhodnotil. Důležité je však analyzovat, zda zmíněné nad- a podhodnocování vykazuje nějaký druh systematickosti v čase, je proto nutné analyzovat dlouhodobé odchylky odhadovaných úrovní úmrtnosti od empirických hodnot úmrtnosti.

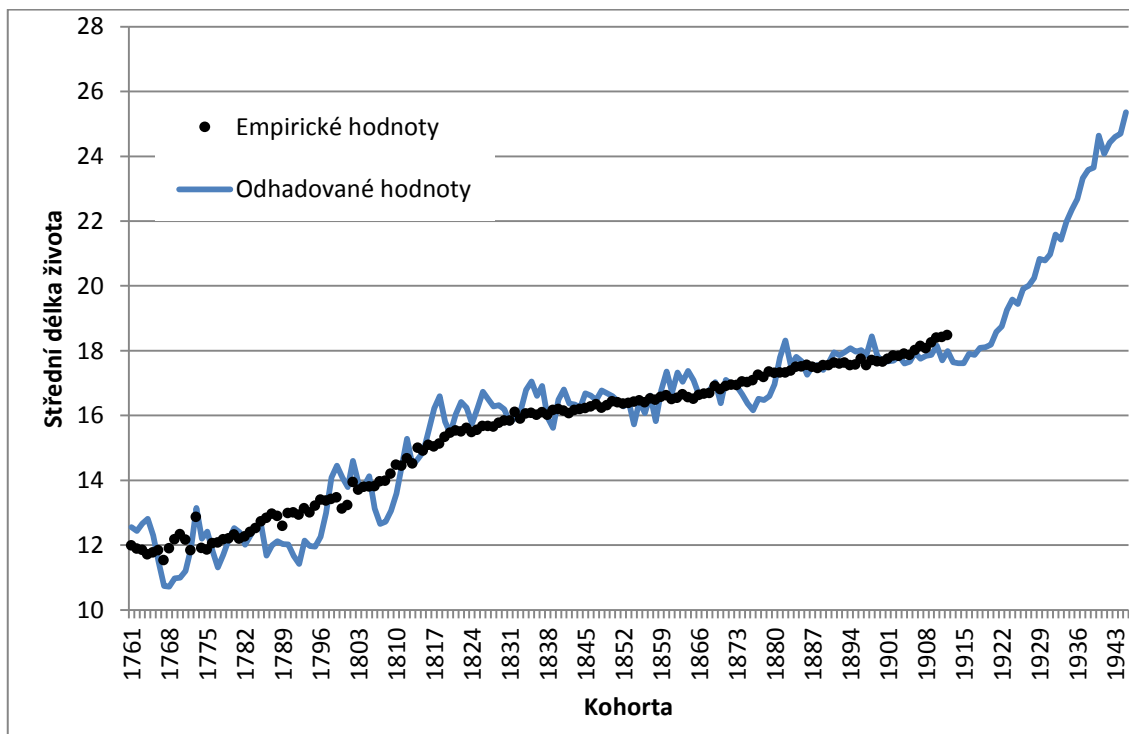
Jednoduché srovnání v čase je možné např. prostřednictvím souhrnné charakteristiky úmrtnosti, jako je střední délka života osoby v přesném věku x . Analyzovat průběh specifických

měr úmrtnosti podle věku u všech kohort by vyžadovalo mnohem více prostoru. Hodnoty odhadovaných a empirických specifických měr úmrtnosti jsou (i s hodnotami střední délky života) uvedeny v příloze 2.

Obr. 21 a 22 zobrazuje vývoj empirických i odhadovaných středních délek života osoby 60leté ve Švédsku, pro kohorty narozené v letech 1761–1945, pro muže a ženy. Empirické hodnoty kohortní střední délky života je možné, jak již bylo uvedeno dříve, zjistit pouze u kohort již vymřelých. V tomto případě je poslední empirická hodnota kohortní střední délky života 60leté osoby za kohortu narozenou v roce 1912. Následující hodnoty střední délky života jsou již pouze odhadované na základě navrhovaného modelu a není tedy tyto odhady možnost srovnání s empirickými hodnotami.

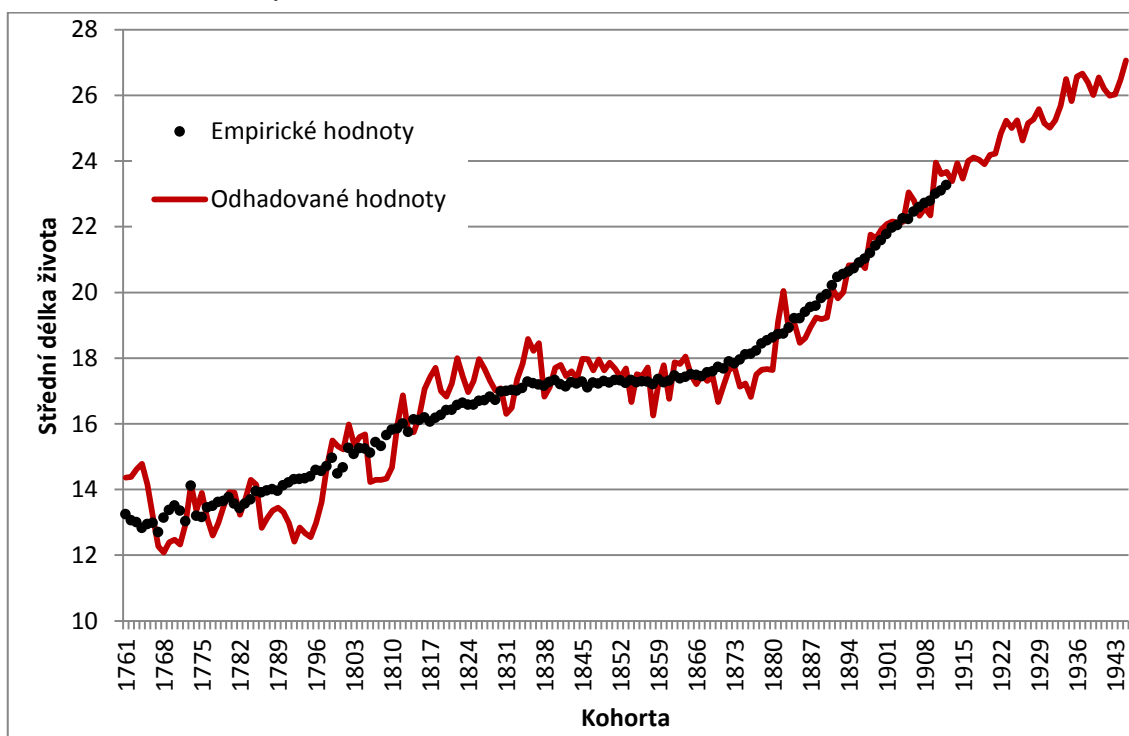
Při pohledu na Obr. 21 a 22 je zřejmé, že navrhovaný model vystihuje realitu velmi dobře. Výraznější odchylky od empirických hodnot střední délky života jsou zejména u nejstarších kohort zahrnutých do výběru a toto období je delší než u mužů, trvá přibližně až do poloviny 19. století. Větší odchylky u starších kohort jsou dány odhadem založeným na velmi starých kohortách, kde byly jednak celkové úmrtnostní poměry velmi odlišné a dostupná data navíc nemusí být úplně spolehlivá. Vývoj empirických hodnot kohortní střední délky života pro osoby v přesném věku 60 let je v případě mužů i žen v čase rostoucí, růst však není monotónní a hodnoty kohortní střední délky života žen dosahují pro všechny sledované kohorty vyšších hodnot než hodnoty u mužů. Všechny tyto základní vývojové tendence se pomocí aplikované metody odhadu podařilo zachytit.

Obr. 21: Srovnání empirických a odhadovaných hodnot střední délky života v přesném věku 60 let, kohorty 1761–1945, Švédsko, muži



Zdroj: Human Mortality Database (2012)

Obr. 22: Srovnání empirických a odhadovaných hodnot střední délky života 60leté osoby, kohorty 1761–1945, Švédsko, ženy



Zdroj: Human Mortality Database (2012)

Hodnoty kohortní střední délky života švédských mužů v přesném věku 60 let dosahovaly na začátku sledovaného období, tedy u kohort narozených po roce 1761, hodnoty těsně pod úroveň 12 let. Během dalších přibližně 50 let, pro kohorty narozené roku 1809, již byla střední délka života v tomto přesném věku 14 let a pro kohorty narozené roku 1834 překročila kohortní střední délka života 60leté osoby poprvé 16 let. Růst střední délky života 60leté osoby o 2 roky v první části sledovaného období trval přibližně 50 kohort a zvýšení o další 2 roky pouze 25 kohort.

Po tomto rychlém nárůstu kohortní střední délky života dochází v jejím vývoji ke zpomalení, protože zvýšení kohortní střední délky života 60leté osoby na 18 let dosáhla až kohorta narozená po roce 1900, konkrétně v roce 1906. Srovnáme-li tyto intervaly, pak rozhodně kohortní střední délka života 60leté osoby neroste monotónně, první nárůst o 2 roky trval 50 let, druhý 25 let a třetí dokonce 70 let. Přesto je možné vidět, že navržený model vystihuje tyto změny tempa růstu velmi dobře.

Hodnoty kohortní střední délky života švédských žen v přesném věku 60 let dosahovaly na začátku sledovaného období přibližně o jeden rok vyšších hodnot než v případě mužů, tedy 13 let. Úroveň 14 let střední délky života 60leté osoby dosáhly až kohorty žen narozené v letech po roce 1788. Zvýšení hodnoty kohortní střední délky života o další dva roky, na 16 let, dosáhly kohorty narozené po roce 1814, tedy podobně jako u mužů, za přibližně 26 let. Kohorty žen narozené po roce 1875 dosahovaly již kohortní střední délky života 60leté osoby na úrovni 18 let a za dalších 16 let, tedy kohorty narozené v roce 1891 již měly tuto střední délku života

20 let. Poslední empiricky pozorované kohorty (1912) měly kohortní střední délku života 60leté osoby na úrovni 23,2 roku.

V případě žen tedy kohortní střední délka života 60leté osoby vzrostla pro sledované období (1761–1912) o více než 10 let a v případě mužů o 6,5 roku. Navrhovaný model odhadu měr úmrtnosti v obou případech vystihnul změny velmi dobře. Jeho odchylky se postupně s časem snižovaly. Největší odchylky mezi empirickými a odhadovanými hodnotami kohortní střední délky života 60leté osoby za celé sledované období dosáhnul navrhovaný model v případě mužů 1,52 roku (kohorta narozená v roce 1752) negativním směrem (tedy model podhodnotil střední délku života osoby v přesném věku 60 let a nadhodnotil tak specifické míry úmrtnosti) a 1,46 roku pozitivním směrem (kohorta 1818). V případě žen byly maximální odchylky 1,9 roku negativním směrem (kohorta 1792) a 1,96 roku pozitivním směrem (kohorta 1764).

V případě, že vezmeme v úvahu pouze posledních 100 kohort (1813–1912), pak se hodnoty maximálních odchylek mezi empirickými a odhadovanými hodnotami kohortní střední délky života 60leté osoby sníží a činí v případě mužů 0,93 roku negativním směrem (kohorta 1876) a 1,46 roku pozitivním směrem (tato hodnota zůstala stejná, vztahuje se ke kohortě narozené v roce 1818) a v případě žen je to 1,3 roku negativním směrem (kohorta 1876) a 1,52 roku (kohorta 1818) pozitivním směrem (navrhovaný model odhadoval v tomto případě nižší specifické míry úmrtnosti).

Protože hodnoty kohortní střední délky života 60leté osoby v čase rostou a absolutní odchylky mezi empirickými a odhadovanými hodnotami kohortní střední délky života 60leté osoby klesají, je vhodné hodnotit úroveň těchto odchylek relativně, tedy ve vztahu k úrovni kohortní střední délky života 60leté osoby.

$$\frac{e_{x,z}}{e_{x,z}} - 1, \quad (45)$$

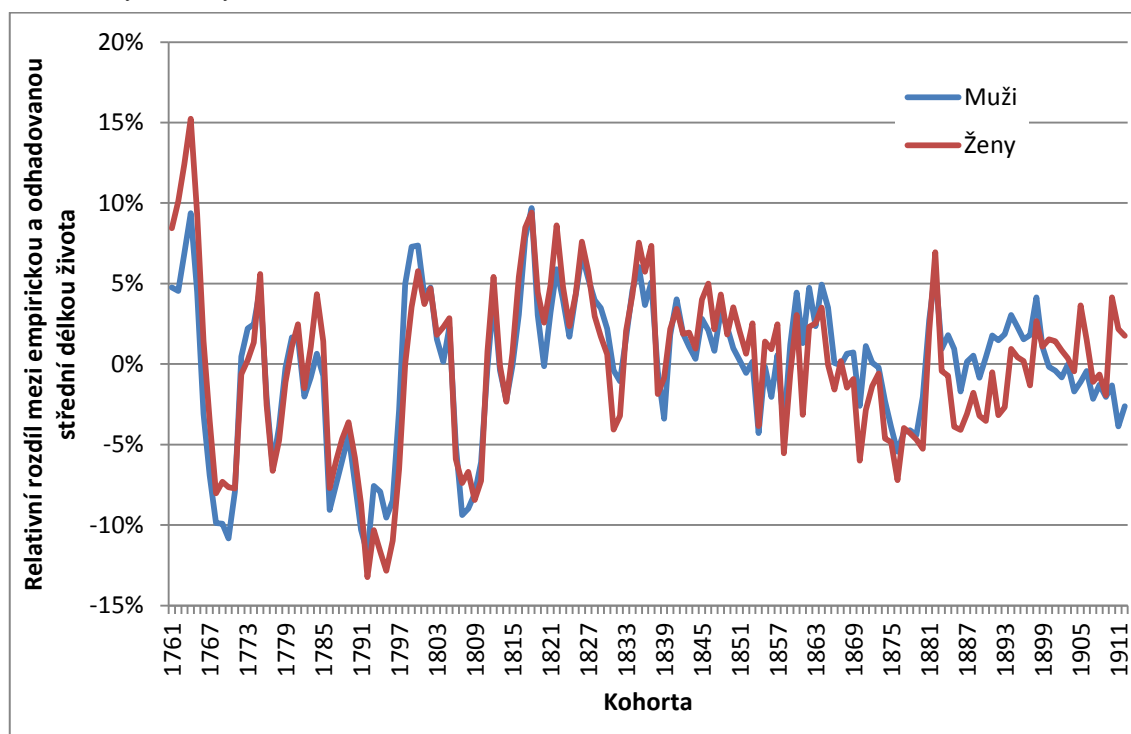
kde $e_{x,z}$ je empirická kohortní střední délka života x leté osoby, kohorty z a $e_{x,z}$ je odhadovaná kohortní střední délka života x leté osoby, kohorty z . Vývoj relativních odchylek je vyneseno do grafu na Obr. 23.

V souladu s předchozími závěry lze pozorovat klesající hodnotu relativní odchylky mezi empirickou a odhadovanou hodnotou kohortní střední délky života 60leté osoby. Počáteční hodnoty na úrovni 10–15% se postupně snižují na hodnoty v rozmezí $\pm 5\%$ a v případě posledních uvažovaných kohort (přibližně kohorty 1885–1912) je tento interval dokonce ještě užší.

Přes rozdílný vývoj v absolutních hodnotách kohortní střední délky života 60leté osoby mezi muži a ženami jsou odchylky navrhovaného modelu predikce úmrtnosti v obou případech téměř identické.

Obr. 21 a 22 neobsahují pouze hodnoty střední délky života pro vymřelé kohorty, ale pro kohorty dosud žijící jsou zde vyneseny do grafu také odhady jejich středních délek života pro osoby v přesném věku 60 let. V případě žen uvažuje model přibližně stejné tempo růstu jako doposud, tedy pro kohorty narozené na konci 2. světové války je odhad střední délky života 60leté osoby přibližně 27 let, což je nárůst o přibližně 3 roky proti kohortám narozeným kolem roku 1910.

Obr. 23: Relativní odchylka mezi empirickou a odhadovanou hodnotou kohortní střední délky života 60leté osoby, kohorty 1761–1912, Švédsko, v %



Zdroj: Human Mortality Database (2012)

V případě mužů se dle navrhovaného modelu předpokládá výrazný nárůst kohortní střední délky života 60leté osoby z úrovně přibližně 18 let (kohorta 1912) na hodnotu o téměř 7 let vyšší (1945). Pro kohorty narozené na konci 2. světové války je tento odhad střední délky života 60leté osoby přibližně 25 let.

Protože se jedná o odhady budoucích hodnot měř úmrtnosti (byť pro kohorty narozené před 60 a více lety) je složité tyto závěry ověřit. Jednou z možností je srovnat predikované hodnoty kohortních středních délek života 60leté osoby s hodnotami transverzálními.

V případě švédských žen byla střední délka života 60leté osoby v roce 2005 přibližně 24,8 roku (Human Mortality Database, 2012), což znamená rozdíl proti odhadované kohortní střední délce života přibližně 2,5 roku.

Švédští muži měli odhadovanou střední délku života 60leté osoby v roce 2005 přibližně 21,4 roku (Human Mortality Database, 2012), což v porovnání s odhadovanou kohortní střední délkou života znamená rozdíl přibližně 3,5 roku. Tento rozdíl je větší než v případě žen, ale i z transverzálních dat je zřejmé, že střední délka života mužů roste rychleji než v případě žen. Do roku 2011 vzrostla transverzální střední délka života (od roku 2006) u mužů o 1,1 roku a o 0,7 roku u žen (tamtéž). Předpoklad rychlého růstu kohortní střední délky života mužů tedy nemusí být tak nereálný, jak na Obr. 21 na první pohled vypadá (pro porovnání rozdílů mezi transverzálními a kohortními středními délkami života viz článek Hulíkové Tesárkové a Mazoucha (2013)). Rozdíly ve střední délce života mezi hodnotami z transverzálních tabulek a odhadnutých kohortních tabulek popsané výše, jsou mj. dány tím, že transverzální tabulky jsou založeny na předpokladu, že v dalších letech zůstanou úmrtnostní poměry neměnné na

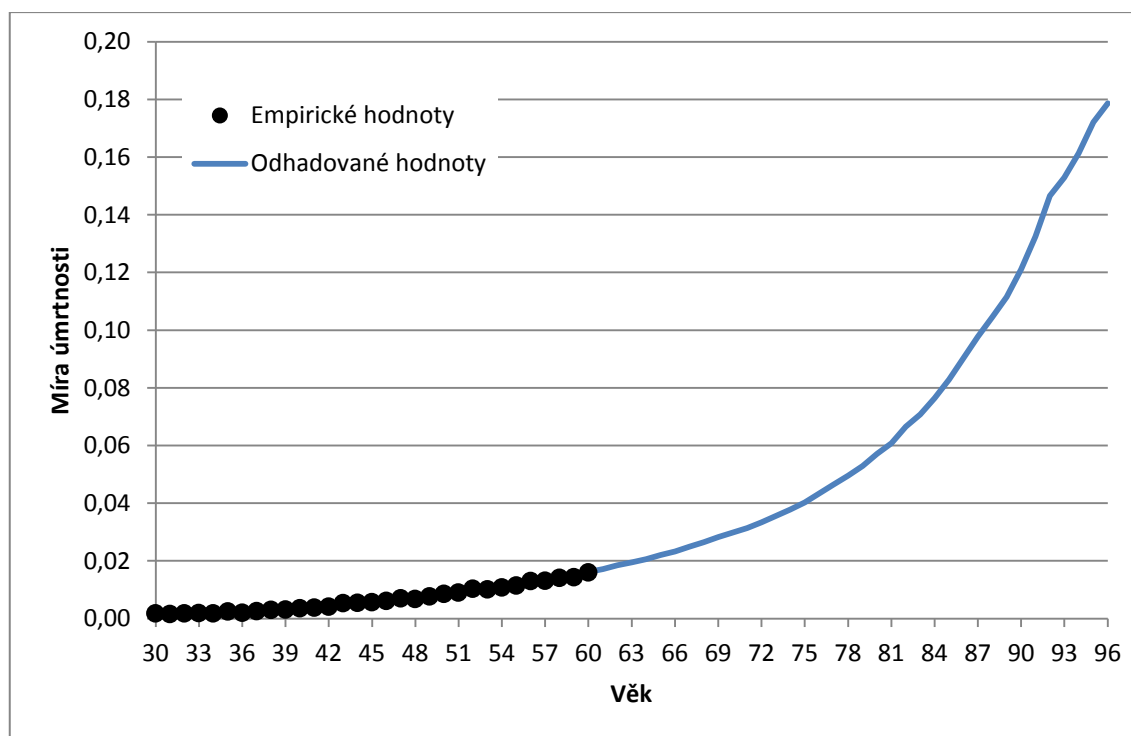
úrovni roku sestavení tabulky. Navrhovaný postup odhadu však vychází z trendů minulého vývoje, které jsou tedy popsáním způsobem v určité míře extrapolovány do dalších let.

Protože byla na začátku kapitoly analyzována také data za Českou republiku, budou zde představeny také výsledky analýzy na těchto kohortních datech. Bohužel ověření modelu je možné pouze v situaci, kdy disponujeme empirickými pozorováními, tedy daty o již vymřelých kohortách. V případě Švédska je dat velké množství a je proto možné ověřit předpoklady navrhovaného modelu odhadu úmrtnosti. V případě České republiky se ale musíme spokojit s omezenou datovou základnou, která nám mnoho možností, jak data ověřit, nedává.

Přestože pro Českou republiku jsou v HMD data již za kohorty narozené v roce 1871, nelze z těchto dat zkonstruovat odhad střední délky života již vymřelých kohort na základě navrhovaného modelu, protože navržený model, díky svému principu (viz výše popis odhadu jednotlivých specifických měr úmrtnosti) vyžaduje velké množství kohortních údajů do minulosti.

Přesto je možné zkonstruovat odhad budoucího vývoje úmrtnosti pro nejmladší kohorty 60letých, tedy narozených těsně po 2. světové válce. Obr. 24 a 25 ukazují, jak vypadá odhad vývoje specifických měr úmrtnosti, odhadnutý na základě navrhované metody.

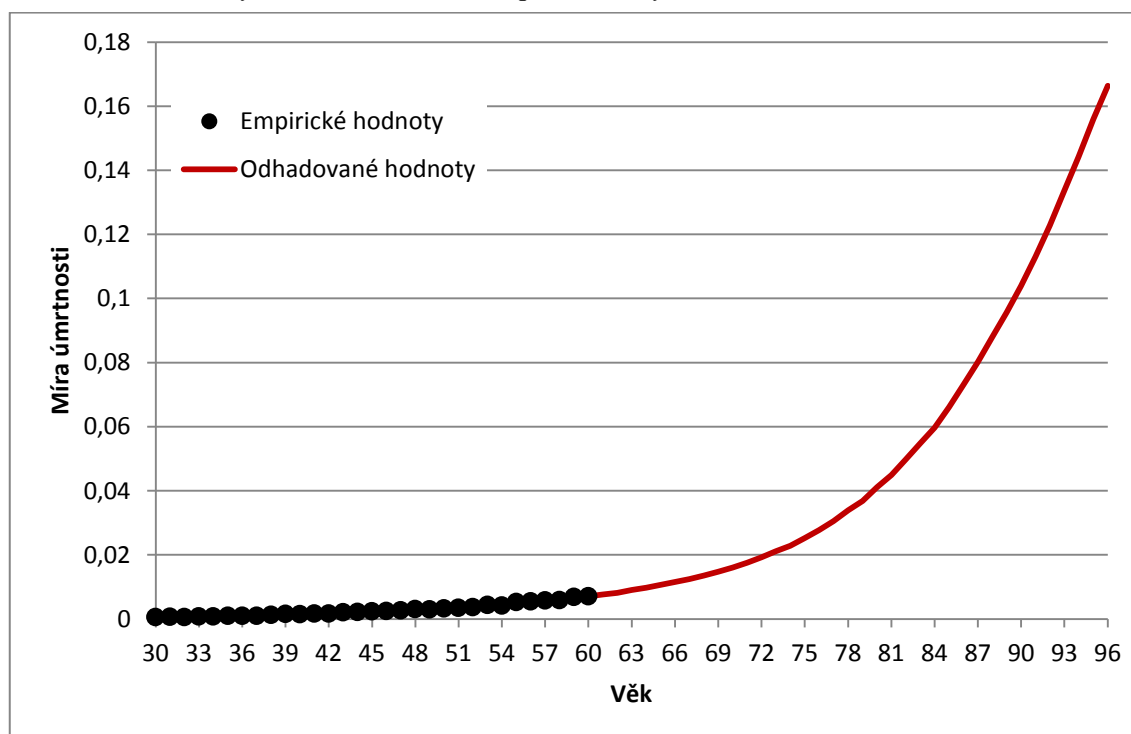
Obr. 24: Vývoj specifických měr úmrtnosti podle věku, 1. hlavní soubory událostí, empirické a odhadované hodnoty, kohorta 1946, Česká republika, muži



Zdroj: Human Mortality Database (2012)

Pozn.: Hodnoty $ar_{x,z,n,s}$ spočteny dle vzorce (27) pro $z = 1946$, $n = 40$, $s = 61$ a $\alpha = 0,9$.

Obr. 25: Vývoj specifických měr úmrtnosti podle věku, 1. hlavní soubory událostí, empirické a odhadované hodnoty, kohorta 1946, Česká republika, ženy



Zdroj: Human Mortality Database (2012)

Pozn.: Hodnoty $ar_{x,z,n,s}$ spočteny dle vzorce (27) pro $z = 1946$, $n = 40$, $s = 61$ a $\alpha = 0,9$.

Pro lepší ilustraci výsledků odhadu specifických měr je možné na jejich základě zkonstruovat kohortní úmrtnostní tabulky pro každé pohlaví zvlášť, výsledky a základní funkce těchto úmrtnostních tabulek jsou součástí Tab. 2.

Odhadovaná kohortní střední délka života 60letého muže je v případě kohorty narozené v roce 1946 na úrovni 21,3 roku. Opět je složité ověřit tento odhad na jakýchkoliv empirických datech kromě srovnání s hodnotou střední délky života 60leté osoby z transverzální úmrtnostní tabulky. Ta byla dle Human Mortality Database (2012) přibližně 18,2 roku v roce 2006 (kohorta 1946 má 60 let v roce 2006). Tento rozdíl je nižší než rozdíl mezi středními délkami života mužů ve Švédsku o půl roku.

V případě žen je odhadovaná kohortní střední délka života 60leté pro kohortu narozenou v roce 1946 na úrovni 25,7 roku. V roce 2006 byla hodnota střední délky života 60leté ženy spočtená transverzálním způsobem 22,3 roku, tedy rozdíl přibližně 3,3 roku, což se blíží rozdílu mezi středními délkami života 60letých mužů ve Švédsku. Na rozdíl od Švédska je v České republice odhadován větší rozdíl střední délky života odhadovaný transverzálním a kohortním způsobem u žen než u mužů.

I přes tyto skutečnosti je nutné poznamenat, že se jedná o jeden z prvních odhadů budoucí kohortní míry úmrtnosti pro Českou republiku vůbec.

Tab. 2. Odhad budoucích specifických měr úmrtnosti a střední délky života, kohorta 1946, Česká republika

Věk	Muži			Ženy		
	Koeficient růstu	Odhadovaná hodnota	Odhadovaná hodnota	Koeficient růstu	Odhadovaná hodnota	Odhadovaná hodnota
x	$ar_{x,z,n,s}$	$m_{x,z}$	$e_{x,z}$	$ar_{x,z,n,s}$	$m_{x,z}$	$e_{x,z}$
60	–	0,01617	21,26	–	0,00706	25,67
61	1,0611	0,01715	20,60	1,0912	0,00768	24,85
62	1,0790	0,01851	19,95	1,0662	0,00816	24,04
63	1,0545	0,01950	19,31	1,1031	0,00901	23,23
64	1,0522	0,02059	18,68	1,0748	0,00976	22,44
65	1,0717	0,02202	18,06	1,0917	0,01056	21,65
66	1,0560	0,02326	17,45	1,0909	0,01150	20,88
67	1,0703	0,02481	16,85	1,0765	0,01244	20,11
68	1,0660	0,02653	16,26	1,0914	0,01352	19,36
69	1,0636	0,02822	15,69	1,0822	0,01468	18,62
70	1,0550	0,02976	15,12	1,0913	0,01601	17,88
71	1,0531	0,03137	14,56	1,0908	0,01748	17,17
72	1,0625	0,03337	14,01	1,0992	0,01923	16,46
73	1,0684	0,03557	13,47	1,0967	0,02103	15,77
74	1,0625	0,03781	12,94	1,0805	0,02270	15,10
75	1,0646	0,04024	12,42	1,1083	0,02529	14,43
76	1,0775	0,04336	11,91	1,0996	0,02776	13,79
77	1,0709	0,04643	11,41	1,0996	0,03054	13,16
78	1,0675	0,04958	10,93	1,1115	0,03398	12,56
79	1,0663	0,05285	10,46	1,0850	0,03684	11,97
80	1,0804	0,05719	10,01	1,1167	0,04114	11,40
81	1,0641	0,06081	9,56	1,0923	0,04491	10,86
82	1,0938	0,06655	9,13	1,1075	0,04973	10,34
83	1,0634	0,07072	8,73	1,1014	0,05487	9,84
84	1,0800	0,07631	8,33	1,0889	0,05960	9,36
85	1,0845	0,08282	7,95	1,1076	0,06614	8,91
86	1,0902	0,09036	7,60	1,1055	0,07305	8,48
87	1,0826	0,09777	7,27	1,0973	0,08013	8,09
88	1,0687	0,10448	6,96	1,0963	0,08797	7,72
89	1,0665	0,11148	6,68	1,0878	0,09565	7,39
90	1,0865	0,12102	6,40	1,0864	0,10389	7,08
91	1,0941	0,13241	6,16	1,0884	0,11309	6,80
92	1,1064	0,14651	5,97	1,0863	0,12284	6,56
93	1,0429	0,15280	5,83	1,0877	0,13350	6,35
94	1,0553	0,16138	5,72	1,0802	0,14437	6,19
95	1,0665	0,17205	5,63	1,0808	0,15599	6,07
96	1,0385	0,17865	5,60	1,0664	0,16630	6,01

Zdroj: Human Mortality Database (2012)

Pozn.: Hodnoty $ar_{x,z,n,s}$ spočteny dle vzorce (27) pro $z = 1946$, $n = 40$, $s = 61$ a $\alpha = 0,9$.

3.4 Shrnutí analýzy dat a výsledků aplikace navržené metodologie predikce úmrtnosti

V této kapitole byla v první části nejprve analyzována základní data o úmrtnosti, která by byla vhodná pro následnou aplikaci navržené metodologie. Šlo zejména o ověření výskytu vlivu některých faktorů, jako je kohortní vliv nebo vliv období na úroveň úmrtnosti. Bylo ověřeno, že v datech týkajících se úmrtnosti jak za Českou republiku, tak i za Švédsko existují výkyvy, způsobené kohortním efektem nebo efektem období. Tyto efekty pak narušují obecné zákony a pravidla úmrtnosti.

V další části kapitoly byly analyzovány vývoje poměrů a rozdílů měr úmrtnosti. Bylo potvrzeno, že přístup modelování poměrů měr úmrtnosti je vhodnější než modelování diferencí měr úmrtnosti. Poměry měr úmrtnosti se jeví jako stabilní v čase, tedy jejich vývoj mezi jednotlivými kohortami lze považovat za konstantní (stacionární). Toto ověření proběhlo také na základě nástrojů analýzy časových řad.

Po empirickém ověření těchto základních předpokladů, které byly definovány v předchozí kapitole, bylo přistoupeno k samotnému procesu konstrukce odhadů budoucích specifických měr úmrtnosti na základě navrhované metodiky. V rámci třetí části kapitoly byly na datech demonstrovány všechny kroky vedoucí k prognóze specifických měr úmrtnosti a dalších tabulkových funkcí.

Analýza výsledků, zejména porovnání odhadovaných hodnot s empirickými pozorováními, ukázala, že navrhovaná metoda odhadu specifických měr úmrtnosti dává velmi dobré výsledky, které se od skutečných empirických hodnot liší jen velmi málo. V rámci prezentace výsledků byly demonstrovány možné přístupy k vizualizaci dat, jako je porovnání vývoje přímo specifických měr úmrtnosti, jejich logaritmů nebo tabulkových funkcí konstruovaných na základě odhadů specifických měr úmrtnosti.

Na základě výsledků lze říci, že navrhovaný model dobře reaguje na výkyvy ve vývoji úmrtnostních poměrů, které v minulosti proběhly. Rozdíly mezi odhadovanou a empirickou hodnotou kohortní střední délky života 60leté osoby nepřesáhly výrazněji 5 % (za posledních přibližně 100 srovnávaných kohort), což lze považovat za velmi dobrý výsledek (v absolutním vyjádření jsou tyto rozdíly pod 1 rok).

V závěrečné části kapitoly byly zkonstruovány také odhady budoucích hodnot měr úmrtnosti pro kohorty dosud žijící. V případě Švédska lze dle navrhovaného modelu odhadu budoucích měr úmrtnosti očekávat střední délku života 60leté osoby, narozené v období konce 2. světové války na úrovni 25,5 roku pro muže a 27 let pro ženy. V případě České republiky jsou tyto hodnoty 21,3 pro muže a 25,7 pro ženu. Tyto hodnoty stojí na předpokladu, že nedojde k náhlé změně dosavadních trendů vývoje úmrtnostních poměrů.

Celkové hodnocení navrhovaného modelu je po aplikaci na reálných datech takové, že navrhovaný model je schopen kvalitní predikce specifických měr úmrtnosti a jeho praktické použití prokázalo jeho snadné použití i v případě omezeného datového zdroje, který nedisponuje dlouhými historickými daty o úrovni úmrtnosti.

V následující kapitole budou diskutovány slabé a silné stránky použití navrhovaného modelu s návrhem možných řešení a zlepšení stávajících postupů.

Kapitola 4

Diskuse silných a slabých stránek navrhovaného modelu a možnosti jeho dalšího rozvoje

Po ukázce praktické aplikace navrhovaného modelu odhadu specifických měř úmrtnosti s komentářem výsledků v minulé kapitole je tato kapitola věnována diskusi jednotlivých částí navrhovaného modelu s důrazem na hodnocení jeho silných a slabých stránek.

Je zřejmé, že model je možné modifikovat v řadě jeho částí, proto by bylo dobré ukázat jejich možnosti a srovnat je s dalšími přístupy, které lze nalézt v odborné literatuře. Cílem kapitoly je zhodnocení celého modelu s cílem navržení možných modifikací navrhovaného modelu do budoucna.

Pozornost bude nejprve soustředěna na problém s variabilitou odhadovaných hodnot, vyjádřenou kolísáním odhadů kohortní střední délky života. Zde budou rozebrány možnosti lepších odhadů některých komponent navrhovaného modelu a jejich dopad na modelované hodnoty. Následně budou diskutovány podobné přístupy uplatňované v rámci jiných metod odhadu, které byly již částečně představeny v první kapitole a jejich možná aplikace do navrhované metodiky.

Závěrem kapitoly bude celkové shrnutí stavu a možných dalších modifikací navrhovaného modelu. Možné slabiny nebo naopak výhody proti dalším používaným metodám budou diskutovány v průběhu celé kapitoly.

4.1 Variabilita výsledků navrhovaného modelu

Jak bylo patrné z Obr. 21 a 22, výsledky dané modelem vykazují určitou variabilitu, která je mírně vyšší než variabilita empirických hodnot, v tomto případě kohortní střední délky života 60leté osoby. Otázkou je, co takovou variabilitu způsobuje a zda je možné ji úpravou metody odstranit.

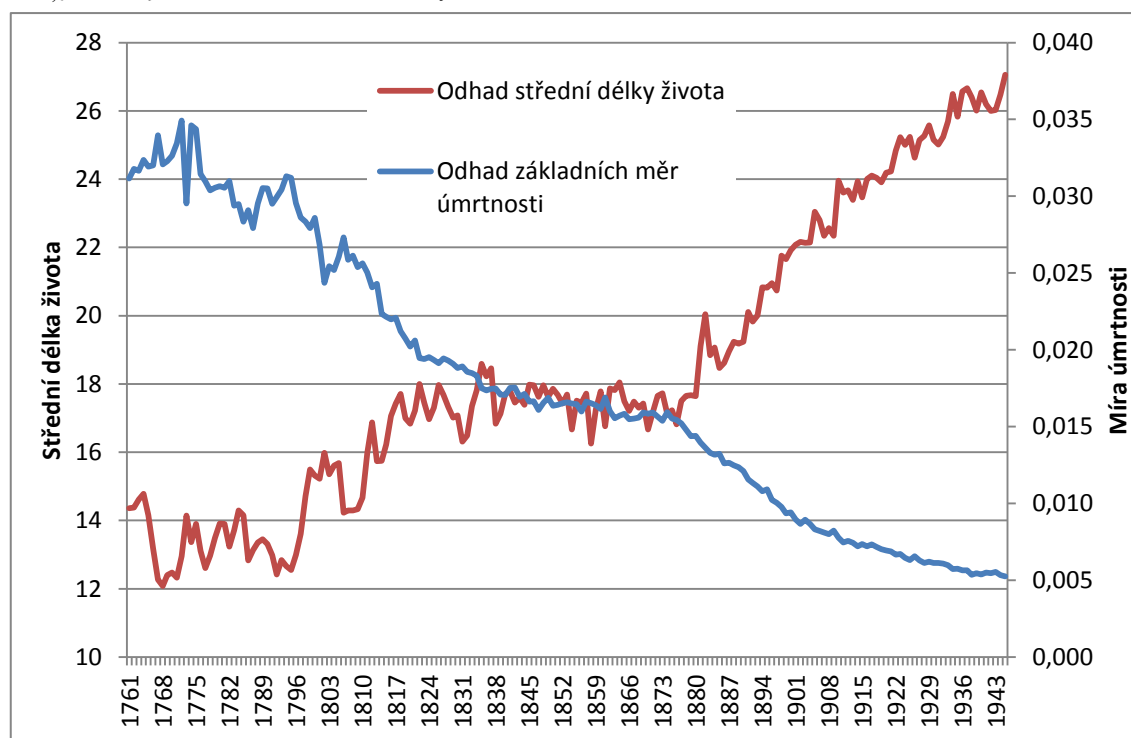
V rámci navrhovaného modelu se využívá empirických dat, která jsou, jak bylo vidět v úvodu předchozí kapitoly, velmi variabilní v čase. Díky působení různých vnějších vlivů se zde projevují rozličné efekty jako je např. efekt kohorty nebo období.

V modelu se tak může variabilita projevit ve dvou oblastech. Jedním z nich je variabilita poměrů měr úmrtnosti $ar_{x,z}$, kterou lze pozorovat na prezentovaných grafických výstupech a druhou možností je stanovování první míry úmrtnosti, na které se celý model staví, tedy $bm_{x,z}$.

4.1.1 Alternativy odhadu základní úmrtnosti $bm_{x,z}$

Nejprve budou analyzovány hodnoty $bm_{x,z}$ v čase a jejich možný vliv na souhrnné ukazatele úmrtnosti. Na Obr. 26 je srovnán vývoj odhadu kohortní střední délky života 60leté osoby (identický s Obr. 22) doplněný o hodnoty odhadů základní míry úmrtnosti $bm_{x,z}$, pro kohorty švédských žen, narozených 1761–1945.

Obr. 26: Srovnání vývoje kohortní střední délky života 60leté osoby a odhadu základní míry úmrtnosti $bm_{x,z}$ kohorty 1761–1945, Švédsko, ženy



Zdroj: Human Mortality Database (2012)

Pozn.: Hodnoty $bm_{x,z}$ spočteny dle vzorce (29). Hodnoty odhadu střední délky života zobrazeny na levé svislé ose, odhady míry úmrtnosti zobrazeny na pravé svislé ose.

Hodnota $bm_{x,z}$ stojí na počátku celého modelu a její úroveň, jak bylo popsáno v kapitole 2, může zásadně ovlivnit celkový odhadovaný průběh úmrtnosti kohorty. Z Obr. 26 je vidět, že jednotlivé výkyvy v hodnotě $bm_{x,z}$ částečně korespondují s výkyvy v odhadu kohortní střední délky života. V tomto případě jsou mnohem důležitější jednotlivé odchylky od trendu než trend samotný. Lze také vidět, že hodnoty $bm_{x,z}$ se nevyvíjí monotónně a vykazují zejména z počátku sledovaného období (přibližně kohorty 1761–1810) vysokou variabilitu (opět se nejspíše jedná o příčiny související s horší kvalitou záznamu událostí nebo evidencí osob).

Jestliže se vrátíme k samotnému postupu odhadu základní míry úmrtnosti $bm_{x,z}$, může být výsledek z Obr. 26 překvapující. Vzorec pro výpočet $bm_{x,z}$ byl definován následovně:

$$\begin{aligned}
 bm_{x,z} &= \text{geomean } m_{x-4,z}, m_{x+4,z} = \\
 &=^9 \frac{m_{x-4,z} * m_{x-3,z} * m_{x-2,z} * m_{x-1,z} * m_{x,z} * m_{x+1,z} * m_{x+2,z} * m_{x+3,z} * m_{x+4,z}}{\phantom{m_{x-4,z} * m_{x-3,z} * m_{x-2,z} * m_{x-1,z} * m_{x,z} * m_{x+1,z} * m_{x+2,z} * m_{x+3,z} * m_{x+4,z}}}, \quad (29)
 \end{aligned}$$

kde $bm_{x,z}$ je základní míra úmrtnosti, x je věk a z je rok narození kohorty.

Jedná se tedy o geometrický průměr s celkem 9 členy. Jeho konstrukce byla navržena tak, aby co nejlépe vystihoval kohortní efekt a navíc eliminoval vlivy období. Proto jsou všechny členy výrazu specifickými měrami úmrtnosti jedné kohorty. Tato skutečnost má však i své nevýhody (viz dále).

Podobný přístup eliminace výkyvů měř úmrtnosti můžeme v literatuře nalézt na několika místech. Van Brokenhoven např. odvodil vyhlazovací metodu určenou pro eliminaci výkyvů měř úmrtnosti v čase (Brokenhoven, 2010). Základ teorie je v odhadu pravděpodobnosti úmrtí q_x ve tvaru

$$q_x = 1 - e^{-e^{f(x)}}, \quad (46)$$

kde q_x je pravděpodobnost úmrtí ve věku x a $f(x)$ je definována jako

$$f(x) = a + bx + cx^2, \quad (47)$$

kde parametry a , b a c jsou odhadnuty metodou nejmenších čtverců z hodnot $q_{x-5}, \dots, q_{x-1}, q_x, q_{x+1}, \dots, q_{x+5}$ (Brokenhoven, 2010).

Princip je tedy založen na výpočtu váženého průměru, jehož váhy jsou dány kvadratickou funkcí $f(x)$. Počet hodnot, které se zahrnují do odhadu této funkce je 11 a Brokenhoven sám uvádí, že jestliže data nejsou příliš variabilní nebo hodnoty q_x blízké 0 nebo 1, dává tato metoda velmi dobré výsledky (Brokenhoven, 2010).

Pokud bychom uvedenou metodu implementovali do navrhované metodiky, je nutné zmínit úskalí jejího výpočtu. To spočívá v počtu zahrnovaných věků. V případě původního vzorce (29) bylo do výpočtu zahrnuto 9 hodnot, tedy při odhadu míry úmrtnosti ve věku 60 let to byly hodnoty měř úmrtnosti za věky 56–64 let.

Tato skutečnost vede k tomu, že i když je základní míra úmrtnosti určována pro věk 60 let, jsou pro její odhady využity míry až do věku 64 let (reálně pak znamená tato skutečnost fakt, že disponujeme-li hodnotami měř úmrtnosti např. za rok 2011 a konstruujeme odhad pro věk 60 a více let, pak poslední možná kohorta, pro kterou lze odhad zkonstruovat není kohorta 1951, ale kohorta narozená v roce 1946, právě z důvodu odhadu základní míry úmrtnosti, pro jejíž odhad je třeba znát hodnoty specifických měř úmrtnosti až do 64 let). V případě metody Brokenhovena by (dle výrazu (46), resp. (47)) bylo nutné zahrnout do výpočtu dokonce věkovou skupinu 65 let.

Určitou obdobou navrhované metody je metodika popsána v kapitole 2.2.1 autorů Habermana a Renshawa (2013), kteří se snaží také odhadovat míry úmrtnosti na základě modelování jejich vztahů. I zde je nutné definovat základní míru úmrtnosti, od které se začnou modelovat další odhady následujících specifických měř.

Haberman a Renshaw odhadují základní míru úmrtnosti jako průměr posledních tří kohortních měr úmrtnosti v konkrétním věku x (Haberman, Renshaw, 2013), zapíšeme-li to vzorcem, vypadal by asi následovně

$$bm_{x,z} = \frac{m_{x,z-2} + m_{x,z-1} + m_{x,z}}{3}, \quad (48)$$

kde $bm_{x,z}$ je základní míra úmrtnosti ve věku x , kohorty z , $m_{x,z}$ je míra úmrtnosti kohorty z ve věku x (Haberman, Renshaw, 2013).

Tento přístup však nerespektuje kohortní faktor, resp. zahrnutím několika kohort do jedné hodnoty ho eliminuje. Na druhou stranu tímto způsobem lze teoreticky eliminovat také nežádoucí faktor období, protože se hodnoty přeci jen částečně vyrovnají (bohužel autoři nepublikovali dílčí výsledky vlastních odhadů těchto základním měr).

Možností odhadu základní míry úmrtnosti je tedy více, ale vždy je nutné hodnotit daný přístup z více hledisek. V následujícím oddíle se zaměří pozornost na variabilitu samotných modelovaných měr úmrtnosti a možnosti, jak ji řešit, ale hned v následující podkapitole se ještě problematika odhadu základní míry úmrtnosti diskutovat bude.

4.1.2 Vyrovnávání odhadovaných měr úmrtnosti

Výsledky prezentované v předchozí kapitole na Obr. 19 nebo 20 ukazují, že odhadované (ale i empirické) hodnoty měr úmrtnosti mohou vykazovat také určitou míru variability. Ta se projevuje tím, že pro určité věky neplatí vztah (25), tedy $m_{x,z} > m_{x-1,z}$.

Tato skutečnost se vyskytuje zejména ve vysokých věcích (přibližně nad 90 let, ale v empirických datech je pozorovatelná i dříve – viz Obr. 19 nebo 20). Některé modely proto aplikují metodu vlastního odhadu budoucích měr úmrtnosti pouze do určitého věku a zbývající míry úmrtnosti ve vysokých věcích modelují různými jinými vyrovnávacími technikami.

Např. software LCFIT (Sprague, 2009), jenž umožňuje online predikci specifických měr úmrtnosti metodou Lee-Cartera, modeluje hodnoty měr úmrtnosti touto metodou pouze do věku 85 let. Od tohoto věku aplikuje na odhadnuté míry vyrovnávací metodu Coale-Guo (Sprague, 2009).

Podobný přístup volí také Haberman a Renshaw (2013), kteří ve svém modelu uvažují pouze věky do 85 let. Zbývající míry úmrtnosti modelují na základě předchozích odhadů (Haberman, Renshaw, 2013).

Jak již bylo zmíněno v úvodní kapitole, vyrovnávání měr úmrtnosti v nejvyšších věcích je aplikováno také v případě Human Mortality Database. Zde je aplikována metoda vyrovnávání i na kohortní data (Wilmoth, 2007), což vede k paradoxním situacím, kdy na stránkách HMD můžeme nalézt kompletní kohortní úmrtnostní tabulky generací, které ještě nevyměřely (Human Mortality Database, 2012).

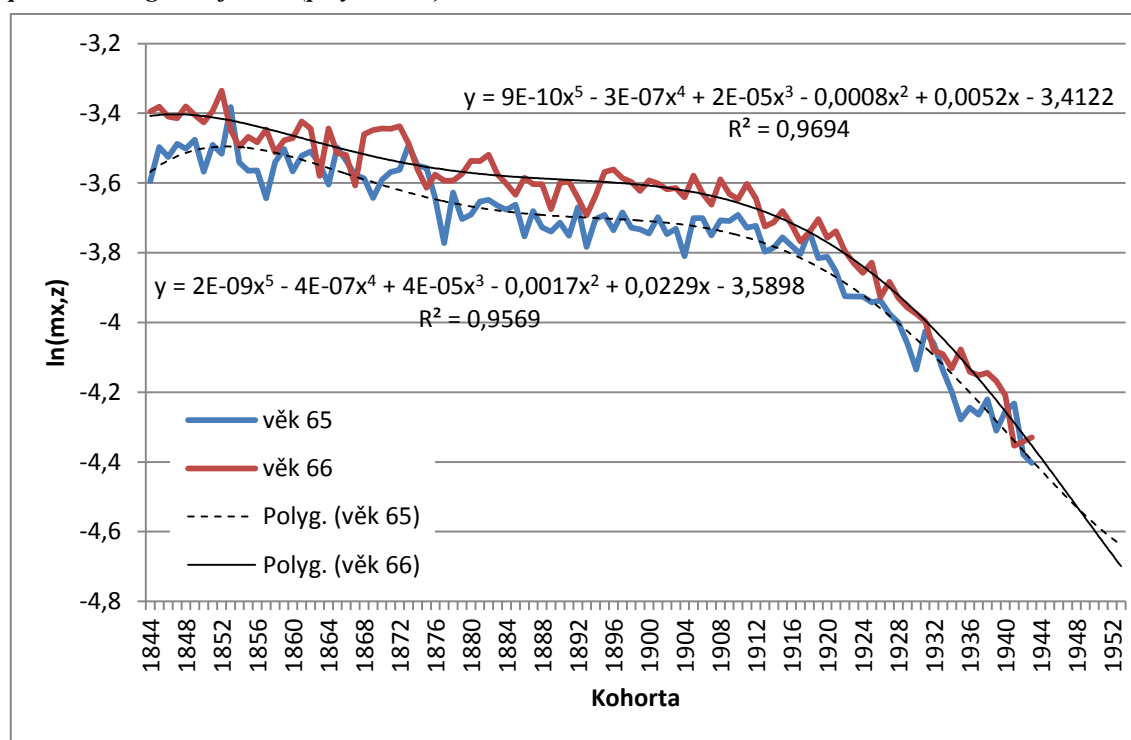
Nutnost vyrovnávat míry úmrtnosti v navrhovaném modelu není příliš nutná. Situace popsaná výše, tedy kdy míra úmrtnosti ve věku x vyšší než míra úmrtnosti ve věku $x + 1$ není v tomto případě příliš častá. Vycházela by ze situace, kdy platí, že hodnota $ar_{x,z,n,s} \leq 1$. V případě Švédska nastala tato situace u mužů pro věkové skupiny 60–90 let a kohorty 1761–1945 celkem dvakrát (pro věk 80 let u kohort 1772 a 1773) a v případě žen čtyřikrát (vždy

pro věkovou skupinu 80 let a pro kohorty 1763 a 1764 a stejně jako u mužů 1772 a 1773). Při zahrnutí celého věkového rozpětí (věk 60–96) tato situace nastala 91krát u mužů a 40krát u žen. Podrobněji viz přílohu 2, kde jsou všechny koeficienty $ar_{x,z,n,s}$, pro $x = 60, \dots, 96$, a $z = 1761, \dots, 1945$ publikovány. Z tohoto hlediska je navrhovaný model možné hodnotit jako stabilní a nutnost vyrovnávání měr úmrtnosti ve vyšších věcích je vhodná nejspíše až v úplně nejvyšších věcích (nad 95 let, což by v případě modelu končícího věkem 96 let bylo již vlastně odhadem a ne vyrovnáváním).

Konstrukce navrhovaného modelu má proti jiným metodám také výhodu v tom, že modelování koeficientů růstu, které jsou (téměř) vždy větší než jedna vede k tomu, že nenastávají situace, kdy by dlouhodobě byly hodnoty $m_{x,z} > m_{x-1,z}$.

V případě modelů, jako je LC model (Lee, Carter, 1992) nebo model navržený Habermanem a Renshawem (2006), které jsou založeny na modelování přímo specifických měr úmrtnosti, může snadno dojít (nejspíše ale v dlouhém období) k situaci, kterou ilustruje Obr. 27.

Obr. 27: Vývoj empirických hodnot $\ln(m_{x,z})$ pro $x = 65$ a 66 , kohorty 1761–1943, Švédsko, ženy, data proložena regresní funkcí (polynomem), odhad budoucích hodnot do roku 1953



Zdroj: Human Mortality Database (2012)

Pozn.: Data proložena polygonem 5. řádu, uvedena rovnice a hodnota indexu determinace R^2 .

Hypotetická situace na Obr. 27 ukazuje, že v případě vyrovnání dat vhodnou funkcí (v tomto případě polynomem 5. řádu) se může stát, že odhad hodnot $m_{x,z}$ do budoucna může v určitém okamžiku skončit situací, kdy např. $m_{66,z} < m_{65,z}$. Je zřejmé, že ojediněle taková situace nastat může (v jednom konkrétním roce), ale zcela určitě nemůže platit dlouhodobě. V tomto případě má navrhovaný model proti metodám využívajících modelování přímo specifické míry úmrtnosti (např. LC model) výhodu, protože jak bylo dokázáno na empirických datech, tento jev se téměř nevyskytl.

4.2 Další možné modifikace navrhovaného modelu

Výše uvedená problematika vybraných částí modelu, jako je volba základní úmrtnosti a vyrovnávání specifických měr úmrtnosti je vhodnou oblastí pro modifikaci stávajícího modelu. V případě vyrovnávání bude nejjednodušší zvolit některou z klasických funkcí a využít její aplikace k modelování odhadnutých hodnot. Pro možný výčet takových metod viz práci Hulíkové Tesárkové (2012).

V oblasti odhadu základní úmrtnosti $bm_{x,z}$ je možné vydat se několika směry. Prvním je dále rozvíjet odhady založené na kohortním způsobu, tedy pro každou kohortu uvažovat individuální hodnotu $bm_{x,z}$ vypočtenou pouze na základě hodnot dané kohorty (viz stávající navrhovaný přístup nebo metodu Brokenhovea (2010)). Tento přístup má výhodu v respektování kohortního faktoru (s akceptováním možného vlivu období). Nevýhoda je často vysoká variabilita hodnot úmrtnosti v rámci jedné kohorty (zmiňovaný vliv období).

Druhou možností je přistoupit na způsob odhadu založený na přístupu Habermana a Renshova (2013), tedy odhadovat hodnotu základní míry úmrtnosti na základě hodnot této základní míry u předchozích kohort. Takový postup by měl výhodu ve využití predikce vývoje konkrétní hodnoty míry úmrtnosti (např. odhad m_{60} o několik let dopředu). Při znalosti hodnot průměrných koeficientů růstu měr úmrtnosti by pak bylo velmi jednoduché odhadnout i ostatní míry úmrtnosti. Tento postup však vyžaduje velmi striktní předpoklady o zachování určitých vztahů na dlouhou dobu, což nemusí být vždy předpoklad udržitelný.

Pokud bychom přistoupili na takovou podmínku, pak pro budoucí odhad základní míry by bylo vhodné predikovat vývoj některé stabilní (stabilní ne ve smyslu konstantní, tedy neměnné ale ve smyslu co nejméně volatilní) míry úmrtnosti. Odhad budoucí úrovně úmrtnosti takové věkové skupiny by umožňoval (při znalosti vztahů mezi jednotlivými úrovněmi úmrtnosti) odhad zbývajících specifických měr úmrtnosti nejen ve věcích vyšších než pro kterou je odhadována základní míra úmrtnosti, ale také odhad míry úmrtnosti ve věcích nižších. Tento postup však úplně odstraňuje z modelu kohortní efekt, který zde hraje velmi výraznou roli.

Jakýmsi kompromisním řešením by byla konstrukce odhadu, který by kombinoval odhad budoucí úrovně míry úmrtnosti konkrétní věkové skupiny (na základě historických hodnot u předchozích kohort) a efekt kohorty. Ideálem by tedy byla kombinace odhadů APC modelů, které by umožnily odhad konkrétních hodnot základní míry úmrtnosti a na jejím základě pak následně odhadovat další specifické míry úmrtnosti již podle navrhovaného modelu, tedy na základě poměrů měr úmrtnosti.

Z matematického hlediska by bylo vhodné pokusit se navrhovaný model vyjádřit ve formě regresní funkce, kde by bylo možné následně odhadovat jednotlivé parametry. Teoretické převedení na regresní funkci bude nejspíše vzdáleně připomínat LC model. Věkově specifické koeficienty (konstantní v čase) lze přirovnat k parametrům β_x z LC modelu. Základní úroveň úmrtnosti $bm_{x,z}$ je analogicky podobné κ_t (neboli časově závislá proměnná).

Je zřejmé, že přepis navrhovaného modelu do regresní funkce by přinesl výrazný posun v možnostech odhadu jeho parametrů a z toho vyplývajících možnostech budoucích odhadů úrovně úmrtnosti.

Jednou z možností modifikace navrhovaného modelu by také byla možnost odhadu vývoje hodnot $ar_{x,z}$ v čase, tedy vyvrácení základního předpokladu stávajícího modelu o neměnnosti

koeficientů růstu pro konkrétní věk x mezi kohortami. Tento krok by však vyžadoval velmi důkladnou analýzu časových řad $ar_{x,z}$ s důrazem na dodržení podmínky, že hodnota $ar_{x,z}$ nemůže být (pro věky vyšší než 30) menší než 1. Přesto lze tento způsob modifikace považovat za žádoucí, protože můžeme očekávat, že při aplikaci navrhovaného modelu na další země můžeme zjistit neplatnost základní podmínky konstantní úrovně $ar_{x,z}$. K podobnému závěru dochází také Haberman a Renshaw (2013), kde v úvodu článku podrobně diskutují podmínky modelů založených na modelu LC, jejichž základem je předpoklad pokračování lineárního trendu zlepšování úmrtnostních poměrů (κ_t) do budoucna, což znamená zlepšování také všech ostatních měr úmrtnosti stejným exponenciálním tempem (základní rovnice je v logaritmické formě, tedy lineární trend přirozeného logaritmu je exponenciální trend původní hodnoty).

Jak Haberman a Renshaw zdůrazňují, na řadě zemí bylo dokázáno, že tento předpoklad je mylný, protože snižování úrovně měr úmrtnosti se v nejnižších věcích zpomaluje a ve vysokých věcích naopak zrychluje (Haberman, Renshaw, 2013).

4.3 Celkové zhodnocení navrhovaného modelu

Je nutné připomenout, že proces úmrtnosti je velmi složitý a ovlivňovaný celou řadou dalších faktorů (ať již je pojmenujeme jako kohortní efekt nebo efekt období) a nalezení přesných vztahů a zákonitostí, které by navíc bylo možné označit za dlouhodobě platné je téměř nemožné (ne-li vůbec nemožné).

I přes výše zmíněné nedostatky navrhovaného modelu a různá úskalí, která je nutné mít na paměti při jeho modifikacích, lze říci, že se podařilo odvodit a následně prakticky aplikovat model, který s dostatečnou přesností dokáže odhadnout budoucí úroveň úmrtnosti.

Proto je vhodné zdůraznit, že model nevyžaduje výrazné technické zázemí a hluboké matematické znalosti, pracuje s jednoduše definovanými předpoklady a časová náročnost jeho aplikace je v porovnání s jinými modely velmi nízká.

Model samotný prokázal, že odhadované hodnoty v porovnání s empirickými hodnotami nedosahují výrazně odlišných výsledků a zejména dobře vystihnul existující změny ve vývoji empirických hodnot.

Lze tedy říci, že v případě, že by se do modelu implementovaly některé z výše zmíněných modifikací, snese navržený model srovnání i s mnohem složitějšími postupy, které často vyžadují speciální software nebo hluboké matematické znalosti. Pro další rozvoj modelu by bylo vhodné ověření některých modifikací na datech dalších zemí a zhodnocení vhodnosti takové případné modifikace modelu.

Kapitola 5

Závěr

Úkolem závěrečné části práce by mělo být shrnutí poznatků prezentovaných v práci s důrazem na zhodnocení naplnění či nenaplnění cíle vytyčeného na samotném začátku práce. Nejinak tomu bude, pochopitelně, i v tomto případě. Základní cíl předkládané diplomové práce byl definován jako navržení takové metody odhadu budoucího vývoje úrovně úmrtnosti ve vyšších věcích, která by byla založená na kohortním pohledu a mohla představovat jistou alternativu k dnes nejčastěji používaným metodám predikce úmrtnosti.

Zejména druhá část cíle, tedy nalezení takové metody, která by představovala alternativu k nejčastěji používaným metodám predikce úmrtnosti, byla velmi ambiciózní a pro naplnění této části bylo nutné realizovat celou řadu malých kroků, které byly také částečně vytyčeny v úvodu práce ve formě dílčích cílů.

Samotné navržení metody, tedy teoretický přepis funkčních vztahů spolu s definováním základních předpokladů by sám o sobě sice naplnil první část hlavního cíle – navržení metody odhadu budoucího vývoje úmrtnosti založené na kohortním pohledu, ale aby byl model aplikovatelný, je nutné nejprve ověřit jeho předpoklady a následně ověřit také jeho výsledky na reálných datech. K naplnění kompletního cíle tak muselo být postupováno analyticky, krok po kroku, které představovaly jednotlivé kapitoly.

Dříve než byly stanoveny základní principy navrhovaného modelu, bylo nutné provést analýzu stávajících přístupů k modelování a odhadování měr úmrtnosti. Ty byly stručně zmíněny v první úvodní kapitole, ale matematický by jednou z podkapitol části druhé. V této podkapitole bylo hlavním, základním cílem stručně popsat vybrané metodické přístupy, které je možné při modelování úmrtnosti využít a stěžejní částí kapitoly pak byla konstrukce vlastního modelu odhadu budoucího vývoje kohortních měr úmrtnosti.

Kromě zmíněné navrhované metody, jejímž principem je odhad budoucích měr úmrtnosti na základě modelování vztahů mezi úmrtnostmi mezi jednotlivými věky, byly podrobněji diskutovány také další metody, zmíněné již v diskusi s literaturou, která byla součástí úvodní části práce. Mezi základní přístupy modelování úmrtnosti patří např. Gompertz-Makehamův přístup exponenciálního růstu úmrtnosti s věkem, Lee-Carterova metoda, předpokládající (v čase neměnnou) závislost věkově specifických měr úmrtnosti na celkovém vývoji úmrtnosti, nebo různé modifikace navazující na Lee-Carterovu metodu, které jsou založené např. na rozšíření modelu o kohortní faktor nebo nahrazení predikce vývoje celkové úrovně úmrtnosti jinými charakteristikami, jako je třeba střední délka života při narození.

Výčet těchto metod měl ilustrovat možné směry, které se objevují (s možnými modifikacemi) v odborné literatuře v poslední dekádě (ale některé také mnohem dříve). Na klasické přístupy k modelování měr úmrtnosti navázalo představení nejnovějších přístupů, mezi které patří modelování úmrtnosti na základě diferencí nebo poměrů měr úmrtnosti, ať již v kohortním nebo transverzálním přístupu.

Popis konstrukce navrhovaného modelu tak navazoval na podrobnou diskusi publikovaných přístupů. Samotný model byl navržen tak, že využívá možnosti predikovat do budoucnosti míry růstu úmrtnosti mezi jednotlivými věky a ne samotné míry úmrtnosti. V metodice byla podrobně popsána teoretická konstrukce modelu, ale byly zde uvedeny také praktické ukázky výpočtu vybraných hodnot a v některých bodech bylo dokonce navrženo více možných postupů.

Po teoretickém vymezení celého modelu by bylo možné uvažovat, že tímto již byla naplněna jedna část základního cíle práce. Náročnější část – ověření modelu na reálných datech však musela být ještě doplněna. V následující kapitole tak byla v první části nejprve analyzována základní data o úmrtnosti, která by byla vhodná pro následnou aplikaci navržené metodologie. Šlo zejména o ověření výskytu vlivu některých faktorů, jako je kohortní vliv nebo vliv období na úroveň úmrtnosti. Bylo ověřeno, že v datech týkajících se úmrtnosti jak za Českou republiku, tak i za Švédsko existují výkyvy, způsobené kohortním efektem nebo efektem období. Tyto efekty pak narušují obecné zákony a pravidla úmrtnosti.

V další části kapitoly byly analyzovány vývoje poměrů a rozdílů měr úmrtnosti. Bylo potvrzeno, že přístup modelování poměrů měr úmrtnosti je vhodnější než modelování diferencí měr úmrtnosti. Poměry měr úmrtnosti se jeví jako stabilní v čase, tedy jejich vývoj mezi jednotlivými kohortami lze považovat za konstantní (stacionární). Toto ověření proběhlo také na základě nástrojů analýzy časových řad.

Po empirickém ověření těchto základních předpokladů, které byly definovány v teoretické části práce, bylo přistoupeno k samotnému procesu konstrukce odhadů budoucích specifických měr úmrtnosti na základě navrhované metodiky. V rámci třetí části kapitoly byly na datech demonstrovány všechny kroky vedoucí k prognóze specifických měr úmrtnosti a dalších tabulkových funkcí.

Analýza výsledků, zejména porovnání odhadovaných hodnot s empirickými pozorováními, ukázala, že navrhovaná metoda odhadu specifických měr úmrtnosti dává velmi dobré výsledky, které se od skutečných empirických hodnot liší jen velmi málo. V rámci prezentace výsledků byly demonstrovány možné přístupy k vizualizaci dat, jako je porovnání vývoje přímo specifických měr úmrtnosti, jejich logaritmu nebo tabulkových funkcí konstruovaných na základě odhadů specifických měr úmrtnosti.

Na základě výsledků lze říci, že navrhovaný model dobře reaguje na výkyvy ve vývoji úmrtnostních poměrů, které v minulosti proběhly. Rozdíly mezi odhadovanou a empirickou hodnotou kohortní střední délky života 60leté osoby nepřesáhly 5 % (za posledních přibližně 100 srovnávaných kohort), což lze považovat za velmi dobrý výsledek (v absolutním vyjádření jsou tyto rozdíly pod 1 rok).

V závěrečné části kapitoly byly dokonce zkonstruovány také odhady budoucích hodnot měr úmrtnosti pro kohorty dosud žijící. V případě Švédska lze dle navrhovaného modelu odhadu budoucích měr úmrtnosti očekávat střední délku života 60leté osoby, narozené v období konce

2. světové války na úrovni 25,5 roku pro muže a 27 let pro ženy. V případě České republiky jsou tyto hodnoty 21,3 pro muže a 25,7 pro ženu. Tyto hodnoty stojí na předpokladu, že nedojde k náhlé změně dosavadních trendů vývoje úmrtnostních poměrů.

Celkové hodnocení navrhovaného modelu bylo po aplikaci na reálných datech takové, že navrhovaný model je schopen kvalitní predikce specifických měr úmrtnosti a jeho praktické použití prokázalo jeho snadné použití i v případě omezeného datového zdroje, který nedisponuje dlouhými historickými daty o úrovni úmrtnosti.

Přestože navržený model ukázal při své aplikaci, že by mohl být vhodným nástrojem k odhadům budoucích trendů vývoje úmrtnosti, bylo vhodné diskutovat také další poznatky o chování modelu při jeho praktické aplikaci. Jednou ze slabých stránek byla variabilita predikovaných hodnot. Ve čtvrté kapitole se proto soustředila pozornost na možné alternativy a modifikace dílčích částí modelu, aby bylo dosaženo stabilnějších výsledků. Protože model neuvažuje žádné další vyrovnávání hodnot měr úmrtnosti, jeví se jako jedna z cest rozšířit model o tuto možnost vyrovnání měr úmrtnosti zejména ve vyšších věcích.

Protože proces úmrtnosti je však komplexní a velmi složitý proces ovlivňovaný celou řadou faktorů, pro nějž je nalezení funkčních vztahů a zákonitostí, které v něm platí, velmi náročné. Nalezení ideálního modelu je tedy jakýmsi nedostižným cílem, ke kterému je však třeba se pokoušet alespoň přiblížit.

I přes nedostatky navrhovaného modelu a různá úskalí zmíněná v poslední kapitole lze říci, že se podařilo odvodit a následně prakticky aplikovat model, který s dostatečnou přesností dokáže odhadnout budoucí úroveň úmrtnosti.

Pokud by se použil příměr, že za člověka mají hovořit činy, tak lze říci, že za model by měly hovořit jeho výsledky. Model samotný prokázal, že odhadované hodnoty v porovnání s empirickými hodnotami nedosahují výrazně odlišných výsledků a zejména dobře vystihnul existující změny ve vývoji empirických hodnot.

Navíc je nespornou výhodnou navrhované metody také to, že model nevyžaduje výrazné technické zázemí ani hluboké matematické znalosti, pracuje s jednoduše definovanými předpoklady a časová náročnost jeho aplikace je v porovnání s jinými modely velmi nízká.

Základní cíle práce (a tím i jejich dílčí části) lze tedy považovat za splněné. Byl odvozen a následně také prakticky ověřen alternativní přístup k odhadům budoucích hodnot specifických měr úmrtnosti v kohortním přístupu. Jeho výsledky navíc vykazují velmi dobré hodnoty a je tedy možné předpokládat další rozvoj popsané metody s rozšířením souboru dat, na něž je možné ho aplikovat.

Seznam použité literatury

- ALAI, D., H.; SHERRIS, M. 2012. Rethinking age-period-cohort mortality trend models. Sydney. 2012. 29 s. *Working paper 2012/11*. CEPAR, Risk and Actuarial Studies, Australian School of Business.
- ANDREEV, K., F.; VAUPEL, J., W. 2006. Forecasts of Cohort Mortality after Age 50. Rostock. 2006. 28 s. *Working paper WP 2006-012*. Max Planck Institute for Demographic Research.
- ARLT, J.; ARLTOVÁ, M.; 2005. *Ekonomické časové řady*. Praha: Grada Publishing, a. s. ISBN 978-80-247-1319-9.
- ARLTOVÁ, M.; ARLT, J. 1995. Grafické metody analýzy ekonomických časových řad. *Statistika*, 1995, roč. 32, č. 11, s. 483–493.
- BÖRGER, M.; ALEKSIC, M., CH. 2013. *Coherent Projections of Age, Period, and Cohort dependent Mortality Improvements* [online]. Ulm. 2013. 31 s. Preliminary version. Institute of Insurance, Ulm University & Institute for Finance and Actuarial Sciences (ifa), Ulm. [cit. 2012-07-25]. Dostupný z WWW: <http://www.ifa-lm.de/downloads/Aleksic_Boerger_Coherent_Mortality_Projections.PDF>
- BROEKHOVEN, H. 2010. Mortality models and longevity risk. Příspěvek prezentovaný na *Emerging Trends in Mortality and Longevity Symposium*, Wawrick, 13.–14. 9. 2010. Dostupný z WWW: <<http://www.actuaries.org.gr/samos2010/presentations/Presentation%20hvb%20final.pdf>>
- BROUHNS, N.; DENUIT, M.; VERMUNT, J., K. 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, č. 31, s. 373–393
- BUETTNER, T. 2002. Approaches and experiences in projecting mortality patterns for the oldest-old. *North American Actuarial Journal*. roč. 6, č.3, s. 14–25.
- BURCIN, B.; ŠÍDLO, L.; TESÁRKOVÁ, K. 2009. Rectangularization process of the survival curve in selected European countries. Příspěvek prezentovaný na *V. Mezinárodním antropologickém kongresu Aleše Hrdličky "Quo vadis homo ...societas humana?"*, Praha, Humpolec, 2.–5. 9. 2009.

- BURCIN, B.; TESÁRKOVÁ, K.; ŠÍDLO, L. 2010. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie*. 2010, roč. 52, č. 2, s. 77–89. ISSN 0011-8265.
- BURCIN, B.; HULÍKOVÁ TESÁRKOVÁ, K.; KOMÁNEK, D.; KUČERA, T. 2011. *DeRaS: software tool for modelling mortality intensities and life table construction*. Charles University in Prague, 2011. <<http://www.deras.natur.cuni.cz/>>.
- CAIRNS, A., J., G.; BLAKE, D.; DOWD, K.; COUGHLAN, G., D.; EPSTEIN, D.; ONG, A.; BALEVICH, I. 2009. A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States. Londýn. 2009. *Discussion paper PI-0701*. The pension institute, City University, London.
- ČADY, O. 2012. *Vliv screeningových programů karcinomu kolorekta na smrtnost a incidenci tohoto onemocnění v ČR modelovaný pomocí APC přístupu*. Praha. 2012. 111 s. Diplomová práce (Mgr.). Univerzita Karlova. Přírodovědecká fakulta. Katedra demografie a geodemografie.
- ČESKÝ STATISTICKÝ ÚŘAD. 2013. *Projekce obyvatelstva České republiky do roku 2010* [online]. Praha: Český statistický úřad. 2013. [cit. 2012-07-25]. Dostupný z WWW: <<http://www.czso.cz/csu/2013edicniplan.nsf/p/4020-13>>.
- ELO, I. T.; PRESTON, S. H. 1992. Effects of Early-Life Conditions on Adult Mortality: A Review. *Population Index*. 1992, roč. 58, č. 2, s. 186–212.
- FIALA, T. 2005. *Výpočty aktuárské demografie v tabulkovém procesoru*. Praha: Oeconomica, 2005, 177 s. ISBN 80-245-0821-4.
- GAVRILOV, L., A.; GAVRILOVA, N., S. 1991. *The biology of life span : A quantitative approach*. Harwood Academic Publishers GmbH, Chur, Switzerland, 1991, 385 s.
- GOMPERTZ, B. 1825. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*. 115: 513–585.
- HABARTOVA, P. 2009. Projekce vývoje úmrtnosti na nemoci oběhové soustavy v České republice s využitím metody Lee-Carter. *Demografie*. 2009, roč. 51, č. 3, s. 173–181. ISSN 0011-8265.
- HABERMAN, S.; RENSHAW, A. 2013. Modelling and projecting mortality improvement rates using a cohort perspective. *Insurance: Mathematics and Economics*, č. 53, s. 150–168.
- HULÍKOVÁ TESÁRKOVÁ, K. 2012. *Selected methods of mortality analysis focused on adults and the rdest age-groups*. Praha. 2012. 261 s. Disertační práce (Ph.D.). Univerzita Karlova. Přírodovědecká fakulta. Katedra demografie a geodemografie.
- HULÍKOVÁ TESÁRKOVÁ, K.; MAZOUCH, P. 2013. Basic cohort mortality analysis at higher ages: an analysis of the rectangularisation process based on cohorts born in 1890-1910 in the Czech Republic and France. *Demografie [online]*, 2013, roč. 55, č. 1, s. 27–46. ISSN 0011-8265. Dostupný z WWW: <[http://www.czso.cz/csu/2013edicniplan.nsf/t/3400464F4A/\\$File/180313q1.pdf](http://www.czso.cz/csu/2013edicniplan.nsf/t/3400464F4A/$File/180313q1.pdf)>
- Human Mortality Database. 2012. [cit. září 2012] <<http://www.mortality.org>>

- KOSCHIN, F. 1981. Gompertz-Makehamův zákon aneb o přesnosti. *Demografie*. 1981, roč. 23, č. 3, s. 207–220. ISSN 0011-8265
- KOSCHIN, F. 1999. Jak vysoká je intenzita úmrtnosti na konci lidského života? *Demografie*. 1999, roč. 41, č. 2, s. 105–119. ISSN 0011-8265
- LEE, R. D.; CARTER, L. R. 1992. Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*. 1992, roč. 87, č. 14, s. 659-675
- MAKEHAM, W. M., 1860. On the Law of Mortality and the Construction of Annuity Tables. *Journal of the Institute of Actuaries and Assurance*. 8: 301–310.
- MAZOUCH, P. 2009. Možnosti modelování úmrtnosti. *Forum Statisticum Slovacum*, 2009, č. 7, s. 110–114. ISSN 1336-7420.
- MAZOUCH, P. 2010. Differences and ratios of mortality rates in Czech Republic. Demänovská Dolina 25.08.2010 – 28.08.2010. In: *AMSE 2010 [CD-ROM]*. Banská Bystrica : Občianske združenie Financ, 2010, s. 250–256. ISBN 978-80-89438-02-0.
- MAZOUCH, P.; HULÍKOVÁ TESÁRKOVÁ, K. 2010. Different ways of mortality modelling. Vídeň 1.9.2010 – 4.9.2010. Konferenční poster. *European population conference*. Vídeň.
- MAZOUCH, P.; HULÍKOVÁ TESÁRKOVÁ, K. 2012. Cohort mortality pattern modeling – model application to Swedish cohort data. Stockholm 13.6.2012 – 16.6.2012. Konferenční poster. *European population conference*. Stockholm.
- MELOUN, M.; MILITKÝ, J. 2002. *Kompendium statistického zpracování dat*. Praha: Academia, 2002, 764 s. ISBN 80-200-1008-4
- PASTOR, K. 2007. A cubic model for graduation of life tables. *Journal of the Applied Mathematics, Statistics and Informatics*. 2007, Roč. 3, Č. 2., pp. 225–236. Dostupný z WWW: <http://jamsi.fpv.ucm.sk/docs/v03n02_12_2007/v03n02_08_PASTOR.pdf>.
- PAVLÍK, Z.; RYCHTAŘÍKOVÁ, J.; ŠUBRTOVÁ, A. 1986. *Základy demografie*. Praha : Academia, 1986. 736 s.
- PITACCO, E.; DENUIT, M.; HABERMAN, S.; OLIVIERI, A. 2009. *Modeling longevity dynamics for pensions and annuity business*. Oxford: Oxford university press, 2009. 416 s. ISBN 978-0-19-954727-2
- RENSHAW, A.; HABERMAN, S. 2006. A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics*, 38: 556–570.
- ROUBÍČEK, V. 1997. *Úvod do demografie*. Praha: Vysoká škola ekonomická, 1997, 352 s. ISBN 80-85963-43-4.
- RYCHTAŘÍKOVÁ, J.; ŘEHÁK, J.; CASELLI, G.; VALLIN, J. 1994. *Analysis of mortality in the Czech Republic: Causal Models of mortality changes in generations and the international comparative analysis*. Final report of the research project of the Central European University, grant č. 879, category G. 67 s.
- SPRAGUE, W., W. 2009. *LCFIT: web based software for estimating the Lee-Carter method for modeling and forecasting mortality, with extensions to Coherent Forecasts of Li and Lee, and Lee-Carter method with sparse and incomplete data*. University of California, Berkeley, 2009. <<http://lcf.it.demog.berkeley.edu/>>

- SRB, V. 1972. Příčiny vzestupu úmrtnosti v letech 1960–1970. *Demografie*, 1972, roč. 14, č. 3, s. 219–227. ISSN 0011-8265.
- THATCHER, A.R.; KANNISTO, V.; VAUPEL, J.W. 1998. *The Force of mortality at ages 80 to 120*. Odense University Press : Odense Monographs on Population Aging, 1998, roč. 5. ISBN 87-7838-381-1.
- WILMOTH, J., R. 2006. Age-Period-Cohort Models In Demography. In CASELLI, G.; VALLIN, J. *Demography: Analysis and Synthesis*. Elsevier, 2006, s. 227–236.
- WILMOTH, J., R.; ANDREEV, K.; JDANOV, D.; GLEI, D., A.; BOE, C.; BUBENHEIM, M.; PHILIPPOV, D.; SHKOLNIKOV, V.; VACHON, P. 2007. *Methods Protocol for the Human Mortality Database*. Human mortality database. 31. května 2007. 5. verze. Dostupný z WWW: <<http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>>.

Přílohy

Příloha č. 1: Korelogramy funkcí ACF a PCF pro jednotlivé věkové skupiny 60–96 let, kohorty narozené 1780–1949, Švédsko

Příloha č. 2: Odhady budoucích specifických měr úmrtnosti a střední délky života (kohorty 1761–1945), srovnání s empirickými hodnotami specifických měr a středné délky života (kohorty 1761–1912)

Přílohy k práci jsou dostupné výhradně v elektronické podobě na přiloženém CD.