VYUŽITÍ METOD ANALÝZY PŘEŽITÍ PRO MODELOVÁNÍ DOBY NEZAMĚSTNANOSTI V ČESKÉ REPUBLICE

Adam Čabla, Ivana Malá*

Abstract

The Use of Survival Analysis Methods for the Modelling of Unemployment in the Czech Republic

Unemployment belongs to the most serious economic and social problems of developed countries. The problem can be described by the unemployment rate or number of the unemployed, in this text a duration of unemployment is of interest. The unemployment duration in the Czech Republic in 2008, 2010 and 2014 is analysed with the use of survival analysis methods and a finite mixture of lognormal distributions is used to describe an overall distribution of unemployment spell as well as the component distributions given by gender and education of the unemployed. Data from the Labour Force Sample Survey (performed by the Czech Statistical Office) are used for the statistical analysis. The unemployment duration is given in the questionnaire of the survey in intervals, we supposed the data to be right or interval censored, exact values of the unemployment duration are not included in the data. The strong positive effect of education on the duration of unemployment is quantified, as well as a less distinctive gender gap. An increase in unemployment duration is quantified for the period of economic crisis with respect to periods before (2008) and after (2014) the crisis.

Keywords: unemployment duration, censored data, finite mixture of probability distributions, survival analysis

JEL Classification: C41, J64, C24

Úvod

Nezaměstnanost je problémem všech tržních ekonomik, proto je jejímu zkoumání věnována pozornost širokého spektra analytiků a výsledky týkající se nezaměstnanosti (například aktuální hodnoty míry nezaměstnanosti) jsou očekávány s velkým zájmem státních orgánů, ekonomických subjektů i laické veřejnosti. Obdobně velkým problémem, jako je počet nezaměstnaných, respektive míra nezaměstnanosti, je také doba, po kterou nezaměstnaní novou práci hledají. V případě nezaměstnanosti delší než jeden rok hovoříme o dlouhodobé nezaměstnanosti, míra dlouhodobé nezaměstnanosti je také sledována jako důležitá charakteristika stavu trhu práce a také ekonomiky jako celku (Krueger a kol., 2014). Počet nezaměstnaných a jejich strukturu, počet volných míst a délku nezaměstnanosti ovlivňuje mnoho nejrozličnějších faktorů. Vyhledávat tyto faktory, popsat a kvantifikovat jejich vliv, případně následně formulovat a realizovat opatření, která by se snažila

* Adam Čabla (adam.cabla@vse.cz), Ivana Malá (malai@vse.cz), Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky.
Článek vznikl v rámci projektu IG410045 *Doba nezaměstnanosti po krizi* poskytnutého Fakultou informatiky a statistiky Vysoké školy ekonomické v Praze.

negativní vlivy potlačovat a pozitivní vlivy podporovat, je stálým úkolem všech orgánů, které jsou za vývoj ekonomiky zodpovědné. Z demografických faktorů jde o pohlaví, vzdělání, věk, bydliště a podobně, silný vliv má systém, výše a doba výplaty dávek v nezaměstnanosti (Hunt, 1995, Røed a kol., 2007, Bover a kol., 2002, Jurajda, Tannery, 2003), daňový systém a také existence a výše minimální mzdy (Daveri, Tabellini, 2000 nebo Hagedorn, Manovskii, 2013). Šance na trhu práce je ovlivněna také zdravotním stavem a případnými zdravotními omezeními; nezaměstnanost (zvláště dlouhodobá) naopak negativně ovlivňuje zdravotní stav nezaměstnaných (Korpi, 2001, Krueger a kol., 2011).

Cílem příspěvku je modelovat rozdělení délky nezaměstnanosti v České republice v letech 2008 (před nástupem ekonomické krize), 2010 (v době ekonomické krize) a 2014 (v době oživení ekonomiky) a charakterizovat rozdíly mezi jednotlivými obdobími. Rozdělení doby nezaměstnanosti je odhadnuto zvlášť pro muže a ženy a dále pro skupiny definované nejvyšším dosaženým vzděláním nezaměstnaného. Rozdělení doby nezaměstnanosti v České republice je pak konstruováno jako konečná směs rozdělení podskupin. Předkládané výsledky poskytují pohled na dynamiku (ve třech obdobích) a rozdělení délky nezaměstnanosti z pohledů, které se běžně nepoužívají, ale mohou poskytnout zajímavou informaci. Sestrojené modely považujeme za přijatelný a užitečný popis situace jednotlivých skupin nezaměstnaných vzhledem k délce nezaměstnanosti.

V tomto textu se zabýváme pouze zmíněnými dvěma charakteristikami nezaměstnaných, závislost na dalších vysvětlujících proměnných je zkoumána například v práci Čably (2014) pro data z prvních dvou období zkoumaných v tomto textu a Čably (2016) pro data posledního období. Články se zabývají vlivem proměnných, jako jsou velikost domácnosti, věk, velikost bydliště, rodinný stav, autor ovšem používá regresní model zrychleného času (AFT), nikoliv v tomto textu používaný model směsí, který je pro tolik proměnných obtížně použitelný. V práci Čably, Malé (2017) je použit regresní model AFT pro všechna tři období na data, která jsou použita i v tomto textu.

Pro zkoumání doby nezaměstnanosti lze využít registr Úřadů práce nebo data z Výběrového šetření pracovních sil (VŠPS), které provádí Český statistický úřad (ČSÚ, 2016). V tomto textu budeme využívat data z VŠPS, která tvoří náhodný výběr v kontrastu s databází ÚP, obsahující všechny nezaměstnané, kteří se na úřadu registrovali. V použitých datech z VŠPS byl překryv (respondenti uvedli, že jsou registrováni na úřadu práce) 74–81 %. Údaje v šetření obsahují délku nezaměstnanosti pouze v intervalech a bez přesných dob délky v případě, že nezaměstnaný práci našel. Pro ty, kteří práci ve zkoumaném období nenašli, je známa doba nezaměstnanosti a informace, že k nalezení práce nedošlo. V takovém případě je pro modelování rozdělení doby vhodné použít metodiku analýzy přežití, která pracuje s neúplnými (cenzorovanými) daty. Neparametrické postupy byly využity například v práci Čably (2014) pro data z prvních dvou období zkoumaných v tomto textu, Wicher a Wike (2008) použili semiparametrický regresní model pro modelování délky nezaměstnanosti v Německu. Regresní model AFT nebo Coxův regresní model byly aplikovány Jarošovou (2006). V tomto textu je použit parametrický model konečných směsí pravděpodobnostních rozdělení (komponenty jsou definovány vysvětlujícími proměnnými a tedy příslušnost ke komponentám je pozorována) navazující na článek Malé (2013).

1. Model a data

1.1 Použitá data a formulace problému v rámci teorie analýzy přežití

V předkládané analýze použijeme data z VŠPS ze tří období obsahujících pět po sobě následujících čtvrtletí (čtvrté čtvrtletí prvního roku až čtvrté čtvrtletí druhého roku z let 2007–2008 a 2013–2014 a první čtvrtletí 2010 až první čtvrtletí 2011). Šetření provádí čtvrtletně Český statistický úřad (ČSÚ) od roku 1992; od roku 2002 jsou obsah a forma použitého dotazníku harmonizovány se standardem Evropské unie a dotazník je tak národní modifikací celoevropského šetření Labour Force Sample Survey (LFSS, 2015) a výsledky šetření jsou využívány hlavně pro mezinárodní srovnání. Výběrovou jednotkou šetření je byt, byty jsou do šetření zařazovány prostřednictvím dvoustupňového náhodného výběru a tvoří rotující panel, ve kterém byty setrvávají vždy jeden rok. Šetření obsahuje základní informace o šetřeném bytu a domácnostech v něm žijících. Pro osoby starší patnácti let obvykle bydlících v bytech udává jejich ekonomické postavení, charakteristiku hlavního, resp. druhého, zaměstnání, předchozí pracovní zkušenost, způsob hledání zaměstnání, vzdělávání a údaj o situaci respondenta před rokem. Podle metodiky VŠPS se za nezaměstnané považují všechny osoby patnáctileté a starší, které v průběhu referenčního týdne (týdne, ve kterém je konáno šetření v daném bytě) nebyly zaměstnané, byly připraveny k nástupu do práce ihned nebo do čtrnácti dnů a v průběhu posledních čtyř týdnů hledaly aktivně práci (ČSÚ, 2016). Ve výběrovém šetření VŠPS jsou v dotazníku nezaměstnaným nabízeny pro délku nezaměstnanosti intervaly: do jednoho měsíce, 1–3 měsíce, 3–6 měsíců, 6–12 měsíců, 1–2 roky, 2–4 roky a déle než čtyři roky, žádné přesné doby do nalezení práce nejsou registrovány.

Dobu do nalezení (znovunalezení) práce budeme považovat za náhodnou veličinu představující dobu do události. Modelováním rozdělení takových veličin se zabývá analýza přežití, jejíž název pochází z medicíny, kde je často sledována doba do úmrtí. Dobu nezaměstnanosti označíme T a použijeme silné nástroje analýzy dat vyvinuté v této teorii k modelování jejího pravděpodobnostního rozdělení. Budeme předpokládat, že T má spojitě pravděpodobnostní rozdělení nabývající pouze nezáporných hodnot. Vzhledem k tomu, že populace nezaměstnaných je velmi nehomogenní vzhledem k postavení na trhu práce, a tedy také k rozdělení doby nezaměstnanosti, budeme předpokládat, že množina nezaměstnaných se skládá z homogennějších podmnožin takových, že je již možné nalézt vhodný model pro dobu nezaměstnanosti v těchto podmnožinách (komponentách). Jako model pro komponentní rozdělení bude použito dvouparametrické logaritmicko-normální rozdělení, které je jednovrcholové a kladně sešikmené. Volba pravděpodobnostního rozdělení je základním předpokladem úspěšného parametrického modelu. Pro zkoumaná data byly porovnány modely využívající také jiná rozdělení (Weibullovo, gama, Dagumovo nebo logaritmicko-logistické), podle Akaiikova informačního kritéria se ukázala být nejlepší volbou rozdělení logaritmicko-normální a velmi podobné logaritmicko-logistické s tím, že pro odhady Dagumova rozdělení nemáme dostatek pozorování.

Hustotu rozdělení celkové doby nezaměstnanosti f najdeme jako směs (K) komponentních logaritmicko-normálních rozdělení ve tvaru (McLachlan, Peel, 2000)

$$f(t; \psi) = \sum_{j=1}^K \pi_j f_j(t; \mu_j, \sigma_j^2), \quad (1)$$

kde $f_j(t; \mu_j, \sigma_j^2)$, $j = 1, 2, \dots, K$ jsou hustoty komponentních logaritmicko-normálních rozdělení, váhy splňují podmínky $0 \leq \pi_j \leq 1$, $\sum_{j=1}^K \pi_j = 1$ a charakterizují podíl jednotlivých složek v celé populaci. Komponentní hustoty závisejí na dvourozměrných vektorech (neznámých) parametrů (μ_j, σ_j^2) , $j = 1, 2, \dots, K$, vektor ψ obsahuje všechny neznámé parametry v modelu, tedy $K - 1$ parametrů π_j , $j = 1, 2, \dots, K - 1$, (K -tá hodnota je určena jako doplněk součtu $K - 1$ hodnot π_j do jedničky) a $2K$ složek parametrů komponentních rozdělení. Parametry logaritmicko-normálního rozdělení mají význam střední hodnoty (parametr μ a rozptylu (parametr σ^2) náhodné veličiny $\ln T$. Pro základní charakteristiky komponentních rozdělení platí (označíme T_j dobu nezaměstnanosti v j -té komponentě ($j = 1, 2, \dots, K$), $E(T_j)$ její střední hodnotu, $D(T_j)$ rozptyl a $t_{j,P}$, 100P% kvantil pro $0 < P < 1$)

$$E(T_j) = e^{\mu_j + \sigma_j^2/2}, t_{j,0.5} = e^{\mu_j} \text{ a } D(T_j) = e^{2\mu_j + \sigma_j^2} (e^{\sigma_j^2} - 1),$$

střední hodnota i rozptyl komponentních rozdělení tedy závisejí na obou parametrech (na rozdílu od mediánu). V modelu konečné směsi můžeme použít vážený průměr (1) komponentních hodnot pro nalezení střední hodnoty a distribuční funkce směsi ve tvaru

$$E(T) = \sum_{j=1}^K \pi_j E(T_j), \quad F(t) = \sum_{j=1}^K \pi_j F_j(t), \quad t > 0, \quad (2)$$

kde F_j je distribuční funkce doby nezaměstnanosti v j -té komponentě (McLachlan, Peel, 2000, Lawless, 2003). V analýze přežití se místo distribuční funkce používá její doplněk funkce přežití S ($S(t) = P(T > t) = 1 - F(t)$). Pro konečnou směs opět platí

$$S(t) = P(T > t) = \sum_{j=1}^K \pi_j S_j(t), \quad t > 0, \quad (3)$$

kde S_j , $j = 1, 2, \dots, K$, jsou funkce přežití komponentních rozdělení. Pro kvantily rozdělení směsi podobné rovnice neplatí a je třeba je hledat pomocí numerických metod jako inverzní funkci k distribuční funkci definované v (2).

Důležitou charakteristikou rozdělení v analýze přežití je riziková funkce (označíme ji $h(t)$) udávající intenzitu událostí v čase t . Pro krátký časový interval $(t, t + \Delta t)$ platí, že pravděpodobnost, že událost nastane během tohoto časového intervalu za podmínky, že k ní nedošlo do času t , je $h(t)\Delta t$. V našem případě popisuje funkce h intenzitu nalezení práce v krátkém časovém intervalu po čase t , jestliže do této doby nezaměstnaný pracovník nenalezl. Funkce rizika je definována jako (Lawless, 2003)

$$h(t) = \frac{f(t)}{S(t)}, \quad t > 0. \quad (4)$$

Použijeme (1)–(3), po dosazení do (4) dostaneme pro konečnou směs

$$h(t; \Psi) = \frac{\sum_{j=1}^K \pi_j f_j(t; \mu_j, \sigma_j^2)}{\sum_{j=1}^K \pi_j S_j(t; \mu_j, \sigma_j^2)} = \frac{\sum_{j=1}^K \pi_j S_j(t; \mu_j, \sigma_j^2) h_j(t; \mu_j, \sigma_j^2)}{\sum_{j=1}^K \pi_j S_j(t; \mu_j, \sigma_j^2)}, t > 0, \quad (5)$$

kde h_j je funkce rizika v j -té komponentě, ve vzorci jsou vyznačeny neznámé parametry. Vzorec (5) definuje funkci rizika směsi opět jako vážený průměr komponentních funkcí, váhy jsou ale v tomto případě závislé také na komponentním rozdělení (prostřednictvím funkce přežití).

Funkce rizika pro logaritmicko-normální rozdělení je rostoucí do maximální hodnoty a dále klesající (Lawless, 2003). Znamená to, že intenzita získávání práce nejdříve roste, nabývá maxima a pak s rostoucí délkou nezaměstnanosti klesá. Výše uvedené vlastnosti odpovídají požadavkům, které na rozdělení doby hledání práce intuitivně klademe. Funkce rizika konečné směsi logaritmicko-normálních rozdělení nemusí mít obecně tento tvar, v případě odhadnutých směsí v tomto textu tomu tak ovšem bude. Uveďme ještě, že odhadnuté Weibullovo rozdělení poskytovalo podobné odhady charakteristik jako je střední hodnota nebo rozptyl, znamenalo ale funkci rizika mající naopak minimum a rostoucí s časem. Tento průběh není v souladu s poznatky týkajícími se nezaměstnanosti, neboť by to znamenalo, že čím delší je doba nezaměstnanosti, tím větší je intenzita nacházení práce.

Pro porovnání rozdělení doby nezaměstnanosti pro různé skupiny nezaměstnaných může sloužit také zbytková doba nezaměstnanosti, definovaná jako $T | T > t$ pro pevná $t \geq 0$. Pro $t = 0$ se jedná o celkovou dobu nezaměstnanosti $E(T)$. Budeme se tedy zajímat o dobu nezaměstnanosti za podmínky, že nezaměstnaný je již nezaměstnaným po dobu t . Pravděpodobnostní rozdělení této doby je podmíněným rozdělením doby T za podmínky, že $T > t$. Funkci přežití $S_i(t')$ tohoto podmíněného rozdělení ($S_i(t') = P(T > t + t' | T > t)$) lze zapsat jako vážený aritmetický průměr komponentních podmíněných funkcí přežití $S_{j,t}$, $j = 1, 2, \dots, K$, $t > 0$, ve tvaru ($S_i(t') = 1, t' \leq 0$).

$$S_i(t') = \frac{\sum_{j=1}^K \pi_j S_j(t + t')}{\sum_{j=1}^K \pi_j S_j(t)} = \frac{\sum_{j=1}^K \pi_j S_j(t) S_{j,t}(t')}{\sum_{j=1}^K \pi_j S_j(t)} = \sum_{j=1}^K \frac{\pi_j S_j(t)}{\left[\sum_{i=1}^K \pi_i S_i(t) \right]} S_{j,t}(t'), t' > 0. \quad (6)$$

Pomocí numerických metod byly určeny charakteristiky polohy zbytkové doby nezaměstnanosti (medián m_i a střední hodnota e_i) jako funkce času t . Pro sledovanou veličinu měly všechny komponentní funkce (i funkce směsí) minimální hodnotu. Čas, ve kterém bylo dosaženo minima střední zbytkové doby e a mediánové zbytkové doby m , a hodnota funkcí v tomto bodě jsou tabelovány jako charakteristiky, které mohou pomoci popsat vlastnosti sledovaných délek nezaměstnanosti.

1.2 Popis použitých modelů

Sestrojíme model rozdělení doby nezaměstnanosti jako směs logaritmicko-normálních rozdělení (1). Budeme uvažovat komponenty definované pohlavím nezaměstnaného (směs dvou rozdělení pro muže a ženy, $K = 2$) a dále nejvyšším dosaženým vzděláním (směs čtyř rozdělení ($K = 4$) pro komponenty definované pro základní vzdělání (a bez vzdělání, Z), středoškolské vzdělání bez maturity (S), úplné středoškolské vzdělání s maturitou (ÚS) a vysokoškolské vzdělání (VŠ)). Zvolený model obsahuje v prvním případě pět neznámých parametrů (π_1 a $\mu_j, \sigma_j^2, j = 1, 2$), ve druhém jedenáct parametrů (π_1, π_2, π_3 a $\mu_j, \sigma_j^2, j = 1, 2, 3, 4$). Zvolený přístup umožňuje získat nejen informaci o rozdělení doby nezaměstnanosti pro všechny nezaměstnané, ale můžeme porovnat také jednotlivé složky nezaměstnaných mezi sebou.

Data, na základě kterých budeme parametry směsi odhadovat, obsahují pouze cenzorované hodnoty (stoprocentní cenzorování). Pro nezaměstnané, kteří práci ve sledované době našli, známe časový interval, ve kterém k nalezení práce došlo (intervalové cenzorování). Pro ty, kteří práci nenašli, známe pouze čas, do kterého k nalezení práce nedošlo, a data považujeme za zprava cenzorovaná. Podle McLachlana, Peelea (2000) lze v případě, že známe příslušnost nezaměstnaných do jednotlivých komponent směsi, rozdělit všechna pozorování do komponent a v nich odděleně odhadnout parametry komponentních rozdělení metodou maximální věrohodnosti. Maximálně věrohodnými odhady pravděpodobností $\pi_j, j = 1, 2, \dots, K$, jsou relativní četnosti pozorování z j -té složky v celém výběru. Na rozdíl od úplných dat, kdy jsou známy explicitní vzorce pro odhady parametrů logaritmicko-normálních rozdělení, je v tomto případě třeba použít numerickou proceduru maximalizující věrohodnostní funkci (Lawless, 2003). Numerické metody byly použity i pro výpočet dalších charakteristik v případě, že nebylo možné použít známé vzorce a vztahy. Všechny výpočty byly provedeny v programu R (R CORE TEAM, 2014), pro numerické hledání maximálně věrohodných odhadů parametrů rozdělení složek byl použit balíček Survival (Therneau, 2015). Pro testování rozdílů mezi složkami směsi byl použit balíček Interval (Fay, Shaw, 2010) a neparametrický logrank test (pro intervalově cenzorovaná data). Při odhadu směrodatných odchylek odhadů kvantilů byla v tabulce 1 a v tabulkách 3 a 4 použita metoda bootstrap, simulace (vždy 1 000 opakování) byly provedeny v balíčku Boot (Canty, Ripley, 2016). Odhad přesnosti určení střední hodnoty byl získán pomocí Taylorova rozvoje a odhadnuté kovarianční matice odhadů parametrů (získané balíčkem Survival).

2. Výsledky studie

2.1. Nezaměstnanost v České republice ve sledovaném období

Český statistický úřad pravidelně publikuje počty nezaměstnaných, kteří nenašli práci, na základě šetření VŠPS v intervalech 0–3 měsíce, 3–6 měsíců, 6 měsíců až jeden rok (6–12), jeden až dva roky (12–24) a více než dva roky (24+). Rozložení četností nezaměstnaných v jednotlivých intervalech je znázorněno na obrázku 1 (pro muže, ženy

a celou populaci) v letech 2007 až 2014. Mezi nezaměstnanými bylo ve sledovaných letech mezi 50 a 55 procenty žen, ačkoliv ze zaměstnaných bylo žen 42 až 45 % (ČSÚ, 2016). Dva nejtmaší sloupce na obrázku 1 obsahují dlouhodobě nezaměstnané (osoby, které jsou bez práce déle než jeden rok). Tato skupina je často zkoumána samostatně, situaci v České republice se zabývá například práce Löstera, Langhamrové (2011).

V roce 2008 vlivem hospodářské krize došlo ke snížení zahraniční poptávky, které přimělo podniky omezit výrobu a propouštět zaměstnance. Dalším důvodem pro snížení poptávky po zboží a službách plynulo z obavy z nepříznivého vývoje v budoucnu, které mělo za následek snížení spotřeby domácností a přimělo obchodní společnosti k odložení investic. Proto společnosti propouštěly zaměstnance, kteří nebyli nepostradatelní, nejprve zaměstnance méně kvalifikované, v době krize došlo k propouštění i zaměstnanců kvalifikovaných (tento postup popisují také rozdělení mezd a jejich vývoj). Současné vzhledem k nedostatku poptávky a při očekáváních nevznikala nová pracovní místa. Propouštění a nedostatek volných míst vedlo krátkodobě ke změně struktury nezaměstnaných (přibývá krátkodobě nezaměstnaných). Pokud ale tito nenalézají nové zaměstnání, jejich doba nezaměstnanosti roste a s ní i doba nezaměstnanosti v celé České republice. Při oživení ekonomiky dochází k zvýšení počtu volných míst, a tedy více osob nachází práci. V letech 2007 a 2008 bylo dlouhodobě nezaměstnaných přes 50 % nezaměstnaných osob, od 2012 se procento ustálilo kolem 48 %. Na obrázku 1 je zřejmé, že rozdíl v délce nezaměstnanosti mezi muži a ženami není příliš velký, tato skutečnost bude pozorována i v dalším textu z jiných pohledů. Podíl dlouhodobě nezaměstnaných žen je téměř stejný (2008, 2011 a 2014) nebo větší než podíl mužů.

Obrázek 1 | Procentní rozložení uchazečů o zaměstnání podle délky hledání práce v letech 2007–2014

Poznámka: Sloupce v pořadí muži, ženy a všichni nezaměstnaní dohromady.

Zdroj: Český statistický úřad

Obrázek 2 | Počet nezaměstnaných podle nejvyššího dosaženého vzdělání (obrázek vlevo) a procentní zastoupení nezaměstnaných s daným vzděláním mezi nezaměstnanými (vpravo)

Zdroj: Český statistický úřad

Na obrázku 2 vlevo je v jednotlivých letech zobrazen počet nezaměstnaných v České republice, nezaměstnaní jsou rozděleni do skupin podle nejvyššího dosaženého vzdělání (od nejnižšího do nejvyššího). Nejmenší (absolutní) počet nezaměstnaných byl v roce 2008, pak rostl do nejvyšší hodnoty v roce 2010. Po kolísání v dalších letech došlo mezi lety 2013 a 2014 k patrnému poklesu počtu nezaměstnaných (celkem 12,3 %, viz obrázek 2; 9,3 % v registru úřadů práce). Na obrázku 2 vpravo byly počty nezaměstnaných z obrázku vlevo převedeny na procentní rozložení mezi všemi nezaměstnanými. Obě části obrázku 2 ukazují, že s ekonomickou krizí přibývalo nezaměstnaných také mezi středoškolsky a vysokoškolsky vzdělanými pracovníky.

2.2 Výsledky modelování doby nezaměstnanosti

Z databáze VŠPS ve sledovaných obdobích byla použita data o všech respondentech ve věku 16–65 let, kteří byli zahrnuti aspoň do jednoho z pěti po sobě následujících šetření VŠPS aspoň v jednom šetření (kterého se zúčastnili třeba i před uvedenou dobou) byli nezaměstnaní a jejich délka nezaměstnanosti byla do dvou let. Tuto informaci poskytuje databáze šetření, neboť obsahuje postavení respondenta (zaměstnaný, nezaměstnaný, neaktivní) v celé historii, kdy byl zahrnut do šetření. Volba zahrnuje krátkodobě nezaměstnané a první rok dlouhodobé nezaměstnanosti (na obrázcích 1 a 2 sloupce bez poslední, nejtmavší části). Vzhledem k tříměsíční době mezi šetřeními byla horní mez intervalu (po všech úpravách) nezaměstnanosti stanovena na 27 měsíců (pro nezaměstnané v intervalu 18–24 měsíců při jedné návštěvě a zaměstnané do jednoho měsíce při další).

Průměrný věk nezaměstnaných v analyzovaném souboru byl v jednotlivých obdobích 36,6 let (směrodatná odchylka 12,9 let), 37,5 let (13,1) a 37,2 (12,7). Průměrný věk mužů byl 37,2 (14,1), 37,6 (13,8) a 37,5 (13,6) a průměrný věk žen 35,9 (11,8), 37,5 (12,3) a 36,8 (11,8). Rozdíly průměrů pro muže a ženy jsou maximálně 1,3 roku, tyto rozdíly jsou však (díky rozsahu výběrů) statisticky významné (dosažené hladiny významnosti (p -hodnoty) t -testu s oboustrannou alternativou jsou menší než 0,0001). Nezaměstnaní, kteří práci našli, byli spíše mladší (35,7 let (12,5), 36,0 (12,6) a 36,1 (12,2)) než ti, kteří práci nenašli (37,2 (13,1), 38,1 (13,2) a 38,2 (13,0)), rozdíly průměrného věku jsou 1,5 roku v prvním období a 2,1 roku v dalších (pro všechna období jsou p -hodnoty testů s jednostrannou alternativou menší než 0,001). Tyto hodnoty odpovídají tezi, že vyšší věk je faktorem, který snižuje šance na nalezení zaměstnání.

Pokud nezaměstnaný po dobu sledování našel zaměstnání a zase ho ztratil, byl započítán pouze jednou jako nezaměstnaný, který našel zaměstnání. Žádný nezaměstnaný, který by našel (a ztratil) ve sledované době maximálně jednoho roku zaměstnání dvakrát, nalezen nebyl.

V tabulce 1 jsou uvedeny počty nezaměstnaných (sloupec n) celkem a v jednotlivých komponentách. V závorce jsou uvedeny počty těch nezaměstnaných, kteří práci ve sledovaném období našli. Relativně malý počet nezaměstnaných, kteří ve sledované době maximálně jednoho roku práci našli, vyplývá z panelového uspořádání dat a patrně způsobuje přecenění doby nezaměstnanosti v odhadech založených na těchto datech. Dále tabulka obsahuje odhady charakteristik polohy (střední hodnoty, mediánu, kvartilů $t_{0,25}$, $t_{0,75}$) a variability (kvantilové odchylky $q = 0,5$ ($t_{0,75} - t_{0,25}$)). Všechny hodnoty jsou uvedeny v měsících a v závorce je uveden odhad jejich směrodatné odchylky (přesnosti). Pomocí Akaikeho kritéria byl jako model pro všechny nezaměstnané vybrán model směsi založený na vzdělání, výsledky dosažené tímto modelem jsou obsaženy vždy v posledním řádku jednotlivých bloků v tabulce s názvem směs.

Použití směsi rozdělení je výhodné v případě, že zkoumaná populace obsahuje takové podmnožiny, že rozdělení analyzovaného znaku je v nich rozdílné. Shoda komponentních rozdělení byla posouzena pomocí neparametrického logrank testu bez předpokladu logaritmicko-normálního rozdělení (Fay, Shaw, 2010). Při volbě komponent definovaných pohlavím nezaměstnaného byl rozdíl v rozdělení doby nezaměstnanosti statisticky významný pouze v období krize (p -hodnoty postupně 0,16, 0,0006, 0,22), pro komponenty definované nejvyšším dosaženým vzděláním byly statisticky významné rozdíly ve všech obdobích (p -hodnoty menší než 0,0001).

V tabulce 1 jsou kvantifikovány rozdíly mezi sledovanými skupiny nezaměstnaných, které byly očekávány. Charakteristiky polohy jsou delší pro ženy než pro muže a vždy je patrný klesající trend s rostoucím vzděláním, vliv vzdělání na délku nezaměstnanosti je silnější než vliv pohlaví. Rozdíly mezi muži a ženami jsou od půl měsíce do dvou, rozdíl pro vzdělání je větší (zhruba dvojnásobné doby nezaměstnanosti pro nezaměstnané se základním vzděláním než pro vysokoškoláky). Rozdíly mezi střední školou bez maturity a s maturitou jsou maximálně půl měsíce, výsledky jsou velmi podobné (a doby pro úplné střední vzdělání vždy kratší). V dalším textu uvidíme, že v době krize se projevil

Tabulka 1 | Odhady charakteristik polohy a variability pro komponenty směsi

Složka	<i>n</i>	$\hat{t}_{0,75}$	medián doby	$\hat{t}_{0,75}$	střední doba	$\hat{q}$
2007/2008						
muži	1 265 (514)	6,07 (0,23)	11,07 (0,41)	20,18 (1,03)	16,45 (3,13)	7,06 (1,59)
ženy	1 628 (613)	6,55 (0,22)	11,88 (0,40)	21,54 (1,02)	17,54 (0,92)	7,50 (0,48)
Z	553 (161)	8,34 (0,49)	15,67 (1,06)	29,44 (2,92)	24,26 (13,04)	10,55 (6,44)
S	1 291 (507)	6,32 (0,22)	11,21 (0,39)	19,84 (0,96)	16,09 (2,50)	6,78 (1,52)
ÚS	826 (306)	5,97 (0,28)	10,75 (0,50)	19,39 (1,24)	15,75 (1,12)	6,71 (0,58)
VŠ	223 (100)	4,51 (0,40)	8,00 (0,65)	14,21 (1,52)	11,50 (2,42)	4,58 (1,05)
směs	2 893 (1127)	6,30 (0,16)	11,43 (0,29)	20,93 (0,72)	17,19 (0,66)	7,30 (0,34)
2010/2011						
muži	2 352 (790)	7,07 (0,19)	13,27 (0,41)	24,92 (1,06)	20,52 (0,99)	8,92 (0,49)
ženy	2 401 (711)	7,93 (0,22)	14,97 (0,50)	28,24 (1,34)	23,31 (1,25)	10,15 (0,62)
Z	713 (135)	11,57 (0,74)	23,68 (2,12)	48,45 (6,29)	41,60 (6,47)	18,44 (2,91)
S	2 246 (704)	7,53 (0,21)	13,84 (0,43)	25,43 (1,09)	20,78 (1,00)	8,95 (0,50)
ÚS	1 447 (469)	6,68 (0,23)	12,32 (0,47)	22,71 (1,93)	18,58 (1,09)	8,01 (0,55)
VŠ	347 (120)	5,61 (0,41)	10,71 (0,85)	20,43 (2,22)	16,94 (2,11)	7,41 (1,02)
směs	4 753 (1501)	7,48 (0,15)	14,03 (0,32)	26,51 (0,84)	22,94 (0,79)	9,53 (0,39)
2013/2014						
muži	1 294 (652)	6,46 (0,20)	11,11 (0,33)	19,09 (0,76)	15,34 (0,65)	6,32 (0,35)
ženy	1 528 (766)	6,66 (0,15)	11,66 (0,34)	20,40 (0,58)	16,45 (0,54)	6,78 (0,34)
Z	371 (122)	9,28 (0,61)	17,21 (1,32)	31,89 (3,59)	26,15 (3,34)	11,30 (1,68)
S	1 245 (602)	6,69 (0,21)	11,42 (0,35)	19,50 (0,78)	15,65 (3,34)	6,41 (0,36)
ÚS	892 (436)	6,56 (0,24)	10,96 (0,38)	18,32 (0,83)	14,64 (0,70)	5,88 (0,38)
VŠ	314 (187)	4,52 (0,31)	8,08 (0,50)	14,43 (1,12)	11,69 (1,00)	4,95 (0,52)
směs	2 822 (1418)	6,57 (0,14)	11,38 (0,24)	19,80 (0,55)	16,26 (0,48)	6,61 (0,25)

Poznámka: Vzdělání Z základní, S střední, ÚS úplné střední (s maturitou) VŠ (vysokoškolské) a směs pro všechna tři sledovaná období (v měsících). V závorce jsou uvedeny počty nezaměstnaných, kteří práci našli (sloupec *n*) a odhad směrodatné odchylky odhadu (ostatní sloupce).

Zdroj: Malá (2013), vlastní výpočty

i pozitivní vliv maturitní zkoušky na délku nezaměstnanosti. Přesnost odhadů je podobná pro jednotlivé skupiny, vždy je nejméně přesně určena hodnota pro nezaměstnané se základním vzděláním. Tato okolnost je také patrná z hodnot kvartilové odchylky, kdy největší variabilita sledovaných dob nezaměstnanosti je vždy pro skupinu nezaměstnaných se základním vzděláním. Dále je patrné prodloužení všech odhadnutých dob v období probíhající krize, hodnoty se následně opět vrátily k hodnotám před krizí (též obrázek 4). K tomuto jevu dochází přesto, že v době krize docházelo k nárůstu nezaměstnanosti a čerstvě nezaměstnaní dobu nezaměstnanosti zkracují. Převládla ovšem skutečnost, že velmi málo nezaměstnaných práci nacházelo. Hodnoty charakteristik pro celou populaci (modelované směsí rozdělení popisujících vzdělání) přibližně odpovídají střednímu vzdělání bez maturity, pro všechna období jsou charakteristiky polohy doby nezaměstnanosti kratší pro úplné střední vzdělání. Rozdíl je ve variabilitě, která je pro směs menší než pro podmnožiny, neboť ty jsou homogennější než celá populace. Je třeba připomenout, že data jsou uvažována pro nezaměstnanost do dvou let, proto hodnoty charakteristik výrazně delší (zvláště dosažené pro střední vzdělání) lze na základě modelu vyčíslit (včetně směrodatné odchylky), ale je třeba si uvědomit, že v takovém případě je vliv volby rozdělení a modelu zásadnější, než v případě hodnot v rozsahu dat.

Na obrázku 3 jsou znázorněny odhadnuté komponentní hustoty pro všechna tři období (od leva doprava), čtyři skupiny nezaměstnaných a pro celou populaci (hustota směsí). Vzhledem ke stejnému měřítku grafů je zřejmý rozdíl mezi prvním a třetím obdobím na straně jedné a druhým obdobím probíhající krize na straně druhé. Obě středoškolské skupiny dobře odpovídají rozdělení směsí již proto, že mají největší váhu (tabulka 1, celkem 73–77 %). V období krize je patrný pozitivní vliv maturitní zkoušky.

Obrázek 3 | Odhadnuté hustoty rozdělení doby do nalezení práce pro skupiny uchazečů rozdělené podle nejvyššího dosaženého vzdělání

Poznámka: Obrázky pro období 2007/2008 (vlevo), 2010/2011 (střed) a 20013/2014 (vpravo).

Zdroj: vlastní výpočty

Na obrázku 4 jsou znázorněny odhady z tabulky 1 tak, aby bylo patrné rozdělení jednotlivých složek a také vývoj v čase. Mediány a střední hodnoty jsou spojeny čarou tak, aby byl znázorněn průběh těchto hodnot ve sledovaných obdobích a také velikost změny.

Obrázek 4 | Charakteristiky doby do nalezení práce pro celou populaci nezaměstnaných (směs) a jednotlivé složky

Poznámka: Vždy tři časová období 2007/2008, 2010/2011 a 2013/2014 vedle sebe. Čarou je znázorněn vývoj v čase mediánu a střední hodnoty.

Zdroj: vlastní výpočty

Tabulka 2 | Hodnoty odhadnutých distribučních funkcí doby do nalezení práce do jednoho (12 měsíců) a dvou let (24 měsíců)

Období	2007/2009		2010/2011		2013/2014	
	12	24	12	24	12	24
muži	0,54	0,81	0,46	0,74	0,54	0,83
ženy	0,50	0,79	0,41	0,69	0,51	0,81
Z	0,53	0,82	0,26	0,52	0,52	0,64
S	0,55	0,81	0,44	0,73	0,55	0,85
ÚS	0,68	0,90	0,49	0,77	0,68	0,83
VŠ	0,51	0,80	0,55	0,80	0,53	0,90
směs	0,54	0,81	0,43	0,71	0,54	0,82

Zdroj: vlastní výpočty

Z těchto rozdělení je také možno odhadnout hodnoty distribuční funkce v čase 12 a 24 měsíců, to znamená pravděpodobnost, že nezaměstnaný najde práci do jednoho a do dvou let. Tyto hodnoty jsou uvedeny v tabulce 2. Do dvou let našlo zaměstnání 90 % vysokoškoláků, v době krize o 10 p. b. méně. V celé populaci (na základě směsi) je to 80 % (70 % v době krize), po dvou letech nezaměstnanosti zbývalo bez práce ještě 20 % nezaměstnaných v prvním období, 29 % v době krize a 18 % v době oživení.

Tabulka 3 | Odhadnuté maximální hodnoty funkce rizika (v měsících), odhadnuté směrodatné odchylky jsou uvedeny v závorce

Období	pohlaví		vzdělání				směs
	muži	ženy	Z	S	ÚS	VŠ	
2007/2008	8,96 (0,69)	9,79 (0,66)	11,41 (2,01)	9,96 (0,72)	9,06 (0,83)	7,10 (1,09)	9,20 (0,50)
2010/2011	9,68 (0,64)	10,72 (0,73)	12,47 (2,33)	10,91 (0,76)	9,60 (0,75)	7,37 (1,43)	9,73 (0,49)
2013/2014	10,84 (0,71)	10,97 (0,69)	13,15 (2,28)	11,55 (2,37)	11,86 (2,19)	7,02 (1,08)	10,66 (0,42)

Zdroj: vlastní výpočty

V tabulce 3 jsou uvedeny doby, ve kterých dochází k největší intenzitě získávání práce, tedy doby nezaměstnanosti, ve kterých je maximální funkce rizika. Tyto doby jsou též znázorněny svislými čarami na obrázcích 5 a 6, intenzita nacházení práce roste k bodu maxima a pak pomalu klesá. Pro všechny skupiny se doba prodlužuje vzhledem ke sledovaným obdobím (celkem o necelé 2 měsíce), k maximu dochází dříve u mužů než u žen a doby maxima jsou seřazeny sestupně se vzrůstajícím vzděláním. Době určené z modelu směsi opět odpovídá doba pro úplné středoškolské vzdělání. Směrodatné odchylky jsou největší ve skupinách pro základní vzdělání (nejvyšší hodnota maxima) a vysokoškolské vzdělání (nejnižší hodnota maxima). V obou případech hodnotu ovlivňuje nižší počet pozorování ve skupinách a dále se projevuje širší spektrum volby vysokoškoláků a problémy osob se vzděláním základním. Na obrázku 5 (pro komponenty definované pohlavím) a obrázku 6 (komponenty definované vzděláním) je vidět, že v období krize byla menší intenzita nacházení práce a hodnoty v maximech jsou na prostředním obrázku nižší.

Těmto výsledkům odpovídá také průběh charakteristik polohy zbytkových dob nezaměstnanosti. Medián a střední hodnota zbytkové doby nezaměstnanosti klesají od hodnoty v bodě 0 (charakteristiky celkové doby nezaměstnanosti uvedené v tabulce 1) k minimální hodnotě a pak opět rostou (a růst je přibližně lineární v čase). Směrnice růstu jsou vždy větší pro střední hodnotu než pro medián (střední hodnota zbytkové doby roste rychleji než její medián). V tabulce 4 jsou odhadnuty délky nezaměstnanosti, ve kterých je dosaženo minima, a dále hodnoty charakteristiky zbytkové doby v tomto bodě. Minimální hodnoty je dosaženo dříve pro medián než pro střední hodnotu, rozdíl byl 1,2–2,2 měsíce. V roce 2010 se všechny charakteristiky posunuly ve smyslu vodorovné i kolmé

osy, v roce 2014 se zkrátily střední hodnoty, nicméně poloha minima se posunula ještě o 2 měsíce doprava (obdobně jako v tabulce 3 doba maxima funkce rizika). Vzhledem k tomu, že po oživení ekonomiky sice více lidí najde práci, ale dlouhodobě nezaměstnaní zůstávají, a proto se doba nezaměstnanosti zkracuje, pro změnu struktury je třeba delšího času než v předkládané analýze.

Obrázek 5 | Odhadnuté funkce rizika pro nezaměstnané muže a ženy a výsledné směsi (založené na vzdělání)

Poznámka: Svislé čáry znázorňují body maxima funkcí. Obrázky pro období 2007/2008 (vlevo), 2010/2011 (střed) a 20013/2014 (vpravo).

Zdroj: vlastní výpočty

Obrázek 6 | Odhadnuté funkce rizika pro skupiny popsané vzděláním a výsledné směsi

Poznámka: Svislé čáry znázorňují body maxima funkcí. Obrázky pro období 2007/2008 (vlevo), 2010/2011 (střed) a 20013/2014 (vpravo).

Zdroj: vlastní výpočty

Tabulka 4 | Odhady charakteristik polohy zbytkových dob nezaměstnanosti

	délka nezaměstnanosti (měsíce)			
složka	$\min m_t$	t	$\min e_t$	t
2007/2008				
muži	8,63 (0,45)	5,82 (0,66)	14,44 (1,00)	4,30 (0,55)
ženy	9,19 (0,46)	6,41 (0,62)	13,31 (0,98)	4,76 (0,55)
Z	12,72 (1,81)	6,98 (1,50)	21,86 (2,71)	5,06 (0,98)
S	8,38 (0,46)	6,79 (0,72)	13,72 (0,92)	5,11 (0,61)
ÚS	8,24 (0,56)	6,00 (0,83)	13,66 (1,14)	4,47 (0,70)
VŠ	5,99 (0,79)	4,83 (1,10)	9,81 (1,59)	3,64 (0,93)
směs	8,89 (0,48)	5,96 (0,32)	15,11 (0,43)	4,35 (0,77)
2010/2011				
muži	10,77 (0,50)	5,93 (0,64)	18,48 (1,16)	4,30 (0,52)
ženy	12,22 (0,59)	6,50 (0,73)	21,08 (1,39)	4,70 (0,59)
Z	20,98 (1,12)	6,25 (1,44)	39,50 (2,45)	4,28 (1,01)
S	10,91 (0,50)	6,98 (0,71)	18,37 (1,11)	5,13 (0,59)
ÚS	9,75 (0,52)	6,10 (0,71)	16,47 (1,16)	4,48 (0,58)
VŠ	8,86 (1,03)	4,36 (1,38)	15,44 (2,54)	3,13 (1,12)
směs	11,49 (0,29)	5,79 (0,46)	20,97 (1,44)	3,95 (0,48)
2013/2014				
muži	7,84 (0,36)	7,88 (0,70)	12,56 (0,70)	6,06 (0,61)
ženy	5,53 (0,38)	7,56 (0,72)	13,83 (0,76)	5,74 (0,61)
Z	13,73 (1,64)	8,26 (2,16)	23,30 (3,78)	6,04 (1,79)
S	7,96 (0,37)	8,39 (0,73)	12,68 (0,70)	6,18 (0,64)
ÚS	7,30 (0,41)	8,89 (0,88)	11,49 (0,75)	6,97 (10,79)
VŠ	6,10 (0,57)	4,73 (1,10)	10,04 (1,23)	3,54 (0,93)
směs	8,21 (0,51)	7,46 (0,25)	13,65 (0,49)	5,48 (0,68)

Poznámka: Jsou uvedeny minimální hodnoty a časové okamžiky, ve kterých k nim dochází.

Zdroj: vlastní výpočty

3. Závěr a diskuse

Model konečné směsi pravděpodobnostních rozdělení umožňuje popsat pravděpodobnostní rozdělení náhodné veličiny v případě, že základní populace je nehomogenní (vzhledem k pravděpodobnostnímu rozdělení zkoumaného znaku) a lze předpokládat, že existují homogennější podmnožiny, ve kterých lze rozdělení lépe modelovat. Dostáváme tak informaci nejen o celé populaci, ale také o jednotlivých komponentách (podmnožinách). V tomto textu je použit model, ve kterém lze příslušnost k podskupinám pozorovat, je tedy známá a popsána vysvětlujícími proměnnými (v našem případě pohlavím nebo nejvyšším dosaženým vzděláním nezaměstnaného). Hustota pravděpodobnosti směsi je popsána jako vážený průměr komponentních hustot (s vahami odpovídajícími zastoupení komponent v základní populaci). Tato formulace umožňuje také přímočarou a intuitivní interpretaci výsledků. Pokud chceme modelovat dobu nezaměstnanosti, přirozeně máme pozorování délky nezaměstnanosti pro ty, kteří práci našli (úplné pozorování), a také informaci o těch, kteří v době šetření práci ještě neměli (neúplné, cenzorované pozorování). Proto se nabízí použít metody, které byly navrženy v analýze přežívání pro cenzorovaná data. V případě dat z registrů úřadů práce jsou hodnoty úplné nebo zprava cenzorované (Jarošová, 2006). V případě dat z VŠPS, používaných v předkládaném textu, místo úplných pozorování máme pouze intervaly, ve kterých nezaměstnaný práci našel. Data jsou intervalově cenzorovaná (pro ty, kteří práci našli) a zprava cenzorovaná (pro ty, kteří jsou stále nezaměstnaní). V tomto textu jsme se pokusili ukázat, že metody jsou dobře využitelné a pohled z hlediska charakteristik používaných v analýze přežívání může být zajímavý a přínosný.

Byly zvoleny vysvětlující proměnné, jejichž vliv na nezaměstnanost a její délku je znám a často diskutován – pohlaví a vzdělání (nicméně postup považujeme za univerzálně použitelný). Lze očekávat delší doby nezaměstnanosti pro ženy než pro muže a klesající délku doby nezaměstnanosti v závislosti na vzdělání. V předkládaném textu jsou tyto závislosti popsány a kvantifikovány z různých úhlů pohledu, rozdíly mezi muži a ženami jsou menší než rozdíly mezi skupinami definovanými pomocí vzdělání.

Model konečné směsi rozdělení s pozorovanými příslušnostmi ke komponentám vyžaduje pro odhad neznámých parametrů dostatečný počet pozorování v každé kombinaci hodnot vysvětlujících proměnných definujících komponenty. Proto byly v této práci použity jen dvě a čtyři komponenty a nepřihlíželi jsme k dalším charakteristikám, jako například věk, kde jsme uvažovali všechny nezaměstnané ve věku 16–65 let. V práci Čably, Malé (2017) je použit regresní model AFT, který umožňuje zařazení více vysvětlujících proměnných, a tedy podrobnější dělení nezaměstnaných. Kromě tří sledovaných období, pohlaví a vzdělání byly použity další proměnné jako věk (uvažovaný v pětiletých kategoriích, první kategorie 16–20 let a poslední 50–54, nezaměstnaní od 55 let jsou zařazení v jedné skupině) a velikost obce, kde nezaměstnaný žije.

Analyzovaná data se týkají tří let, které popisují období těsně před nástupem ekonomické krize, v období vrcholu a nakonec v době ekonomického oživení. Použité postupy směřují k posouzení rozdílů (kvantifikaci rozdílů) v době nezaměstnanosti a jejím

rozdělení. Zvolená období umožňují posoudit vliv ekonomické krize na délku nezaměstnanosti v jednotlivých podskupinách.

V případě, že je sestrojen model pro dobu nezaměstnanosti a jsou odhadnuty jeho parametry (v předkládaném textu byly použity maximálně věrohodné odhady), lze odhadnout hodnoty libovolných charakteristik. Směrodatné odchylky pro všechny odhadované charakteristiky byly určeny numericky pomocí metody bootstrap, časově náročný přístup poskytoval možnost získat i odchylky pro veličiny v tabulce 4, týkající se zbytkových dob nezaměstnanosti (určovaných numericky) a hodnot času, ve kterých bylo dosaženo minima. Všechny výsledky uvedené v práci jsou založeny na numerických výpočtech, proto nejsou uvedeny žádné explicitní vzorce, které ani nelze odvodit.

V analýze podle Akaikova informačního kritéria (které zahrnuje penalizaci za počet parametrů v modelu) byl jako optimální model pro popis délky nezaměstnanosti vybrán model směsi založené na skupinách definovaných nejvyšším dosaženým vzděláním. Podle všech výstupů (tabulky 1–4) však jsou získané hodnoty velmi podobné pro jedno logaritmicko-normální rozdělení a obě uvažované směsi. Nicméně nejen vzhledem k Akaikovu kritériu, ale také k informaci o jednotlivých podskupinách, model směsi se ukazuje být vhodnou volbou.

Analýzovaná data zahrnují dobu před ekonomickou krizí, roky v době krize a potom rok 2014, kdy bylo již patrné oživení ekonomiky. Výsledky ukazují, že doba nezaměstnanosti sice silně závisí na míře nezaměstnanosti, ale zatímco míra nezaměstnanosti odráží více okamžitý stav, doba nezaměstnanosti více zohledňuje minulost v tom smyslu, že nezaměstnaní, kteří o práci přišli v předchozím období (v době nízké nezaměstnanosti pro období krize nebo naopak v době krize pro dobu oživení), ovlivňují délku nezaměstnanosti prostřednictvím svých (delších) délek nezaměstnanosti. Bereme v úvahu osoby s délkou nezaměstnanosti do dvou let, proto se doby jen dotýkají (vzhledem k uvažovaným letem pozorování).

Pro analýzu byli zvoleni všichni nezaměstnaní s dobou nezaměstnanosti do dvou let (v souladu s volbou v Jarošové, 2006), tato doba obsahuje i první rok dlouhodobé nezaměstnanosti. Pokud bychom tento předpoklad neudělali, při stejné metodice by podmíněné rozdělení (za podmínky, že doba nezaměstnanosti je delší než rok) popisovalo rozdělení dlouhodobé délky nezaměstnanosti. Bohužel ale počet nezaměstnaných, kteří práci našli později než po 24 měsících nezaměstnanosti, je v analyzovaných datech malý, a proto ani odhady by nebyly pro dlouhé doby dostatečně přesné. Pro popis rozdělení dlouhodobé nezaměstnanosti by bylo třeba zvolit jiný přístup.

Literatura

- Bover, O., Arellano, M., Bentolila, S. (2002). Unemployment Duration, Benefit Duration and the Business Cycle. *The Economic Journal*, 112(479), 223–265, <https://doi.org/10.1111/1468-0297.00034>
- Canty, A., Ripley, B. (2016). *A Package for Bootstrap Functions*. Version 1.3-18. Dostupné z: <https://cran.r-project.org/web/packages/boot/boot.pdf>.

- Čabla, A. (2014). Unemployment Duration before and during The Economic Crisis in the Czech Republic. *Acta Aeria Publici*, 11, 19–26.
- Čabla, A. (2016). Minimal Adequate Model of Unemployment Duration in the Post-Crisis Czech Republic. *Statistika*, 96, 50–62.
- Čabla, A., Malá, I. (2017). Modelling of Unemployment Duration in the Czech Republic. *Prague Economic Papers* – Articles first published online, <https://doi.org/10.18267/j.pep.620>
- ČSÚ (2016). Český statistický úřad. Dostupné z: <http://www.czso.cz> www.czso.cz
- Daveri, F., Tabellini, G. (2000). Unemployment and Taxes – do Taxes Affect the Rate of Unemployment? *Economic Policy*, 30, 47–88.
- Fay, M. P., Shaw P. A. (2010). Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The Interval R Package. *Journal of Statistical Software*, 36(2), 1–34, <https://doi.org/10.18637/jss.v036.i02>
- Hagedorn, M., Manovskii, I. (2013). Job Selection and Wages over the Business Cycle. *The American Economic Review*, 103(2), 771–803, <https://doi.org/10.1257/aer.103.2.771>
- Hunt, J. (1995). The Effect of Unemployment Compensation on Unemployment Duration in Germany. *Journal of Labor Economics*, 13(1), 88–120, <https://doi.org/10.1086/298369>
- Jarošová, E. (2006). Modelování délky trvání nezaměstnanosti. *Statistika*, 86(3), 240–251.
- Jurajda, Š., Tannery, F. J. (2003). Unemployment Durations and Extended Unemployment Benefits in Local Labor. *Industrial and Labor Relations Review*, 56(2), 324–348, <https://doi.org/10.2307/3590941>
- Korpi, T. (2001). Accumulating Disadvantage: Longitudinal Analyses of Unemployment and Physical Health in Representative Samples of the Swedish Population. *European Sociological Review*, 17(3), 255–273, <https://doi.org/10.1093/esr/17.3.255>
- Krueger, A. B., Mueller, A., Davis S. J., Ayşegül, Ş. (2011). Job Search, Emotional Well-Being, and Job Finding in a Period of Mass Unemployment: Evidence from High Frequency Longitudinal Data [with Comments and Discussion]. *Brookings Papers on Economic Activity*, 1–81.
- Krueger, A. B., Cramer, J., Cho, D. (2014). Are the Long-Term Unemployed on the Margins of the Labor Market? *Brookings Papers on Economic Activity*, 229–280.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed. Hoboken: John Wiley & Sons. ISBN 978-0-471-37215-8.
- LFSS (2015). Eurostat. Dostupné z: <http://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey>
- Löster, T., Langhamrová, J. (2011). Analysis of Long-term Unemployment in the Czech Republic. Praha 22.12.2011 - 23.12.2011. In: LÖSTER, T., PAVELKA, ed., *International Days of Statistics and Economics*. Slaný: Melandrium, 228–234.
- Malá, I. (2013). Použití konečných směrů pravděpodobnostních rozdělení pro modelování rozdělení doby nezaměstnanosti v České republice. *Acta Oeconomica Pragensia*, 21(5), 47–63, <https://doi.org/10.18267/j.aop.415>
- McLachlan, G. J., Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section, New York.
- R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Dostupné z: <http://www.R-project.org/>.

- Røed, K., Jensen, P., Thoursie, A. (2008). Unemployment Duration and Unemployment Insurance: a Comparative Analysis Based on Scandinavian Micro Data. *Oxford Economic Papers*, 60(2), 254–274, <https://doi.org/10.1093/oep/gpm021>
- Therneau, T. (2015). *A Package for Survival Analysis in S*. version 2.38. Dostupné z: <http://CRAN.R-project.org/package=survival>
- Wicher, L., Wilke, A. R. (2008). Simple Non-parametric Estimators for Unemployment Duration Analysis. *Appl. Statist.*, 57(1), 117–126.