
MODELOVÁNÍ ÚMRTNOSTI VE VYSOKÝCH VĚCÍCH V ČESKÉ REPUBLICE S VYUŽITÍM KOHORTNÍCH DAT ZE SOUSEDNÍCH ZEMÍ

Jan Fojtík¹⁾ – Tomáš Karel¹⁾ – Martin Matějka¹⁾ – Pavel Zimmermann¹⁾

MODELLING THE MORTALITY OF OLDER PEOPLE IN THE CZECH REPUBLIC BASED ON COHORT DATA FROM SURROUNDING COUNTRIES

Abstract

The article presents one possible way in which to reduce uncertainty in modelling the mortality of older age groups. The lack of reliable data is usually one of the most important problems of old-age mortality modelling. The uncertainty of the estimates is reduced in this article by using multiple sources of mortality data and using the credibility approach to mixing data from several countries. For the oldest age, the extinct cohort method is applied to increase the reliability of the population data. Using Czech data mixed with data from surrounding countries, results in narrower confidence intervals of estimates and consequently more accurate predictions than when only Czech data were used.

Keywords: Old-age mortality, credibility data model, multi-population mortality models

Demografie, 2017, 59: 23–32

ÚVOD

Modelování a vývoj úmrtnosti ve vysokých věcích je v současné době stále častěji diskutovaným tématem například v souvislosti s důchodovou reformou či komerčním penzijním pojištěním. Otázka budoucího vývoje úmrtnosti a střední délky života ve vyšších věcích je v těchto analýzách naprosto klíčová. Špatné nastavení těchto parametrů může v budoucnu vést k selhání celého penzijního systému. Jedním z důležitých předpokladů pro kvalitní odhad intenzity úmrtnosti je dostatečné množství spolehlivých po-

zorování. Nejčastějším problémem při modelování intenzity úmrtnosti ve vysokých věcích jsou právě malé počty pozorování, mnohdy i chybějící či chybné údaje v jednotlivých věcích, tudíž i vysoká variabilita a heteroskedasticita²⁾ měř úmrtnosti. Tyto problémy způsobují snížení predikční kvality modelu z důvodu širokých intervalů spolehlivosti.

Úmrtnost ve vysokých věcích bývá modelována samostatně od úmrtnosti v dospělých věcích. Malý počet pozorování znemožňuje efektivní využití modelů s velkým počtem parametrů, například

1) Vysoká škola ekonomická v Praze.

2) Rozptýl je nehomogenní a závislý na vysvětlujících proměnných, např. věk.

modelů, které uvažují pro každý věk jiný parametr. Proto se pro modelování závislosti úmrtnosti na věku většinou využívají jednorozměrné modely s malým počtem parametrů. Původní modely vycházely z předpokladu o exponenciálním nárůstu intenzity úmrtnosti s věkem (Gompertz, 1825). Později byl do modelu přidán i člen nezávislý na věku za účelem lepšího postihnoutí úmrtnosti způsobené externími příčinami (Makeham, 1860). S přibývajícím daty počtu zemřelých ve vysokých věcích byl exponenciální nárůst úmrtnosti zpochybněn a byly navrženy alternativní modely. Populární alternativou k modelům s exponenciálním nárůstem se staly modely založené na logistické funkci. Historicky první logistický model použil Perks (1932). K tomuto modelu se dále přiklonili ve svých aplikacích Heligman et al. (1980), Thatcher et al. (1998) nebo Thatcher (1999). Koschin (1999) vycházel z Gompertz-Makehamova modelu, ale exponenciální nárůst zpomalil přidáním dalšího parametru do modelu. Přehled parametrických modelů úmrtnosti ve vysokých věcích je možné nalézt například v Pitacco et al. (2009) nebo Burcin a kol. (2010). Mimo nárůstu úmrtnosti s rostoucím věkem je také zkoumán vývoj této závislosti v čase. Věkově specifické posuny logistického modelu lze nalézt například v Hulíková Tesárková (2012). Dlouhodobý vývoj trendů v úmrtnosti ve vysokých věcích byl zkoumán například v Gavrilov et al. (2011). Problematickou časové změny úmrtnosti ve vysokých věcích v souvislosti s Gompertzovým a Kannistovým modelem pro různé datové soubory se zabývala Gavrilova et al. (2014).

V našem článku se více než na predikce samotné zaměřujeme na možnost snížení nejistoty v predikcích využitím dodatečných datových zdrojů. Námí aplikovaný přístup k zúžení intervalů spolehlivosti odhadů úmrtnosti ve vysokých věcích, založený na směřování dat z více populací, byl navržen v článku Ahcan et al. (2014). Byl aplikován na modelování úmrtnosti z průřezových dat v dospělých věcích při použití Lee-Carterova modelu (Lee et al., 1992). Náš článek je zaměřen na využití této metodologie pro modelování kohortní úmrtnosti ve vysokých věcích v České republice při využití informace ze sousedních zemí. Jsou použita kohortní data o počtu zemřelých a o velikosti expozice k riziku smrti v dokončených

jednoletých věcích. Data o expozici jsou zpřesněna metodou vymřelých kohort (Wilmoth et al., 2007). Metodologie navržená v Ahcan et al. (2014) je dvoukrokový postup. V prvním kroku je vytvořena optimální lineární kombinace specifických měr úmrtnosti ze sousedních populací, která vysvětluje maximum variability pozorované v českých datech. Pomocí tzv. kredibilitního přístupu, používaného především v aktuárských aplikacích, jsou následně mixována česká data s touto lineární kombinací úmrtnosti v sousedních státech. Využití této metody umožňuje zahrnutí informace o vývoji úmrtnosti z geograficky a ekonomicky blízkých populací a vede tak k zúžení intervalů spolehlivosti odhadů úmrtnosti. Alternativně by šlo využít pro zahrnutí informace z okolních států čistě bayesovských metod. Příkladem aplikace na dospělé věky je například Antonio et al. (2015). Aplikace bayesovských metod pro odhad úmrtnosti ve vysokých věcích byla navržena například v Bjak et al. (2015), kde byl aplikován zobecněný aditivní model v kombinaci s Gompertzovým, respektive Kannistovým parametrickým modelem. Využití dodatečné informace, například z okolních států, v tomto přístupu však není zahrnuto.

V následující kapitole je zavedeno značení, předpoklady a specifikace modelu. Dále jsou popsána vstupní data a jejich úprava metodou vymřelých kohort. Její aplikace, podmínky použití a metodologie je popsána ve třetí kapitole. Následně je popsána aplikace metody tzv. kredibilitního mixování dat pro vytvoření datového souboru obsahujícího informace z více populací. V další kapitole jsou popsány odhady parametrů modelu úmrtnosti pomocí metody maximální věrohodnosti a způsob odhadů intervalů spolehlivosti. Poslední část článku tvoří numerické výstupy, provedené analýzy včetně grafických výstupů a závěr.

1. ZNAČENÍ A PŘEDPOKLADY

Uvažujeme jednoleté kohorty, které budeme značit indexem c . Expozici k riziku z kohorty c v dokončeném věku x označíme jako $E_{x,c}$. Počet zemřelých pak analogicky $D_{x,c}$. Dále označíme $L_{x,c}$ spodní trojúhelník Lexisova diagramu pro počet zemřelých osob v dokončeném věku x v okamžiku $c + x$ a $U_{x,c}$ značí horní trojúhelník Lexisova diagramu pro po-

čet zemřelých v dokončeném věku x a okamžiku $c + x + 1$. Platí tedy

$$D_{x,c} = L_{x,c} + U_{x,c}.$$

Počet obyvatel v dokončeném věku x z kohorty c v okamžiku $x + c + 1$ označíme jako $P_{x,c}$. Výpočet těchto veličin a jejich vztah je popsán v kapitole 3.

Specifickou míru úmrtnosti označujeme $m_{x,c}$. Příslušné státy (populace) pak označujeme pravým horním indexem. Parametrických funkcí, které lze předpokládat pro vývoj úmrtnosti s věkem, je celá řada. Výčet třinácti parametrických funkcí, lišících se počtem parametrů i modelovanou mírou úmrtnosti, lze nalézt v *Burcin a kol.* (2010). Vzhledem k tomu, že počet pozorování ve vysokých věcích je v České republice omezený a kolísavost měr úmrtnosti je značná, omezili jsme se na dvouparametrické funkce. Z publikovaných dvouparametrických funkcí lze volit mezi exponenciálním průběhem či logistickým průběhem úmrtnosti s rostoucím věkem. Protože se k logistickému průběhu přiklání celá řada moderních aplikací, například *Thatcher et al.* (1998), *Wilmoth et al.* (2007), *Zeng* (2003) nebo *United Nations* (2015), a obě funkce se do věku okolo 100 let příliš neliší, omezili jsme se pouze na logistickou specifikaci. Námi zvolená metodologie je však univerzálně aplikovatelná na libovolnou parametrickou specifikaci modelu úmrtnosti ve vysokých věcích.

Předpokládáme tedy, že pro danou kohortu c se specifická míra úmrtnosti řídí logistickým modelem, tj.

$$\log\left(\frac{m_x}{1 - m_x}\right) = a + bx, \quad (1)$$

neboli

$$m_x = \frac{1}{1 + e^{-(a+bx)}},$$

kde a_c a b_c jsou parametry. Odhady parametrů, respektive predikce úmrtnosti, jsou označeny jako \hat{a}_c a \hat{b}_c .

2. VSTUPNÍ DATA

Zdrojem dat pro všechny populace je The Human Mortality Database (*Wilmoth et al.*, 2012), dále jen HMD. Vycházíme z kohortně řazených dat o počtech zemřelých a expozicích k riziku pro jednotlivé kohorty, země a obě pohlaví. Mezi uvažované země je zařazeno

Maďarsko, Východní Německo, Česká republika a Slovensko. Výběr zemí byl řízen dostupností dat (kohorty od roku 1886 do roku 1901) dané země a její geografickou, historickou i ekonomickou blízkostí. Polsko a Rakousko nemohly být do analýzy zařazeny kvůli neúplné datové základně. Datový soubor pro Polsko neobsahoval kohortní informace o počtu zemřelých, respektive spodní a horní trojúhelníky Lexisova diagramu, pro roky 1970 až 1986. V případě Rakouska data neobsahovala všechny požadované věkové kategorie pro vybrané kohorty. Požadované věkové kategorie, tj. od 70 let jsou dostupné až pro kohortu 1901. V případě Rakouska by pro odhad parametrů modelu zůstala pouze jedna kohorta, a proto jsme se jej z důvodu zachování dostatečného množství kohort pro odhad parametrů rozhodli vyřadit.

Pro odhad expozice k riziku smrti v jednotlivých kohortách předpokládáme rovnoměrné rozdělení počtu zemřelých v rámci Lexisova diagramu, který jsme ověřili na použitých datových souborech. Dle *Chin Long Chiang* (1984) by pro rovnoměrné rozdělení počtu zemřelých měl být poměr

$$h_x = \frac{L_{x,c}}{L_{x,c} + U_{x,c}}$$

přibližně roven hodnotě 0,5. Při analýze datových souborů se hodnoty tohoto poměru pohybují okolo této konstanty (viz graf 1a a 1b), proto považujeme tento předpoklad o rovnoměrném rozdělení počtu zemřelých za splněný. Pro analýzu používáme kohorty, které jsme pozorovali až do věku $\omega = 110$ let. Dostupná jsou data za kohortu od roku 1886 do roku 1901.

3. ODHAD EXPOZICE K RIZIKU ÚMRTÍ

Odhad expozice k riziku úmrtí pro kohortu c v dokončeném věku x jsme provedli zvlášť pro věk 70–80 let a pro věk nad 80 let. Pro dokončené věky od 70 do 80 let vycházíme z předpokladu, že úmrtí jsou rozdělena během roku rovnoměrně. Pak je počet zemřelých v každém věku x z každé kohorty c ve spodním trojúhelníku $L_{x,c}$ roven počtu zemřelých v horním trojúhelníku $U_{x,c}$. V tomto případě platí rovnost $E_{x,c} = P_{x,c}$. Pro odhad expozice k riziku úmrtí pro kohortu c v dokončeném věku x jsou tedy použity přímo počty obyvatel $P_{x,c}$ v dokončeném věku x a čase $c + x + 1$ publikované HMD. Pro ilustraci je výše

uvedený postup zobrazen pomocí Lexisova diagramu v grafu 2.

Pro zlepšení kvality dat o expozici k riziku úmrtí v dokončených věcích nad 80 let byla použita metoda vymřelých kohort. Metoda předpokládá nulovou migraci osob ve vysokém věku. Díky absenci migrace dochází k úbytku obyvatelstva pouze úmrtím, a proto lze zpětným nasčítáním zemřelých osob získat odhady počtu obyvatel z dané kohorty v daném věku. Odhad počtu obyvatel z kohorty c v dokončeném věku x v okamžiku $x + c + 1$ metodou vymřelých kohort tedy provedeme součtem všech budoucích úmrtí v dané kohortě:

$$P_{x,c} = \sum_{i=0}^{\omega-x} (U_{x+i,c} + L_{x+1+i,c}).$$

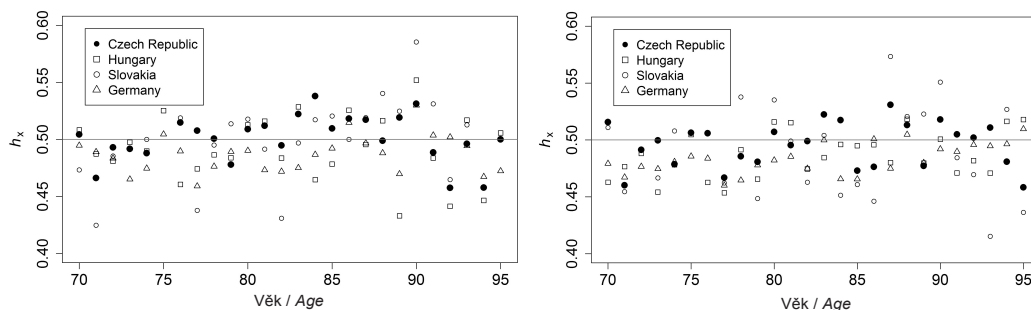
Vzhledem k předpokladu o rovnoměrném rozdělení počtu zemřelých opět předpokládáme, že expozice k riziku smrti bude přibližně rovna odhadnutým počtům obyvatel, tj. expozici odhadneme pro dokončené věky nad 80 let jako $E_{x,c} = P_{x,c}$. Míry úmrtnosti spočítané za pomoci metody vymřelých kohort ve věku 80 let plynule navazují na míru úmrtnosti spočítanou přímo ze vstupních dat, proto není potřeba žádných dodatečných úprav pro řešení návaznosti obou použitých metod.

4. VÝPOČET KREDIBILITNÍCH DAT

Na takto upravená data aplikujeme metodu „směšování dat“ z více populací, která byla původně vyvinuta

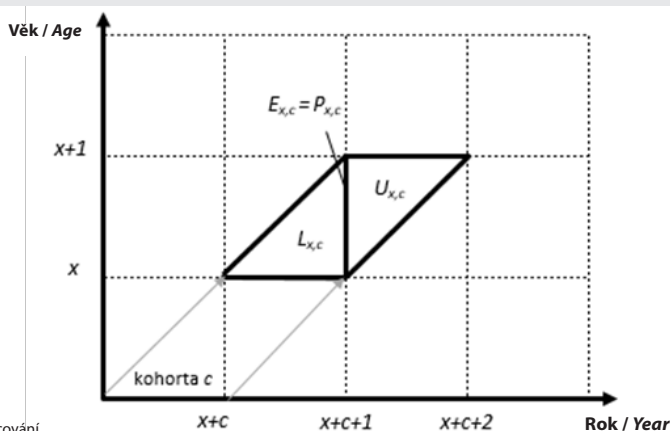
Graf 1a a 1b: Porovnání poměru h_x – kohorta 1890, muži (vlevo) a ženy (vpravo)

A comparison of the ratio h_x – cohort 1890, males (left) and females (right)



Zdroj: The Human Mortality Database; vlastní výpočty.
Source: The Human Mortality Database; authors' calculations.

Graf 2: Lexisův diagram | Lexis diagram



Zdroj: Vlastní zpracování.
Source: Authors' calculation.

v *Ahcan et al.* (2014) pro modelování úmrtnosti v dospělých věcích s využitím Lee-Carterova modelu. Prvním krokem této metodologie směřování dat je určení optimální lineární kombinace („váženého průměru“) specifické míry úmrtnosti z okolních států, kterou značíme $m_{x,c}^{[AVE]}$. Na tuto míru lze nahlížet jako na informaci relevantní pro úmrtnost v České republice obsaženou v datech z okolních států. Tato míra je definována jako

$$m_{x,c}^{[AVE]} = \sum_{k=1}^n w_k m_{x,c}^{[k]},$$

kde $m_{x,c}^{[k]}$ je specifická míra úmrtnosti v zemi k pro dokončený věk x pro kohortu c . Dále n reprezentuje počet zemí použitých v modelu (tj. $n = 3$: Východní Německo, Maďarsko a Slovensko). Optimální váhy jednotlivých okolních zemí w_k jsou získány pomocí minimalizace součtu čtverců odchylek váženého průměru specifických měr úmrtnosti $m_{x,c}^{[AVE]}$ od specifických měr úmrtnosti v Česku $m_{x,c}^{[0]}$

$$w_k = \arg \left(\min_{w_k} \sum_x \sum_c \left(m_{x,c}^{[0]} - w_k m_{x,c}^{[k]} \right)^2 \right)$$

při splnění omezujících kritérií

$$w_k \geq 0, \quad k = 1, \dots, n,$$

$$\sum_{k=1}^n w_k = 1.$$

Druhým krokem metodologie směřování je pak výpočet výsledné specifické míry úmrtnosti $m_{x,c}^{[z]}$, kterou budeme dále označovat jako kredibilitní. Její výpočet vychází z teorie kredibility, známé z aplikací v pojistné matematice např. *Boland* (2007) nebo *Cipra* (2006). Teorie kredibility zahrnuje mnoho odhadovacích postupů založených na směřování interních dat o vlastním portfoliu pojišťovny s externími daty například z celého pojistného trhu, či z jiných portfolií. Využívá se zejména v případech, kdy velikost interních dat není dostatečná pro určení spolehlivého odhadu, což je také případ dat o úmrtnosti ve vysokých věcích v malých populacích, například v České republice.

Kredibilitní specifická míra úmrtnosti $m_{x,c}^{[z]}$ je vypočtena opět jako vážený průměr. Nyní je však průměr počítán z $m_{x,c}^{[AVE]}$ představující informaci z okolních populací a z $m_{x,c}^{[0]}$, která obsahuje pouze informaci

z České republiky. Vzorec pro tento vážený průměr je v pojistné matematice znám jako tzv. kredibilitní formule (např. *Boland*, 2007) a má tvar:

$$m_{x,c}^{[z]} = m_{x,c}^{[0]} z_{x,c} + m_{x,c}^{[AVE]} (1 - z_{x,c}),$$

kde $z_{x,c}$ je tzv. kredibilitní koeficient spočítaný podle následujícího vzorce převzatého z *Ahcan et al.* (2014, s. 18):

$$z_{x,c} = \frac{E_{x,c}^{[0]}}{E_{x,c}^{[0]} + \sum_{k=1}^n w_k E_{x,c}^{[k]}},$$

kde $E_{x,c}^{[k]}$ je expozice k riziku úmrtí v k -tém státě. Čím větší je tedy expozice k riziku v populaci $[0]$, tím vyšší je váha specifické míry úmrtnosti této populace ($m_{x,c}^{[0]}$) ve výsledné kredibilitní specifické míře úmrtnosti $m_{x,c}^{[z]}$.

5. ODHAD PARAMETRŮ MODELU POMOCÍ POISSONOVSKÉ LOGISTICKÉ REGRESE

Specifické míry úmrtnosti $m_{x,c}^{[0]}$ České republiky a kredibilitní specifické míry úmrtnosti $m_{x,c}^{[z]}$ jsou následně modelovány za účelem extrapolace do vyšších věků. Metodologie je pro data z České republiky i pro kredibilitní data v každé kohortě totožná, proto pro přehlednost v této a následující sekci neuvádíme horní indexy $[0]$ a $[z]$, ani index kohorty c . Parametry logistického modelu (1), a a b , jsou odhadnuty metodou maximální věrohodnosti za předpokladů převzatých z *Brouhns et al.* (2002) a využitím statistického softwaru R. Předpokládáme, že počet zemřelých D_x má Poissonovo rozdělení se střední hodnotou $m_x E_x$, kde m_x je logistická funkce věku (1), tedy

$$D_x \sim \text{Poisson}(m_x E_x).$$

Odhady parametrů jsou získány maximalizací logaritnické věrohodnostní funkce (bez konstantních členů):

$$l(a, b) = \sum_x D_x \ln(m_x(a, b) E_x) - E_x m_x(a, b),$$

kde $m_x(a, b)$ je logistická funkce (1), u níž pro přehlednost naznačujeme závislost na parametrech a a b . V případě kredibilitních dat je pozorovaný počet

zemřelých získán z kredibilitního odhadu specifické míry úmrtnosti jako $E_{x,c}^{[0]}m_{x,c}^{[z]}$. Alternativně by bylo možné použít iterativní váženou metodu nejmenších čtverců používanou například systémem DeRaS (<http://deras.natur.cuni.cz>).

6. ODHAD INTERVALŮ SPOLEHLIVOSTI PREDIKCE

Vhodnost a přesnost modelů odhadnutých z $m_{x,c}^{[0]}$ a $m_{x,c}^{[z]}$ je vyhodnocena na základě porovnání intervalů spolehlivosti. Jelikož analytická vyjádření nejsou při naší specifikaci modelu možná, intervaly spolehlivosti byly pro každou kohortu odhadnuty simulační metodou, tzv. reziduálním bootstrapováním. Detaily o metodách bootstrapu lze nalézt například v *Efron et al.* (1994). Postup bootstrapu je shodný pro kredibilitní i pouze český datový soubor a je aplikován ve všech kohortách. V dalším textu je tedy index $[0]$, resp. $[z]$ a c opět vynechán. Pro danou kohortu a datový soubor je postup bootstrapu následující:

Predikované specifické míry úmrtnosti označíme \hat{m}_x . Předpoklad Poissonova rozdělení je spjat s předpokladem o heteroskedasticitě, který je konzistentní s pozorovanými daty. V případě Poissonova rozdělení je rozptyl roven střední hodnotě. Predikce \hat{m}_x je tedy

zároveň i predikcí rozptylu. Proto jsou pro simulace využita Pearsonova rezidua definovaná jako

$$e_x = \frac{m_x - \hat{m}_x}{\sqrt{\hat{m}_x}},$$

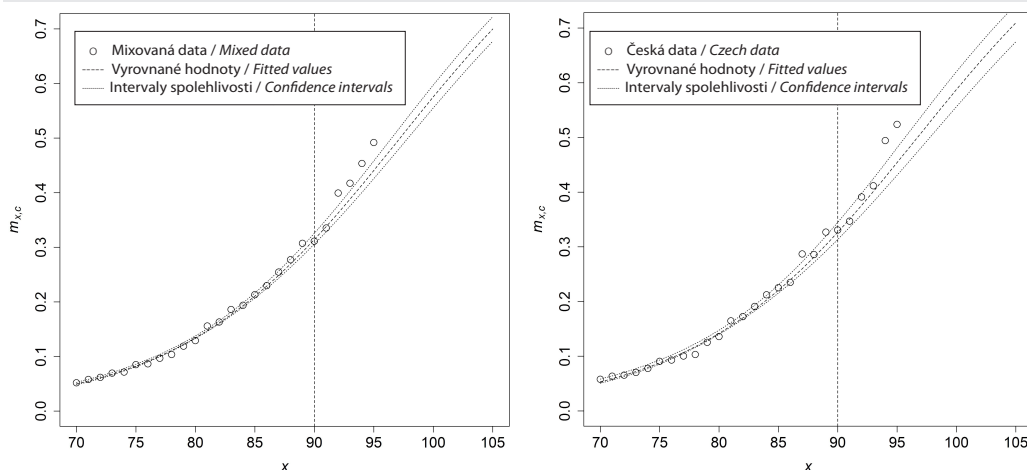
jejichž rozptyl je vzhledem k věku konstantní, neboť jsou standardizována odmocninou z predikce rozptylu \hat{m}_x . Z těchto reziduí jsou pak generovány náhodným výběrem s vrácením náhodné výběry. Tyto výběry se většinou označují jako tzv. bootstrapové výběry. Rezidua z i -tého bootstrapového výběru označíme e_x^i . Z každého bootstrapového výběru jsou pak inverzní funkcí k funkci (1) vytvářena tzv. pseudo data m_x^i

$$m_x^i = e_x^i \sqrt{\hat{m}_x} + \sqrt{\hat{m}_x}$$

a následně bootstrapové počty zemřelých $E_x m_{x,c}^i$. Tyto počty zemřelých simulují jakousi „alternativní realitu“, tj. simulují situaci, kdy bychom mohli uplynulé roky sledovat znovu a získat alternativní realizace počtu zemřelých. Z každé takové alternativní realizace získáme opět metodou maximální věrohodnosti „bootstrapové“ odhady parametrů a a b a „bootstrapové“ predikce specifických měr úmrtnosti logistického modelu (1). Odhady intervalů spolehlivosti jsou pak počítány jako 2,5%, resp. 97,5% empirický

Graf 3a a 3b: Porovnání intervalů spolehlivosti – kohorta 1886, muži

A comparison of confidence intervals – cohort 1886, males



Zdroj: Vlastní zpracování.
Source: Authors' calculation.

kvantil odhadů parametrů a příslušných predikcí spočtených v alternativních realizacích.

$$s_c = \sum_{x=90}^{105} u_{x,c} - l_{x,c},$$

7. NUMERICKÉ VÝSLEDKY

Pro odhad parametrů logistických křivek (1) byly použity všechny kohorty od 1886 do 1901 a věky od 70 do 90 let pro obě pohlaví. Využití pozorování z nižších věků by příliš snižovalo relativní váhu dat z vyšších věků, které jsou pro model nejdůležitější. Naopak použití vyšších věků je zatíženo již příliš velkou chybou. Odhad byl proveden pro česká i kredibilitní data. Odhad modelu a extrapolace do věku 105 let je pro první uvažovanou kohortu $c=1886$ pro česká a kredibilitní data pro ilustraci zobrazen v grafu 3a a 3b pro muže, respektive v grafu 4a a 4b pro ženy. Za pomoci statistického programu R bylo provedeno 15 000 bootstrapových replikací, z kterých byly vypočteny 95% intervaly spolehlivosti pro odhady parametrů a pro extrapolace od 90 do 105 let.

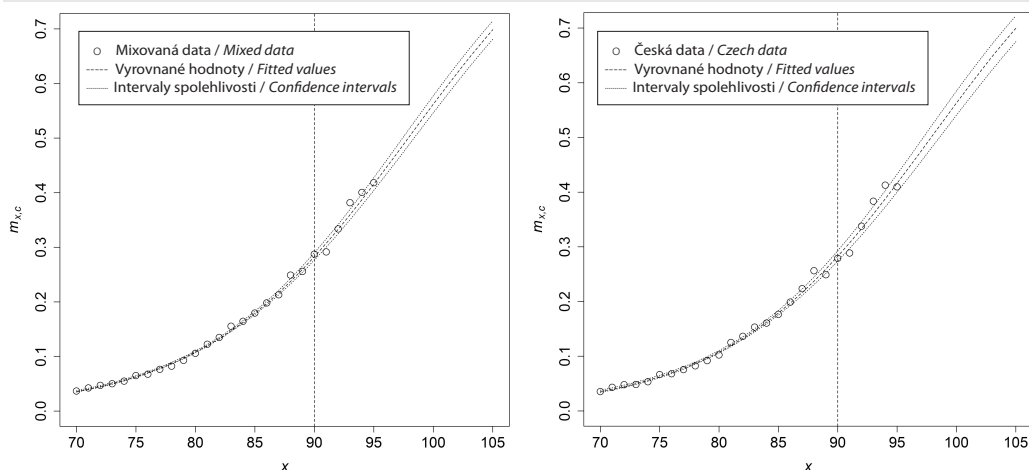
Intervaly spolehlivosti predikcí konstruovaných z českých a kredibilitních dat byly pro všechny kohorty porovnány na základě kritéria, které představuje přibližnou šířku intervalu spolehlivosti sečtenou přes všechny extrapolované věky. Toto kritérium označíme pro danou kohortu c jako s_c a definujeme ho jako

kde $u_{x,c}$ a $l_{x,c}$ značí odhad 97,5% a 2,5% meze intervalu spolehlivosti ve věku x pro kohortu c . Relativní rozdíl s_c , tedy relativní zúžení intervalů spolehlivosti při použití kredibilitních dat místo českých dat, je pro všechny kohorty zobrazeno v grafu 5. Je patrné, že díky využití externí informace došlo k významnému snížení nejistoty v předpovědích ve všech uvažovaných kohortách. V průměru o 33 % pro muže, respektive o 21 % za ženy.

Zároveň se však nepatrně liší i předpovědi. Kredibilitní data jsou na jedné straně zatížena zkreslením vzniklým zahrnutím okolních populací, ale na druhé straně jsou v nich potlačena (více diverzifikována) chybná pozorování. U malé populace je tedy i částečně zavádějící považovat změnu předpovědi díky použití kredibilitních dat za zkreslení. Česká data jsou sesbírána pouze pro českou populaci, ale malý počet pozorování způsobuje výraznější vliv chybných záznamů. V grafu 6a a 6b jsou zobrazeny relativní rozdíly v extrapolacích v maximálním predikovaném věku 105 let pro všechny kohorty (vlevo pro muže, vpravo pro ženy). Je zřejmé, že rozdíl je minimální a je vysoce kompenzován zúžením intervalu spolehlivosti.

Graf 4a a 4b: Porovnání intervalů spolehlivosti – kohorta 1886, ženy

A comparison of confidence intervals – cohort 1886, females



Zdroj: Vlastní zpracování.
 Source: Authors' calculation.

ZÁVĚR

Informace o počtu obyvatel ve vysokém věku jsou nepřesné a k jejich zpřesnění může vést použití metody vymřelých kohort. U malých populací se pak nabízí i možnost využití dat z geograficky a ekonomicky blízkých populací. Jednou z možností jak tyto informace využít je kredibilitní přístup aplikovaný v tomto článku. V článku byla zkoumána extrapolace specifické míry úmrtnosti od věku 90 do věku 105

let pro 16 kohort pomocí Poissonovské regrese za předpokladu logistického růstu úmrtnosti s věkem. Ve všech kohortách došlo k výraznému snížení rozptylu odhadu parametrů a šířky intervalů spolehlivosti. Zúžení se pohybovalo kolem 33 %. Dosažené výsledky naznačují, že při použití kredibilitního přístupu k mixování dat lze při minimální změně předpovědi dosáhnout velmi výrazného zvýšení přesnosti odhadů.

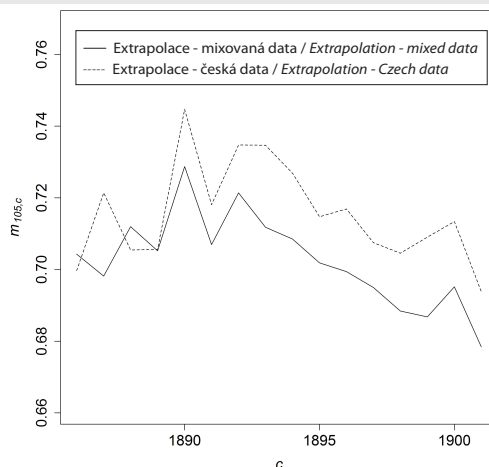
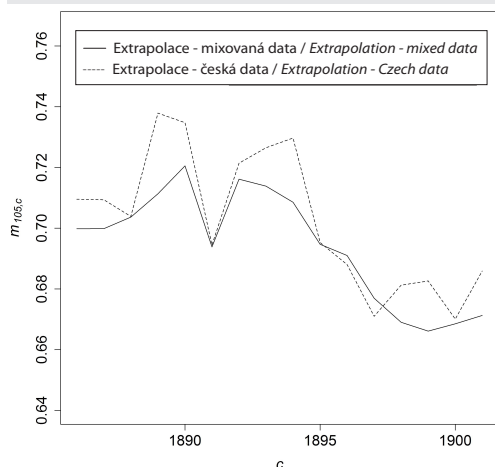
Graf 5: Relativní zúžení intervalů spolehlivosti pro všechny sledované kohorty

The relative decrease in the confidence intervals for all cohorts



Zdroj: Vlastní zpracování.
Source: Authors' calculation.

Graf 6a a 6b: Relativní rozdíl v extrapolacích ve věku 105 let pro všechny kohorty (vlevo pro muže, pravou pro ženy) | The relative difference in extrapolation at age 105 for all cohorts (left – males, right – females)



Zdroj: Vlastní zpracování.
Source: Authors' calculation.

Poděkování

Tento článek byl podpořen grantem Interní grantové agentury VŠE IGS 6/2015 “Využití bayesovských metod pro modelování úmrtnosti”.

Literatura:

- Ahcan, A. – Medved, D. – Olivieri, A. – Pitacco, E. 2014. Forecasting Mortality for Small Populations by Mixing Mortality Data. *Insurance Mathematics & Economics*, 54, s. 12–27.
- Antonio, K. – Bardoucos, A. – Ouburg, W. 2015. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, 5(2), s. 245–281.
- Bijak, J. – Dodd, E. – Forster, J. J. – Smith, P. W. F. 2015. *English Life Tables No. 17 Methodology*. Southampton: University of Southampton & Office for National Statistics, Titchfield.
- Boland, P. J. 2007. *Statistical and Probabilistic Methods in Actuarial Science*. Chapman & Hall/CRC Interdisciplinary Statistics. ISBN 9781584886952.
- Brouhns, N. – Denuit, M. – Vermunt, J. K. 2002. A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance Mathematics & Economics*, s. 373–393.
- Burcin, B. – Tesárková, K. – Šidlo, L. 2010. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie*, 52(2), s. 77–89.
- Cipra, T. 2006. *Pojistná matematika: teorie a praxe*. Praha: Ekopress. ISBN 80-86929-11-6.
- Efron, B. – Tibshirani, R. J. 1994. *An introduction to the bootstrap*. Chapman and Hall/CRC. ISBN 9780412042317.
- Gavrilov, L. – Gavrilova, N. 2011. Ageing and longevity: Mortality laws and mortality forecasts for ageing populations. *Demografie*, 53(2), s. 109–128.
- Gavrilova, N. – Gavrilov, L. 2014. *Mortality Trajectories at Extreme Old Ages: A Comparative Study of Different Data Sources on US Old-Age Mortality*. Living to 100 monograph.
- 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, s. 513–583.
- Heligman, L. – Pollard, J. 1980. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01), s. 49–80.
- Koschin, F. 1999. Jak vysoká je intenzita úmrtnosti na konci lidského života. *Demografie*, (41)2, s. 105–119.
- Lee, R. D. – Carter, L. R. 1992. Modeling and forecasting us mortality. *Journal of the American statistical association*. 1992, 87(419), s. 659–671.
- Long Chiang, Ch. 1984. *The life table and ITS applications*. Florida: University of California, Berkeley, 1984, s. 316. ISBN 0-89874-570-5.
- Makeham, W. M. 1860. On the Law of Mortality and the Construction of Annuity Tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 1860, 8(6), s. 301–310.
- Perks, W. 1932. On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63(1) s. 12–57.
- Pitacco, E. – Denuit, M. – Haberman, S. 2009. *Modelling longevity dynamics for pensions and annuity business*. Oxford: Oxford University Press.
- Hulíková Tesárková, K. 2012. *Selected methods of mortality analysis focused on adults and the old-age*. Disertační práce, Praha: Univerzita Karlova.
- Thatcher, A. R. 1999. The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1), s. 5–43.
- Thatcher, A. R. – Kannisto, V. – Vaupel, J. W. 1998. *The force of mortality at ages 80 to 120* Odense: Odense University Press Odense.
- United Nations, Department of Economic and Social Affairs, Population Division. 2015. *World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections*. Working Paper No. ESA/P/WP.242. New York: UN.
- Wilmoth, J. R. – Shkolnikov, V. – Barbieri, M. 2012. Human mortality database 2012.
- Wilmoth, J. R. – Andreev, K. – Jdanov, D. – Gleij, D. A. 2007. *Methods Protocol for the Human Mortality Database*.
- Zeng, Y. – Vaupel, J. W. 2003. Oldest old mortality in China. *Demographic Research*, 2003, 8, s. 215–244.

MARTIN MATĚJKA

je absolventem oboru statisticko-pojistné inženýrství na Vysoké škole ekonomické v Praze, kde v roce 2014 nastoupil na prezenční formu doktorského studia. Mezi jeho hlavní oblast zájmu se řadí využití zobecněných dynamických lineárních modelů v pojištnictví a demografii. Od roku 2011 pracuje také jako analytik v oblastech finanční a pojistné matematiky.

TOMÁŠ KAREL

je absolventem oboru statisticko-pojistného inženýrství na Vysoké škole ekonomické v Praze, je studentem doktorského studia na katedře statistiky a pravděpodobnosti. Zabývá se především bayesovskou statistikou a její aplikací v makroekonomii.

JAN FOJTÍK

je absolventem oboru statisticko-pojistného inženýrství na Vysoké škole ekonomické v Praze, kde v roce 2016 nastoupil na prezenční formu doktorského studia. Mezi jeho hlavní oblast zájmu se řadí pojištnictví, zejména moderní přístupy v životním pojištění. V oblasti pojištnictví se od roku 2014 pohybuje také jako konzultant.

PAVEL ZIMMERMANN

je od roku 2005 odborným asistentem na katedře statistiky a pravděpodobnosti Vysoké školy ekonomické v Praze. Zabývá se statistickými modely a jejich aplikacemi v neživotním pojištění a pro účely modelování úmrtnosti.

SUMMARY

The predictions of old-age models typically suffer from a high level of uncertainty as data are often lacking or are of very low quality. This article presents one possible way of reducing uncertainty in modelling. The mortality model is based on cohort data on the number of deaths and the population exposed to risk of death. In this article the authors improve the quality of the data on the population exposed to risk of death using the extinct cohort method for the oldest age. They reduce the uncertainty of the estimates by using multiple sources of mortality data. They use the credibility approach, where data

from several countries are mixed. In this case Czech data are mixed with data from surrounding countries (Hungary, Slovakia, and former East Germany) with weights determined using an optimisation method. The maximum likelihood method is used to fit the logistic model to the old-age mortality data. The uncertainty in the model is compared with a model based on a single data source from the Czech Republic. Using the credibility approach and mixed mortality data results in narrower confidence intervals of the estimates and consequently more accurate predictions.